

Unsupervised Learning of True Ranking Estimators using the Belief Function Framework

Andrea Argentini Enrico Blanzieri

June 2011

Technical Report # DISI-11-480

This work has been presented at 11th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU) as poster

Unsupervised Learning of True Ranking Estimators using the Belief Function Framework

Andrea Argentini*, Enrico Blanzieri

Dipartimento di Ingegneria e Scienza dell'informazione
Università di Trento

{argentini,blanzier}@disi.unitn.it

Abstract

A variant of the ranking aggregation problem is considered in this work. The goal is to find an approximation of an unknown true ranking given a set of rankings. We devise a solution called Belief Ranking Estimator (BRE), based on the belief function framework that permits to represent beliefs on the correctness of the rankings position as well as uncertainty on the quality of the rankings from the subjective point of view of the expert. The results of a preliminary empirical comparison of BRE against baseline ranking estimators and state-of-the-art methods for ranking aggregation are shown and discussed.

1 Introduction

The ranking aggregation task originates from the necessity to aggregate different and possibly contradictory rankings of a finite set of items. Given several rankings as input, the goal is to find a new ranking that minimize the distance to each of the ranking in the input set [2]. The problem of ranking aggregation emerges when it is necessary to combine the opinion of experts with different background, such as the combination of movie preferences, or

*Corresponding Author

the combination of web engine results of a queries. Ranking aggregation solution takes also in consideration the uncertainty associated to the data when this is available, as in the context of sensor network data or tracking moving objects [6].

In our work we consider a variant of the problem of ranking aggregation, that is when the goal is to find a satisfying approximation of an unknown true ranking given a usually-low number of rankings that are assumed to be diverse-quality estimations of the true ranking. The main difference with respect to the usual ranking aggregation problems met in case of preferences or queries results combination is that we assume that the true ranking over the set of items do exists. We claim that this is the case when the rankings comes from bioinformatics rankers because of the underlying physical reality of the unknown biological phenomenon at hand. Moreover and with the same motivation we assume the rankings to be permutations of a big but finite set of known items. To treat this problem we have to consider data fusion aspects where rankings from different experts are combined to produce a new ranking and also the uncertainty related to the elements of the ranking. Finally, the evaluation of the solution methods has to take into account correlation and distance with respect to the true ranking.

In this paper we devise a solution to the ranking aggregation problem based on belief function theory. The belief function theory provides a robust framework for reasoning with imprecise and uncertainty data, allowing the modeling of subjective knowledge in a non Bayesian way. Belief function has been applied on machine learning problems such as clustering, classification [7] and data fusion under uncertainty. To the best of our knowledge, application of belief function on a ranking aggregation problem has not yet been proposed. Our solution, called Belief Ranking Estimator (BRE), estimates in an unsupervised way the true ranking given a set of estimating ranked permutations. Our approach models the uncertainty relying only on the ranking values, without using any other external information such as the confidence of the expert. We test the methods over synthetic data, comparing our method against some baseline and competitor methods and we provide also a discussion of the related work. The paper is organized as follows. In the following section we briefly present belief function framework. One section is devoted to the presentation of our method and another to the empirical results. Finally, we draw some conclusions.

2 The Belief Functions Theory

The theory of the belief functions, also known as Dempster-Shafer theory, is based on the pioneering work of Dempster [1] and Shafer [4]. More recent advances of this theory has been introduced in the *Transferable Belief Model* (TBM), proposed by Smets [5]. The three main ingredients of the framework are the *basic belief assignment* (bba), *belief* functions and *plausibility* functions. The *basic belief assignment* is the way to represent the believes held by an expert regarding the values of a variable in a finite domain. We define Θ , called the *frame of discernment*, as the set of propositions exclusive and exhaustive in a certain domain. Let 2^Θ denote the set of the subsets of Θ . A function $m : 2^\Theta \rightarrow [0, 1]$ is called *basic belief assignment* (bba) if it satisfies : $\sum_{A \subseteq \Theta} m(A) = 1$ In order to combine distinct sources m_1, \dots, m_n on Θ , the belief function provides several combination rules, such as conjunctive rules, disjunctive rules and caution rules [5]. One of the most used rules, is the conjunctive rules defined as: $m_{1 \odot 2}(A) = \sum_{B \cap C = A} m_1(B)m_2(C)$ $A \subseteq \Theta$. The conjunctive rule is justified when all the sources of belief are supposed to assert the truth and to be independent. For belief function theory is a generalization of probability in a Bayesian setting, homologous of the conditioning and marginalization rules are expressed in the framework. In order to make decision dealing with belief functions, in the TBM is used the pignistic transformation [5] which transforms a mass function into a probability function *Betp* defined as: $Betp(\omega) = \sum_{A \subseteq \Theta} \frac{m(A)}{1 - m(\emptyset)|A|} \quad \forall \omega \in \Theta$.

3 Method

3.1 Notation and Definition of the Problem

Let $X = \{x_1, \dots, x_n\}$ a set of elements to be ranked by an expert opinion. We denote as $\tau = (\tau(1), \dots, \tau(n))$ a ranking associated to X , where $\tau(i)$ is the rank associated to the element x_i . We suppose to have $\tau^{Rank} = \{\tau^{Rank}(1), \dots, \tau^{Rank}(n)\}$, the golden Truth Ranking on the element of X , and we denote as R_j the expert involved in the ranking, so for each expert we have a corresponding ranking $\tau^{R_j} = \{\tau^{R_j}(1), \dots, \tau^{R_j}(n)\}$. We suppose also that the most important element for a ranker R_j receives a ranking value near to n instead of 1. The problem in its general form is stated as follows. Given N rankings τ^{R_j} of length n of the n items $X = \{x_1, \dots, x_n\}$, namely

Algorithm 1 Belief Ranking Estimator

```
input  $I = \tau^1, \dots, \tau^N$  {a vector of N Rankings}  
input  $T$  {Numbers of iterations}  
   $k = 1$   
   $BE = \text{ComputeBelief\_From\_Rankings}(I)$   
  while  $k \neq T + 1$  do  
     $W = \text{ComputeWeight}(I)$   
     $BE = \text{ApplyWeight}(W, BE)$   
     $FinalRank_k = \text{Combination}(Bmod)$   
     $I[\text{pos}(\max(W))] = FinalRank_k$   
     $k++$   
  end while  
output  $FinalRank_k$ 
```

permutations, that estimate with unknown quality the unknown true ranking $\tau^{T_{rank}}$ find a ranking that estimates the true ranking.

3.2 Belief Ranking Estimator

The Belief Ranking Estimator (BRE) method is an unsupervised algorithm that iteratively learns an estimation of an unknown Truth Ranking, given a sample of rankings that are permutations of the items of the Truth Ranking assumed to be approximations of the Truth Ranking of unknown quality. As showed in Alg. 1 the input parameters are the rankings and the number of iterations. The algorithm is structured in the following parts: the mapping of the rankings into belief functions (ComputeBelief_From_Ranking), the weight computation from the Truth estimator (ComputeWeight), the application of the weight to current belief model of the rankers (ApplyWeight), finally the output ranking is produced by the combination of all the bba of the rankings. After the first iteration, the method replaces the ranking with the maximum weight, with the combined ranking produced in the previous iteration. At each iteration, this replacement drives the methods to combine rankings that are supposed to be better than the original worst ranking, and to produce as output a better estimator. With a number of iterations T , please note that the number of replacements is $T - 1$. A not weighted version of our method *BRE - NW*, is also evaluated in this work. *BRE - NW* combines the belief distribution of the rankers given in input, without the application of the weights. We refer as weighting schema to the *BRE - 1T* version.

BBA from the rankings: For our problem, we consider a simple frame of discernment $\Theta = \{P, \neg P\}$, where P , $\neg P$ are the hypothesis that a ranked element is in the right position or not respectively. Given a set of N rankings

$\tau^1, \dots, \tau^j, \dots, \tau^N$ of the same n elements, the bba of the j -th ranking on the i -th element is assigned as:

$$\begin{aligned} m_{ji}(P) &= \frac{\tau^j(i)}{n} \\ m_{ji}(\neg P) &= 0 \\ m_{ji}(\Theta) &= 1 - \frac{\tau^j(i)}{n} \end{aligned} \quad (1)$$

The bba definition reflects the fact that the high-ranking elements should have more belief to be in the right position from the point of view of the expert who expressed the ranking. Since the lack of external information about the correctness of the ranking we are not able to assert if an element is not in the right position ($\neg P$), the remain belief is assigned consequently to the uncertainty about the two possible hypothesis, namely to Θ .

Weight Computation: The weight of the rankings are computed as the normalized values of the Spearman footrule distance [2] between the rankings and the mean ranking as in the following formula: $W_j = \frac{D(\tau^j, M)}{\frac{1}{2}n^2} \quad \forall j \in 1..N$.

Where $D(\cdot, \cdot)$ is the Spearman distance defined over two rankings τ, σ as $D(\tau, \sigma) = \sum_{i=1}^n |\tau(i) - \sigma(i)|$ and M is the mean ranking. Since weight values are in the interval $[0, 1]$, rankings are estimated to be similar or dissimilar to the estimator when the weight values go to 0 or tend to 1 respectively. According with this, the worst rankers will have the higher weight values.

Discounting of the BBA: The application of the weights to the correspondent mass distribution of the rankings is computed in the ApplyWeight method, as following:

$$\begin{array}{ll} \text{if } W_j = \min_j W & \text{if } W_j \neq \min_j W \\ m'_{ji}(P) = m_{ji}(P) + (W_j * m_{ji}(\Theta)) & m'_{ji}(P) = 1 - m'_{ji}(\Theta) \\ m'_{ji}(\neg P) = 0 & m'_{ji}(\neg P) = 0 \\ m'_{ji}(\Theta) = 1 - m'_{ji}(P) & m'_{ji}(\Theta) = m_{ji}(\Theta) + (W_j * m_{ji}(P)) \end{array}$$

where m'_{ji} are the discounted bba of the j -th ranking on the i -th element. The idea is to reduce the uncertainty, proportionally to the correspondent weight for the best rankings (namely, the ranking with minimum weight), and to increase the uncertainty for all the other rankings.

Combination: All the masses of the rankings are combined, using the Conjunction rule (Sec. 2) obtaining a new mass distribution that contains the belief expressed in all the rankings. The use of the conjunctive rule is justified when all the sources of belief are supposed to tell the truth and to be

independent. These requirements are fully satisfied here, since we suppose that the rankings are independent and all rankings are totally reliable because the unsupervised context does not allow to make other assumptions on their quality. Finally, we apply the Pignistic Transformation (Sec. 2) to the combined belief distribution in order to produce the output ranking. Final ranking $O = (O(1), \dots, O(i), \dots, O(n))$ is produced by ranking all the elements with respect to the $BetP_i(P)$. Although there is no theoretical constraint about the number of iterations, we propose as number of iteration at least $T = \frac{N}{2} + 1$ (referred as BRE-MAXT). The rationale of this rule of thumb is that replacing more than one half of the original rankings can possibly lead to poor performance due to the replacement of some of the best rankings with information affected by the worse ones.

4 Experimental Results on Synthetic Data

In this section we describe the results of an evaluation test on synthetic data that suits perfectly the problem at hand. The aim of this series of experiments is to evaluate the ranking produced by our method with respect to the mean and the median of the rankings in input and the optimal aggregated list [2]. The performance are measured with Spearman correlation coefficient (referred to as ρ , [3]) and with normalized Spearman footrule distance (referred to as D) computed with respect to the Truth Ranking (T_{rank}). For all the synthetic data experiments, the input rankings are randomly generated under the constraint to exhibit fixed values of ρ with a fixed ranking T_{rank} of 300 elements. In order to have more reliable results we performed 10 independent replicas of the procedure using the same generation parameters. We generate our data in order to combine a number of rankings N equals to 3, 10, 30. For each N value, we propose different cases that correspond to different values of ρ of the ranking with respect to the T_{rank} . With $N = 3$ we define 4 different cases: (*case 1*) 1 ranker extremely good ($\rho = .80$) with respect to the other ($\rho = .06, \rho = .01$), (*case 2*) two good rankers ($\rho = .60, .40$) and a very poor one ($\rho = .01$), (*case 3*) 3 rankers with high correlation ($\rho = .80, .60, .10$), (*case 4*) 3 rankers with poor correlation ($\rho = .01, .06, .03$). With $N = 10$ and $N = 30$, three cases *good*, *equal* and *poor* are defined. In the case *good* the 80% of the rankers are highly informative ($\rho \in [0.95, 0.70]$) and the remaining 20% are low informative ($\rho \in [0.30, 0.1]$). In the case *equal* the rankers are equally distributed among the three types: highly, medium ($\rho \in [0.70, 0.30]$) and

Table 1: The result of BRE and the other competitors w.r.t. the Truth Ranking for all cases of combination of 3, 10 and 30 rankers.

Method	Evaluation measure ρ									
	3 Rankers Cases MAXT=4T				10 Rankers Cases MAXT=6T			30 Rankers Cases MAXT=16T		
	1	2	3	4	good	equal	poor	good	equal	poor
Mean	.4781	.5419	.7958	.0782	.9621	.8760	.7793	.9856	.9543	.8802
Median	.4257	.5106	.7678	.0693	.9748	.8656	.7641	.9941	.9546	.8550
Optimal list	.4065	.4953	.7515	.0594	.9754	.8686	.7781	.9957	.9663	.8787
BRE-NW	.4888	.5254	.7799	.0804	.9383	.8453	.7723	.9409	.8941	.8074
BRE-1T	.4322	.5717	.8214	.0733	.9751	.915	.8751	.9896	.9686	.9176
BRE-MAXT	.4237	.5786	.8314	.0701	.9780	.9250	.8750	.9800	.9725	.9273
Evaluation measure: D										
Mean	.4584	.4341	.2817	.6225	.1199	.2156	.2903	.0734	.1265	.2035
Median	.4628	.4492	.2846	.6270	.0798	.2309	.2974	.0305	.1313	.2298
Opt.list	.4531	.4399	.2808	.6311	.0538	.1780	.2523	.0144	.0672	.1660
BRE-NW	.4511	.4393	.2912	.6235	.1444	.2326	.3129	.1406	.1919	.2669
BRE-1T	.4697	.4147	.2591	.6282	.0953	.1749	.2115	.0614	.1036	.1672
BRE-MAXT	.4719	.4107	.2477	.6286	.0891	.1649	.2131	.0824	.1023	.1600

low informative. Finally, the case *poor* is similar to the *good* one, but with inverted percentages of highly and low informative rankers. For $N = 10, 30$ ρ of the generated rankings with respect to T_{rank} are randomly chosen within the defined intervals. Taking into account ρ in Tab 1, BRE with weighting schema outperforms all the three competitors in the case 2,3 (N=3) and in all the three cases for N=10. For N=30 (good case) BRE outperforms the mean but not the median and the optimal list that show slightly better values. With regard to the iterative process, BRE-MAXT outperforms all the competitors with a notable difference in most of the cases, except in the good case (N=30) and cases 1,2 (N=3). Taking into account D , BRE-1T and BRE-MAXT shows the same performance for N=3 as with ρ . For N=30 (all cases) BRE-1T outperforms only the median and the mean but not the optimal list that shows always low D values. These preliminary results show the significant performance of BRE w.r.t the competitors using both evaluation measures. We point out that BRE-MAXT in the poor cases is the best estimator of the Truth Ranking.

5 Conclusion

In this work we have presented preliminary results on the Belief Ranking Estimator, an unsupervised method that estimates a true ranking given a set of estimating ranked permutations. BRE, through the use of the belief func-

tion framework, models the uncertainty of each ranking and combine them according to weights computed as distances from the mean of the rankings. From results on synthetic data, with low-quality input a variant of BRE has provided better estimation of the true ranking with respect to the mean the median and optimal aggregated lists used as competitor method. As future work, we plan to extend BRE to the aggregation of top-k lists.

References

- [1] A. P. Dempster. Upper and lower probabilities generated by a random closed intervals. *The Annals of Mathematical Statistics*, 39:957–966, 1968.
- [2] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10 th WWW conference*, pages 613–622, 2001.
- [3] M.G. Kendall and J.D. Gibbons. *Rank correlation methods*. Oxford University Press, 1990.
- [4] G. Shafer. *A Mathematical Theory of Evidence*. 1976.
- [5] P. Smets and R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66(2):191–234, 1994.
- [6] M.A. Soliman, I.F. Ilyas, and S. Ben-David. Supporting ranking queries on uncertain and incomplete data. *The VLDB Journal*, 19(4):477–501, 2010.
- [7] Z. Younes, F. Abdallah, and T. Dencœux. An evidence-theoretic k-nearest neighbor rule for multi-label classification. *Scalable Uncertainty Management*, pages 297–308, 2009.