UNIVERSITÀ DEGLI STUDI
DI TRENTO

DEPARTMENT OF INFORMATION ENGINEERING AND COMPUTER SCIENCE
ICT International Doctoral School

Predictive Modeling of Human Behavior:

Supervised Learning from Telecom Metadata

by

Andrey Bogomolov

Advisor
Prof. Fabio Pianesi,
Fondazione Bruno Kessler, EIT Digital

Industrial Co-Advisor
Michele Caraviello,
Telecom Italia SpA

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE

This thesis is dedicated to developing
Artificial Intelligence for Social Good

Acknowledgements

The work of Andrey Bogomolov is partially supported by Telecom Italia Semantics and Knowledge Innovation Laboratory (SKIL) research grant “T” and EIT Digital Doctoral School grant.

For happiness and stress recognition problems the data was collected using FUNF Open Sensing Framework based on Android Developer API in compliance with the Android Software Development Kit License Agreement issued by Google Inc., a Delaware corporation with principal place of business at 1600 Amphitheatre Parkway, Mountain View, CA 94043, United States.

The experiments involving United States individuals or conducted on the territory of United States were set up in full comliance with National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research regulations concerning basic ethical principles that should underlie the conduct of biomedical and behavioral research involving human subjects and guidelines assuring that such research is conducted in accordance with those principles (“The Common Rule”, 45 CFR pt. 46).

For crime hotspot prediction problem the source data processing, reported data transformations and feature selection, which were derived from anonymised and aggregated mobile network dataset, provided by Telefónica Digital Limited, a corporation existing under the laws of England, with company number 07884976 and a registered office or regular place of business located at 260 Bath Road, Slough, Berkshire, SL1 4DX, were done during the public competition – “Datathon for Social Good” organized by Telefónica, The Open Data Institute and the MIT during the Campus Party Europe 2013 at the O2 Arena in London during 2-7 September 2013, in full compliance with the competition rules and legal limitations imposed by the “Terms and Conditions” document.

For energy prediction problem the source data was processed in full compliance with the rules of the public competition – Telecom Italia Big Data Challenge 2014.

During the my graduate research career I have benefited from the help, support, advice and suggestions of many people. I am gratefull to my thesis advisor, Fabio Pianesi, provided years of technical and methodological guidance, advice, and support. I would like to thankall the people from Telecom Italia Semantics and Knowledge Innovation Laboratory (SKIL) team and my industrial advisor Michele Caraviello for productive collaboration, wise management and highly professional software engineering support. I am truely gratefull to my friends and team members who supported me with positive emotions and fantastic high mountain sports experience.

Abstract

Big data, specifically Telecom Metadata, opens new opportunities for human behavior understanding, applying machine learning and big data processing computational methods combined with interdisciplinary knowledge of human behavior.

In this thesis new methods are developed for human behavior predictive modeling based on anonymized telecom metadata on individual level and on large scale group level, which were studied during research projects held in 2012-2016 in collaboration with Telecom Italia, Telefonica Research, MIT Media Lab and University of Trento. It is shown that human dynamics patterns could be reliably recognized based on human behavior metrics derived from the mobile phone and cellular network activity (call log, sms log, bluetooth interactions, internet consumption).

On individual level the results are validated on use cases of detecting daily stress and estimating subjective happiness. An original approach is introduced for feature extraction, selection, recognition model training and validation. Experimental results based on ensemble stochastic classification and regression tree models are discussed.

On large group level, following big data for social good challenges, the problem of crime hotspot prediction is formulated and solved. In the proposed approach we use demographic information along with human mobility characteristics as derived from anonymized and aggregated mobile network data. The models, built on and evaluated against real crime data from London, obtain accuracy of almost 70% when classifying whether a specific area in the city will be a crime hotspot or not in the following month. Electric energy consumption patterns are correlated with human behavior patterns in highly nonlinear way. Second large scale group behavior prediction result is formulated as predicting next week energy consumption based on human dynamics analysis derived out of the anonymized and aggregated telecom data, processed from GSM network call detail records (CDRs). The proposed solution could act on energy producers/distributors as an essential aid to smart meters data for making better decisions in reducing total primary energy consumption by limiting energy production when the demand is not predicted, reducing energy distribution costs by efficient buy-side planning in time and providing insights for peak load planning in geographic space.

All the studied experimental results combine the introduced methodology, which is efficient to implement for most of multimedia and real-time applications due to highly reduced low-dimensional feature space and reduced machine learning pipelines. Also the indicators which have strong predictive power are discussed opening new horizons for computational social science studies.

Keywords: telecom metadata, call detail records, supervised learning, individual characteristics, affective states, crime hotspots, electric energy consumption, human behavior understanding, social good, human dynamics, big data, machine learning, artificial intelligence

Contents

1 Introduction **1**
1.1 Motivation - The Problem of Human Behavior Computation 2
1.2 Research Goals 3
1.3 Contribution and Structure 4
1.4 Main Publications List 6

2 Supervised Learning of Individual Characteristics: Affective States **7**
2.1 Happiness Recognition from Telecomunications Metadata 8
2.2 Daily Stress Recognition from Mobile Phone Data 24

3 Predictive Modeling of Large Scale Group Behavior **40**
3.1 Predicting Crime Hotspots Based on Aggregated Anonymized People Dynamics 41
3.2 Predicting Electric Energy Consumption Using Aggregated Telecom Data 51

4 Summary and Future Work **54**
4.1 Computing Human Behavior: The New Kind of Science 55
4.2 Supervised Learning of Individual Characteristics 55
4.3 Supervised Learning of Large Scale Group Behavior 56
4.4 Challenges 57

A Gradient Boosting Implementation for Stress Recognition Problem **60**

B Additional Visualizations for Energy Consumption Prediction Use Cases **62**

Bibliography **68**

List of Figures

1 Recorded Happiness Scores Density 11
2 Within Person and Between-subject Variance for Happiness Recognition Problem 12
3 Misclassification Error vs Number of Tree Classifiers in Ensemble Model for Happiness Recognition Problem 18
4 Final Model ROC Curve for Happiness Recognition Problem 23
5 Recorded Daily Stress Scores Histogram 29
6 Within- and Between-Subject Variance for Daily Stress Recognition Problem 29

7 Sample Visualization of Anonymized Data by Smartsteps Platform . 44
8 Ground Truth of Crime Hotspots 48
9 Predicted by Algorithm Crime Hotspots 48

10 Spatially Mapped Max Energy Consumption vs Telecom Patterns for a Weekday on a SET grid 63
11 Energy Consumption Time Series Decomposition for a Residential Area 64
12 Energy Consumption Time Series Decomposition for a Touristic Area 64
13 Energy Consumption Time Series Decomposition for a City Center and Industrial Areas 64
14 Spectral Characteristics of Typical Energy Consumption Response Variables 65

List of Tables

2.1 Recorded Daily Happiness 11
2.2 List of Basic Features for Happiness Recognition Problem 15
2.3 List of Basic Bluetooth Proximity Features for Happiness Recognition Problem 16
2.4 Top-30 Features for Happiness Recognition 17
2.5 SVM Kernel Functions 17
2.6 10-fold CV Model Comparison for Happiness Recognition Problem: Accuracy 19
2.7 10-fold CV Model Comparison for Happiness Recognition Problem: Kappa 20
2.8 10-fold Cross Validation Happiness Recognition Metrics for Each Fold Comparison 20
2.9 Final Classifier Happiness Recognition Model Confusion Matrix for Training Set 20
2.10 Final Classifier Happiness Recognition Model Confusion Matrix for Test Set 21
2.11 Final Classifier Happiness Recognition Model Performance Metrics Comparison 21
2.12 Selected Features Ranked by Mean Decrease in Accuracy for Daily Stress Recognition Problem 28
2.13 List of Basic Features for Daily Stress Recognition Problem 30
2.14 List of Basic Bluetooth Proximity Features for Stress Recognition Problem 32
2.15 Stress Recognition Model Performance Metrics 36
2.16 10-fold Cross-Validation Metrics for Stress Recognition Problem . . . 36
2.17 Stress Recognition Model Metrics Comparison for Feature Subsets . . 36

3.1 SmartSteps Data Description 44
3.2 Number of Crime Hotspots in January 46
3.3 Crime Hotspots Prediction Top-20 Selected Features Ranked by Mean Decrease in Accuracy 48
3.4 Crime Hotspots Prediction Problem Metrics Comparison 49
3.5 Mean Daily Consumption Model Metrics for Energy Prediction Problem 53
3.6 Peak Daily Consumption Model Metrics for Energy Prediction Problem 53

B.1 Mean Daily Energy Consumption Predictor Variables 66
B.2 Peak Daily Energy Consumption Predictor Variables 67

Chapter 1

Introduction

1.1 Motivation - The Problem of Human Behavior Computation

Since the beginnings in the 1930s, computation has been primarily approached from the tradition of engineering (which seeks to build practical systems using computations) and mathematics (which seeks to prove theorems about computation). And starting from 1970s, computing has been described as being at the crossroads of mathematical, engineering, and empirical traditions[Denning, 2003].

Recent advancements in artificial intelligence and machine learning open new frontiers for computation, specifically for human behavior understanding on personal and group level, thus fuel up *Computational Social Science* and *Social physics* – the new interdisciplinary fields of science which attract growing academic, government and business interest.

Big data, specifically Telecom Metadata, opens new opportunities for human behavior understanding, combining machine learning and big data processing computational methods multiplied by the interdisciplinary knowledge of human behavior.

The computer hardware capable for terabyte-scale computing is becoming affordable for mid-sized companies and universities. Cloud computing and hardware virtualization delivers simplicity, availability and scalability at a fraction of the cost. *High Performance Computing* is shifting from expensive supercomputers and specialized HPC clusters to commodity servers, running as GPU-accelerated and heterogeneous computing.

How human could create companies and government agencies that are efficient, cooperative and creative? These are the questions of social physics, and they are especially important in global competition, environmental challenges, economic and political crises and government failures. The first engine that drives social physics is big data. Specifically, *data driven modeling* is the second wave of *New Kind of Science*, exploiting the nature of computation to be explored experimentally. The results of these experiments have great relevance to understanding the world than traditional modelling by mathematical functions or simulation.

In modern science *Social physics* refers to using *big data analysis* and the mathematical laws to understand the behavior of human crowds [George et al., 2014]. The core idea is that data about human activity, such as phone call records, sms records, geo location data from mobile phone, credit card transactions, social networks and web activity logs, contain mathematical patterns that are characteristic of the way how social interactions evolve in time and space. These mathematical models could serve as a filter for analysis of behavior change and for detecting emerging behavioral patterns for a bunch of filed of science, such as Economics, Finance, Psychology, City Science, Architecture, Law, Public Security, Criminology, Politics, Ecology, Energy Efficiency, Biology, Medicine, etc.

The new ubiquitous digital data from the smart phone and telecommunication network is becoming available for understanding all the aspects of human life. By using these data to build predictive models, computational theory of human behavior we could engineer better social systems.

1.2 Research Goals

At the highest level, our goal is to evaluate and improve upon the state-of-the-art scholar results of software-enabled human behavior understanding derived from behavioral big data. We focus specifically on telecom metadata, since this is obviously cheap and ubiquitous source of behavioral information.

Since we are interested in practical and applied *data driven* results rather than theoretical solutions, we implement our machine learning algorithms in a state-of-the-art way, featuring the possibility to define and measure the impact of the actual predictive variables in a *human interpretable way.* For that, deep learning, convolutional neural networks and automated feature autoencoders techniques are out of scope of this research.

We are interested in *predicting* human behavior, overreaching *characterization* or *recognition.* This approach requires that the response variable that we predict for our use case problems is shifted in the future. Concerning this point, the key research goal is to test, if human behavior could be predicted on historical data and develop appropriate algorithmic methods for solving this type of problems.

Our approach starts with *individual behavior* recognition and prediction and extends to *large scale group behavior* prediction.

We explore use cases in *affective computing*, such as daily happiness and stress recognition, and extend our methodolody and computational approach to *crime hotspots prediction* and *energy consumption prediction* based on telecom and multimodal data.

These applications in machine learning are interesting both because they tend to suffer from noise, limited data quality, unpredictable latency during data collection, data privacy limitations, due to the large size of the arrays of source data with wide separation between different data providers, also because the data logging could be not regular and predictable based on internal policies of telecom companies in different countries.

By handling all of these cases, our methodology could cover a significant fraction of scientific and engineering limitations.

The next key research goal is *maximizing the performance* of computation for the cases that are covered. Since feature engineering involves huge labour costs as well as a benefit of human interpretability, we balance our solution minimizing these overheads while maximizing the time-to-solution benefits and computational speed to achieve the best overall performance for deploying such systems in real life setting.

1.3 Contribution and Structure

In this thesis we address several important problems related to the processing and analysis of complex, multi dimentional, structured and semi-structured data.

First, we do algorithmic recognition of individuals' affective states based on the data collected on their smart phones.

Second, we are the first who formulated and algorithmically solved the problem of predicting crime hot spots from telecom metadata in a city.

Third, we propose new methods for learning the intrinsic relationships between the characteristics given observations (logs) framed as machine learning problems with reduced and highly compactified feature space for energy consumption prediction problem.

Finally, the research effort presented in this thesis provides insight into developing novel signal processing and machine learning techniques for human behavior understanding.

The main scientific contributions of the thesis include the computational algorithms for supervised learning of human individual characteristics based on affective computing use cases, such as (1) happiness recognition from telecom metadata, and (2) daily stress recognition from mobile phone data and algorithms for modeling of large scale group behavior, such as (3) predicting crime hotspots based on aggregated and anonymized people dynamics, and (4) electric energy consumption using aggregated telecom network data. These contributions include feature engineering, problem formulation using machine learning approach and optmization algorithms development for non-linear classification and regression problems.

Also we show the benefit to the real world applications in social psychology, city science, criminology, energy efficiency and telecommunications.

The outline of this thesis is as follows.

In chapter 1 we summarize motivation for this research, show the nature of the problem of Human Behavior Computation. Next we show the research goals, scientific contribution and thesis structure. Finally, we list the main publications published during this research work.

In chapter 2 we go into details of state-of-the-art literature on affective states recognition, investigate the automatic recognition of people's happiness and daily stress from three different sets of data, such as people activity, as detected through their smartphones (data pertaining to transitory properties of individuals), weather conditions (data pertaining to transitory properties of the environment); and personality traits (data concerning permanent dispositions of individuals). Taken together, and despite the discussed limitations, our results not only provide evidence that individual happiness and daily stress can be reliably predicted, but they also point to the necessity of considering at the same time people's transitory properties (smartphone activity), transitory properties of the environment and information about stable individual characteristics. For the sake of transitory individual properties, mobile phone usage patterns have important advantage over alternative methods: they are less unobtrusive and raise limited privacy problems as compared to, e.g., voice analysis or the exploitation of data from physiological sensors. Moreover, and importantly, automatic stress detection based on mobile phone data can take advantage of the extensive usage and diffusion of these devices, it can be applied in several real world situations and it can be exploited for a variety of applications that are delivered by means of the same mobile device.

In chapter 3 we formulate the machine learning problems and discuss the solutions for predictive modeling of large scale group behavior, specifically we are (1) predicting crime hotspots based on aggregated anonymized people dynamics and (2) predicting electric energy consumption using aggregated telecom data.

We show that daily average and peak energy consumption prediction metrics by our approach for the next 7 days are much better than the normal time series forecasting baselines used in the industry. Our results prove that human dynamics, extracted from aggregated and anonymized mobile phone data, are good proxies for modelling crime hostpots and energy consumption. Also we discuss the practical implications for our results.

In chapter 4 we summarize the research results, propose future work and analyse challenges connected with predictive human behavior understanding based on telecom metadata and multimodal data.

1.4 Main Publications List

This thesis is a summary of computer science research conducted during 2013-2017 and widely published in the following conference and journal articles:

1. Andrey Bogomolov, Bruno Lepri, and Fabio Pianesi. Happiness recognition from mobile phone data. In *Social Computing (SocialCom), 2013 International Conference on*, pages 790–795. IEEE, 2013.

2. Andrey Bogomolov, Bruno Lepri, Michela Ferron, Fabio Pianesi, and Alex Sandy Pentland. Daily stress recognition from mobile phone data, weather conditions and individual traits. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 477–486. ACM, 2014.

3. Andrey Bogomolov, Bruno Lepri, Michela Ferron, Fabio Pianesi, and Alex Sandy Pentland. Pervasive stress recognition for sustainable living. In *Pervasive Computing and Communications (PERCOM), 2014 IEEE International Conference on*, pages 345–350. IEEE, 2014.

4. Andrey Bogomolov, Bruno Lepri, Jacopo Staiano, Nuria Oliver, Fabio Pianesi, and Alex Pentland. Once upon a crime: towards crime prediction from demographics and mobile data. In *Proceedings of the 16th international conference on multimodal interaction*, pages 427–434. ACM, 2014.

5. Andrey Bogomolov, Bruno Lepri, and Fabio Pianesi. Generalized compression dictionary distance as universal similarity measure. In *Big Data from Space (BiDS'14)*, number doi: 10.2788/1823, pages 63–66. Publications Office of the European Union, 2014.

6. Andrey Bogomolov, Bruno Lepri, Jacopo Staiano, Emmanuel Letouzé, Nuria Oliver, Fabio Pianesi, and Alex Pentland. Moves on the street: Classifying crime hotspots using aggregated anonymized data on people dynamics. *Big Data*, 3(3):148–158, 2015.

7. Andrey Bogomolov, Bruno Lepri, Roberto Larcher, Fabrizio Antonelli, Fabio Pianesi, and Alex Pentland. Energy consumption prediction using people dynamics derived from cellular network data. *EPJ Data Science*, 5(1):13, 2016.

Chapter 2

Supervised Learning of Individual Characteristics: Affective States

Supervised learning in Computer Science is an approach of creating a model from data, i.e. “learn from data”, such that is implies the existance of prior etalon labels or values, which directly characterize the explored process or states. Supervised Learning of Individual Characteristics of a human is a complex interdisciplinary problem, involving knowledge of machine learning, psychology, neuroscience and software engineering.

2.1 Happiness Recognition from Telecomunications Metadata

“Life, Liberty and the pursuit of Happiness” are described by Thomas Jefferson, the principal author of the Declaration of Independence, and the third President of the United States, as ”self-evident” and is among the basic rights of every person. Also the term “happiness” appears in several drafts of the European Union constitution. To be happy is one of the major goals, if not the ultimate goal, of human beings. During the last decades, the consequences and benefits of happiness have also come into the focus of research. Happiness might not only be a goal of life but also a means for reaching other goals and for facilitating desirable behaviors and outcomes. In a recent review, Lyubomirsky et al. [Lyubomirsky et al., 2005] showed that happy people are successful in many life domains and that this success is at least partly due to their happiness. Happy people are more social, altruistic, active, like themselves and others more, have strong bodies and immune systems, and have better conflict resolution skills. Moreover, pleasant moods promote creative thinking. For this reason, there was an explosion of research on happiness and subjective well-being. Several theories have been proposed to identify the causes of this elusive state (see for a general review [Myers and Diener, 1995]). At the same time, several studies dealt with the problem of measuring happiness. Researchers have generally relied on self-reports which are sometimes coupled with informant data, interviews by trained clinicians, unobtrusive observations of nonverbal expressions, and physiological assessments (see [Diener, 1994]). Respondents typically are asked to rate their levels of positive and negative affect over a particular period of time or to make a judgment of their overall life quality. Moreover, Zajonc and McIntosh [Zajonc and McIntosh, 1992] showed the current inadequateness of using psychophysiological measures or brain techniques for this task.

The development of the mobile phones and other mobile smart devices in recent years created new alternatives. Some digital applications have started to replace paper-based questionnaires to collect happiness data. For example, Track your Happiness [app, 2013b] is a web based platform for tracking and visualizing user’s happiness. It was created as part of Matt Killingsworths doctoral research at Harvard University. Three times a day the user is notified by email or text message to report on his smart phone what he is doing and how he is feeling during that activity. After 50 samples, a happiness report is created with details about how user’s happiness varies depending on a variety of factors, such as weekday or location. The main weakness of this tool is that the measuring method at the end is still based on self-reported surveys.

Instead, our approach proposes using smartphone usage patterns and the social interactions captured by these devices to automatically recognize users’ daily happiness. Nowadays, smartphones allow for unobtrusive and cost-effective access to previously inaccessible sources of data related to daily social behavior [Lane et al., 2010a]. These devices are, in fact, able to sense a wealth of behavioral data: i) location, ii) other devices in physical proximity through Bluetooth scanning, iii) communication data, including both metadata (logs of who, when, and duration) of phone calls and text messages (sms) as well as their actual contents, iv) scheduled events, v) operational status, vi) movement patterns, vii) usage information, etc. Recently, the social psychologist Geoffrey Miller wrote "The Smartphone Psychology Manifesto" in which he argues that the smartphone should be taken seriously as new research tools for psychology. In his opinion, these tools could revolutionize all fields of psychology and other behavioral sciences making these disciplines more powerful, sophisticated, international, applicable, and grounded in real-world behavior [Miller, 2012]. Muaremi et al. [Muaremi et al., 2012] suggested to use smartphones to measure the level of happiness of a person and of an entire community. Along similar lines, recent works have started using smartphone data to automatically infer users' personality traits [Chittaranjan et al., 2013], [Staiano et al., 2012b], [Montjoye et al., 2013].

In this work we formulated the automatic recognition of daily happiness as a 3-class classification problem based on information concerning: a) peoples activity, as detected through their smartphones; b) weather conditions; c) personality traits. The first information type is represented by features extracted from call and sms logs and from Bluetooth hits, able to capture (i) amount of call, sms and proximity; (ii) diversity of call, sms, and proximity; and (iii) regularity in user behaviors. Concerning weather conditions, the commonsense and the literature suggest, that they can have an important impact on daily happiness states. Finally, individual dispositions (as captured by personality traits) are expected to interact with situational and contextual aspects (including weather conditions) to play a role in determining daily happiness states.

Classification experiments are performed through a variety of approaches, including support vector machines, neural networks and random forests, yielding an accuracy up to 80.81%.

2.1.1 Related Work

The majority of works in affect recognition focuses on the recognition of emotional states using the visual and acoustic signals found in the speech [Schuller et al., 2005] in the gestures and in the faces of people [Cohn, 2007] [Pantic, 2009] (see [Zeng et al., 2009] for a survey of acoustic and visual approaches). While the authors sometimes use emotion, mood, and happiness interchangeably, they usually measure transient affective states and not moods or daily happiness. One of the most recent examples is the Mood Meter system that detects smiles using video cameras distributed around the MIT campus [Hernandez et al., 2012]. These momentary smiles, the momentary facial expressions or the momentary spoken expressions, usually studied in affective computing, tell little about long-term states such as daily happiness.

Recently, some researchers have started to investigate the relationships between social media data and dimensions related to happiness and social well-being (e.g. Satisfaction with Life). In particular, Quercia et al. [Quercia et al., 2012] studied the relationship between the sentiments expressed in tweets and community socio-economic well-being and they found that the higher the normalized sentiment score of a community's tweets, the higher the community's socio-economic well-being. Again, Quercia et al. [Quercia, 2013] collected Satisfaction With Life (SWL) test results from a Facebook application and showed that aggregate country-level results significantly vary across twelve rich countries and strongly correlate with official well-being scores.

To our knowledge the automatic recognition of happiness using smart phones has not been the main target of the research community so far. Some previous works tried to estimate only some possible determinants of happiness. BeWell [Lane et al., 2012] monitors three daily behaviors (activity, social interaction, and sleep patterns), describes the effect of the behavior on well-being, and provides some feedback to the user. Another similar example was proposed by Rabbi et al. [Rabbi et al., 2011] that tries to estimate the mental and physical health based on the amount of speech occurring in natural settings and the overall physical activity performed during a day. Moreover, Muaremi et al. [Muaremi et al., 2012] in a position paper proposed to use smartphones to automatic measure the happiness level of an individual and of an entire community.

Some previous papers tried to focus on mood assessment using mobile phones. Moturu, et al. [Moturu et al., 2011] explored the associations between sleep, mood and sociability by analyzing mobile-phone-generated social communication data and self-reported mood and sleep data. Rachuri et al. [Rachuri et al., 2010] proposed EmotionSense, a novel system for social psychology study of user emotion based on mobile phones. They also showed how the information collected by EmotionSense can be used by social scientists in order to understand the patterns of interaction and the correlation of emotions with places, groups, and activity. More recently, Ma et al. [Ma et al., 2012] proposed a novel framework called MoodMiner for assessing and analyzing mood in daily life. MoodMiner uses mobile phone datamobile phone sensor data and communication data (including acceleration, light, ambient sound, location, call log, etc.) to extract human behavior pattern and assess daily mood. The authors reported a not impressive performance of 50% of accuracy. A significant work is the one conducted this year by LiKamWa et al. [Robert LiKamWa, 2013], in which the authors reported a smartphone software system, called MoodScope, able to infer the mood of its user based on how the smartphone is used.

2.1.2 Dataset: Living Laboratory

We exploited a dataset, called "Friends and Family", capturing more than eight complete weeks in the lives of 117 subjects living in a married graduate student residency of a major US university, collected between 21 February, 2010 and 16 July, 2011. During this period, each participant was equipped with an Android-based cellular phone incorporating a sensing software explicitly designed for collecting mobile data. Such software runs in a passive manner and does not interfere with the normal usage of the phone [Aharony et al., 2011].

The data collected consisted of: i) call logs, ii) SMS logs, iii) proximity data, obtained by scanning near-by phones and other Bluetooth devices every five minutes, and iii) data from surveys administered to participants, which provided self-reported information about personality ("Big Five") and self reported information about daily happiness.

More specifically, social interactions were derived from Bluetooth proximity detection data in a manner similar to those in previous reality mining studies

[Eagle and Pentland, 2006], [Madan et al., 2010]. The FUNF phone sensing platform [app, 2013a] was used to detect Bluetooth devices in the user's proximity. The Bluetooth scan was performed periodically, every five minutes, in order to keep from draining the battery while achieving a high resolution for social interaction.

With this approach, the Bluetooth log of a given smartphone would contain the list of devices in its proximity, sampled every 5 minutes. Knowing the Bluetooth identifiers of each smartphone in the study, we could thus infer when 2 participants' phones were in proximity.

In total, the dataset consisted of 33497 phone calls, 22587 SMS, and 1460939 Bluetooth hits.

2.1.3 Happiness data

The participants were also asked to fill daily surveys about their self-perceived happiness. The happiness information was reported by the participants filling a seven items scale with 1 = "being very unhappy", 4 = "being neither unhappy, nor happy", and 7 = "being very happy". The distribution of the happiness score in the sample is provided in Table 2.1. The distribution of daily happiness is visualized in Fig. 1. We see that it has a negative skew – the density is moved to the higher region of happiness score. The distribution has negative excess kurtosis which is called platykurtic. In our case, that means that the focus group reported specific daily happiness score more often, than neutral. In our sample 91 subjects have the highest score, 43 – low score, 112 – neutral scores at least one day. Within person and between-subject variance for happiness scores (Fig. 2) show, that within-person daily happiness vary much higher than between-person. This property shows the possibility to reliably understand daily emotional states of the same person for each variety ground-truth class from the data.

As can be seen, most of the people for most of the time described themselves as having been from moderately to quite happy during the day.

Table 2.1: Recorded Daily Happiness

number of records	12991
mean	4.84
standard deviation	1.26
median	5.00
mean average deviation	1.48
min	1.00
max	7.00
range	6.00
skew	-0.39
kurtosis	-0.07

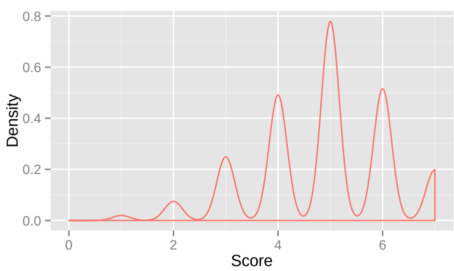

Figure 1: Recorded Happiness Scores Density

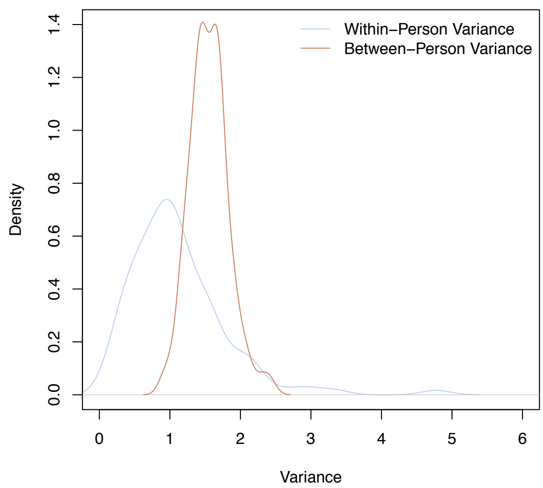

Figure 2: Within Person and Between-subject Variance for Happiness Recognition Problem

2.1.4 Personality Data

Several studies in social psychology highlighted the impact of personality traits on happiness: in particular, these studies showed the role played by extraversion and neuroticism [Vitters, 2001], [Lyubomirsky et al., 2006].

Costa and McCrae have also found evidence of a positive relationship between extraversion and happiness, whereas negative affect was more strongly correlated with neuroticism [Costa and McCrae, 1980]. In an effort to more fully explicate the relationship between extraversion and happiness, several studies have emerged. One of them conducted by Gray suggests that extraverts are more sensitive to rewarding stimuli (social and not) than introverts [Gray, 1991]. Numerous studies have shown that subjective well-being is related to the Five-Factor Model (FFM) of personality, especially the domains of Neuroticism, Extraversion, and Conscientiousness, and that, although subjective well-being is not subsumed by personality, the two constructs are reliably correlated [DeNeve and Cooper, 1998]. At a psychological level, several plausible mechanisms have been proposed to explain the relationship between personality and subjective well-being. For example, some researchers [Sanderson, 1999], [Carver and Scheier, 1990] have emphasized the roles of Extraversion and Neuroticism in reward and punishment systems, respectively. Others have proposed that the relationship arises from indirect, instrumental effects of personality on the experiences an individual encounters [McCrae and Costa, 1991].

Big Five personality traits were measured by asking subjects to use 1-5 point scale to answer the online version of the 44 questions Big Five questionnaire developed by John et al. [John and Srivastava, 1999] . The Big Five questionnaire owes its name to the five traits, that it takes as a constitutive of people's personality: Extraversion vs. Introversion (sociable, assertive, playful vs. aloof, reserved, shy); Emotional stability vs. Neuroticism (calm, unemotional vs. insecure, anxious); Agreeableness vs. Disagreeable (friendly, cooperative vs. antagonistic, faultfinding); Conscientiousness vs. Unconscientiousness (self-disciplined, organized vs. inefficient, careless); Openness to experience (intellectual, insightful vs. shallow, unimaginative). The scores of the five traits were computed by summing the (inverted when needed) raw scores of the items, i.e. questions, pertaining to each trait.

2.1.5 Weather data

Weather is widely believed to influence people's mood and daily happiness. For example, the majority of people think they feel happier on days with a lot of sunshine as compared to rainy days. However, the number of studies investigating the associations between daily weather and happiness or between daily weather and mood is relatively small [Keller et al., 2005].

More recently, Denissen et al. [Denissen et al., 2008] investigated the effects of six daily weather variables (temperature, wind power, sunlight, precipitation, air pressure, photoperiod) on three mood variables (positive affect, negative affect, and tiredness). Furthermore, they examined also individual differences in sensitivity to weather fluctuations including personality and demographic characteristics. Their results revealed main effects of temperature, wind power, and sunlight on negative affect. Sunlight had a main effect on tiredness and mediated the effects of precipitation and air pressure on tiredness. Finally, the authors found that individual differences in weather sensitivity could be not explained by the personality traits, the gender or the age.

More specifically on happiness, Tsutsui [Tsutsui, 2012] found that subjective happiness is related to temperature. The effects of other meteorological variables – humidity, wind speed, precipitation, and sunshine – were not significant. Moreover, happiness is more strongly affected by current temperature than by average temperature over the day. While enjoyment (a positive affect measure) is affected by weather in a similar way to happiness, sadness and depression (negative affect measures) behave somewhat differently.

In our experiments we used the following weather parameters: mean temperature, pressure, total precipitation, humidity, visibility and wind speed metrics.

2.1.6 Feature Extraction

Several cross-sectional studies have documented that happy people are usually involved in more social and active behavior. Moreover, several works highlighted an association between happiness and the actual number of friends or companions people report they can rely on [Lee and Ishii-Kuntz, 1987].

Friendship has been found to have one of the highest positive correlations with self-rated happiness [Campbell et al., 1976]. For example, the happiest college students (the top 10%) have been shown to have high quality social relationships [Diener and Seligman, 2002]. Happy people also report being more satisfied with their friends and their social activities [Lyubomirsky et al., 2006]. Not surprisingly, loneliness is negatively correlated with happiness, especially in older adults [Lee and Ishii-Kuntz, 1987]. As shown by Eagle et al. [Eagle et al., 2009a] and by Dong et al. [Dong et al., 2011a], data collected from mobile phones (call logs, sms and Bluetooth hits) can provide insights into the relational dynamics of individuals, in particular those concerning friendship. It makes, therefore, sense to address the discussed relationships between happiness and friendship by addressing social communication as mediated by, or detected through, the smartphone.

Based on these considerations, on previous works attempting to characterize social interactions by means of mobile phone data and taking inspirations from works on personality prediction using mobile phone data [Chittaranjan et al., 2013], [Montjoye et al., 2013], we derived the 25 call and sms features reported in Table 2.13 and the 9 proximity features reported in Table 2.14.

For each feature from our basic subset we calculated mean, median, min, max, quantiles with a linear step of 0.05, quantiles for the cases of 0.5, 1, 1.5 and 2 standard deviations from the mean (applying Chebyshev's inequality), variance and standard deviation functions. Moreover, to capture the possible influence of the previous happiness states on the current one we computed each basic subset of features using backward moving windows of different size, more specifically of 2, 3, 4 and 5 days.

Finally, in order to get bias corrected empirical entropy estimates we applied *Miller-Madow correction* for entropy calculation [Miller, 1955], which is explained in Equation 2.13.

$$\hat{H}_{MM}(\theta) \equiv -\sum_{i=1}^{p} \theta_{ML,i} \log \theta_{ML,i} + \frac{\hat{m}-1}{2N}, \tag{2.1}$$

where $\hat{m}$ is a number of bins with nonzero θ-probability. The likelihood function is given as the product of probability density functions $P(\theta) = f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta)$ for a random sample $X_1, \cdots, X_n$. θ_{ML} is the maximum likelihood estimate of θ, which maximizes $P(\theta)$.

2.1.7 Call and SMS Features

The features reported in Table 2.13 fall under four broad categories: (i) general phone usage, (ii) diversity, (iii) active behaviors and (iv) regularity.

Features for general phone usage (Table 2.13) capture the total number of outgoing, incoming and missed calls, and the total number of sent and received sms. Moreover, they also capture the duration of the calls the user has performed, reporting about the total duration of the calls, the total duration of outgoing calls, the total duration of incoming calls and the average of each feature. Finally, we also extracted the Outgoing to Incoming Calls Ratio, Missed to (Outgoing + Incoming) Calls Ratio, and SMS Sent/Received Ratio.

Concerning regularity features, we measured the time elapsed between calls, the time elapsed between sms exchanges and the time elapsed between call and sms. More precisely, we consider both the average and variance of the inter-event time of ones' call, sms, call+sms. A thing to note is that even though two users have the same inter-event time for both call and sms, their mean inter-event times for call+sms can be very different.

Diversity measures how evenly an individual's time is distributed among others. In our case, the diversity of user behavior is addressed by means of three kinds of features: (i) entropy of contacts, (ii) unique contacts to interactions ratio, (iii) number of unique contacts. We compute the diversity features both for calls and sms.

In particular, the entropy of an individual is the ratio between his/her total number of contacts and the relative frequency at which he/she interacts with them. The more one interacts equally often with a large number of contacts the higher the entropy will be.

2.1.8 Proximity Features

Starting from the Bluetooth hits collected inside the Friends and Family dataset, we filtered the events where $RSSI = 0$, assuming this case being the best proxy for close social proximity in space. RSSI is an 8-bit integer that denotes whether the received power level is within the Golden Receiver Power Range (GRPR). The lower and upper thresholds of GRPR are loosely bound, leaving them to be device specific. That means

Table 2.2: List of Basic Features for Happiness Recognition Problem

General Phone Usage
1. Total Number of Calls (Outgoing+Incoming) 2. Total Number of Incoming Calls 3. Total Number of Outgoing Calls 4. Total Number of Missed Calls 5. Number of SMS received 6. Number of SMS sent
Diversity
7. Number of Unique Contacts Called 8. Number of Unique Contacts who Called 9. Number of Unique Contacts Communicated with (Incoming+Outgoing) 10. Number of Unique Contacts Associated with Missed Calls 11. Entropy of Call Contacts 12. Call Contacts to Interactions Ratio 13. Number of Unique Contacts SMS received from 14. Number of Unique Contacts SMS sent to 15. Entropy of SMS Contacts 16. Sms Contacts to Interactions Ratio
Active Behaviors
17. Percent Call During the Night 18. Percent Call Initiated 19. Sms response rate 20. Sms response latency 21. Percent SMS Initiated
Regularity
22. Average Inter-event Time for Calls (time elapsed between two events) 23. Average Inter-event Time for SMS (time elapsed between two events) 24. Variance Inter-event Time for Calls (time elapsed between two events) 25. Variance Inter-event Time for SMS (time elapsed between two events)

that RSSI is merely a relative parameter. The RSSI parameter is intended to be used for power control purpose. The receiver sends increase or decrease transmission power request to the transmitting side depending on whether the perceived RSSI level. A positive or negative RSSI, measured in dB, means the received power level is above or below the GRPR, respectively. A zero implies that it is ideal.

From the filtered Bluetooth proximity data we extracted the following basic Bluetooth proximity features (Table 2.14).

In this case, the extracted features fall under three broad categories: (i) general proximity information, (ii) diversity, and (iii) regularity.

2.1.9 Methodology and Experimental Results

We formulated the automatic recognition of daily happiness as a classification problem with 3 classes (happy, neutral, not happy). The ground truth labels for classification problem were set to -1 for "not happy" class, when the ($score < 4$), 0 for "neutral" ($score = 4$) and 1 for "happy" daily mood ($score > 4$).

We separated all the data at random following uniform distribution in a training and a control test set in proportion of 80:20. To let optimization algorithms converge

Table 2.3: List of Basic Bluetooth Proximity Features for Happiness Recognition Problem

General Bluetooth Proximity
1. Number of Bluetooth IDs 2. Times most common Bluetooth ID is seen 3. Bluetooth IDs accounting for n% of IDs seen 4. Bluetooth IDs seen for more than k time slots 5. Time interval for which a Bluetooth ID is seen 6. Entropy of Bluetooth contacts
Diversity
7. Contacts to interactions ratio
Regularity
8. Average Bluetooth interactions inter-event time (time elapsed between two events) 9. Variance of the Bluetooth interactions inter-event time (time elapsed between two events)

more efficiently the feature matrix was normalized by each column to $[-1, 1]$ interval and centered to 0. Then, we trained a three families of classifiers: (i) Support Vector Machines, (ii) Neural Networks and (iii) Random Forests.

For *support vector machines model*[Vapnik, 1995] we used the following *decision function* for each class:

$$D(x) = sign(\sum_{i=1}^{l} y_i \alpha_i K(x_i, x) + b) \tag{2.2}$$

Our multi-class classification problem is solved as a one-versus-all case by a winner-takes-all strategy, in which the classifier with the highest output function is assigned to the class.

We used the kernel functions provided in Table 2.1.9.

Linear kernel did not provide accurate enough results thus only Gaussian radial basis [Buhmann and Buhmann, 2003] kernel results are reported in this chapter.

The second family of statistical models used to solve our classification problem is based on *random forest algorithm.*

Random forests is a combination of tree predictors, such that each tree is dependent on the values of a random vector sampled independently with the same distribution for all the classification trees in the forest[Breiman, 2001]. The decision boundary is formed according to the margin function. Given an ensemble of tree classifiers $h_1(\mathbf{x}), h_2(\mathbf{x}), ..., h_K(\mathbf{x})$ and if the training set is drawn at random from the empirical distribution of the random vector $Y, \mathbf{X}$ the margin function is defined as:

$$\begin{aligned} mg(\mathbf{X}, Y) = avg_k I(h_k(\mathbf{X}) = Y) - \\ \max_{j!=Y} avg_k I(h_k(\mathbf{X}) = j), \end{aligned} \tag{2.3}$$

where $I(\cdot)$ is the characteristic function. The margin function measures the distance between the average votes at $(\mathbf{X}, Y)$ for the right class and the average vote for any other class. For this model the generalization error function is:

$$PE^* = P_{\mathbf{X},Y}(mg(\mathbf{X}, Y) < 0), \tag{2.4}$$

Table 2.4: Top-30 Features for Happiness Recognition

	-1	0	1	MeanDecreaseAccuracy	MeanDecreaseGini
meanTemperature	0.0099	0.0033	0.0094	0.0082	154.8379
humidity	0.0047	-0.0033	0.0115	0.0074	149.7668
pressure	-0.0002	0.0008	0.0015	0.0011	149.0302
windSpeed	0.0028	0.0017	0.0051	0.0040	142.7727
visibility	0.0024	-0.0009	0.0051	0.0034	120.8683
neuroticism	0.0690	0.0288	0.0399	0.0419	90.5721
conscientiousness	0.0659	0.0480	0.0668	0.0627	90.2708
extraversion	0.0511	0.0357	0.0472	0.0454	76.9467
openness	0.0656	0.0340	0.0406	0.0429	73.9181
totalPrecipitation	0.0007	-0.0000	0.0012	0.0009	73.5273
agreeableness	0.0536	0.0282	0.0235	0.0289	70.7261
bluetoothQ95TimeForWhichIdSeen	0.0233	0.0120	0.0149	0.0155	22.4018
bluetoothQ90TimeForWhichIdSeen	0.0161	0.0082	0.0141	0.0131	19.8364
smsRepliedEventsLatencyMedian	0.0093	0.0085	0.0096	0.0093	18.7719
bluetoothIdsMoreThan04TimeSlotsSeen	0.0114	0.0066	0.0124	0.0110	15.6738
bluetoothMaxTimeForWhichIdSeen	0.0141	0.0072	0.0095	0.0097	15.2546
bluetoothTotalEntropyMillerMadow	0.0058	0.0017	0.0035	0.0035	13.9000
bluetoothTotalEntropyShannon	0.0065	0.0009	0.0060	0.0050	13.1826
callMeanInterEventTimePerDay	-0.0003	-0.0000	0.0014	0.0009	13.1180
incomingAndOutgoingCallsPerDay	-0.0000	-0.0002	0.0024	0.0015	12.3962
bluetoothQ50TimeForWhichIdSeen	0.0104	0.0104	0.0091	0.0095	12.2270
callStandardDeviationInterEventTimePerDay	0.0004	-0.0005	0.0007	0.0004	10.1723
bluetoothIdsMoreThan19TimeSlotsSeen	0.0087	0.0033	0.0089	0.0077	9.7388
incomingCallsPerDay	0.0003	-0.0005	0.0009	0.0005	9.6572
outgoingContactsToInteractionsRatioPerDay	0.0005	-0.0004	0.0013	0.0009	9.2016
callsInitiatedRatioPerDay	-0.0001	-0.0001	0.0014	0.0008	9.0245
entropyMillerMadowCallsOutgoingWindow3Days	0.0001	-0.0008	0.0014	0.0008	8.7199
bluetoothIdsMoreThan09TimeSlotsSeen	0.0074	0.0046	0.0040	0.0046	8.6006
bluetoothQ75TimeForWhichIdSeen	0.0027	0.0018	0.0032	0.0028	8.4454
outgoingCallsPerDay	-0.0001	-0.0001	0.0012	0.0007	8.3368

Table 2.5: SVM Kernel Functions

Name	Kernel	Parameters
Linear	$\mathbf{u}^\top \mathbf{v}$	-
Radial Basis Function	$\exp\{-\gamma\lvert\mathbf{u}-\mathbf{v}\rvert^2\}$	γ

where $P_{\mathbf{X},Y}$ is the probability over $\langle \mathbf{X}, Y \rangle$ space. For any event $A \subset \Omega$ of the feature space the characteristic function $I(\cdot)$ of A is:

$$I_A(x) = \begin{Bmatrix} 1 & \iff (x \subset A) \\ 0 & otherwise \end{Bmatrix} \begin{Bmatrix} 1 & \iff \exists x \\ 0 & otherwise \end{Bmatrix} \tag{2.5}$$

Random Forests classifiers were trained with a stepwise increase of the number of trees equal to the upper limit of 2^{11}. Optimal number of trees for model generalization as measured by mean misclassification rate for 10-fold cross validation strategy is estimated to be 100 (Fig. 3).

The third approach for a solution is based on multi-layer perceptron *neural network*. In a parametric functional form it is expressed as:

$$f_\theta(\mathbf{x}) = S(\langle \mathbf{w}, \tilde{\mathbf{x}} \rangle + b), \tag{2.6}$$

where

$$\tilde{\mathbf{x}} = S(\mathbf{W}^{hidden}\mathbf{x} + \mathbf{b}^{hidden}), \tag{2.7}$$

with parameters

$$\theta = \left\{\mathbf{W}^{hidden}, \mathbf{b}^{hidden}, \mathbf{w}, b\right\}, \tag{2.8}$$

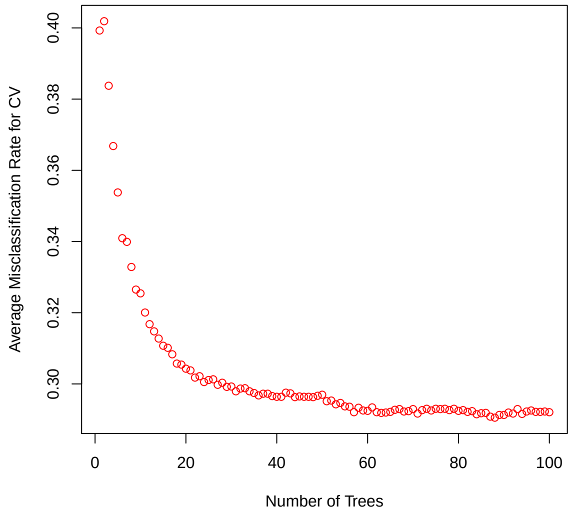

Figure 3: Misclassification Error vs Number of Tree Classifiers in Ensemble Model for Happiness Recognition Problem

and the decision function is a sigmoid function

$$S = \frac{1}{1 + e^{-x}} \tag{2.9}$$

The happiness recognition problem was solved as the following optimization problem:

$$\hat{\theta} = \arg\min_{\theta} \hat{R}_{\theta}(f_{\theta}, D_n), \tag{2.10}$$

where the sum of empirical risk and regularization term is expressed

$$\hat{R}_{\theta}(f_{\theta}, D_n) = \sum_{i=1}^{n} L(f_{\theta}(\mathbf{x}^{(i)}), t^{(i)}) + \lambda\Omega(\theta). \tag{2.11}$$

Neural network training was performed using classical backpropagation algorithm. Neural network topology was iteratively searched, based on the assumptions inferred from the data, for the best balance between empirical error minimization and maximum κ metric validated using 10-fold cross validation technique.

In order to select the final approach and the model, we compared these three families by means of a set of accuracy and κ metrics. The κ measures pairwise agreement among a set of functions which are making classification decisions with correction for an expected chance agreement [Cohen, 1960]:

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)} \tag{2.12}$$

$\kappa = 0$ if there is no agreement more than expected by chance following the empirical distribution. $\kappa = 1$ when there is a *max* agreement. κ is a state-of-the-art statistics about how significantly the classification model is different from chance. More importantly, it makes the interpretation of the scale of what the model has learned as an intuitive task. Also there is some critique in the literature of this metric, but we found it highly relevant for our classification model search and comparison task. κ statistic has the properties of more robust measure than the simple percent agreement measure taking into account the agreement occurring by chance.

During the learning and model selection process we used a random sampling with replacement to generate a new set of data for each fold from the basic training set following leave-one-out 10-fold cross validation scheme. We adopted this strategy in order to deal with potential data loss in cases when calls, sms and Bluetooth proximities existed in the real world but were not registered in the database.

Given the unbalanced classes of the ground truth daily happiness labels (that people were more happy than unhappy during the experiment) we found a solution to prevent data overfitting by the way we trained the models and searched in the possible classifier's space. Our *structural risk minimization* as opposed to empirical risk minimization solution, to prevent the data overfitting, was incorporated by working with a regularization penalty into the learning process, by balancing the model's complexity against its fitting the training data and by sampling of the model training sets in the way they mimic the empirical distributions without most probable erroneous outliers.

Accuracy and κ metrics comparison for each model type using 10-fold cross validation strategy are provided in Table 2.6 and Table 2.7. Random Forest classifier showed an average accuracy for this classification problem better than the best SVM-based model with radial basis kernel for 10.09% and better than neural network classifier for 5.56%. The conservative measure of agreement κ comparison explains the problem of unbalanced classes for an SVM classifier. SVM did good only for majority classes. As κ has a tendency to take the observed categories' frequencies as given, which causes the effect of underestimating agreement for a category that is most commonly used. Neural network classifier captured relations between the variables in a pretty stable manner among the unbalanced classes, comparing to SVM. This best tradeoff between generalization capabilities of a neural network and high overfitting for a neural network is still less efficient in comparison to the Random Forest classifier results.

We found that distribution of the estimated performance metrics does not vary substantially among each fold, leading to a good generalization despite of possible existence of heterogeneous data in each fold and the "noise" coming from the resampling procedure.

Given the reason discussed above, as we trained the models on artifitially generated data from each fold of the training sample, we have lower accuracy and κ results for the cross validation reported results (Table 2.8) than we get on the native data (training and test feature subsets).

Table 2.6: 10-fold CV Model Comparison for Happiness Recognition Problem: Accuracy

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
SVM	0.6404	0.6417	0.6434	0.6432	0.6442	0.6462
NeuralNetwork	0.6529	0.6606	0.6718	0.6708	0.6810	0.6865
RandomForest	0.6891	0.7000	0.7093	0.7081	0.7162	0.7247

The final model is based on a Random Forests algorithm and uses a 111-dimensional feature vector. We identified the variables that have more predictive power for the Random Forest model by calculating mean decrease in Gini index (Table 2.4).

This model shows 80.81% accuracy on the training set and 80.36% accuracy on

Table 2.7: 10-fold CV Model Comparison for Happiness Recognition Problem: Kappa

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
SVM	0.0012	0.0056	0.0133	0.0115	0.0162	0.0226
NeuralNetwork	0.1381	0.1708	0.2055	0.1963	0.2173	0.2570
RandomForest	0.3094	0.3358	0.3565	0.3533	0.3733	0.3950

Table 2.8: 10-fold Cross Validation Happiness Recognition Metrics for Each Fold Comparison

	Accuracy	Kappa	AccuracySD	KappaSD
1	0.6489	0.0370	0.0041	0.0154
2	0.7077	0.3424	0.0054	0.0141
3	0.7081	0.3533	0.0111	0.0275
4	0.7057	0.3525	0.0131	0.0322
5	0.7055	0.3527	0.0114	0.0274
6	0.7022	0.3458	0.0131	0.0325
7	0.7026	0.3479	0.0136	0.0339
8	0.6996	0.3432	0.0118	0.0263
9	0.6994	0.3417	0.0131	0.0313

the test set (Table 2.11). It this table the final happiness recognition model performance metrics and statistical estimates comparison are provided. In contrast, to show generalization power of the model, we provide detailed metrics for each fold during 10-fold cross validation process (Table 2.8).

The confusion matrices for this model for each training and test sets are provided in Table 2.9 and Table 2.10. These matrices show in details that there is a major agreement between classes. But Table 2.10 shows that there are still a lot of misclassified samples as “unhappy” (152 out of 2597 samples) which are, in fact, the happy state for the actual class. 243 samples are classified as “neutral” being actually “happy”. The best result we have for the majority “happy” class (1616 correctly classified). These tables prove us good generalization capabilities of the model – for each the training and the test sets we have similar results in detail.

To show our result visually in a simple form, that actually the final proposed model delivers better results than “by chance”, we report the multi-class area under the ROC curve [Wandishin and Mullen, 2009], [Hand and Till, 2001]. The final model ROC curve plotted for the test subset is provided in Fig. 4. The multi-class area under the ROC curve (AUC) = 0.844 means a good reduction of missclassification error for each class combination [Krzanowski, 2005]. The area between the ROC curve and the diagonal is, in fact, what our model does better then “by change”. The area between the ROC curve and the perpendicular lines going to the upper left corner $[1.0, 1.0]$ is the measure of what is not explained by our model, what causes the misclassification error.

Table 2.9: Final Classifier Happiness Recognition Model Confusion Matrix for Training Set

	-1	0	1
-1	782	119	75
0	153	1170	145
1	600	903	6448

Table 2.10: Final Classifier Happiness Recognition Model Confusion Matrix for Test Set

	-1	0	1
-1	197	30	14
0	34	274	37
1	152	243	1616

Table 2.11: Final Classifier Happiness Recognition Model Performance Metrics Comparison

	Training set	Test set
Accuracy	0.8081	0.8036
Kappa	0.5879	0.5743
AccuracyLower	0.8004	0.7878
AccuracyUpper	0.8156	0.8187
AccuracyNull	0.6415	0.6419
AccuracyPValue	2.139e-303	8.826e-73
McnemarPValue	5.647e-208	1.738e-57

2.1.10 Discussion

The present study provides the first evidence that daily happiness of individuals can be automatically predicted using a set of indicators obtained from mobile phone data (call logs, sms logs and Bluetooth interactions) and additional indicators related to weather factors and individuals' personality traits. Our final model based on a Random Forest classifier using 111-dimensional feature vector obtains an accuracy measure of 80.81% for a 3-class classification problem.

An investigation of the most important features to predict daily happiness revealed interesting associations. Indicators linked to the weather factors were really useful to predict daily happiness. Our results confirm some of the results obtained by a previous study done by [Tsutsui, 2012]. In our study, we found a confirmatory association between subjective happiness and temperature. Additionally, we found significant effects of other meteorological variables – humidity, wind speed, pressure, total precipitation and visibility – for predicting daily happiness.

We also found interesting associations between personality traits and daily happiness. All the traits contribute significantly the prediction of daily happiness response variable. These results are interesting because the previous studies in social psychology focus their analyses mainly on the associations between happiness or subjective well-being and Extraversion or Neuroticism. Instead, our work shows the important contribution played also by Conscientiousness, Openness and Agreeableness for the automatic classification of daily happiness. Moreover, these results open us the possibility of creating a multi-step stochastic model in which first we estimate the personality and then we use the estimated variables as independent variables for a happiness recognition problem.

Regarding the role played by mobile phone data in predicting daily happiness, it's interesting to note the significant contribution of the proximity features. Among the top 30 features used for happiness recognition, 10 features are proximity ones calculated from the Bluetooth data. In particular, an interesting predictive role is played by the time intervals for which an id is seen.

Social proximity, measured by Bluetooth interactions, has strong predictive power for daily happiness recognition based on median, 0.75, 0.9 and 0.95 quantiles from the distribution of the time intervals, measured in seconds, for which a Bluletooth identifier is seen.

Moreover, features capturing the diversity in co-location interactions are in the top 30 list (e.g. entropy of proximity contacts). This result seem to confirm previous studies in social psychology that found associations between people's happiness and the richness in terms of the amount and the diversity of people's social interactions.

For call interactions, we can infer the role played by general phone usage features such as the number of incoming calls and the number of outgoing calls. This result is consistent with previous studies on happiness showing that happy people are usually more social and have more interaction and exchange. In addition, a role of the regularity in call patterns (average and standard deviation values of the time elapsed between two calls) and active behavior of an individual (the ratio of calls initiated by the individual) also play predictive role. The latter indicator could be easily explained making reference to the standard definition of happy people as more active people given by social psychologists.

The role of sms interactions is less evident from our investigation. The only feature related to sms interactions among the top 30 predictive features is a feature related to an active behavior of the individual and more specifically is the latency in replying to a text message that we define as the median value that people take to answer a text. The predictive power of the sms data needs further investigations.

Among the limitations of the present study we list the following: our sample comes from a population living in the same environment (our subjects were all married graduate students living in a campus facility of a major US university); the non-availability of proximity data concerning the interaction with people not participating in the data collection, a fact that is common to many other studies of this type. The first two problems are at least partially attenuated by the large variability of the sample in terms of provenance and cultural background, which can be expected to correspond to a wide palette of interaction behaviors that efficaciously counterbalance the effects of living-place homogeneity. Moreover, the Funf framework used for collecting this data did not account for the data loss in cases when calls, sms and Bluetooth proximities existed in the real world but were not registered in the database as not available. The source data loss, when the battery is out, was also not registered and was unpredictable during the data collection. In order to solve these potential problems, we proposed and used a random sampling with replacement for each fold of our learning task and we used it to generate better datasets from the separate training set.

2.1.11 Conclusion

The goal of this chapter is to investigate the feasibility of automatic recognition of people's daily happiness from mobile phone data. To achieve this goal, we formulated the automatic recognition of daily happiness as a 3-class classification problem based on the information concerning: a) peoples activity, as detected through their smartphones; b) the weather conditions; c) personality traits. The first information type is represented by features extracted from call and sms logs and from Bluetooth hits, able to capture (i) amount of call, sms and proximity; (ii) diversity of call, sms, and proximity; (iii) user active behaviors; and (iv) regularity in user behaviors.

Despite the limitations of this study discussed above, we believe that our results (80.81% for a 3-class classification problem) have provided compelling evidence that individual daily happiness can be reliably predicted from smartphone usage data and

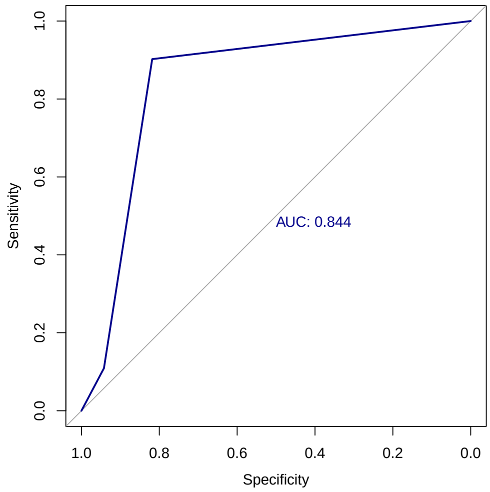

Figure 4: Final Model ROC Curve for Happiness Recognition Problem

from additional indicators related to the weather factors and individual dispositions (personality traits). Hence, on a practical side our results are a first important step towards automatic systems able to predict people's daily happiness and towards engineering a construct that is not only a goal of people's life but also a means for reaching other goals and for facilitating the desirable behaviors and outcomes.

2.2 Daily Stress Recognition from Mobile Phone Data

2.2.1 Introduction

Nowadays, the number of mobile phones in use worldwide is about 5 billion, with millions of new subscribers everyday [1]. Mobile phones allow for unobtrusive and cost-efficient access to huge streams of previously inaccessible data related to daily social behavior [Lane et al., 2010b]. These devices are able to sense a wealth of behavioral data such as (i) location, (ii) other devices in physical proximity through Bluetooth scanning, (iii) communication data, including both metadata (logs of who, when, and duration) of phone calls and text messages (sms), etc. Correspondingly, the availability is continuously growing of huge streams of *personal data* related to activities, routines and social interactions [Lane et al., 2010b, Dong et al., 2011a] which represent a novel opportunity to address fundamental problems of our societies in different fields, such as mobility and urban planning [Gonzalez et al., 2008], finance [Singh et al., 2013], healthy living and subjective well-being [Madan et al., 2012, LiKamWa et al., 2013].

In this work, we focus on one of the most widespread problem of subjective well-being and society – the stress. Stress is a well-known condition in modern life and research has shown that the amount of cumulative stress plays a role in a broad range of physical, psychological and behavioural conditions, such as anxiety, low self-esteem, depression, social isolation, cognitive impairments, sleep and immunological disorders, neurodegenerative diseases and other medical conditions [Cohen et al., 1997], while also significantly contributing to healthcare costs. Hence, measuring stress in daily life situations has become an important challenge [Plarre et al., 2011]. Today, the availability of huge and diverse streams of pervasive data produced by and about people allows for automatic, unobtrusive, and fast recognition of daily stress levels. An early prediction of stress symptoms can indeed help to prevent situations that are risky for human life[Healey and Picard, 2005].

Several studies have produced interesting results that support the feasibility of detecting stress levels through physiological sensors (see [Healey and Picard, 2005], [Jovanov et al., 2003]). However, the use of physiological sensors is limited by several shortcomings. Stress detection systems based on physiological measurement such as heart-rate variability or skin conductance are intrusive and need to be easily wearable to be exploited in natural settings; the data they produce can be confounded by daily life activities such as speaking or drinking; they exhibit important between-person differences [Plarre et al., 2011].

Recently, social psychologist Miller wrote "The Smartphone Psychology Manifesto" in which he argued that the smartphones should be seriously considered as new research tools for psychology. In his opinion, these tools could revolutionize all fields of psychology and other behavioral sciences making these disciplines more powerful, sophisticated, and grounded in real-world behavior [Miller, 2012] and [Lathia et al., 2013]. Indeed, several works have started to use smartphone activity data in order to detect and predict personality traits [de Oliveira et al., 2011, Staiano et al., 2012b, Chittaranjan et al., 2013, Montjoye et al., 2013], mood states [LiKamWa et al., 2013], and daily happiness [Muaremi et al., 2012]. Stopczynski et al. [Stopczynski et al., 2014] described the Copenhagen Networks Study, a large-scale

[1] http://www.ericsson.com/ericsson-mobility-report

study designed to measure human interactions spanning multiple years.

Smartphones data can be used to detect stress levels as well. Indeed, stress levels are associated with the type of activities people engage in, including those executed at/through their smartphone (for instance, a high number of phone calls and/or e-mails from many different people could be associated with higher stress levels). Weather conditions – an environmental transitory property – in turn, have been argued [Howarth and Hoffman, 1984], [Sanders and Brizzolara, 1982] to be often associated with stress, acting either directly (as stressors) or indirectly (by affecting individual sensitivity to stressors). Finally, the impact of all these transitory factors – (smartphone) activities and weather conditions – on stress induction can be expected to be modulated by personal characteristics and differences [Suls et al., 1998], [Vollrath and Torgersen, 2000]. For example, a neurotic person could react with higher levels of stress to a high number of interactions (call, sms or proximity interactions) than an emotionally stable person; an extrovert or agreeable person, in turn, might well find him/herself at ease with a high number of interactions.

In this chapter, we approach the automatic recognition of daily stress as a 2-class classification problem (non-stressed vs stressed) based on information concerning different types of data: a) people activities, as detected through their smartphones; b) weather conditions; c) personality traits. The information about people activities is represented by features extracted from call and sms logs and from Bluetooth hits, are able to capture (i) the amount of calls, of sms and of proximity interactions; (ii) the diversity of calls, of sms, and of proximity interactions; and (iii) regularity in user behaviors. In addition, we use weather conditions (environmental and transitory factors) along with personality traits (internal and stable factors); the latter are mediating factors that can modulate people responses to stressors (e.g., weather, daily activity). This multifactorial approach will be compared to approaches based only on a family of features (personality, weather conditions, mobile phone features) or simpler combinations of families of features (personality and weather conditions; personality and mobile phone features; weather conditions and mobile phone features).

Classification experiments are performed using a variety of approaches and the best solution for our classification problem was found using an ensemble of tree classifiers based on a Random Forest algorithm. Our multifactorial approach obtains an accuracy score of 72.28% for a 2-class daily stress recognition problem, providing evidence that individual daily stress can be reliably predicted from the combination of smartphone usage data, weather conditions and individual dispositions (personality traits). Interestingly, if one of these information sources is dropped, the recognition performances decrease drastically.

The main contributions of this chapter are as follows:

1. We propose a multi-factorial data-driven approach to the prediction of individual daily stress;

2. We validate our approach with a seven-months dataset collected from 111 subjects;

3. We provide a comprehensive analysis of the predictive power of the proposed approach and a comparison with approaches based only on single families of features (personality, weather conditions, mobile phone features) or pairwise combinations thereof (personality and weather conditions, personality and mobile phone features, weather conditions and mobile phone features).

2.2.2 Related Work

A large body of research on stress detection focused on physiological measurements to infer stress levels (see [Healey and Picard, 2005], [Majoe et al., 2010], [Plarre et al., 2011]). Heart-rate variability, galvanic skin response, respiration, muscle activity and temperature are among the most relevant features. However, despite providing reliable insights on stress levels, this approach has major limitations because it comprises wearable sensors that need to be carried at all times to allow for continuous monitoring.

Among the different changes in physiological parameters that happen during stressful situations, variation in speech production has inspired a number of studies using acoustic sensing on smartphones. Research on stress detection based on voice analysis considered different speech characteristics such as pitch, glottal pulse, spectral slope and phonetic variations. For example, Lu and colleagues [Lu et al., 2012] proposed StressSense, an Android application for stress detection from human voice in real-life conversation, and they achieved 81% and 76% accuracy for indoor and outdoor environments.

However, these methods depend on sound quality, which is not granted in natural settings (e.g., crowded public places, noisy outdoor), and the correlation between speech and emotion is subjected to large individual differences [Scherer et al., 2002]. Hence, our performance of 72.28% is a good and reliable alternative to stress detection. Other studies focused on the video analysis of behavioural correlates of psychological stress [Giakoumis et al., 2012]. These systems, despite providing an unobtrusive method for stress monitoring, cannot be employed in a large variety of real world and mobile environments and pose privacy concerns related to the recording of people's behaviour.

A promising approach that can overcome the major shortcomings of stress detection based on physiological measures and on audio/video analysis is activity recognition from smartphone usage patterns. Studies in this field have been mainly focused on the understanding of relational dynamics of individuals [Eagle et al., 2009b]. Recently studies have started to investigate how smartphone usage habits can provide insights into users' affective state [LiKamWa et al., 2013] and stress levels [Bauer and Lukowicz, 2012]. LiKamWa and colleagues [LiKamWa et al., 2013] proposed MoodScope, a mobile software system that recognizes the users' mood, but not stress states, from smartphone usage analysis. They collected usage data and self-reported mood in a two months longitudinal study and used them to train mood models. Smartphone usage data consisted in phone calls, SMSes, e-mail messages, application use, web browsing histories and location changes, while self-reported mood was collected from users' input at least four times a day. MoodScope reached a 66% accuracy of participants' daily-average mood, with phone calls and categorized applications as the most useful features for mood discrimination.

Bauer and Lukowicz [Bauer and Lukowicz, 2012] focused on mid-term stress detection, monitoring 7 students during a two week exam session followed by two weeks of non-stressful period. The recorded data were related to participants' mobility patterns and social interactions, and included users' location, Bluetooth proximity, phone calls and SMSes. These features allowed to detect an average behaviour modification of 53% for each user during the exam session. A limitation of this study is the small number of subjects. Our multifactorial approach outperforms the approach proposed by [Bauer and Lukowicz, 2012] although a direct comparison may be not adequate given the different focus: our approach tends to classify people daily as "not stressed" or "stressed", while Bauer and Lukowicz try to detect stressful situations.

In 2013, Sano and Picard [Sano and Picard, 2013] reported an accuracy performance in stress recognition of 75% using a combination of features obtained from mobile phones and wearable sensors. However, the limited number of subjects used in their experiments (18) and the limited number of days (5) make preliminary the results of this study.

2.2.3 Data Collection

From November 12, 2010 to May 21, 2011, we collected a dataset capturing the lives of 117 subjects living in a married graduate student residency of a major US university. Our sample of subjects has a large variety in terms of provenance and cultural background: we have subjects from 16 countries such as USA, China, Israel, India, Iran, Russia, etc. During this period, each participant was equipped with an Android-based cellular phone incorporating a sensing software explicitly designed for collecting mobile data. Such software runs in a passive manner and does not interfere with the every day usage of the phone. The data collected consisted of: (a) call logs, (b) sms logs, (c) proximity data, obtained by scanning nearby phones and other Bluetooth devices every five minutes, and (d) data from surveys administered to participants, which provided self-reported information about personality traits ("Big Five") and self reported information about daily stress.

Proximity interaction data were derived from Bluetooth hits in a similar way as in previous reality mining studies [Eagle and Pentland, 2006]. Bluetooth scans were performed every 5 minutes in order to keep the battery from draining while achieving a high enough temporal resolution. The Bluetooth log of a given smartphone was then used to extract the list of the other participants' phones which were in proximity.

In total, the dataset consisted of 33497 phone calls, 22587 SMS, and 1460939 Bluetooth hits.

2.2.4 Stress data

At the evening, the participants were also asked to fill daily surveys about their daily self-perceived stress level. The stress information was reported by the participants filling a seven items scale with 1 = "not stressed", 4 = "neutral" and 7 = "extremely stressed". In our experiments we used the data only for the subjects (111 subjects) who had provided at least 2 weeks of consecutive data.

The distribution of daily stress is visualized in Fig. 5. We see that it has a small negative skew – the density is moved to the higher region of stress score. The distribution has negative excess kurtosis, which in our case means that the sample reported a specific daily stress score more often than the neutral. Fig. 6 shows that within-person daily stress variance is more spread than between-person, but the density of between-person variance is higher.

2.2.5 Personality Data

Several studies in social psychology investigated the relationships between personality traits and psychological stress. Personalities that tend to be more negative are usually associated with greater distress, while outgoing and positive personalities generally experience less distress [Suls et al., 1998], [Vollrath and Torgersen, 2000]. The majority of the studies that have examined the relationship between personality

Table 2.12: Selected Features Ranked by Mean Decrease in Accuracy for Daily Stress Recognition Problem

Rank	Feature	0	1	Mean Decrease in Accuracy	Mean Decrease in Gini Index
1	personality.Conscientiousness	13.65	18.04	23.35	159.96
2	personality.Agreeableness	14.22	19.73	22.92	167.30
3	personality.Neuroticism	15.96	21.04	22.56	183.87
4	personality.Openness	14.20	14.18	21.38	139.23
5	personality.Extraversion	15.75	15.02	21.07	158.51
6	weather.MeanTemperature	14.50	6.34	17.44	322.27
7	sms.RepliedEvents.Latency.Median	8.83	13.85	15.63	48.74
8	weather.Humidity	15.33	2.10	15.45	298.13
9	sms.AllEventsPerDay	8.61	0.56	10.50	42.91
10	bluetooth.Q95TimeForWhichIdSeen	4.99	6.05	9.94	32.47
11	bluetooth.MaxTimeForWhichIdSeen	6.24	7.23	9.47	32.12
12	sms.IncomingAndOutgoingPerDay	7.45	1.26	9.38	41.59
13	weather.Visibility	9.94	1.26	9.22	251.27
14	weather.WindSpeed	8.77	1.30	8.67	282.10
15	bluetooth.Q90TimeForWhichIdSeen	4.24	6.75	8.64	28.41
16	bluetooth.TotalEntropyShannon	5.04	3.51	8.56	31.37
17	call.EntropyMillerMadowOutgoingTotal	4.25	4.10	8.54	27.49
18	call.EntropyShannonOutgoingAndIncomingTotal	4.23	4.86	8.53	26.28
19	bluetooth.TotalEntropyMillerMadow	5.06	4.22	8.50	32.09
20	bluetooth.IdsMoreThan09TimeSlotsSeen	6.11	5.85	8.43	27.88
21	bluetooth.IdsMoreThan04TimeSlotsSeen	6.34	4.59	8.04	24.64
22	call.EntropyShannonMissedOutgoingTotal	3.13	4.92	7.85	24.34
23	bluetooth.IdsMoreThan19TimeSlotsSeen	2.97	5.16	7.78	20.87
24	call.EntropyShannonOutgoingTotal	3.10	6.45	7.78	24.79
25	bluetooth.Q75TimeForWhichIdSeen	5.16	4.70	7.76	22.07
26	call.EntropyMillerMadowMissedOutgoingTotal	4.09	5.45	7.55	24.64
27	call.EntropyMillerMadowOutgoingAndIncomingTotal	3.87	6.29	7.51	28.63
28	sms.OutgoingAndIncomingTotalEntropyMillerMadow	4.68	3.84	7.19	17.63
29	sms.OutgoingTotalEntropyMillerMadow	5.22	1.49	7.19	18.88
30	bluetooth.Q50TimeForWhichIdSeen	1.53	7.29	7.08	18.91
31	bluetooth.Q68TimeForWhichIdSeen	2.36	5.96	6.68	19.05
32	sms.OutgoingTotalEntropyShannon	2.53	2.77	5.13	17.59

and distress focused on the Big Five traits [John and Srivastava, 1999], a personality model owing its name to the five traits it takes as a constitutive of people's personality: Extraversion, Neuroticism, Agreeableness, Conscientiousness, and Openness to experience. Researchers showed significant associations between psychological stress, on the one hand, and Neuroticism, Extraversion and Conscientiousness, on the other. Duggan et al. [Duggan et al., 1995] found that individuals high in Neuroticism may be more vulnerable to experiencing distress as they respond more negatively to daily stressors and report more daily stressful events and higher levels of daily stress. When people with high scores in Neuroticism encounter stressful events, they tend to experience them as more aversive than those low in this trait [Bolger and Schilling, 1991], [Gunthert, 1999]. Finally, in a study with university students, Volrath and Torgensen [Vollrath and Torgersen, 2000] showed that students with more adaptive personalities such as Extraversion and Conscientiousness are more positive and sociable and hence less affected by daily stress.

In our study, Big Five personality traits were measured by asking subjects to answer the online version of the 44 questions Big Five questionnaire developed by John et al. [John and Srivastava, 1999], by means of 5-point likert scales. The scores on the five traits were the average over the raw scores (inverted when needed) of the items pertaining to each trait.

2.2.6 Weather data

The question about the relationship between mood, health and weather has been extensively debated [Hardt and Gerbershagen, 1999],[Sanders and Brizzolara, 1982]. Studies in environmental psychology investigated the role of weather as a stressor and showed significant effects of temperature, hours of sunshine and humidity on

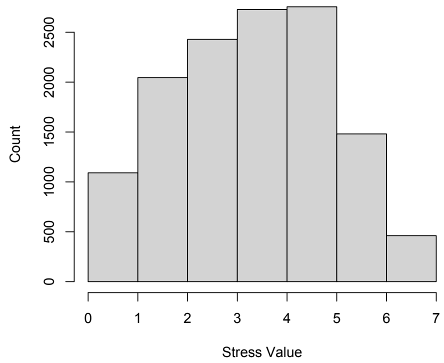

Figure 5: Recorded Daily Stress Scores Histogram

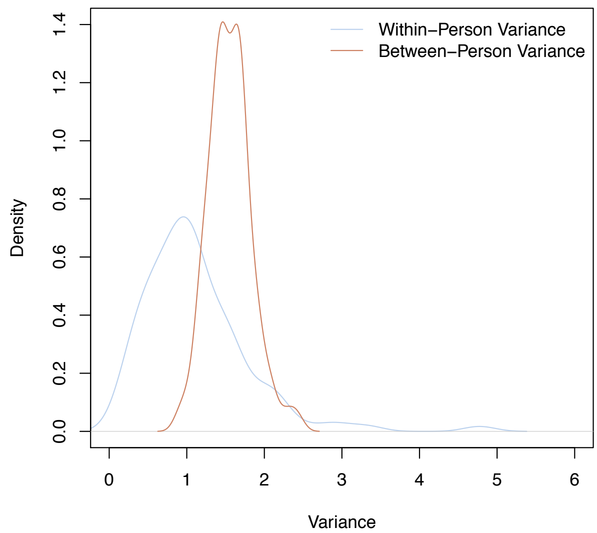

Figure 6: Within- and Between-Subject Variance for Daily Stress Recognition Problem

mood [Howarth and Hoffman, 1984], [Sanders and Brizzolara, 1982]. A large-scale study by Faust and colleagues [Faust et al., 1974] on 16,000 students in Switzerland showed an association between weather and sleep disorders, depressed mood and irritability. More recently, Denissen et al. [Denissen et al., 2008] investigated the effects of six daily weather variables (temperature, wind power, sunlight, precipitation, air pressure, photoperiod) on three mood variables (positive affect, negative affect, and tiredness). Their results revealed main effects of temperature, wind power, and sunlight on negative affect, while sunlight had also a main effect on tiredness and mediated the effects of precipitation and air pressure on tiredness.

In our experiments, we used the following weather variables: (i) mean temperature, (ii) pressure, (iii) total precipitation, (iv) humidity, (v) visibility and (vi) wind speed (measured in m/s) metrics. The source weather data were collected from Wolfram Alpha [2]. All weather metrics are computed on a daily scale for the same day that is under investigation and the source data are extracted from the Boston area weather stations (e.g. KBOS) located on the same relative elevation as the campus where the data collection was performed.

[2]http://www.wolframalpha.com

2.2.7 Feature Extraction

Based on previous works that characterize social interactions by means of mobile phone data and use social interactions data to predict people's behaviors, states [Bogomolov et al., 2013a, LiKamWa et al., 2013], and traits [de Oliveira et al., 2011, Chittaranjan et al., 2013, Montjoye et al., 2013], we derived the 25 call and sms basic features reported in Table 2.13 and the 9 proximity basic features reported in Table 2.14.

For each basic feature, we calculated second order features, such as mean, median, min, max, 99%, 95% quantiles, quantiles corresponding to 0.5, 1, 1.5 and 2 standard deviations (applying Chebyshev's inequality), variance and standard deviation functions. Moreover, for each basic feature we calculated the same functions as above for 2 and 3 days backward-moving window to account for the possibility that past events influenced the current stress state.

In the following subsections we will describe more in detail the 25 call and sms basic features and the 9 proximity basic features.

2.2.8 Call and Sms Features

The features reported in Table 2.13 fall under four broad categories: (i) general phone usage, (ii) active behaviors, (iii) regularity, and (iv) diversity.

Table 2.13: List of Basic Features for Daily Stress Recognition Problem

General Phone Usage
1. Total Number of Calls (Outgoing+Incoming)
2. Total Number of Incoming Calls
3. Total Number of Outgoing Calls
4. Total Number of Missed Calls
5. Number of SMS received
6. Number of SMS sent
Diversity
7. Number of Unique Contacts Called
8. Number of Unique Contacts who Called
9. Number of Unique Contacts Communicated with (Incoming+Outgoing)
10. Number of Unique Contacts Associated with Missed Calls
11. Entropy of Call Contacts
12. Call Contacts to Interactions Ratio
13. Number of Unique Contacts SMS received from
14. Number of Unique Contacts SMS sent to
15. Entropy of SMS Contacts
16. Sms Contacts to Interactions Ratio
Active Behaviors
17. Percent Call During the Night
18. Percent Call Initiated
19. Sms response rate
20. Sms response latency
21. Percent SMS Initiated
Regularity
22. Average Inter-event Time for Calls (time elapsed between two events)
23. Average Inter-event Time for SMS (time elapsed between two events)
24. Variance Inter-event Time for Calls (time elapsed between two events)
25. Variance Inter-event Time for SMS (time elapsed between two events)

Features for *general phone usage* consist of: the total number of outgoing, incoming and missed calls and the total number of sent/received sms. Moreover, we also computed the following ratios: outgoing to incoming calls, missed to (outgoing + incoming) calls, sent to received sms.

Then, we captured the *active behaviors* of an individual computing the following features: (i) percentage of calls done during the night, (ii) percentage of initiated calls during the night, (iii) the sms response rate, (iv) the sms response latency, and (v) the percentage of initiated sms. In particular, we consider a text from a user (A) to be a response to a text received from another user (B) if it is sent within an hour after user A received the last text from user B. The response rate is the percentage of texts people respond to. The latency is the median time it takes people to answer a text. Note that by definition, latency will be less or equal to one hour.

Diversity and *regularity* have been shown to be important for the characterization of different facets of human behavior. In particular, entropy, used as a measure of diversity, has been successfully applied to predict mobility [Song et al., 2010], spending patterns [Krumme et al., 2013, Singh et al., 2013], online behavior [Sinatra and Szell, 2014] and personality traits [Montjoye et al., 2013]. Concerning *regularity* features, we measured the time elapsed between calls, the time elapsed between sms exchanges and the time elapsed between call and sms. More precisely, we consider both the average and the variance of the inter-event time of one's call, sms and sum thereof (call+sms). Noticeably, in fact, even when two users have the same inter-event time for both call and sms, that quantity can be different for their sum.

Diversity measures how evenly an individual's time is distributed among others. In our case, the diversity of user behavior is addressed by means of three kinds of features: (i) entropy of contacts, (ii) unique contacts to interactions ratio, (iii) number of unique contacts, all computed both on calls and on sms. In particular, the entropy of an individual is the ratio between his/her total number of contacts and the relative frequency at which he/she interacts with them. The more one interacts equally often with a large number of contacts, the higher the entropy will be. For entropy calculation, we applied *Miller-Madow correction*[Miller, 1955], which is explained in Equation 2.13.

$$\hat{H}_{MM}(\theta) \equiv -\sum_{i=1}^{p} \theta_{ML,i} \log \theta_{ML,i} + \frac{\hat{m}-1}{2N}, \tag{2.13}$$

where $\hat{m}$ is a number of bins with nonzero θ-probability. The likelihood function is given as the product of probability density functions $P(\theta) = f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta)$ for a random sample $X_1,\cdots,X_n$. θ_{ML} is the maximum likelihood estimate of θ, which maximizes $P(\theta)$. Miller-Madow correction was applied, dealing with the data quality problems, to get bias-corrected empirical entropy estimate.

2.2.9 Proximity Features

Starting from the Bluetooth hits collected, we filtered out all the cases with $RSSI < 0$. From the filtered Bluetooth proximity data we extracted the following basic Bluetooth proximity features (Table 2.14). In this case, the extracted features fall under three broad categories: (i) general proximity information, (ii) diversity, and (iii) regularity. As for call and sms, we applied Miller-Madow correction for entropy calculation.

2.2.10 Methodology

We formulated the automatic recognition of daily stress as a binary classification problem ("not stressed" vs "stressed"), with labels 0 for "not stressed" and label 1 for

Table 2.14: List of Basic Bluetooth Proximity Features for Stress Recognition Problem

General Bluetooth Proximity
1. Number of Bluetooth IDs
2. Times most common Bluetooth ID is seen
3. Bluetooth IDs accounting for n% of IDs seen
4. Bluetooth IDs seen for more than k time slots
5. Time interval for which a Bluetooth ID is seen
6. Entropy of Bluetooth contacts
Diversity
7. Contacts to interactions ratio
Regularity
8. Average Bluetooth interactions inter-event time (time elapsed between two events)
9. Variance of the Bluetooth interactions inter-event time (time elapsed between two events)

"stressed". The two classes included all the cases with scores <= 4 and scores > 4, respectively. The sizes of the resulting two classes are 36.16% for "stressed" and 63.84% for "not stressed". The inclusion of the cases with stress=4 in the 0 class meant to provide a more clearcut distinction between the "stressed" and the "non-stressed" cases.

The data set was then randomly split into a training (80% of data) and a testing (20% of data) dataset, carefully avoiding that data for the same subjects appeared in both the training- and in the test-set. In order to accelerate the convergence of the models, we *normalized* each dimension of the feature vector [Box and Cox, 1964].

Additionaly, we also used a leave-one-subject-out cross-validation strategy. Hence, 111 models for each personality trait were trained on 110-subject subsets, evaluating them against the remaining ones and finally averaging the results. The results obtained are not significantly different from the ones obtained using the random split 80% vs 20%. In the rest of this chapter, we will discuss only the results obtained with the random split 80% vs 20%.

2.2.11 Feature Selection

As an initial step, we carried out a *Pearson correlation analysis* to visualize and better understand the relations between variables in the feature space. We found quite a large subset of features with strong mutual correlations and another subset of uncorrelated features. Hence, there was room for feature space reduction. We excluded using *principal component analysis* (PCA) because the transformation it is based on produces new variables that are difficult to interpret in terms of the original ones making the interpretation of the results more complex.

Therefore, we turned to a pipelined *variable selection* approach, based on *feature ranking* and *feature subset selection*, which was perfomed using only data from the training set. The metric used for feature ranking was the mean decrease in the *Gini coefficient of inequality*. This choice was motivated because it outperformed other metrics, such as mutual information, information gain and chi-square statistic with an average improvement of approximately 28.5%, 19% and 9.2% respectively [Singh et al., 2010]. The Gini coefficient ranges between 0, expressing perfect equality in predictive power and 1, expressing maximal inequality in predictive power. The feature with maximum mean decrease in Gini coefficient is expected to have the maximum influence in minimizing the out-of-the-bag error. It is known in the literature that minimizing the out-of-the-bag error results in maximizing common performance metrics used to evaluate models (*e.g.* accuracy, F1, AUC, etc.) [Tuv et al., 2009].

The feature selection procedure produced a reduced subset of 32 features from an initial pool of about 500 features. Hence, we obtained a low-dimensional feature space that makes our approach efficient to implement into mobile and multimedia applications.

2.2.12 Model Building

We formulated the automatic recognition of daily stress as a classification problem with two classes ("not stressed" or "stressed"). The ground truth labels for classification problem were set to 0 for "not stressed", where label score <= 4 and 1 for "stressed", where label score > 4.

We separated all the data at random, following an uniform distribution, in a training set and in a control test set fixing the proportion of 80:20. To let optimization algorithms converge more efficiently the feature matrix was centered and normalized by each column [Box and Cox, 1964].

Applying a grid search approach we trained a number of sets of classifiers: support vector machines, neural networks, ensemble of tree classifiers based on a Breiman's Random Forest (RF) and Friedmans Generalized Boosted Model (GBM) [Freund and Schapire, 1997] algorithms with different parameters.

Multiple regression models, *support vector machines model*[Vapnik, 1995] with linear and Gaussian radial basis [Buhmann and Buhmann, 2003] kernels and multi-layer perceptron *neural network* did not provide good classification results or required building separate models for each personality type and weather conditions.

Random forest algorithm produces a combination of tree predictors, such that each tree is dependent on the values of a random vector sampled independently with the same distribution for all the classification trees in the forest[Breiman, 2001]. The decision boundary is formed according to the margin function. Given an ensemble of tree classifiers $h_1(\mathbf{x}), h_2(\mathbf{x}), ..., h_K(\mathbf{x})$ and if the training set is drawn at random from the empirical distribution of the random vector $Y, \mathbf{X}$ the margin function is defined as:

$$\begin{aligned} mg(\mathbf{X}, Y) = avg_k I(h_k(\mathbf{X}) = Y) - \\ \max_{j!=Y} avg_k I(h_k(\mathbf{X}) = j), \end{aligned} \tag{2.14}$$

where $I(\cdot)$ is the characteristic function. The margin function measures the distance between the average votes at $(\mathbf{X}, Y)$ for the right class and the average vote for any other class. For this model the generalization error function is:

$$PE^* = P_{\mathbf{X},Y}(mg(\mathbf{X}, Y) < 0), \tag{2.15}$$

where $P_{\mathbf{X},Y}$ is the probability over $\langle \mathbf{X}, Y \rangle$ space. For any event $A \subset \Omega$ of the feature space the characteristic function $I(\cdot)$ of A is:

$$I_A(x) = \begin{Bmatrix} 1 & \iff (x \subset A) \\ 0 & otherwise \end{Bmatrix} \begin{Bmatrix} 1 & \iff \exists x \\ 0 & otherwise \end{Bmatrix} \tag{2.16}$$

For the second ensemble model (GBM-based) we adopted a greedy function approximation and the stochastic gradient boosting strategy, which are described in [Friedman, 2001] and [Friedman, 2002]. The optimization problem was formulated as finding a function, $\hat{f}(\mathbf{x})$, that minimizes the loss function $\Psi(y, f)$:

$$\hat{f}(\mathbf{x}) = \arg\min_{f(\mathbf{x})} {}_{y|\mathbf{x}} [\Psi(y, f(\mathbf{x}))|\mathbf{x}] \tag{2.17}$$

The algorithmic implementation for this chapter solution is described in Algorithm 1.

In order to find the best model, we trained a number of models and selected the best one based on κ metrics for the 10-fold validation strategy. The Cohen's κ measures pairwise agreement among a set of functions which are making classification decisions with correction for an expected chance agreement [Cohen, 1960]. $\kappa = 0$ if there is no agreement more than expected by chance following the empirical distribution. $\kappa = 1$ when there is a *max* agreement. κ is a state-of-the-art statistics about how significantly the classification model is different from chance. More importantly, it makes the interpretation of the scale of what the model has learned to be an intuitive task. κ statistic has the properties of more robust and conservative measure to show what we have learned from the data than F1 and area under the ROC curve metrics.

During the learning and model selection process we followed a leave-one-out 10-fold cross validation strategy. We adopted this strategy in order to prevent data overfitting and to deal with potential data loss in cases when calls, sms and Bluetooth proximities existed in the real world but were not registered by the mobile application.

The performance metrics used to evaluate the models are: accuracy, sensitivity, specificity and Cohen's κ. For detailed analysis we provide model confusion matrix [Stehman, 1997]. The significance of the results is supported by the proper statistical tests [Garcia and Herrera, 2008] and is shown in the tables.

We trained a variety of classifiers: (i) an ensemble of tree classifiers based on a Random Forest algorithm [Breiman, 2001], (ii) a Generalized Boosted Model (GBM) [Freund and Schapire, 1997], (iii) Support Vector Machines with linear and Gaussian radial basis kernels, and (iv) Neural Networks. The best solution of the classification problem was found using an ensemble of tree classifiers based on *Random Forest* algorithm. In the rest of this chapter, we report the performance results only for Random Forest.

Random forest algorithm produces a combination of simple decision tree predictors, such that each tree is dependent on the values of a random vector sampled independently with the same distribution for all the classification trees in the forest [Breiman, 2001]. The decision boundary is formed according to the margin function. Given an ensemble of tree classifiers $h_1(\mathbf{x}), h_2(\mathbf{x}), ..., h_K(\mathbf{x})$ and if the training set is drawn at random from the empirical distribution of the random vector $Y, \mathbf{X}$ the margin function is defined as:

$$\begin{aligned} mg(\mathbf{X}, Y) = avg_k I(h_k(\mathbf{X}) = Y) - \\ \max_{j != Y} avg_k I(h_k(\mathbf{X}) = j), \end{aligned} \tag{2.18}$$

where $I(\cdot)$ is the characteristic function. The margin function measures the distance between the average votes at $(\mathbf{X}, Y)$ for the right class and the average vote for any other class. For this model the generalization error function is:

$$PE^* = P_{\mathbf{X},Y}(mg(\mathbf{X}, Y) < 0), \tag{2.19}$$

where $P_{\mathbf{X},Y}$ is the probability over $\langle \mathbf{X}, Y \rangle$ space. For any event $A \subset \Omega$ of the feature space the characteristic function $I(\cdot)$ of A is:

$$I_A(x) = \begin{Bmatrix} \begin{cases} 1 & \iff (x \subset A) \\ 0 & otherwise \end{cases} \end{Bmatrix} \begin{Bmatrix} \begin{cases} 1 & \iff \exists x \\ 0 & otherwise \end{cases} \end{Bmatrix} \tag{2.20}$$

Random Forests classifiers were trained with a stepwise increase of the number of trees equal to the upper limit of 2^{11}. Optimal number of trees for model generalization as measured by mean misclassification rate for 10-fold cross-validation strategy is estimated to be 112 trees.

In order to find the final model, we trained a number of models and selected the best one based on κ metrics for the 10-fold validation strategy. The Cohen's κ measures pairwise agreement among a set of functions which are making classification decisions with correction for an expected chance agreement [Cohen, 1960]:

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)} \tag{2.21}$$

$\kappa = 0$ if there is no agreement more than expected by chance following the empirical distribution; while $\kappa = 1$ when there is a *max* agreement. κ is a state-of-the-art statistics about how significantly the classification model is different from chance. Importantly, κ is a more robust measure than the simple percent agreement, given that it takes into account chance agreement occurring without being too conservative.

During the learning and model selection process we used a random sampling with replacement to generate a new set of data for each fold from the basic training set, and followed leave-one-out 10-fold cross validation scheme. We adopted this strategy in order to prevent data overfitting and to deal with potential data loss in cases where calls, sms and Bluetooth proximities existed in the real world but were not registered by the smartphone logger software. Our *structural risk minimization*, as opposed to empirical risk minimization solution to prevent data overfitting, was incorporated by working with a regularization penalty into the learning process, balancing the model's complexity against training data fitting and by sampling the model training sets in such a way that they mimic the empirical distributions without most probable erroneous outliers.

Model parameter estimation selection was done iteratively on the basis of our exploratory analysis, inferred knowledge of the relationships between variables and model performance metrics (κ and Accuracy). Confounding variables are identified but not removed from the dataset during training and test phases.

2.2.13 Experimental Results

The performance metrics used to evaluate our approach are: accuracy, κ, sensitivity, and specificity. The recognition model based on random forest algorithm shows 90.68% accuracy on the training set and 72.28% accuracy on the test set. In Table 2.15 we provide the final stress recognition model performances on the test set along with their statistical significance [Garcia and Herrera, 2008].

Information about accuracy and κ metrics distribution using 10-fold cross validation strategy is provided in Table 2.16. As we can see, the distribution of the estimated performance metrics does not vary substantially among folds, signaling a good generalization despite the possible existence of heterogeneous data in each fold and the "noise" coming from the resampling procedure.

We also compared our approach based on combining multiple indicators with simpler approaches using as predictors (i) only personality traits, (ii) only weather conditions, (iii) only activities inferred from mobile phone data, (iv) a combination of personality traits and weather conditions, (v) a combination of personality traits and activities inferred from mobile phone data, (vi) a combination of weather conditions and activities inferred from mobile phone data. Table 2.17 reports accuracy, κ, sensitivity, specificity and F1 for each approach. In this table we also report the

Metric	Value
Accuracy	0.7228
95% CI	(0.7051, 0.7399)
No Information Rate	0.6384
P-Value [Acc > NIR]	¡ 2.2e-16
Kappa	0.3752
Sensitivity	0.5272
Specificity	0.8335
'Positive' Class	"stressed"

Table 2.15: Stress Recognition Model Performance Metrics

	Accuracy	Kappa
Min.	0.6959	0.2995
1st Qu.	0.7156	0.3535
Median	0.7282	0.3817
Mean	0.7232	0.3684
3rd Qu.	0.7312	0.3869
Max.	0.7404	0.4010

Table 2.16: 10-fold Cross-Validation Metrics for Stress Recognition Problem

performance of (vii) a simple majority classifier, which always returns the majority class as prediction (accuracy = 63.84%). Finally, we also ran experiments with three classes ("not stressed", "neutral", "stressed"), with labels -1 for "not stressed", label 0 for "neutral", and label 1 for "stressed". The class "not stressed" included all the cases with scores < 4, the class "neutral" included all the cases with scores $= 4$, and the class "stressed" included all the scores > 4. The sizes of the resulting three classes are 42.83% for "not stressed", 20.98% for "neutral", and 36.15% for "stressed". The global accuracy obtained by our multifactorial model, 59.57%, significantly outperformed the performance of simple majority classifier, which always returns the class "not stressed" as prediction.

2.2.14 Discussion

The comparison among the performance of the various models in Table 2.17 provides convincing evidence that none of the features sets (personality, weather, smart-

Table 2.17: Stress Recognition Model Metrics Comparison for Feature Subsets

Model	Accuracy	Kappa	Sensitivity	Specificity	F1
Our Multifactorial Model	72.28	37.52	52.72	83.35	57.89
Baseline Majority Classifier	63.84	0.00	100.00	0.00	0.00
Weather Only	36.16	0.00	100.00	0.00	0.00
Personality Only	36.16	0.00	100.00	0.00	0.00
Bluetooth+Call+Sms	48.59	6.80	73.80	34.32	50.94
Personality+Weather	43.55	2.96	81.90	21.83	51.20
Personality+Bluetooth+Call+Sms	46.40	7.01	83.17	25.57	52.88
Weather+Bluetooth+Call+Sms	49.60	-5.45	38.45	55.91	35.55

phone activity) considered alone is endowed with a good enough predictive power. This conclusion applies also to pairwise combinations of the same features sets to the extent that neither personality+smartphone activity, nor personality+weather, nor weather+smartphone activity do any better than the majority classifier (accuracy=63.84%). Interestingly, significant improvements over the latter can only be obtained by the simultaneous usage of the three features sets: our final model based on a Random Forest classifier using 32-dimensional feature vectors obtained a 72.28% accuracy for our 2-class classification problem.

As pointed out in Section 2, some recent works have used mobile phones data for stress recognition [Bauer and Lukowicz, 2012, Sano and Picard, 2013]. Bauer and Lukowicz [Bauer and Lukowicz, 2012] reported a 53% of accuracy in detecting the transition from stressful periods (a two week exam session) to non-stressful periods (two weeks after the exam session). Our multifactorial approach outperforms the approach proposed by [Bauer and Lukowicz, 2012] altough a direct comparison may be not adequate given the different task. More recently, Sano and Picard reported an accuracy performance of 75% using a combination of features from mobile phones and more obtrusive wearable sensors. However, the limited number of subjects (only 18) and the limited number of days (only 5) make the results preliminary. Other approaches used video and audio features for stress recognition [Giakoumis et al., 2012, Lu et al., 2012]. For instance, StressSense, an application for stress detection from human voice, achieved a 76% of accuracy in outdoor environments. However, this method depends on sound quality and it may pose privacy concerns for people perceiving voice recording and analysis as a threat to their privacy. Hence, our performance of 72.28 shows that the proposed multifactorial approach is a reliable and less obtrusive alternative.

An investigation of the most important predictors of daily stress reveales interesting associations. Table 1 reports the 32 features selected and used in our model ranked by their mean reduction in accuracy. All the personality traits contributes significantly in predicting the daily stress variable. These results are interesting because the previous studies in social psychology focused their analyses mainly on the associations between stress and Neuroticism, Extraversion and Conscientiousness. Instead, our work shows the important contribution played also by Agreeableness and Openness to Experience to the automatic classification of daily stress. Moreover, these results open us the possibility of creating a multi-step stochastic model in which we first estimate the personality and then we use those estimates as independent variables for the daily stress recognition problem. Our current approach uses self-reported information on personality and this strategy could be a limitation for scaling to larger sample of users. However, recent studies showed that personality traits may be reliable recognized from mobile phone data [de Oliveira et al., 2011, Staiano et al., 2012b, Chittaranjan et al., 2013, Montjoye et al., 2013].

With regard to weather, we found confirmation for the association between temperature and stress. Moreover, significant effects of other meteorological variables – humidity, visibility, and wind speed – for predicting daily stress were also found.

Regarding the mobile phone data, it is interesting to note the contribution of proximity features. Out of the selected 32 features, 11 features are proximity ones, 6 comes from call data and 6 from sms data. In particular, an interesting predictive role is played by the number of time intervals for which an id is seen. The results obtained using proximity features seem to confirm previous findings in social psychology: in particular, the relevant role played by face-to-face interactions and by interactions with strong ties in determining the stress level of a subject

[Krackhardt, 1992]. For sure, this result requires further investigation. In addition, two features capturing the entropy in proximity interactions are among the selected ones. This finding seems to confirm results available in the social psychological literature about the associations between stress and the richness/diversity of social interactions [Goode, 1960]. Further confirmation to this conclusion comes from the similalry important role played by entropy-based call and sms features.

The remaining selected features related to sms interactions are (i) the latency in replying to a text message, defined as the median time to answer a text message and (ii) the amount of sms communications (outgoing+incoming).

2.2.15 Implications and Limitations

Stress has become a major problem in our society. Ubiquitous connectivity, information overload, increased mental workload and time pressure are all elements contributing to increase general stress levels. While in some cases people may realize that they are undergoing stressful situations, severe and chronic stress may be more difficult to detect. Moreover, stress may be considered the norm in a modern and demanding society. Nonetheless, while slightly increased stress levels may be functional for productivity, prolonged and severe stress can be at the source of several physical dysfunctions like headache, sleep or immunological disorders, unhealthy behaviours such as smoking and bad eating habits, as well as of psychological and relational problems. Beside manifest social costs, stress also entails considerable financial costs for our economies, which are estimated by the World Health Organization in 300 billion dollars a year for American enterprises, and 20 billion euro for Europe ones, in terms of absenteeism and low productivity.

Our technology provides a cost-effective, unobtrusive, widely available and reliable tool for stress recognition. It detects daily stress levels with a 72.28% accuracy combining real life data from different sources, such as personality traits, social relationships (in terms of calls, sms and proximity interactions), and weather data. The development of a reliable stress recognition system is a first but essential step toward wellbeing and sustainable living, and its scope can be extended to different areas of applicability. Providing people with a tool capable of gathering rich data about real life, and transforming them into meaningful insights about stress levels, paves the way to a new generation of context-aware technologies that can target therapists, enterpises and common citizens.

This technology can inform the design of automatic systems for the assessment and treatment of psychological stress. With such a tool, therapists could monitor and record patients' daily stress levels, access longitudinal data, identify recurrent or significant stressors and modulate treatment accordingly.

In work environments, where stress has become a serious problem affecting productivity, leading to occupational issues and causing health diseases, our system could be extended and employed for early detection of stress-related conflicts and stress contagion, and for supporting balanced workloads. Awareness is a first but crucial step to motivate people to change their behaviour and take informed and concrete steps toward a healthy lifestyle and appropriate stress coping strategies. Mobile applications developed on the basis of our technology could provide feedback to increase people's awareness of their stress levels, alerts when they reach a warning threshold, and suggest stress management and relaxation techniques when appropriate.

However, our study has also some limitations. We can list the following ones: (i) our sample comes from a population living in the same environment. Our subjects were all married graduate students living in a campus facility of a major US university; and (ii) the non-availability of proximity data concerning the interaction with people not participating in the data collection, a fact that is common to many other relevant studies and that has been also pointed out by [Quercia et al., 2012]. The first problem is at least partially attenuated by the large variability of the sample in terms of provenance and cultural background (in our sample we have subjects from 16 countries and from all the continents), which can be expected to correspond to a wide palette of interaction behaviors that efficaciously counterbalance the effects of living-place homogeneity.

Conclusion

The goal of this chapter was to investigate the automatic recognition of people's daily stress from three different sets of data: a) people activity, as detected through their smartphones (data pertaining to transitory properties of individuals); b) weather conditions (data pertaining to transitory properties of the environment); and c) personality traits (data concerning permanent dispositions of individuals). The problem was modeled as a 2-way classification one. The results convincingly suggest that all the three types of data are necessary for attaining a reasonable predictive power. As long as one of those information sources is dropped, performances drop below those of the baselines. Moreover, the distributional data for accuracy and κ show the robustness and generalization power of our multifactorial approach.

Taken together, and despite the limitations discussed above, our results not only provide evidence that individual daily stress can be reliably predicted, but they also point to the necessity of considering at the same time people's transitory properties (smartphone activity), transitory properties of the environment and information about stable individual characteristics. For the sake of transitory individual properties, mobile phone usage patterns have important advantage over alternative methods: they are less unobtrusive and raise limited privacy problems as compared to, e.g., voice analysis or the exploitation of data from physiological sensors. Moreover, and importantly, automatic stress detection based on mobile phone data can take advantage of the extensive usage and diffusion of these devices, it can be applied in several real world situations and it can be exploited for a variety of applications that are delivered by means of the same device. For example, applications used to inform the design of clinical decision support systems or self-monitoring applications of stress levels in work settings and in other daily life situations, which allows people to identify personal stressors and enforces their proactive role in stress prevention and management.

Chapter 3

Predictive Modeling of Large Scale Group Behavior

3.1 Predicting Crime Hotspots Based on Aggregated Anonymized People Dynamics

3.1.1 Introduction

Crime, in all its facets, is a well-known social problem affecting the quality of life and the economic development of a society. Studies have shown that crime tends to be associated with slower economic growth at both the national level [Mehlum et al., 2005] and the local level, such as cities and metropolitan areas [Cullen and Levitt, 2009]. Crime-related research always attracted attention of criminal law and sociology scholars. Dating back to the beginning of the 20-th century, studies have focused on the behavioral evolution of criminals and its relations with specific characteristics of the neighborhoods in which they grew up, lived, and acted, and on behavioral development of factors like exposure to specific peer networks, neighborhood characteristics (*e.g.* presence/absence of recreational/educational facilities) and poverty indexes, has provided a wealth of knowledge from both individual and collective standpoints [Weinberg, 1954]. Existing works in the fields of criminology, sociology, psychology and economics tend to mainly explore relationships between criminal activity and socio-economic variables such as education [Ehrlich, 1975], ethnicity [Braithwaite, 1989], income level [Patterson, 1991], and unemployment [Freeman, 1999].

Several studies in criminology and sociology have provided evidence of significant concentrations of crime at micro geographical areas [Brantingham and Brantingham, 1999, Weisburd and Green, 1994]. It is important to note that such clustering of crime in small geographic areas (*e.g.* streets), commonly referred to as *hotspots*, does not necessarily align with trends that are occurring at a larger geographic level, such as communities. Research has shown, for example, that in what are generally seen as good parts of town there are often streets with strong crime concentrations, and in what are often defined as bad neighborhoods, many places are relatively free of crime [Weisburd and Green, 1994].

In 2008, criminologist David Weisburd proposed to switch the popular people-centric paradigm of police practices to a place-centric one [Weisburd, 2008], thus focusing on geographical topology and micro-structures rather than on criminal profiling. In our result, *crime prediction* is used in conjunction with a place-centric definition of the problem and with a data-driven approach: we specifically investigate the predictive power of aggregated and anonymized human behavioral data derived from a multimodal combination of mobile network activity and demographic information to determine whether a geographic area is likely to become a *scene of crime* or not.

As the number of mobile phones actively in use worldwide approaches the 6.8 billion mark[1], they become a very valuable and unobtrusive source of human behavioral data: mobile phones can be seen as sensors of aggregated human activity [Dong et al., 2011b, Laurila et al., 2013] and have been used to monitor citizens' mobility patterns and urban interactions [Gonzalez et al., 2008, Zheng et al., 2014a], to understand individual spending behaviors [Singh et al., 2013], to infer people's traits [Montjoye et al., 2013, Staiano et al., 2012a] and states [Bogomolov et al., 2013b], to map and model the spreading of diseases such

[1] http://www.itu.int

as malaria [Wesolowski et al., 2012] and H1N1 flu [Frias-Martinez et al., 2011], and to predict and understand socio-economic indicators of territories [Eagle et al., 2010, Soto et al., 2011, Smith-Clarke et al., 2014]. Recently, Zheng *et al.* proposed a multi-source approach, based on human mobility and geographical data, to infer noise pollution [Zheng et al., 2014b] and gas consumption [Shang et al., 2014] in large metropolitan areas.

In this chapter, we use human behavioral data derived from a combination of mobile network activity and demography, together with open data related to crime events to predict crime *hotspots* in specific neighborhoods of a European metropolis: London. The main contributions of this work are:

1. The use of human behavioral data derived from anonymized and aggregated mobile network activity, combined with demographics, to predict crime hotspots in a European metropolis;

2. A comprehensive analysis of the predictive power of the proposed model and a comparison with a state-of-the-art approach based on official statistics;

3. A discussion of the theoretical and practical implications of our proposed approach.

3.1.2 Related Work

Researchers have devoted attention to the study of criminal behavior dynamics both from a people- and place- centric perspective. The people-centric perspective has mostly been used for individual or collective criminal profiling. Wang *et al.* [Wang et al., 2013] proposed *Series Finder*, a machine learning approach to the problem of detecting specific patterns in crimes that are committed by the same offender or group of offenders. In [Short et al., 2008], it is proposed a biased random walk model built upon empirical knowledge of criminal offenders behavior along with spatio-temporal crime information to take into account repeating patterns in historical crime data. Furthermore, Ratcliffe [Ratcliffe, 2006] investigated the spatio-temporal constraints underlying offenders' criminal behavior.

An example of a place-centric perspective is crime hotspot detection and analysis and the consequent derivation of useful insights. A novel application of quantitative tools from mathematics, physics and signal processing has been proposed by Toole *et al.* [Toole et al., 2011] to analyse spatial and temporal patterns in criminal offense records. The analyses they conducted on a dataset containing crime information from 1991 to 1999 for the city of Philadelphia, US, indicated the existence of multi-scale complex relationships both in space and time. Using demographic information statistics at community (town) level, Buczak and Gifford [Buczak and Gifford, 2010] applied fuzzy association rule mining in order to derive a finite (and consistent among US states) set of rules to be applied by crime analysts. Other common models are the ones proposed by Eck *et al.* [Eck et al., 2005] and by Chainey *et al.* [Chainey et al., 2008] that rely on kernel density estimation from the criminal history record of a geographical area. Similarly, Mohler *et al.* [Mohler et al., 2011] applied the self-exciting point process model (previously developed for earthquake prediction) as a model of crime. The major problem of all these approaches is that they relies on the prior occurrence of crimes in a particular area and thus cannot generalize to previously unobserved areas.

More recently, the proliferation of social media has sparked interest in using this kind of data to predict a variety of variables, including electoral outcomes

[Tumasjan et al., 2010] and market trends [Bollen et al., 2011]. In this line, Wang *et al.* [Wang et al., 2012] proposed the usage of social media to predict criminal incidents. Their approach relies on a semantic analysis of tweets using natural language processing along with spatio-temporal information derived from neighborhood demographic data and the tweets metadata.

In this chapter, we tackle the crime hotspot forecasting problem by leveraging mobile network activity as a source of human behavioral data. Our work hence complements the above-mentioned research efforts and contributes to advance the state-of-the-art in quantitative criminal studies.

3.1.3 Datasets

The datasets we exploit in this chapter were provided during a public competition - the Datathon for Social Good - organized by Telefónica Digital, The Open Data Institute and MIT during the Campus Party Europe 2013 at the O2 Arena in London in September 2013.

Participants were provided access to the following data, among others:

- Anonymized and aggregated human behavioral data computed from mobile network activity in the London Metropolitan Area. We shall refer to this data as the Smartsteps dataset, because it was derived from Telefonica Digital's Smartsteps product[2]. A sample visualization of the Smartsteps product can seen in Figure 7.
- Geo-localised Open Data, a collection of openly available datasets with varying temporal granularity. This includes reported criminal cases, residential property sales, transportation, weather and London borough profiles related to homelessness, households, housing market, local government finance and societal well-being (a total of 68 metrics).

We turn now to describing the specific datasets that we used to predict crime hotspots.

3.1.4 Smartsteps Dataset

The Smartsteps dataset consists of a geographic division of the London Metropolitan Area into cells whose precise location (lat,long) and surface area is provided. Note that the actual shape of the cell was not provided. In total, there were 124119 cells. We shall refer to these cells as the Smartsteps cells.

For each of the Smartsteps cells, a variety of demographic variables were provided, computed every hour for a 3-week period, from December 9th to 15th, 2012 and from December 23rd, 2012 to January 5th, 2013.

In particular:

(1) *Footfall*, or the estimated number of people within each cell. This estimation is derived from the mobile network activity by aggregating every hour the total number of unique phonecalls in each cell tower, mapping the cell tower coverage areas to the Smartsteps cells, and extrapolating to the general population – by taking into account the market share of the network in each cell location; and

(2) an estimation of gender, age and home/work/visitor group splits.

[2] http://dynamicinsights.telefonica.com/488/smart-steps

That is, for each Smartsteps cell and for each hour, the dataset contains an estimation of how many people are in the cell, the percentage of these people who are at home, at work or just visiting the cell and their gender and age splits in the following brackets: 0-20 years, 21-30 years, 31-40 years, etc..., as shown in Table 3.1.

Table 3.1: SmartSteps Data Description

Type	Data
4*Origin-based	total # people
	# residents
	# workers
	# visitors
2*Gender-based	# males
	# females
6*Age-based	# people aged up to 20
	# people aged 21 to 30
	# people aged 31 to 40
	# people aged 41 to 50
	# people aged 51 to 60
	# people aged over 60

Figure 7 shows a sample visualization of the information made available from the SmartSteps platform by the Challenge Organizers. The data refers to 1-hour intervals and to each Smartsteps cell. By combining aggregated and anonymized demographics and mobility data, fine-grained spatio-temporal dynamics can be exploited to derive valuable insights for the scenario of interest.

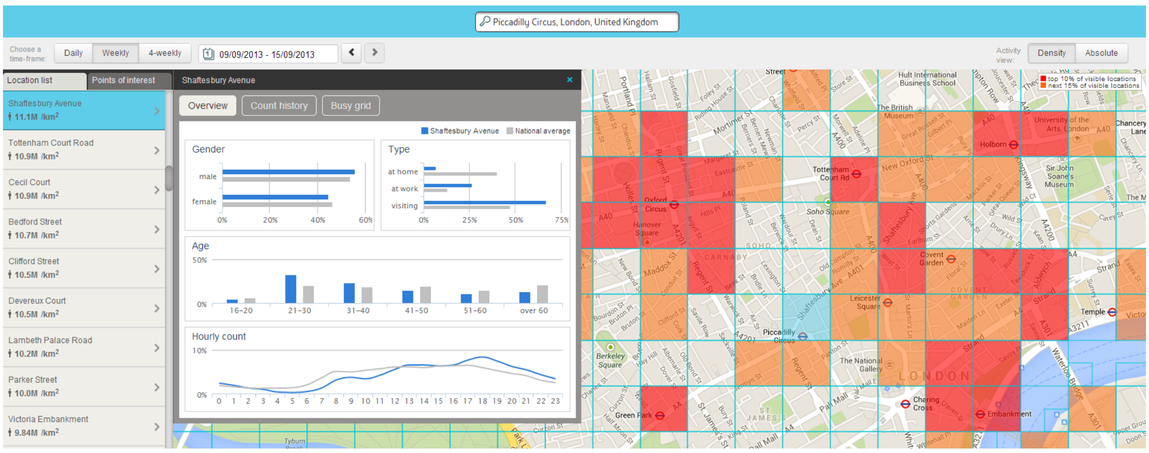

Figure 7: Sample Visualization of Anonymized Data by Smartsteps Platform

3.1.5 Criminal Cases Dataset

The criminal cases dataset includes the geo-location of all reported crimes in the UK but does not specify their exact date, just the month and year. The data provided in the public competition included the criminal cases for December 2012 and January 2013.

In detail, the crime dataset includes: the crime ID, the month and year when the crime was committed, its location (longitude, latitude, and address where the crime took place), the police department involved, the *lower layer super output area* code, the *lower layer super output area* name and its type out of 11 possible types (*e.g.* anti-social behavior, burglary, violent crime, shoplifting, etc.).

Crimes reported for the UK include geolocation but no specific date (just month and year). The crime data included: crime ID, month when the crime was commited, police department, longitude, latitude, address where the crime happend, lower layer super output area code, lower layer super output area name, crime type and last outcome type.

Super Output Areas (SOAs) are geographical boundaries for the collection and publication of small area statistics, which represent spatial neighbourhoods. They are a statistical geography and their key aspects are stability and uniformity of size. SOAs avoid the problems caused by the inconsistent and unstable population geography. They are considered to be better for statistical comparison as they have consistent size and each layer has a specified minimum population to avoid the risk of personal reidentification. SOAs are not subject to frequent boundary changes, and therefore are more suitable for comparison over time. In addition they build on the existing availability of data for output areas (OAs). The SOA layers form a hierarchy based on aggregations of Output Areas. Lower Layer Super Output Areas (LSOAs) were first built using 2001 Census data from groups of Output Areas and have been updated following the 2011 Census. Middle Layer Super Output Areas (MSOAs) are larger areas created by zone-design software using census data from groups of LSOAs. They fit within local authority boundaries. Upper Layer Super Output Area (USOAs) were not used in the datasets analyzed. There are 181,408 OAs and 34,753 LSOAs in England and Wales. Only the LSOAs that correspond to London's Metropolian area were used in the experiments reported in this chapter.

Their aim is to define spatial areas, based on population levels, whose boundaries would not change over time. LSOAs are the smallest type of output areas used for official statistics, have a mean population of 1,500 and a minimum population threshold of 1,000.

The ground truth crime data used in our experiments corresponds to the crimes reported in January 2013 for each of the Smartsteps cells.

3.1.6 London Borough Profiles Dataset

The London borough profiles dataset is an official open dataset containing 68 different metrics about the population of a particular geographic area. The spatial granularity of the bourough profiles data is at the LSOA level.

In particular, the information in this dataset includes statistics about the population, households (census), demographics, migrant population, ethnicity, language, employment, NEET (Not in Education, Employment, and Training) people, earnings, volunteering, jobs density, business survival, house prices, new homes, greenspace, recycling, carbon emissions, cars, indices of multiple deprivation, children in out-of-work families, life expectancy, teenage conceptions, happiness levels, political control (*e.g.* proportion of seats won by Labour, Liberal Democrats and Conservatives), and election turnout.

3.1.7 Methodology

We cast the problem of crime hotspot forecasting as a binary classification task. For each Smartsteps cell, we predict whether that particular cell will be a crime hotspot or not in the next month. In this section we provide details of the experimental setup that we followed.

3.1.8 Data Preprocessing and Feature Extraction

Starting from the bulk of data described in Section 3.1.3, we performed the following preprocessing steps.

Referencing all geo-tagged data to the Smartsteps cells

As the SmartSteps cell IDs, the borough profiles and the crime event locations are not spatially linked in the provided datasets, we first geo-referenced each crime event by identifying the Smartsteps cell which is the closest to the location of the crime. We carried out a similar process for the borough profile dataset. As a result, each crime event and the borough profile information were linked to one of the Smartsteps cells.

Smartsteps features

Diversity and *regularity* have been shown to be important in the characterization of different facets of human behavior and, in particular, the concept of entropy has been applied to assess the predictability of mobility [Song et al., 2010] and spending patterns [Krumme et al., 2013, Singh et al., 2013], the socio-economic characteristics of places and cities [Eagle et al., 2010] and some individual traits such as personality [Montjoye et al., 2013]. Hence, for each Smartsteps variable (see Table 3.1) we computed the mathematical functions which characterize the distributions and information theoretic properties of such variables, *e.g.* mean, median, standard deviation, min and max values and Shannon entropy.

In order to be able to also account for temporal relationships inside the Smartsteps data, the same computations as above were repeated on sliding windows of variable length (1-hour, 4-hour and 1 day), producing *second-order features* that help reduce computational complexity and the feature space itself, while preserving useful data properties.

London borough profile features

No data preprocessing was needed for the London borough profiles. Hence, we used the original 68 London borough profile features.

Crime hotspots ground-truth data

The distribution of the criminal cases data is reported in Table 3.2.

Table 3.2: Number of Crime Hotspots in January

Min.	Q1	**Median**	Mean	Q3	Max.
1	2	**5**	8.2	10	289

Given the high skewness of the distribution (skewness = 5.88, kurtosis = 72.5, mean = 8.2, median = 5; see Table 3.2) and based on previous research on urban crime patterns [Boggs, 1965], we split the criminal dataset with respect to its median into two classes: a *low* crime (class '0') when the number of crimes in the given cell was less or equal to the median, and a *high* crime (class '1') when the number of crimes in a given cell was larger than the median.

Following the empirical distribution, the two resulting classes are approximately balanced (53.15% for the *high* crime class).

3.1.9 Feature Selection

We randomly split all data into training (80% of data) and testing (20% of data) sets. In order to accelerate the convergence of the models, we *normalized* each dimension of the feature vector [Box and Cox, 1964].

Due to similarity of human behavior understanding tasks we followed the same methodology for feature selection as we introduced and discussed in the previous chapters.

The feature selection process produced a reduced subset of 68 features (from an initial pool of about 6000 features), with a reduction in dimensionality of about 90 times with respect to the full feature space. The top 20 features selected by the model are included in Table 3.3.

3.1.10 Model Building

We trained a variety of classifiers on the training data following 5-fold cross validation strategy: logistic regression, support vector machines, neural networks, decision trees, and different implementations of ensembles of tree classifiers with different parameters.

The decision tree classifier based on the Breiman's Random Forest (RF) algorithm yielded the best performance when compared to all other classifiers. Hence, we report the performance results only for the the best model, based on this algorithm.

We took advantage of the well-known performance improvements that are obtained by growing an ensemble of trees and voting for the most frequent class. Random vectors were generated before the growth of each tree in the ensemble, and a random selection without replacement was performed [Breiman, 1996].

The *consistency* of the random forest algorithm has been proven and the algorithm adapts to sparsity in the sense that the rate of *convergence* depends only on the number of strong features and not on the number of noisy or less relevant ones [Biau, 2012].

3.1.11 Experimental Results

In this section we report the experimental results obtained by the Random Forest trained on different subsets of the selected features and always on the test set, which was not used during the training phase in any way.

The performance metrics used to evaluate our approach are: accuracy, F1, and AUC score. As can be seen on Table 3.4, the model achieves almost 70% accuracy when predicting whether a particular Smartsteps cell will be a crime hotspot in the following month or not. Table 3.4 includes all performance metrics obtained by our model.

A spatial visualisation of our results is reported on a map of the London metropolitan area in Figure 9 and compared with a similar visualisation of the ground truth labels in Figure 8. In the maps, green represents "low crime level" and red "high crime level".

Second order features, which we introduced to capture intertemporal dependencies for our problem, not only made the feature space more compact, but also yielded a significant improvement in model performance metrics.

In order to understand the value added by the Smartsteps data, we compared the performance of the Random Forest using all features with two different models trained with (i) only the subset of selected features derived from the borough

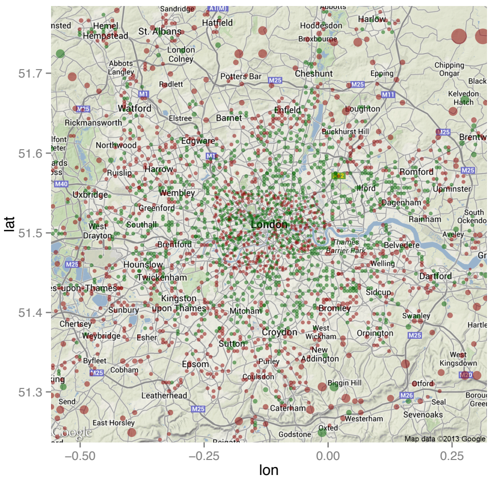

Figure 8: Ground Truth of Crime Hotspots

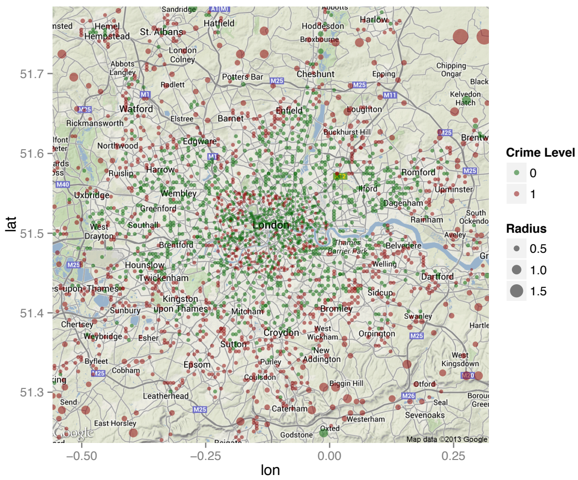

Figure 9: Predicted by Algorithm Crime Hotspots

profiles dataset (Borough Profiles) and (ii) only the subset of selected features derived from the Smartsteps dataset (Smartsteps).

Table 3.4 reports accuracy, F1, and the area under the ROC curve metrics for each of the models. In this Table, we also report the performance of (iii) a simple majority classifier, which always returns the majority class ("High Crime") as prediction (accuracy=53.15%).

As can be seen on Table 3.4, the borough-only model yields an accuracy of 62.18%, over 6% lower than the accuracy obtained with the Smartsteps model (68.37%). The Smartsteps+Borough model yields an increase in accuracy of over 7% when compared with the borough profiles model (69.54% vs 62.18% accuracy) while using the same number of variables.

Table 3.3: Crime Hotspots Prediction Top-20 Selected Features Ranked by Mean Decrease in Accuracy

Rank	Features	0	1	MeanDecreaseAccuracy	MeanDecreaseGini
1	smartSteps.daily.ageover60.entropy.empirical.entropy.empirical	4.48	5.43	9.02	18.75
2	smartSteps.daily.athome.mean.sd	3.20	7.60	8.91	27.13
3	smartSteps.daily.age020.sd.entropy.empirical	5.69	3.97	8.85	16.88
4	smartSteps.daily.age020.mean.entropy.empirical	3.09	5.88	8.85	17.26
5	smartSteps.daily.age020.mean.sd	4.50	5.27	8.65	16.03
6	smartSteps.daily.athome.min.entropy.empirical	6.39	2.32	8.61	15.99
7	smartSteps.daily.athome.sd.sd	3.22	8.58	8.60	45.82
8	smartSteps.daily.athome.sd.mean	3.35	5.83	8.57	24.93
9	smartSteps.daily.ageover60.entropy.empirical.sd	4.62	4.95	8.56	20.45
10	smartSteps.daily.athome.sd.median	5.41	5.04	8.50	26.48
11	smartSteps.daily.age3140.entropy.empirical.max	2.33	5.79	8.44	16.24
12	smartSteps.daily.age3140.min.sd	6.81	4.06	8.31	36.52
13	smartSteps.daily.athome.min.sd	4.36	6.85	8.29	34.26
14	smartSteps.daily.athome.sd.max	4.13	6.87	8.27	34.89
15	smartSteps.monthly.athome.max	3.92	5.42	8.26	29.86
16	smartSteps.monthly.athome.sd	4.43	4.17	8.21	39.70
17	smartSteps.daily.age5160.entropy.empirical.entropy.empirical	4.74	4.11	8.13	16.64
18	smartSteps.daily.age020.sd.sd	3.67	5.88	8.12	16.86
19	smartSteps.daily.athome.entropy.empirical.entropy.empirical	5.13	4.82	8.08	18.55
20	smartSteps.daily.athome.max.sd	2.83	6.29	8.07	26.85

3.1.12 Discussion

The results discussed in the previous section show that usage of human behavioral data (at a daily and monthly scale) significantly improves prediction accuracy when compared to using rich statistical data about a borough's population (households census, demographics, migrant population, ethnicity, language, employment, etc...). The borough profiles data provides a fairly detailed view of the nature and living conditions of a particular area in a city, yet it is expensive and effort-consuming to collect. Hence, this type of data is typically updated with low frequency (*e.g.* every few years). Human behavioral data derived from mobile network activity and demographics, though less comprehensive than borough profiles, provides significantly finer temporal and spatial resolution.

Next, we focus on the most relevant predictors of crime level which show interesting associations. We first take a look at the top-20 variables in our model, which are sorted by their mean reduction in accuracy (see Table 3.3).

Table 3.4: Crime Hotspots Prediction Problem Metrics Comparison

Model	Acc.,%	Acc. CI, 95%	F1,%	AUC
Baseline Majority Classifier	53.15	(0.53, 0.53)	0	0.50
Borough Profiles Model (BPM)	62.18	(0.61, 0.64)	57.52	0.58
Smartsteps	68.37	(0.67, 0.70)	65.43	0.63
Smartsteps + BPM	69.54	(0.68, 0.71)	67.23	0.64

The naming convention that we used for the features shown in Table 3.3 is: the original data source (*e.g.* "smartSteps") is followed by the temporal granularity T (*e.g.* "daily"), the semantics of the variable (*e.g.* "athome"), and its statistics at T (*e.g.* "mean"). Note that *second order* features where we computed statistics across multiple days appear *after* the first statistics. For example, feature 2 in the Table, "smartSteps.daily.athome.mean.sd", is generated by computing the standard deviation of the daily means of the percentage of people estimated to be at home.

As shown in the Table, the Smartsteps features have more predictive power than official statistics coming from borough profiles. No features listed in the top-20 are actually obtained using borough profiles. Moreover, Table 3.3 shows that higher-level features extracted over a sequence of days from variables encoding the daily dynamics (all features with the label smartSteps.daily.*) have more predictive power than features extracted on a monthly basis. For example, feature 2 in Table 3.3. This finding points out at the importance of capturing the temporal dynamics of a geographical area in order to predict its levels of crime.

Furthermore, features derived from the percentage of people in a certain cell who are at home (all features with *.athome.* in their label) both at a daily and monthly basis seem to be of extreme importance. In fact, 11 of the top 20 features are related to the *at home* variable.

It is also interesting to note the role played by diversity patterns captured by Shannon entropy features [Shannon, 1948]. The entropy-based features (all features with *.entropy.* in their label) in fact seem useful for predicting the crime level of places (8 features out of the top 20 are entropy-based features). In our study, the Shannon entropy captures the predictable structure of a place in terms of the types of people that are in that area over the course of a day. A place with high entropy would have a lot of variety in the types of people visiting it on a daily basis, whereas a place with low entropy would be characterised by regular patterns over time. In this case, the daily diversity in patterns related to different age groups, different use (home vs work) and different genders seems a good predictor for the crime level in a given area. Interestingly, Eagle *et al.* [Eagle et al., 2010] found that Shannon entropy used to capture the social and spatial diversity of communication ties within an individual's social network was strongly and positively correlated with economic development.

As previously described, borough profile features (official statistics) have lower predictive power with respect to accuracy than features extracted from aggregated mobile network activity data. Six borough profile features were selected in the final feature vector, including the proportion of the working age population who claim out of work benefits, the largest migrant population, the proportion of overseas nationals entering the UK and the proportion of resident population born abroad –metrics based on 2011 Census Bureau data. The predictive power of some of these variables is in line with previous studies in sociology and criminology. For example, several studies show a positive association among unemployment rate and crime level of an area [Raphael and Winter-Ebmer, 2001]. Still under debate is the positive association among number of immigrants and crime level [Ellis et al., 2009]. However, our experimental results show that the static nature of these variables makes them less useful in predicting crime level's of a given area when compared with less detailed but daily information about the types of people present in a same area throughout the day.

3.1.13 Implications and Limitations

We have outlined and tested a multimodal approach to automatically predict with almost 70% accuracy whether a given geographical area will have high or low crime levels in the next month. The proposed approach could have clear practical implications by informing police departments and city governments on how and where to invest their efforts and on how to react to criminal events with quicker response times. From a proactive perspective, the ability to predict the safety of a geographical area may provide information on explanatory variables that can be used to identify underlying causes of these crime occurrence areas and hence enable officers to intervene in very narrowly defined geographic areas.

The distinctive characteristic of our approach lies in the use of features computed from aggregated and anonymized mobile network activity data in combination with some demographic information. Previous research efforts in criminology have tackled similar problems using background historical knowledge about crime events in specific areas, criminals' profiling, or wide description of areas using socio-economic and demographic indicators. Our findings provide evidence that aggregated and anonymized data collected by the mobile infrastructure contains relevant information to describe a geographical area in order to predict its crime level.

The first advantage of our approach is its predictive ability. Our method predicts crime level using variables that capture the dynamics and characteristics of the demographics and nature of a place rather than only making extrapolations from previous crime histories. Operationally, this means that the proposed model could be used to predict new crime occurrence areas that are of similar nature to other well known occurrence areas. Even though the newly predicted areas may not have seen recent crimes, if they are similar enough to prior ones, they could be considered to be high-risk areas to monitor closely. This is an important advantage given that in some areas people are less inclined to report crimes [Tarling and Morris, 2010]. Moreover, our approach provides new ways of describing geographical areas. Recently, some criminologists have started to use *risk terrain modeling* [Caplan and Kennedy, 2010] to identify geographic features that contribute to crime risk, *e.g.* the presence of liquor stores, certain types of major stores, bars, etc. Our approach can identify novel risk-inducing or risk-reducing features of geographical areas. In particular, the features used in our approach are dynamic and related to human activities.

Our study has several limitations due to the constraints of the datasets used.

First of all, we had access only to 3 weeks of Smartsteps data collected between December 2012 and the first week of January 2013. In addition, the crime data provided was aggregated on a monthly basis. Having access to crime events aggregated on a weekly, daily or hourly basis would enable us to validate our approach with finer times granularity, predicting crimes in the next week, day or even hour.

3.1.14 Conclusion

In this chapter we have proposed a novel approach to predict crime hotspots from human behavioral data derived from mobile network activity, in combination with demographic information. Specifically, we have described a methodology to automatically predict with almost 70% of accuracy whether a given geographical area of a large European metropolis will have high or low crime levels in the next month. We have shown that our approach, while using a similar number of variables, significantly improves prediction accuracy (6%) when compared with using traditional, rich – yet expensive to collect – statistical data about a borough's population. Moreover, we have provided insights about the most predictive features (*e.g.* home-based and entropy-based features) and we have discussed the theoretical and practical implications of our methodology. Despite the limitations discussed above and the additional investigations needed to validate our approach and the robustness of our indicators, we believe that our findings open the door to exciting avenues of research in computational approaches to deal with a well-known social problem such as crime.

3.2 Predicting Electric Energy Consumption Using Aggregated Telecom Data

Energy efficiency is a key challenge for building sustainable societies. Due to growing populations, increasing incomes and the industrialization of developing countries, the world primary energy consumption is expected to increase annually by 1.6%. Hence, this scenario raises issues related to the increasing scarcity of natural resources, the accelerating pollution of the environment, and the looming threat of global climate change.

Recently, several computational works have started in focusing on energy and sustainability problems. In particular, they have targeted energy issues with two main objectives: (i) modeling and predicting energy consumption behaviors [Kolter and Ferreira Jr, 2011] and (ii) inducing behavioral changes in energy consumption [Allcott and Mullainathan, 2010].

This chapter targets the energy consumption prediction task but adopting a novel approach. We use aggregated and anonymized human behavioral data, derived from mobile network activity, in order to predict the energy consumption of a given geographical area. More specifically, we deal with two different tasks: (i) *average daily energy consumption* and (ii) *peak daily energy consumption*. The method we propose has several advantages: (i) it is cheap in the sense that it uses existing data generated by telecom infrastructure. No modification to mobile phones or telecom equipment is required, and (ii) it has good scalability and so it is suitable for large populations, which is desirable for city planning and energy management.

3.2.1 Methodology

We built our approach on a dataset spanning over a period of 2 months and a territory of 6000 square kilometers in the Northern Italy. The telecommunication and the energy consumption datasets have the same spatio-temporal aggregation. The temporal aggregation is ten minute intervals while the spatial one is obtained by partitioning the territory using a regular square grid, called *partitioning grid*. Each square of the grid measures approximately 1 square kilometer.

The *telecommunication dataset* is obtained from the Call Detail Records (CDRs) generated by the cellular network of a telecommunication operator offering its services on the territory under analysis. For the generation of this dataset CDRs recording sent SMSs, received SMSs, issued calls, received calls, and events related with Internet connections have been considered. The *energy consumption dataset* provides information about the structure of the electrical grid and the electrical current flowing through the 180 primary distribution lines serving the majority of the users living in the territory under analysis. Primary lines (medium voltage) are managed by a local company and bring energy from the national grid (high voltage) in order to distribute it among all the users. The dataset is composed by two sub-datasets: (i) the *structure of the electrical grid*, each line is described by providing the number of customer sites it serves in each square of the *partitioning grid* (customer sites often provide energy to more than one customer and they can also serve structures like condominiums, businesses and government organizations), and (ii) the *line measurement grid* that provides the instantaneous current flowing through each primary line every ten minutes.

Looking at the amount of electric current passing a point in an electric circuit per unit of time for each power line, we found that it has a number of cyclic characteristics and trends. We found predictable changes that repeat over daily and weekly periods. Based on these regularities we separated all power lines into 3 clustered areas: residential, touristic and city center/industrial area (see Fig. 14).

Then, we hypothesised that cellular communication patterns, which represent human dynamics in space and time, could be a good proxy for energy consumption prediction. To this end, we computed, from the anonymized and aggregated mobile network activity, a number of features characterizing diversity, regularity and general mobile phone usage in each part of the territory spatially aggregated by square grid. The discovered regularities, described above, were explicitly coded into the feature space by extracting of the number of hour in a day and the number of a weekday for each data source being processed.

The prediction tasks were solved for the next 7 days interval for each electric line ID and were designed as non-linear multiple regressions. We used the Random Forest algorithm proposed by Breiman [Breiman, 1999]. The *consistency* of this algorithm has been proven: the algorithm adapts to sparsity in the sense that the rate of *convergence* depends only on the number of strong features and not on the number of noisy or less relevant ones.

To solve the problem of computational complexity of the huge amount of data samples (>600 millions of CDRs) we moved from time domain of communication patterns to the frequency domain, applying fast Fourier transform algorithm for each group of daily time series. Then, we computed second-order features by mathematical functions, such as mean, median, sum, variance, skewness, kurtosis, entropy, characterising the distributions and it's properties. Also we found that only a small set of harmonics in Fourier domain explains the response variable variance for each type of first-order feature space time series, which reduces the computational complexity by

Table 3.5: Mean Daily Consumption Model Metrics for Energy Prediction Problem

Metric	Baseline	Model
MAE	20.8468	12.3683
RAE	98.9169	58.6869
MSE	790.6041	325.2679
RMSE	28.1177	18.0352
RSE	100.9551	41.5346
R2	-0.0096	0.5847

Table 3.6: Peak Daily Consumption Model Metrics for Energy Prediction Problem

Metric	Baseline	Model
MAE	186.1440	17.3112
RAE	621.7292	57.8201
MSE	36062.7851	601.7531
RMSE	189.9020	24.5307
RSE	2551.8602	42.5810
R2	-24.5186	0.5742

a number of orders. Finally, the computed second-order feature space for each spatial had a reduced number of dimensions (>3000), but still represented the temporal and spatial characteristics (i.e. diversity and regularity) of communication patterns.

A feature selection step was performed before the model building. The feature selection was done on a reduced sample of the training data, which was one week long. The metric used was the decrease in Gini index, which is the impurity measure of a decision tree node. This method is proven to be superior than correlation-based measures or information gain criteria. The feature selection step reduced the feature space to 32 dimensions for each of the two models without losing much accuracy.

For each variable from CDRs we computed the mathematical functions, which characterize the distributions and measure the information theoretic and statistical properties of such variables, e.g. mean, median, standard deviation, min and max values and Shannon entropy. Moreover, the same computations were repeated on sliding windows of variable length (1-hour, 4-hour and 1 day), producing second-order features that capture temporal relationships (see Table B.1 and Table B.2).

3.2.2 Results and Discussion

Prediction metrics for daily average energy consumption for the next 7 days are 2.43 times better than the baseline, which is the training set mean value (MSE = 325.2679 compared to the baseline MSE = 790.6041). Prediction metrics for daily peak energy consumption model for the next 7 days prediction interval are 59.93 times better than the baseline, which is the training set maximum value (MSE = 601.7531 vs baseline MSE = 36062.7851).

In Table 3.5 and Table 3.5 we report Mean Absolute Error (MAE), the Mean Squared Error (MSE), the Root Mean Squared Error (RMSE), the Relative Squared Error (RSE), the Relative Absolute Error (RAE), R^2.

Our results prove that human dynamics, extracted from aggregated and anonymized mobile phone data, are good proxies for modelling energy consumption. The introduced models technically do not account for seasonality on a yearly scale due to the 2-months limitation of the data sample. However, this limitation could easily be solved by training the model on more data.

Our results have several practical implications. Our approach could help to optimize the economy of energy producers/distributors value chain, also acting as an efficient tool for meeting peak electrical energy demand. Hence, could help to reduce total primary energy consumption and ecological footprint including climate change still meeting the people's energy needs.

Chapter 4

Summary and Future Work

4.1 Computing Human Behavior: The New Kind of Science

An important trend is the availability of more and more data, often collected with mobile phones, or internet of things sensors, or generated by other automated way. In many areas of *Computational Social Science* and *Social Physics* experiments on personal data and large group aggregated data are tantalizing regularities in the data. But the challenge that exists is to choose the right data and find good models out of it. Sometimes these models are important for theoretical science. More often they have practical purpose of human behavior understanding, i.e. characterization, prediction, anomaly detection, regularities extraction and algorithmization. These techniques open new business models, new competitive advantages and the new way how traditional offline processes are changed for telecommunications, energy distribution, government, etc.

In the past, one might find a model from data by using statistics or machine learning in order to fit parameters of a mathematical formula or an algorithm. The *new kind of science* that we believe suggests that instead of trying to find in the computational universe the parameters of a mathematical formula but a set of underlying algorithmic rules that can be run to simulate the essence of whatever generates the data. At present, the methods we have for finding models in the computational universe are still fairly ad hoc, we are working with the consequences but not with the source reasons, not with the core generating processes. With the future development of artificial intelligence and machine learning, it will be possible to streamline such process, and to develop some kind of highly systematic methodology, some algorithmic representation of the source processes that drive the data generation that we tend to analyze now.

The second important consequence of the AI-inspired new kind of science, is the effect on everyday world perception by humans. Existing mathematical approach to science and the way of systematic thinking is having a profound effect on how we think about all thing in the world, i.e. now we are using words like "linear dependency", "exponential growth", "correlation", "game theory", which come directly from such an approach. Inside the paradigm of the new kind of science we will be using the new way of thinking, the keywords like "computationally irreducible", "intrinsically random", "heterogeneous parallel computations", "compressed representation", "deep neural representation", "antifragile systems". And as such concepts will become more widespread they will inform thinking about more and more deep and intrinsic things – whether they are describing the operation of an organization, a society, a technical process, or working out what could conceivably be predictable for purposes of human behavior understanding in general sense.

4.2 Supervised Learning of Individual Characteristics

Understanding affective and social behavior of humans with computational machine learning tools is receiving increased interest. Researchers seek to analyze patterns emanating from interactions between humans, as well as between humans and computers or mobile phones, with the goal of designing more responsive and natural interfaces, better product and services offered to individuals on personalized basis. Supervised Learning of human behavior in science involves understanding of body motion, gestures and signs recognition, analysis of facial expressions, and interpretation of affective signals. These signals are usually integrated with the contextual properties of an application specific domain. Social signal processing deals with interactions between humans. Social interaction could be tracked and measured in a relatively cheap and concise way by telecom patterns. It integrates verbal cues with rich sets of non-verbal behavioral cues to deep analysis of social interactions. Ambient intelligence deals with smarter environments, which are more costly to build. In ambient environments, the living space is equipped with many sensors that observe the behavior of humans. More specific, perceptual user interfaces are concerned with more responsive human-computer interfaces. In this domain the computer is given the capacity to detect behavioral changes of its user. The analysis of spatio-temporal dynamics of human actions, observed through different sensory modalities, allows inference and customization on many levels.

Which types of messages are communicated by the data driven behavioral signals? This question is related to psychological issues pertaining to the nature of behavioral signals and the best way to interpret them from computer science perspective. Which human communication patterns and human dynamics cues convey information about a certain type of behavioral signals? Our findings shape the choice of different modalities to be included into an automatic analysis of human behavior. Supervised learning of individual characteristics of human behavior today is the best practical way to formulate the problems of human behavior understanding, given the modern maturity level of machine learning and inferential statistics. How are various behavioral characteristics could be combined in order to optimize inference about the known human behavior? This question is related to issues such as how to distinguish between different types of features, how best to integrate information across modalities, which learning pipeline and which machine learning models to choose, and what data representations and algorithms should we take into account in order to realize context-aware interpretations of individual human behavior?

We show that emotions are important modifiers of human behavior, serving to enrich the response patterns, allowing more predictable and contextualized decisions. Part of the importance of emotions also comes from the fact that humans are adept at recognizing emotional influence by others and it forms the core of a social communication. Our study confirms that social interaction between humans cannot ignore the affective dimension derived from communication patterns.

Specifically, we test our machine learning approach on *affective states recognition*, such as happiness and daily stress recognition, and show that it is practically applicable to create intrinsically data driven models of individual human behavior based on communication patterns extracted from telecom metadata.

4.3 Supervised Learning of Large Scale Group Behavior

Behavior recognition literature mostly focuses on network dynamics or on simple actions, performed by a single actor. A broad class of actions, however, are social in nature, and require either the detailed analysis of multiple actors performing in tandem, or even when the distinction of very fine personal behavior that can not be tracked individually based on legal limitations, privacy of data concerns, or semantically uninterpretable, or technically too costly to compute. Social signal processing arose from the need of intelligent systems interacting with humans to interpret and reproduce social signals, and to increase the sensitivity of the human behavior understanding performed by computer. Group social signals are characterized by what "directly or indirectly provides information about social facts", such as regularity and diversity of group behavior, group social interactions, social emotions, social attitudes, social relations, social identities.

Our results prove that human dynamics, extracted from aggregated and anonymized mobile phone data, are good proxies for modelling energy consumption and crime hotspots prediction. Multimodal data adds the value to the human dynamics models that we find to perform the best. We show that large scale group behavior predictive characteristics could be learned from data in human interpretable way without using autoencoding techniques or neural feature representations and are good in implementing computational approaches to deal with a well-known social problems such as crime and energy efficiency, specifically, our approach could help to optimize the economy of energy producers/distributors value chain, also helping to meet peak electrical energy demand.

We show that supervised learning of large scale group behavior is feasible from telecom metadata which is widely available and cheap to process in the most of urban areas with existing GSM network coverage (2G, 3G and 4G as well).

4.4 Challenges

We found that the state-of-the-art research and industrial applications of human behavior understanding from human communications data is facing the following challenges.

1. *Multimodal nature of human behavior understanding* on individual and group level is the most logical conclusion from the existing research and the challenge by itself, because it requires collecting, processing, and learning from unstructured and semi-structured noisy multimodal data. The new computation methods, inspired by artificial intelligence and robotics, are to come in the near future. Learning of the optimal compactified feature representation is another part of this challenge. Existing deep learning methods provide a lot of opportunities to solve it, but are still uninterpretable by humans, thus could not be a part of decision making process or policy making.

2. *Adversarial nature of human behavior*

Researchers found that humans could change their actions based on previous known action of the adversary person, like in game theory, also there could be misleading purposefully deceptive data, called adversarial examples. This could lead to false discovery of data driven human dynamics. This problem is studied in many other close to computer science disciplines, such as game theory, control theory, operations research, information theory, simulation-based optimization, multi-agent systems, statistics, and genetic algorithms. Thus, the techniques, such as adversarial planning, multi-agent pathfinding, heuristic search for combat, opponent modelling and exploitation plan, goal recognition, adversarial build-order optimization, game balancing, state and action abstraction, including dealing with imperfect information and learning states evaluations and playing policies, most likely will be studied in the context of human dynamics on individual and group level.

Reinforcement learning open an efficient way to approach these set of problems, specifically, under stochastic, partially observable and adversarial conditions.

This is a type of machine learning of how software agents take actions in the modelled environment in order to maximize some notion of cumulative reward (policy). The possible research topics will include adaptive methods working with large-scale empirical evaluations, fewer or zero parameters under a large number of conditions, addressing the exploration problem in large Markov decision processes, learning and acting under partial information like predictive state representation, modular and hierarchical reinforcement learning, improving existing value-function and policy search methods, algorithms that work well with large action spaces, transfer learning. Multi-agent or distributed reinforcement learning will be the main interest for the group level human behavior research.

Second technique is called *inverse reinforcement learning*, this is a type of learning when no reward function is given. Instead, one tries to extract the reward function given an observed behavior from an expert. The idea is to mimic the observed behavior which is often optimal or close to optimal.

Another emerging technique of machine-learning to tackle this class of problems is a technique known as *generative adversarial networks*. Specifically, this is an algorithmic way of estimating generative models via an adversarial process, in which two models are trained simultaneously – a generative model, that captures the data distribution, and a discriminative model, that estimates the probability that a sample came from the training data rather than the first one. The training procedure for generative model is to maximize the probability of discriminative model making a mistake, corresponding to a minimax two-player game from game theory.

3. *Partially observable data*

The human behavior metrics are often stochastic and noisy. We apply state-of-the-art signal processing, data aggregation and intelligent optimization techniques to solve these problems. But the observations of human behavior typically involve the scalar immediate reward associated with the last transition. In many works, the agent is also assumed to observe the current state, which in practice is not true, and assuming full observability we fail to predict a subset of human behavior. In such cases we need to deal with partial observability. Sometimes the set of actions available to the agent is logically restricted, thus could be filtered (the speed of changing locations recorded as the base stations change on a telecom network for a pedestrian). Given that a model of conditions, or the environment, is known, but an analytic solution is not available, or only a simulation model of the environment could be given, – the only way to collect information about such an environment is by interacting with it. The first two of these problems could be considered planning problems (since some form of the model is available), while the last one could be considered as a genuine learning problem. However, under a reinforcement learning methodology both planning problems would be converted to machine learning problems.

4. *Big Data Privacy* is a new challenge inspired by European legislation on personal privacy. The technical solutions include homomorphic encryption, verifiable computation, and multi-party computation that can be used to achieve the goals of data privacy, but still are lacking standartization, speed and ease of computation and wide recognition among human behavior researchers.

Big data is analysed for bits of knowledge that leads to better decisions, and better strategic moves, thus empowering government and businesses. Yet only a small fraction of data is actually analysed. Privacy challenge in big data could be solved by the trade-off between identifying privacy requirements and defining sufficiency of the existing privacy-preserving techniques for specific type of data processing. Privacy challenges in each phase of big data life cycle are presented along with the advantages and disadvantages of existing privacy-preserving technologies in the context of concrete machine learning and analytical applications.

Now concepts of identity based anonymization and differential privacy focus on scalable anonymization methods within the MapReduce framework – the ability to be easily scaled up by horizontal increasing the number of mappers and reducers as a data preprocessing pipeline. In practice, the quantity of data is big but the quality is low, the pattern could not exist or there could be more random data than the searched signal, also the data is heterogeneous, as structured, semi structured, and unstructured. This poses new privacy challenges and open research issues.

We believe that cryptographic and semi-cryptographic techniques of making the data private are in the context of cloud computing, heterogeneous computing highlight the differences in performance cost associated with each. Thus we thing that it should be a *part of new intrinsic algorithms of machine learning*, but not the data preprocessing as is popular nowadays. So, new methods for privacy preserving data mining should be studied and implemented in future; as such, there exists a huge scope for further research in privacy preserving methods in supervised learning of human behavior from big data.

Ethics of automated human behavior understanding could be considered as part of personal privacy or as a separate ethical and regulatory problem. Humans or machines are controlling human behavior? What if one group of people could accurately predict human behavior of others? What are the socially oriented value systems?

5. *Reverse engineering of the new kind of science* – finding the possible simple programs. or algorithms, that are in the basement of complex behavior of individuals and the crowd is the major challenge for the future research in human behavior understanding. As a young field of AI-enabled new kind of science has the potential to take some novel approach in understanding the deep causal relations of human behavior which could be represented in algorithmic form.

Appendix A

Gradient Boosting Implementation for Stress Recognition Problem

Algorithm 1: Gradient Boosting Implementation for Stress Recognition Problem

Input: $\vec{X} = \vec{x}_{i}{}_{n=1}^{N}$ such that $\vec{x}1SDrfFps\mathbb{H}$;
$\vec{Y} = \vec{y}_{1}{}_{n=1}^{N}$ such that $\vec{y} \in \mathbb{F}_2$
Output: $\hat{f}(\vec{x})$ such that $\hat{f}(\vec{x}) = \arg\min_{\rho} \sum_{i=1}^{N} \Psi(y_i, \rho)$.

begin
    $\Psi \longleftarrow$ Bernoulli
    $T \longleftarrow$ number of trees
    $K \longleftarrow$ terminal nodes limit
    $p \longleftarrow$ subsampling rate
    $\hat{f}(\vec{x}) \longleftarrow \emptyset$
    for $i \in T$ **do**
        Compute gradient: $z_i = -\frac{\partial}{\partial f(\vec{x}_i)} \Psi(y_i, f(\vec{x}_i))\Big|_{f(\vec{x}_i)=\hat{f}(\vec{x}_i)}$
        Select $p \times N$ observations from the feature space
        Fit the tree, limited to K terminal nodes
        Compute the optimal terminal node:
        $\rho_k = \arg\min_{\rho} \sum_{\vec{x}_i \in S_k} \Psi(y_i, \hat{f}(\vec{x}_i) + \rho)$
        for $j \in length(\vec{x})$ **do**
            Update $\hat{f}(\vec{x})$, such that
            $$\hat{f}(\vec{x}) \leftarrow \hat{f}(\vec{x}) + \lambda \rho_{j(\vec{x})}$$
        end
    end
    return $\hat{f}(\vec{x})$
end

Appendix B

Additional Visualizations for Energy Consumption Prediction Use Cases

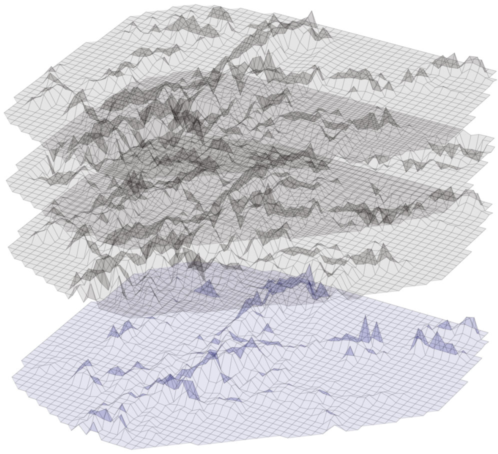

Figure 10: Spatially Mapped Max Energy Consumption vs Telecom Patterns for a Weekday on a SET grid

The SET customers are spatially aggregated into the grid squares and the energy consumption is uniformly divided among the customers, hiding their different type (e.g. houses, condominiums, business activities, industries). The following visualization is an example of spatially mapped max energy consumption vs Telecom network patterns for a randomly sampled weekday.

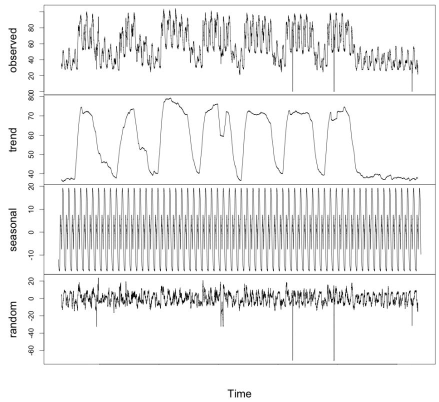

Figure 11: Energy Consumption Time Series Decomposition for a Residential Area

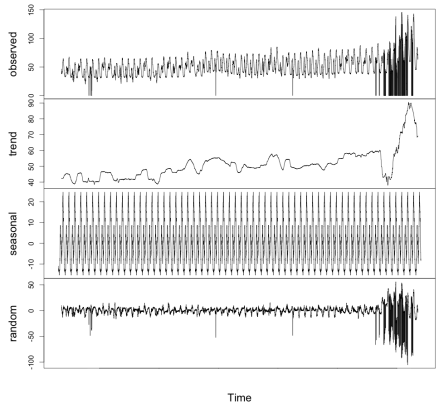

Figure 12: Energy Consumption Time Series Decomposition for a Touristic Area

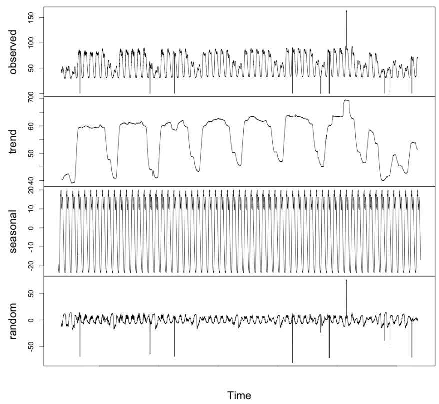

Figure 13: Energy Consumption Time Series Decomposition for a City Center and Industrial Areas

In Figure 11, Figure 12 and Figure 13 the x-axis shows the temporally sampled electric current flowing through the distribution lines recorded every 10 minutes. The y-axis reports the electric current passing through a given power line measured in Ampere at a given time. Monday is the first day of the week.

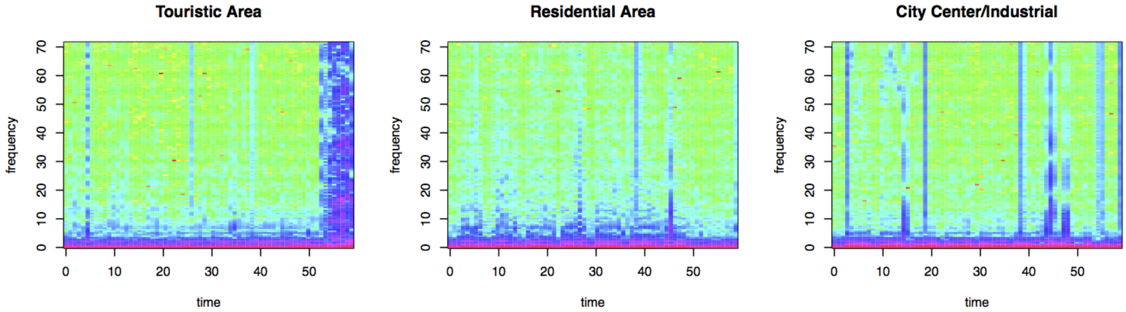

Figure 14: Spectral Characteristics of Typical Energy Consumption Response Variables

In Figure 14 the units reported in the temporal x-axis are days (the temporal scaling was done starting from the 100 milliseconds initial resolution), while the vertical axis represents frequency. The amplitude of a particular frequency at a given time is given by the intensity and the color of each point in the plot.

Table B.1: Mean Daily Energy Consumption Predictor Variables

Rank	Feature	Decrease in MSE	Decrease in Node Impurity
1	Number of consumers per grid powerline	19.64	69109.19
2	Variance of real numbers part of Fourier transform of area codes	4.43	6534.18
3	Variance of real numbers part of Fourier transform of outgoing sms activity	3.89	6811.37
4	Variance of calling direction area codes	3.56	7964.44
5	Variance of entropy of outgoing call activity in time domain	3.47	11568.62
6	Entropy of first harmonic of outgoing calls	3.47	3671.82
7	Entropy of internet activity summed in time domain	3.21	2750.14
8	Kurtosis of entropy distribution of outgoing calls	3.14	1472.87
9	Variance of skewness of temporal distribution of outgoing calls	3.10	2111.95
10	Entropy of fundamental frequency (first harmonic) of internet activity	3.03	1668.04
11	Standard deviation of frequencies distribution skewness of incoming sms	3.03	2848.24
12	Entrophy of sum in time domain of outgoing calls	2.98	3651.77
13	Variance of the kurtosis of outgpoung calls in time domain	2.98	2136.39
14	Kurtosis of time entropy of mobile internet activity	2.96	4409.30
15	Median of outgoing calls variance in frequiency domain	2.92	1317.35
16	Sum of 4 harmonic of incoming calls	2.88	5035.26
17	Sum of outgoing calls skewness in frequency domain	2.88	825.13
18	Median of outgoing sms temporal distribution kurtosis	2.88	1260.23
19	Kurtosis of outgoing calls 32 harmonic	2.81	947.15
20	Median of 11 harmonic of internet activity	2.72	185.14
21	Variance of 29 harmonic of outgoing sms	2.71	1243.93
22	Variance of outgoing sms fundamental frequnecy	2.69	5418.83
23	Sum of intertemporal mean of outgoing sms	2.67	182.93
24	Sum of calling direction area codes	2.63	4075.28
25	Kurtosis of skewness of internet activity frequencies	2.62	2290.58
26	Median of temporal distribution kurtosis of internet activity	2.61	1510.62
27	Sum of 5 harmonic of outgoing calls	2.60	4752.84
28	Sum of incoming calls tetmporal entropy	2.59	2494.76
29	Variance of internet activity temporal entropy	2.58	3569.25
30	Sum of 7 harmonic of outgoing calls	2.58	201.58
31	Skewness of 13 harmonic of incoming calls	2.55	751.57
32	Sum of 11 harmonic of incoming calls	2.55	1431.18

Table B.2: Peak Daily Energy Consumption Predictor Variables

Rank	Feature	Decrease in MSE	Decrease in Node Impurity
1	Number of consumers per grid powerline	21.22	132035.53
2	Entropy of temporal sum of outgoung calls	5.86	24013.99
3	Kurtosis of temporal entropy of internet activity	4.33	15479.71
4	Entropy outgoing calls fundamental frequnecy	4.10	14650.13
5	Sum of 4 harmonic of incoming calls	3.96	11449.25
6	Skewness of 5 harmonic of incoming calls	3.66	9595.91
7	Skewness of temporal entropy of internet activity	3.51	9498.52
8	Sum of spectral distribution skewness of outgoing calls	3.47	2108.19
9	Sum of temporal distribution kurtosis of internet activity	3.35	2267.17
10	Spatial variance of spectral variance of incoming sms	3.30	3971.29
11	Sum of 5 harmonic of incoming calls	3.27	9468.86
12	Sum of calling direction area codes	3.25	6815.90
13	Spectral varince of outgoing sms activity total in time	3.20	13899.48
14	Spatial skewness spectral skewness distribution of internet activity	3.19	3512.42
15	Kurtosis of temporal mean distribition of outgoing sms	3.17	27704.10
16	Spectral varince of temporal median distribition of outgoing sms	3.02	15366.03
17	Spatial median of incoming temporal sum	2.99	2642.34
18	Spectral varince of outgoing sms fundamental frequency	2.96	11624.28
19	Skewness of incoming calls temporal entropy	2.92	4174.09
20	Sum of outgoing calls 5 harmonic	2.89	10251.36
21	Variance of internet activity temporal entropy	2.86	5382.14
22	Median of calling direction area codes	2.85	3591.92
23	Sum of outgoing calls 4 harmonic	2.84	4440.13
24	Median of incoming calls 16 harmonic	2.74	943.98
25	Spatial standard deviation of outgoing call temporal skewness	2.73	2502.05
26	Spatial standard deviation of outgoing call temporal kurtosis	2.72	2888.08
27	Skewness of incoming calls 4 harmonic	2.71	8041.38
28	Spectral distribution skewness of mean internet activity	2.69	1759.88
29	Spatial standard deviation of temporal internet activity entropy	2.68	4233.89
30	Variance of outgoing calls temporal distribution kurtosis	2.67	3283.40
31	Kurtosis of spectral distribution skewness.	2.67	5160.36
32	Sum of internet activity 4 harmonic	2.65	1695.49

Bibliography

[app, 2013a] (Access date: 12-JUN-2013a). Funf open sensing framework. `http://www.funf.org`.

[app, 2013b] (Access date: 12-JUN-2013b). Track your happiness web application. `http://www.trackyourhappiness.org`.

[Aharony et al., 2011] Aharony, N., Pan, W., Ip, C., Khayal, I., and Pentland, A. (2011). Social fmri: Investigating and shaping social mechanisms in the real world. *Pervasive and Mobile Computing*, 7(6):643–659.

[Allcott and Mullainathan, 2010] Allcott, H. and Mullainathan, S. (2010). Behavior and energy policy. *Science*, 327(5970):1204–1205.

[Bauer and Lukowicz, 2012] Bauer, G. and Lukowicz, P. (2012). Can smartphones detect stress-related changes in the behaviour of individuals? In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*, pages 423–426. IEEE.

[Biau, 2012] Biau, G. (2012). Analysis of a random forests model. *The Journal of Machine Learning Research*, 98888(1):1063–1095.

[Boggs, 1965] Boggs, S. L. (1965). Urban crime patterns. *American Sociological Review*, 30(6):pp. 899–908.

[Bogomolov et al., 2013a] Bogomolov, A., Lepri, B., and Pianesi, F. (2013a). Happiness recognition from mobile phone data. *Proceedings of the 2013 International Conference on Social Computing (SocialCom 2013)*, pages 790–795.

[Bogomolov et al., 2013b] Bogomolov, A., Lepri, B., and Pianesi, F. (2013b). Happiness recognition from mobile phone data. In *SocialCom 2013*, pages 790–795.

[Bolger and Schilling, 1991] Bolger, N. and Schilling, E. A. (1991). Personality and the problems of everyday life: The role of neuroticism in exposure and reactivity to daily stressors. *Journal of Personality*, 59(3):355–386.

[Bollen et al., 2011] Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8.

[Box and Cox, 1964] Box, G. E. P. and Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2):211–252.

[Braithwaite, 1989] Braithwaite, J. (1989). *Crime, Shame and Reintegration*. Ambridge: Cambridge University Press.

[Brantingham and Brantingham, 1999] Brantingham, P. L. and Brantingham, P. J. (1999). A theoretical model of crime hot spot generation. *Studies on Crime & Crime Prevention*.

[Breiman, 1996] Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123–140.

[Breiman, 1999] Breiman, L. (1999). Random forests-random features. Technical report, Technical Report 567, Department of Statistics, UC Berkeley, 1999. 31.

[Breiman, 2001] Breiman, L. (2001). Random forests. *Mach. Learn.*, 45(1):5–32.

[Buczak and Gifford, 2010] Buczak, A. L. and Gifford, C. M. (2010). Fuzzy association rule mining for community crime pattern discovery. In *ACM SIGKDD Workshop on Intelligence and Security Informatics*, page 2. ACM.

[Buhmann and Buhmann, 2003] Buhmann, M. D. and Buhmann, M. D. (2003). *Radial Basis Functions*. Cambridge University Press, New York, NY, USA.

[Campbell et al., 1976] Campbell, A., Converse, P. E., and Rodgers, W. L. (1976). *The quality of American life : perceptions, evaluations, and satisfactions / Angus Campbell, Philip E. Converse, Willard L. Rodgers*. Russell Sage Foundation New York.

[Caplan and Kennedy, 2010] Caplan, J. and Kennedy, L. (2010). *Risk terrain modeling manual: Theoretical framework and technical steps of spatial risk assessment for crime analysis*. Rutgers Center on Public Security.

[Carver and Scheier, 1990] Carver, C. S. and Scheier, M. F. (1990). Origins and Functions of Positive and Negative Affect: A Control-Process View. *Psychological Review*, 97(1):19–35.

[Chainey et al., 2008] Chainey, S., Tompson, L., and Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal*, 21:4–28.

[Chittaranjan et al., 2013] Chittaranjan, G., Blom, J., and Gatica-Perez, D. (2013). Mining large-scale smartphone data for personality studies. *Personal Ubiquitous Comput.*, 17(3):433–450.

[Cohen, 1960] Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46.

[Cohen et al., 1997] Cohen, S., C., K. R., and Gordon, L. U. (1997). *Measuring stress: A guide for health and social scientists*. Oxford University Press, USA.

[Cohn, 2007] Cohn, J. F. (2007). Foundations of human computing: facial expression and emotion. In *Proceedings of the ICMI 2006 and IJCAI 2007 international conference on Artifical intelligence for human computing*, ICMI'06/IJCAI'07, pages 1–16, Berlin, Heidelberg. Springer-Verlag.

[Costa and McCrae, 1980] Costa, P. T. and McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology*, 38:668–678.

[Cullen and Levitt, 2009] Cullen, J. and Levitt, S. (2009). Crime, urban flight, and the consequences for the cities. *Review of Economics and Statistics*, (81):159–169.

[de Oliveira et al., 2011] de Oliveira, R., Karatzoglou, A., Concejero Cerezo, P., Armenta Lopez de Vicuña, A., and Oliver, N. (2011). Towards a psychographic user model from mobile phone usage. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 2191–2196. ACM.

[DeNeve and Cooper, 1998] DeNeve, K. and Cooper, H. (1998). The happy personality: a meta-analysis of 137 personality traits and subjective well-being. *Psychol Bull*, 124(2):197–229.

[Denissen et al., 2008] Denissen, J., Butalid, L., Penke, L., and Van Aken, M. (2008). The effects of weather on daily mood: A multilevel approach. *Emotion Researcher*, 8(5):662–667.

[Denning, 2003] Denning, P. J. (2003). Great principles of computing. *Communications of the ACM*, 46(11):15–20.

[Diener, 1994] Diener, E. (1994). Assessing subjective well-being: Progress and opportunities. *Social Indicators Research*, 31(2):103–157.

[Diener and Seligman, 2002] Diener, E. and Seligman, M. E. (2002). Very happy people. *Psychological Science*, 13(1):81–84.

[Dong et al., 2011a] Dong, W., Lepri, B., and Pentland, A. (2011a). Modeling the coevolution of behaviors and social relationships using mobile phone data. In *MUM*, pages 134–143.

[Dong et al., 2011b] Dong, W., Lepri, B., and Pentland, A. (2011b). Modeling the coevolution of behaviors and social relationships using mobile phone data. In *MUM 2011*.

[Duggan et al., 1995] Duggan, C., Sham, P., Lee, A., Minne, C., and Murray, R. (1995). Neuroticism: a vulnerability marker for depression evidence from a family study. *Journal of Affective Disorders*, 35(3):139 – 143.

[Eagle et al., 2010] Eagle, N., Macy, M., and Claxton, R. (2010). Network diversity and economic development. *Science*, 328(5981):1029–1031.

[Eagle and Pentland, 2006] Eagle, N. and Pentland, A. (2006). Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4):255–268.

[Eagle et al., 2009a] Eagle, N., Pentland, A. S., and Lazer, D. (2009a). Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, 106(36):15274–15278.

[Eagle et al., 2009b] Eagle, N., Pentland, A. S., and Lazer, D. (2009b). Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, 106(36):15274–15278.

[Eck et al., 2005] Eck, J., Chainey, S., Cameron, J., and Wilson, R. (2005). Mapping crime: understanding hotspots. *National Institute of Justice: Washington DC.*

[Ehrlich, 1975] Ehrlich, I. (1975). On the relation between education and crime.

[Ellis et al., 2009] Ellis, L., Beaver, K., and Wright, J. (2009). *Handbook of crime correlates*. Academic Press.

[Faust et al., 1974] Faust, V., Weidmann, M., and Wehner, W. (1974). The influence of meteorological factors on children and youths: A 10% random selection of 16,000 pupils and apprentices of basle city (switzerland). *Acta Paedopsychiatrica: International Journal of Child & Adolescent Psychiatry*.

[Freeman, 1999] Freeman, R. B. (1999). The economics of crime. *Handbook of labor economics*, 3:3529–3571.

[Freund and Schapire, 1997] Freund, Y. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119 – 139.

[Frias-Martinez et al., 2011] Frias-Martinez, E., Williamson, G., and Frias-Martinez, V. (2011). An agent-based model of epidemic spread using human mobility and social network information. In *Social Computing (SocialCom), 2011 International Conference on*, pages 57–64. IEEE.

[Friedman, 2001] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, pages 1189–1232.

[Friedman, 2002] Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4):367–378.

[Garcia and Herrera, 2008] Garcia, S. and Herrera, F. (2008). An Extension on "Statistical Comparisons of Classifiers over Multiple Data Sets" for all Pairwise Comparisons. *Journal of Machine Learning Research*, 9:2677–2694.

[George et al., 2014] George, G., Haas, M. R., and Pentland, A. (2014). Big data and management. *Academy of Management Journal*, 57(2):321–326.

[Giakoumis et al., 2012] Giakoumis, D., Drosou, A., Cipresso, P., Tzovaras, D., Hassapis, G., Gaggioli, A., and Riva, G. (2012). Real-time monitoring of behavioural parameters related to psychological stress. *Studies in health technology and informatics*, 181:287.

[Gonzalez et al., 2008] Gonzalez, M., Hidalgo, C., and Barabasi, L. (2008). Understanding individual mobility patterns. *Nature*, 453(7196):779–782.

[Goode, 1960] Goode, W. J. (1960). A theory of role strain. *American Sociological Review*, 25(4):pp. 483–496.

[Gray, 1991] Gray, J. A. (1991). Neural systems, emotion, and personality. In *Neurobiology of learning, emotion, and affect / editor, John Madden IV*, pages 273–306. Raven Press New York.

[Gunthert, 1999] Gunthert, Kathleen Cimbolic; Cohen, L. H. A. S. (1999). The role of neuroticism in daily stress and coping. *Journal of Personality and Social Psychology*, 77:1087–1100.

[Hand and Till, 2001] Hand, D. and Till, R. (2001). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning*, 45:171–186.

[Hardt and Gerbershagen, 1999] Hardt, J. and Gerbershagen, H. (1999). No changes in mood with the seasons: observations in 3000 chronic pain patients. *Acta Psychiatrica Scandinavica*, 100(4):288–294.

[Healey and Picard, 2005] Healey, J. and Picard, R. (2005). Detecting stress during real-world driving tasks using physiological sensors. *Intelligent Transportation Systems, IEEE Transactions on*, 6(2):156–166.

[Hernandez et al., 2012] Hernandez, J., Hoque, M. E., and Picard, R. W. (2012). Mood meter: large-scale and long-term smile monitoring system. In *ACM SIGGRAPH 2012 Emerging Technologies*, SIGGRAPH '12, pages 15:1–15:1, New York, NY, USA. ACM.

[Howarth and Hoffman, 1984] Howarth, E. and Hoffman, M. S. (1984). A multidimensional approach to the relationship between mood and weather. *British Journal of Psychology*, 75(1):15–23.

[John and Srivastava, 1999] John, O. P. and Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In Pervin, L. A. and John, O. P., editors, *Handbook of Personality: Theory and Research*, pages 102–138. Guilford Press, New York, second edition.

[Jovanov et al., 2003] Jovanov, E., O'Donnell Lords, A., Raskovic, D., Cox, P., Adhami, R., and Andrasik, F. (2003). Stress monitoring using a distributed wireless intelligent sensor system. *Engineering in Medicine and Biology Magazine, IEEE*, 22(3):49–55.

[Keller et al., 2005] Keller, M. C., Fredrickson, B. L., Ybarra, O., Cote, S., Johnson, K., Mikels, J., Conway, A., and Wager, T. (2005). A Warm Heart and a Clear Head: The Contingent Effects of Weather on Mood and Cognition. *Psychological Science*, 16(9):724–731.

[Kolter and Ferreira Jr, 2011] Kolter, J. Z. and Ferreira Jr, J. (2011). A large-scale study on predicting and contextualizing building energy usage.

[Krackhardt, 1992] Krackhardt, D. (1992). The strength of strong ties: The importance of philos in organizations. pages 216–239.

[Krumme et al., 2013] Krumme, C., Llorente, A., Cebrian, M., Pentland, A., and Moro, E. (2013). The predictability of consumer visitation patterns. *Scientific Reports*, (1645).

[Krzanowski, 2005] Krzanowski, W. J. (2005). Misclassification rates. In *Encyclopedia of Statistics in Behavioral Science*.

[Lane et al., 2012] Lane, N., Mohammod, M., Lin, M., Yang, X., Lu, H., Ali, S., Doryab, A., Berke, E., Choudhury, T., and Campbell, A. (2012). Bewell: A smartphone application to monitor, model and promote wellbeing. IEEE.

[Lane et al., 2010a] Lane, N. D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., and Campbell, A. T. (2010a). A survey of mobile phone sensing. *Comm. Mag.*, 48(9):140–150.

[Lane et al., 2010b] Lane, N. D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., and Campbell, A. T. (2010b). A survey of mobile phone sensing. *IEEE Communications Magazine*, 48(9):140–150.

[Lathia et al., 2013] Lathia, N., Rachuri, K. K., Mascolo, C., and Rentfrow, P. J. (2013). Contextual dissonance: Design bias in sensor-based experience sampling methods. *Proceedings of Ubicomp 2013*, pages 183–192.

[Laurila et al., 2013] Laurila, J., Gatica-Perez, D., Aad, I., Blom, J., Bornet, O., Do, T., Dousse, O., Eberle, J., and Miettinen, M. (2013). From big smartphone data to worldwide research: The mobile data challenge. *Pervasive and Mobile Computing*, 9:752–771.

[Lee and Ishii-Kuntz, 1987] Lee, G. R. and Ishii-Kuntz, M. (1987). Social Interaction, Loneliness, and Emotional Well-Being among the Elderly. *Research on Aging*, 9(4):459–482.

[LiKamWa et al., 2013] LiKamWa, R., Liu, Y., Lane, N. D., and Zhong, L. (2013). Moodscope: building a mood sensor from smartphone usage patterns. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, MobiSys '13, pages 389–402, New York, NY, USA. ACM.

[Lu et al., 2012] Lu, H., Frauendorfer, D., Rabbi, M., Mast, M. S., Chittaranjan, G. T., Campbell, A. T., Gatica-Perez, D., and Choudhury, T. (2012). Stresssense: detecting stress in unconstrained acoustic environments using smartphones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, pages 351–360, New York, NY, USA. ACM.

[Lyubomirsky et al., 2005] Lyubomirsky, S., King, L., and Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success? *Psychological Bulletin*, 131(6):803–855.

[Lyubomirsky et al., 2006] Lyubomirsky, S., Tkach, C., and Dimatteo, R. M. (2006). What are the differences between happiness and self-esteem? *Social Indicators Research*, 78:363–404.

[Ma et al., 2012] Ma, Y., Xu, B., Bai, Y., Sun, G., and Zhu, R. (2012). Daily mood assessment based on mobile phone sensing. In *Wearable and Implantable Body Sensor Networks (BSN), 2012 Ninth International Conference on*, pages 142–147.

[Madan et al., 2010] Madan, A., Cebrián, M., Lazer, D., and Pentland, A. (2010). Social sensing for epidemiological behavior change. In *UbiComp*, pages 291–300.

[Madan et al., 2012] Madan, A., Cebrián, M., Moturu, S. T., Farrahi, K., and Pentland, A. (2012). Sensing the "health state" of a community. *IEEE Pervasive Computing*, 11(4):36–45.

[Majoe et al., 2010] Majoe, D., Bonhof, P., Kaegi-Trachsel, T., Gutknecht, J., and Widmer, L. (2010). Stress and sleep quality estimation from a smart wearable sensor. In *Pervasive Computing and Applications (ICPCA), 2010 5th International Conference on*, pages 14–19.

[McCrae and Costa, 1991] McCrae, R. R. and Costa, P. T. (1991). Adding liebe und arbeit: The full five-factor model and well-being. *Personality and Social Psychology Bulletin*, 17(2):227–232.

[Mehlum et al., 2005] Mehlum, H., Moene, K., and Torvik, R. (2005). Crime induced poverty traps. *Journal of Development Economics*, (77):325–340.

[Miller, 1955] Miller, G. (1955). Note on the bias of information estimates. *Information theory in psychology: Problems and methods*, 2(95):100.

[Miller, 2012] Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science*, 7(3):221–237.

[Mohler et al., 2011] Mohler, G., Short, M., Brantingham, P., Schoenberg, F., and Tita, G. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, (106):100–108.

[Montjoye et al., 2013] Montjoye, Y.-A., Quoidbach, J., Robic, F., and Pentland, A. (2013). Predicting personality using novel mobile phone-based metrics. In Greenberg, A., Kennedy, W., and Bos, N., editors, *Social Computing, Behavioral-Cultural Modeling and Prediction*, volume 7812 of *Lecture Notes in Computer Science*, pages 48–55. Springer Berlin Heidelberg.

[Moturu et al., 2011] Moturu, S. T., Khayal, I., Aharony, N., Pan, W., and Pentland, A. (2011). Using social sensing to understand the links between sleep, mood, and sociability. In *SocialCom/PASSAT*, pages 208–214.

[Muaremi et al., 2012] Muaremi, A., Arnrich, B., and Tröster, G. (2012). A survey on measuring happiness with smart phones. In *6th International Workshop on Ubiquitous Health and Wellness (UbiHealth 2012)*.

[Myers and Diener, 1995] Myers, D. G. and Diener, E. (1995). Who is happy? *Psychological Science*, 6(1):10–19.

[Pantic, 2009] Pantic, M. (2009). Machine analysis of facial behaviour: Naturalistic and dynamic behaviour. *Philosophical Transactions of Royal Society B*, 364:3505–3513.

[Patterson, 1991] Patterson, E. B. (1991). Poverty, income inequality, and community crime rates. *Criminology*, 29(4):755–776.

[Plarre et al., 2011] Plarre, K., Raij, A., Hossain, S., Ali, A., Nakajima, M., Al'absi, M., Ertin, E., Kamarck, T., Kumar, S., Scott, M., Siewiorek, D., Smailagic, A., and Wittmers, L. (2011). Continuous inference of psychological stress from sensory measurements collected in the natural environment. In *Information Processing in Sensor Networks (IPSN), 2011 10th International Conference on*, pages 97–108.

[Quercia, 2013] Quercia, D. (2013). Dont worry, be happy: The geography of happiness on facebook. `http://profzero.org/publications/websci13_dontworry.pdf`.

[Quercia et al., 2012] Quercia, D., Ellis, J., Capra, L., and Crowcroft, J. (2012). Tracking "gross community happiness" from tweets. In *CSCW*, pages 965–968.

[Rabbi et al., 2011] Rabbi, M., Ali, S., Choudhury, T., and Berke, E. (2011). Passive and in-situ assessment of mental and physical well-being using mobile sensors. In *Proceedings of the 13th international conference on Ubiquitous computing*, UbiComp '11, pages 385–394, New York, NY, USA. ACM.

[Rachuri et al., 2010] Rachuri, K. K., Musolesi, M., Mascolo, C., Rentfrow, P. J., Longworth, C., and Aucinas, A. (2010). Emotionsense: a mobile phones based adaptive platform for experimental social psychology research. In *UbiComp*, pages 281–290.

[Raphael and Winter-Ebmer, 2001] Raphael, S. and Winter-Ebmer, R. (2001). Identifying the effect of unemployment on crime. *Journal of Law and Economics*, 44(1).

[Ratcliffe, 2006] Ratcliffe, J. H. (2006). A temporal constraint theory to explain opportunity-based spatial offending patterns. *Journal of Research in Crime and Delinquency*, 43(3):261–291.

[Robert LiKamWa, 2013] Robert LiKamWa, Yunxin Liu, N. D. L. L. Z. (2013). Moodscope: Building a mood sensor from smartphone usage patterns. `http://niclane.org/pubs/mobisys_moodscope.pdf`.

[Sanders and Brizzolara, 1982] Sanders, J. L. and Brizzolara, M. S. (1982). Relationships between weather and mood. *The Journal of General Psychology*, 107(1):155–156.

[Sanderson, 1999] Sanderson, C.A., C. N. (1999). A life task perspective on personality coherence: Stability versus change in tasks, goals, strategies, and outcomes. In *D. Cervone & Y. Shoda (Eds.), The Coherence of Personality: Social-cognitive Bases of Consistency, Variability, and Organization*, pages 372–392, New York, NY, USA. Guilford Press.

[Sano and Picard, 2013] Sano, A. and Picard, R. W. (2013). Stress recognition using wearable sensors and mobile phones. *Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 671–676.

[Scherer et al., 2002] Scherer, K. R., Grandjean, D., Johnstone, T., Klasmeyer, G., and Bänziger, T. (2002). Acoustic correlates of task load and stress. In *INTERSPEECH*.

[Schuller et al., 2005] Schuller, B., Villar, R., Rigoll, G., and Lang, M. (2005). Metaclassifiers in acoustic and linguistic feature fusion-based affect recognition. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, volume 1, pages 325–328.

[Shang et al., 2014] Shang, J., Zheng, Y., T. W., and Chang, E. (2014). Inferring gas consumption and pollution emission of vehicles throughout a city. In *KDD 2014*.

[Shannon, 1948] Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656.

[Short et al., 2008] Short, M. B., D'Orsogna, M. R., Pasour, V. B., Tita, G. E., Brantingham, P. J., Bertozzi, A. L., and Chayes, L. B. (2008). A statistical model of criminal behavior. *Mathematical Models and Methods in Applied Sciences*, 18(supp01):1249–1267.

[Sinatra and Szell, 2014] Sinatra, R. and Szell, M. (2014). Entropy and the predictability of online life. *Entropy*, (16):543–556.

[Singh et al., 2010] Singh, S. R., Murthy, H. A., and Gonsalves, T. A. (2010). Feature selection for text classification based on gini coefficient of inequality. *Journal of Machine Learning Research-Proceedings Track*, 10:76–85.

[Singh et al., 2013] Singh, V. K., Freeman, L., Lepri, B., and Pentland, A. S. (2013). Predicting spending behavior using socio-mobile features. In *Social Computing (SocialCom), 2013 International Conference on*, pages 174–179. IEEE.

[Smith-Clarke et al., 2014] Smith-Clarke, C., Mashhadi, A., and Capra, L. (2014). Poverty on the cheap: Estimating poverty maps using aggregated mobile communication networks. In *32nd ACM Conference on Human Factors in Computing Systems (CHI2014)*.

[Song et al., 2010] Song, C., Qu, Z., Blumm, N., and Barabasi, A. (2010). Limits of predictability in human mobility. *Science*, (327):1018–1021.

[Soto et al., 2011] Soto, V., Frias-Martinez, V., Virseda, J., and Frias-Martinez, E. (2011). Prediction of socioeconomic levels using cell phone records. In *UMAP 2011*, pages 377–388.

[Staiano et al., 2012a] Staiano, J., Lepri, B., Aharony, N., Pianesi, F., Sebe, N., and Pentland, A. (2012a). Friends don't lie - inferring personality traits from social network structure. In *ACM Ubicomp 2012*, pages 321–330.

[Staiano et al., 2012b] Staiano, J., Lepri, B., Aharony, N., Pianesi, F., Sebe, N., and Pentland, A. (2012b). Friends don't lie: inferring personality traits from social network structure. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 321–330. ACM.

[Stehman, 1997] Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, 62(1):77–89.

[Stopczynski et al., 2014] Stopczynski, A., Sekara, V., Sapiezynski, P., Cuttone, A., Madsen, M. M., Larsen, J. E., and Lehmann, S. (2014). Measuring large-scale social networks with high resolution. *PLoS ONE*, (9).

[Suls et al., 1998] Suls, J., Green, P., and Hillis, S. (1998). Emotional reactivity to everyday problems, affective inertia, and neuroticism. *Personality and Social Psychology Bulletin*, 24(2):127–136.

[Tarling and Morris, 2010] Tarling, R. and Morris, K. (2010). Reporting crime to the police. *British Journal of Criminology*, (50):474–479.

[Toole et al., 2011] Toole, J. L., Eagle, N., and Plotkin, J. B. (2011). Spatiotemporal correlations in criminal offense records. *ACM Trans. Intell. Syst. Technol.*, 2(4):38:1–38:18.

[Tsutsui, 2012] Tsutsui, Y. (2012). Weather and individual happiness. *Weather, Climate, and Society*, 5(1):70–82.

[Tumasjan et al., 2010] Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. *ICWSM*, 10:178–185.

[Tuv et al., 2009] Tuv, E., Borisov, A., Runger, G., and Torkkola, K. (2009). Feature selection with ensembles, artificial variables, and redundancy elimination. *The Journal of Machine Learning Research*, 10:1341–1366.

[Vapnik, 1995] Vapnik, V. N. (1995). *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA.

[Vitters, 2001] Vitters, J. (2001). Personality traits and subjective well-being: emotional stability, not extraversion, is probably the important predictor. *Personality and Individual Differences*, 31(6):903 – 914.

[Vollrath and Torgersen, 2000] Vollrath, M. and Torgersen, S. (2000). Personality types and coping. *Personal. Individ. Differ.*, 29:367–378.

[Wandishin and Mullen, 2009] Wandishin, M. S. and Mullen, S. J. (2009). *Multiclass ROC Analysis*. American Meteorological Society.

[Wang et al., 2013] Wang, T., Rudin, C., Wagner, D., and Sevieri, R. (2013). Learning to detect patterns of crime. In *Machine Learning and Knowledge Discovery in Databases*, pages 515–530. Springer.

[Wang et al., 2012] Wang, X., Gerber, M. S., and Brown, D. E. (2012). Automatic crime prediction using events extracted from twitter posts. In *Social Computing, Behavioral-Cultural Modeling and Prediction*, pages 231–238. Springer.

[Weinberg, 1954] Weinberg, S. K. (1954). Theories of criminality and problems of prediction. *The Journal of Criminal Law, Criminology, and Police Science*, 45(4):412–424.

[Weisburd, 2008] Weisburd, D. (2008). Place-based policing. *Ideas in American Policing*, 9:1–16.

[Weisburd and Green, 1994] Weisburd, D. and Green, L. (1994). Defining the street-level drug market.

[Wesolowski et al., 2012] Wesolowski, A., Eagle, N., Tatem, A., Smith, D., Noor, R., and Buckee, C. (2012). Quantifying the impact of human mobility on malaria. *Science*, 338(6104):267–270.

[Zajonc and McIntosh, 1992] Zajonc, R. B. and McIntosh, D. N. (1992). Emotions research: Some promising questions and some questionable promises. *Psychological Science*, pages 70–74.

[Zeng et al., 2009] Zeng, Z., Pantic, M., Roisman, G., and Huang, T. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):39–58.

[Zheng et al., 2014a] Zheng, Y., Capra, L., Wolfson, O., and Yang, H. (2014a). Urban computing: concepts, methodologies, and applications. *ACM Transaction on Intelligent Systems and Technology*.

[Zheng et al., 2014b] Zheng, Y., Liu, T., Wang, Y., Liu, Y., Zhu, Y., and Chang, E. (2014b). Diagnosing new york city's noises with ubiquitous data. In *ACM Ubicomp 2014*.