Reasoning With Streamed Information from Unreliable Sources

Ladislav Beránek, Radim Remeš

Department of Applied Mathematics and Informatics, Faculty of Economics, University of South Bohemia, CZECH REPUBLIC, Ceske Budejovice, 13 Studentska str., email: beranek@ef.jcu.cz, inrem@ef.jcu.cz

Abstract: **At present, we are increasingly struggling with the need to make decisions based on information from data streams from different sources which are often unreliable. When deciding, we need to process the this observed information, and we must estimate their reliability. In this paper, we propose a framework that allows us to derive information from unreliable sources and to estimate their trustworthy. This framework is fully implemented on data streams with the aim to derive of new facts from incoming information. This information is coming as unstructured messages that are transmitted from heterogeneous and potentially untrustworthy sources. This information is processed using a natural language and belief function theory. The trustworthy of processed information is estimated based on their internal conflict. The proposed framework is evaluated using an experiment that quantifies the efficiency of our solution with respect to accuracy and overhead of the proposed framework.**

Keywords: **streamed data, reasoning, belief function theory, uncertainty.**

I. Introduction

Users share and process all sorts of data within various applications based on the Internet infrastructure. Users evaluate various products and express their opinions to different events. Tripadviser.com server might be an example. Users can evaluate here certain hotel on the base of their satisfaction with its services. Wikipedia provides feedback tool to engage the reader in a review of article quality based on four criteria, i.e., "trustworthy", "objective", "complete" and "well written". Such activity is referred also as crowdsourcing. Many users are used for evaluation or classification of certain product or services. It is sometimes used in science. An example might be the website Galaxy Zoo, where users classify astronomical images.

As this method is useful, organizers usually have little control over quality of users' activity. Reaction of individual users may vary substantially, and in some cases, they may even be controversial. The question is then how to integrate feedback from multiple users to get an objective opinion. Commonly used heuristics such as "majority voting" and "take the average" ignore individual user experience and can fail, for example in an environment where there are users with malicious intent. The aim of this paper is to propose and to test a method to determine the grand truth without knowing the previous experience of users. For this purpose, it is used an approach based on the Dempster-Shafer theory.

Within this theory, the operation discounting is defined. At this operation, the value of belief function varies in dependence on certain additional information or if the pieces of information, to be integrated, are contradictory. When it is necessary to decides to implement discounting process the following questions are to be solved: What resources are to be discounted? Up to what extent these resources should be discounted? The model used in this paper introduces an iterative method which automatically determines the discount rate on the base of the reliability of sources. The advantage of this approach is that it does not require any additional meta-information about the reliability of sources. The method assumes only that the more specific source of information conflicts with the majority opinion, the stronger this source must be discounted.

The rest of this paper is organized as follows. Section 2 provides an overview of related work. Section 3 formulates the problem and introduces a belief function framework with the proposed model. Section 4 presents experimental results on synthetic data. Conclusions are composed in Section 5.

II. Related Work

Currently, several studies deal with the setting involving multiple labelers. For example, the work such as [4, 7, 9, 19, 20, 22] focus on the estimating the error rates of observers. Authors [4] deal with selecting the best set of all available information from users for model training. These works focus on learning classifiers directly from user data instead of estimating ground truth. Work [14, 15] uses a probabilistic framework for solving classification, regression and ordinal regression problem with multiple annotators. This framework assumes that the expertise of each annotator does not depend on these data. Works [23, 25, 26] develop this approach, but do not build fully on this premise. There are some other related works, which focuses on a different setting [3, 24]. Recent work [8] pays attention to regression problem under multiple observers, with the use of less parametric methods for modeling and designing observers regression function.

III. Methodology

Belief function theory framework

Our model is an application of the Dempster-Shafer theory. The Dempster-Shafer theory [16] is designed to deal with the uncertainty and incompleteness of available information. It is a powerful tool for combining evidence and changing prior knowledge in the presence of new evidence. The Dempster-Shafer theory can be considered as a generalization of the Bayesian theory of subjective probability.

In the following paragraphs, we give a brief introduction to the basic notions of the Dempster-Shafer theory (frequently called theory of belief functions or theory of evidence).

Basic Notions

Considering a finite set referred to as *the frame of discernment* Ω, a *basic belief assignment (BBA)* is a function *m:* $2^{\Omega} \rightarrow [0,1]$ so that

$$\sum_{A \subseteq \Omega} m(A) = 1 \tag{1}$$

where $m(\emptyset) = 0$, see [16]. The subsets of 2^{Ω} which are associated with non-zero values of *m* are known as *focal elements* and the union of the focal elements is called *the core*. The value of *m(A)* expresses the proportion of all relevant and available evidence that supports the claim that a particular element of Ω belongs to the set *A* but not to a particular subset of *A*. This value pertains only to the set *A* and makes no additional claims about any subsets of *A*. We denote this value also as a *degree of belief* (or *basic belief mass - BBM*).

Shafer further defined the concepts of *belief* and *plausibility* [16] as two measures over the subsets of Ω as follows:

$$Bel(A) = \sum_{B \subseteq A} m(B), \tag{2}$$

$$Pl(A) = \sum_{B \cap A \neq \varphi} m(B). \tag{3}$$

A *BBA* can also be viewed as determining a set of probability distributions *P* over Ω so that $Bel(A) \leq P(A) \leq Pl(A)$. It can be easily seen that these two measures are related to each other as $Pl(A) = 1 - Bel(\neg A)$. *Moreover, both are equivalent to m.* Thus, one needs to know only one of the three functions *m*, *Bel*, or *Pl* to derive the other two. Hence, we can speak about belief function using corresponding *BBAs* in fact.

Dempster's rule of combination can be used for pooling evidence represented by two belief functions Bel_1 and Bel_2 over the same frame of discernment coming from independent sources of information. The Dempster's rule of combination for combining two belief functions Bel_1 and Bel_2 defined by (equivalent to) *BBAs* m_1 and m_2 is defined as follows (the symbol ⊕ is used to denote this operation):

$$(m_1 \oplus m_2)(A) = \frac{1}{1-k} \sum_{B \cap C = A} m_1(B) \cdot m_2(C), \tag{4}$$

where

$$k = \sum_{B \cap C = \emptyset} m_1(B) \cdot m_2(C). \tag{5}$$

Here *k* is frequently considered to be a *conflict measure* between two belief functions m_1 and m_2 or a measure of conflict between m_1 and m_2 [16]. Unfortunately, this interpretation of *k* is not correct, as it includes also internal conflict of individual belief functions m_1 and m_2 [5, 6]. Demspter's rule is not defined when $k = 1$, i.e. when cores of m_1 and m_2 are disjoint. This rule is commutative and associative; as the rule serves for the cumulation of beliefs, it is not idempotent.

Belief Function Correction

When receiving a piece of information represented by a belief function, some metaknowledge regarding the quality or reliability of the source that provides the information, can be available. In the following paragraphs, we describe briefly some possibilities how to correct the information according to this metaknowledge.

Discounting

To handle the lower reliability of information sources, a discounting scheme has been introduced by Shafer [24]. It is expressed by equations:

$$^{\alpha}m(A) = \begin{cases} (1-\alpha) \times m(A) & \text{if } A \subset \Omega \\ \alpha + (1-\alpha) \times m(\Omega) & \text{if } A = \Omega \end{cases} \tag{6}$$

where $\alpha \in [0,1]$ is a discounting factor and $^{\alpha}m(A)$ denotes the discounted mass of *m*(*A*). The larger α is, the more masses are discounted from $A \subset \Omega$, while the more mass is assigned to the frame of discernment Ω.

IV. Results and Discussion

The idea of discounting mechanism is a weakening of a given belief function (*BBA*). Thus, the principle of the discounting is transferring of parts of basic belief masses (*BBMs*) of all focal elements which are proper subsets of the frame of discernment to the entire frame. This process is the result of some additional information saying that the source is not entirely reliable. The transfer of *BBMS* from a source to the framework reflects an increase of the degree of uncertainty regarding the data that the source produces.

Use of belief function theory for ground truth estimation

Traditional data fusion processing based on Dempster/Shafer theory consists of obtaining of *BBA*s due to some mathematical model in the first step. The second step is the discounting of some BBAs which we know about that they are less reliable (6). The final step is the integration of BBAs using a Demspter's rule (4) or using some other suitable combination rule [11, 17, 18, 21]. As it was described above discounting process is used when we have meta-information about the reliability of some contextual sources of information (BBA) and it is necessary to have some approach how to express the value of discounting factor [1, 2].

In the most cases, the discount rate is adjusted manually, but some authors have suggested several methods how to obtain them automatically. In [18], Smets calculates the discount factor by minimizing the error function. This method focuses on the classification of data and requires a set of labeled data. In [12], Martin et al. establish the discount rate evaluation method that is based only on the values of BBA themselves. Similar approach which is the basis of our work is presented in [10].

Defining what the majority opinion means within the Dempster-Shafer theory is not easy. Murphy [13] for example suggested using average BBAs and argued that the average properties are better suited for the fusion of contradictory evidence:

$$m_{mean} = \frac{1}{M}\sum_{i=1}^{M} m_i \quad (7)$$

This opinion is valid considering the fact that if subset s1 from S corresponds to the cluster of concordant BBAs and if this subset contains more BBAs than any other cluster, then m_{mean} will probably be closer to BBAs forming the s1. Hence m_{mean} can be used as an estimate of the majority opinion [13]. We therefore propose to review the first set of discount factors by the following way:

$$\alpha_i^0 = d_{BPA}(m_i, m_{mean}) \quad (8)$$

where d_{BPA} is defined subsequently [12]:

$$d_{BPA}(m_1, m_2) = \sqrt{1/2(\vec{m_1} - \vec{m_2})^t D(\vec{m_1} - \vec{m_2})} \quad (9)$$

Here, the $\vec{m}$ is BBA expressed in the form of vector and D is the matrix which has dimensions $2^N \times 2^N$ with elements $D(A,B) = |A \cap B|/|A \cup B|$.

Equation (10) gives low values of discounting factor for BBAs near to the mean (they are in accordance with the opinion of the majority) and a high degree of discounting factor for BBAs that differ considerable from the mean (the ones that are the cause of disagreement).

In this paper, we use an iterative method for calculating of discounting factors. In the first step, discounting factors are calculated for each member of the initial settings using equation (8). Then this iteration process is applied on the BBAs set S_1. New values of discounting factors are obtained. This iteration is repeated and the value of discounting factors increases but more and more slowly. To determine the optimal set of discount factors among those computed at each iteration step a posteriori analysis is employed.

We investigate the conjunctive combinations obtained at each step and compare them with categorical BBAs by distance dBPA. Iteration that gives minimum distance is optimal number of iteration i_{opt}.

Relative values of discount factors in single steps affect the result of the result of information fusion process as much as the absolute value. In other words, it is not sufficient to have a high degree of value on unreliable sources, it is also necessary that the measure of the difference between reliable and unreliable sources be large enough. Therefore, we perform the optimum setting of values α_i using iteration. We calculate a discounting factor of the initial set of BBAs and then recalculate new values of BBAs of this set. This process is repeated as described in the previous paragraph. Consecutive values of discount factors are calculated by these iterations process and are further analyzed to determine the best setting according to the predefined criteria which is minimum distance.

An iterative procedure involves the gradual discounting the original BBAs. The term m^{α^0,α^1} indicates BBA discounted value of α^1. Successive values of discounting factors $\{\alpha^0, \ldots, \alpha^K\}$ can be summarized:

$$\left.\begin{aligned} \alpha^K &= 1 - \prod_{i=0}^{K}(1-\alpha^i) \\ \alpha^K &= \beta^{K-1}(1-\alpha^K) + \alpha^K \end{aligned}\right\} \quad (10)$$

Stop condition is distance dBPA. Iteration that gives minimum distance is optimal number of iteration i_{opt}. Important here is that we can also find a source that differs mostly from the average value. It may be omitted from the calculations and it may be explored independently. The advantage of this described approach is that it does not need any meta-information about the reliability of sources.

TABLE 1. THE BBA SET AND THE RESULTS OF AGGREGATION

	{a}	{b}	{c}	{a, b}	{a, c}	{b, c}	{Ω}
m_1	0.5	0.2	0.1	0.1	0.05	0.025	0.025
m_2	0.52	0.12	0.08	0.05	0.1	0.06	0.07
m_3	0.6	0.08	0.12	0.025	0.1	0.025	0.05
m_4	0.2	0.1	0.6	0.05	0.025	0.025	0
m_5	0.48	0.15	0.09	0.13	0.04	0.02	0.09
m_6	0.45	0.21	0.11	0.09	0.06	0.05	0.03
m*	0.3989	0.1022	0.2762	0.0598	0.0785	0.0621	0.0223
m	0.5112	0.1899	0.1169	0.0895	0.049	0.0234	0.0201

The responses of various sources (observers) are represented by the values of belief functions in Table 1. The six different sources are modeled ($m_1 - m_6$). Ground truth has the same values as the values $m_1(\cdot)$. The value of $m^*(\cdot)$ is calculated using equation (4). The value of $m(\cdot)$ in the last but one row of the table is calculated according to the process outlined in the previous section. Source 4 (m_4) is modeled as adversarial, because its reaction is opposite to the ground truth. The discount factor calculated for this source reaches the highest values. The table shows that discounting process overrides the impact of this source and as a result the result of the integration of information sources will be close to grand truth (m*).

V. CONCLUSION

This article examines the problem of multiple observers which provide answers that are not entirely accurate. The problem concerns the use of model that is based on belief function theory and no additional information about the reliability of observers are known. Our approach provides an estimate of the ground truth and predicts the response of each observer of the new instance. Experiments show that the proposed method outperforms several core values and leads to a performance close to the model trained with ground truth. There are many opportunities for further research. One possible direction is to extend our model with more cores learning. The aim is to choose an algorithm or a composite different covariance functions instead of fixing the combination in advance. Consequently, the algorithm may be difficult to learn fits observer selecting multiple cores in data-dependent manner. In addition, it would be very useful to design efficient sampling methods for selection that instance and the response should be taught more. Our aim is to test further the described algorithm on real data and further to verify the model described in this paper.

References

[1] Beranek, L. Nydl, V., The Use of Belief Functions for the Detection of Internet Auction Fraud. In: *Proceedings of the 31st International Conference Mathematical Methods in Economics* 2013. Jihlava: College of Polytechnics Jihlava, 2013, pp. 31 - 36.

[2] Beranek, L., Knizek, J., The Usage of Contextual Discounting and Opposition in Determining the Trustfulness of Users in Online Auctions. *Journal of Theoretical and Applied Electronic Commerce Research, Vol. 7, No. 1*, 2012, pp. 34-50.

[3] Chen, S., Zhang, J., Chen, G., Zhang, C., What if the irresponsible teachers are dominating? *In Proc. 24th AAAI 2010*, 2010.

[4] Crammer, K., Kearns, M., Wortman, J., Learning from multiple sources. *JMLR, Vol. 9, No. 4*, 2008, pp. 1757 1774.

[5] Daniel, M., Several Notes on Belief Combination, In *Proceedings of the Theory of Belief Functions Workshop. Brest: ENSIETA*, 2010. pp. 1–5.

[6] Daniel, M., Conflicts within and between Belief Functions. In Proceeding from IPMU 2010 – *Computational Intelligence for Knowledge-Based Systems Design*, Berlin, 2010, s. 696–705, Lecture Notes in Artificial Intelligence. 6178.

[7] Dawid, A., Skene, A., Maximum likelihood estimation of observer error-rates using the EM algorithm. *Applied Statistics, Vol. 28, No 1*, 1979, pp. 20-28.

[8] Han, X., Huang, X., Eckert, C., Learning from Multiple Observers with Unknown Expertise, In *Proceedings of 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining, Gold Coast*, Australia, 2013, pp. 233–241

[9] Hui, S., Walter, S., Estimating the error rates of diagnostic tests. *Biometrics, Vol. 5,* No , 1980, pp. 167171.

[10] Klein, J., Colot, O., Automatic discounting rate computation using a dissent criterion, In *Proceedings of the Theory of Belief Functions Workshop.* Brest: ENSIETA, 2010. pp. 151–156.

[11] Lysek, J., Stastny, J., Motycka, A., Object Recognition by Means of Evolved Detector and Classifier Program. In MENDEL 2012, *18th International Conference on Soft Computing. Brno University of Technology,* 2012, p. 82–87.

[12] Martin, A., Jouselme, A.L., Osswald, C., Conflict measure for the discounting operation on belief functions. In *IEEE Int. Conf. on Information Fusion FUSION* 2008, Madrid, 2008, pp. 1-8.

[13] Murphy, C., Combining belief functions with evidence conflicts. *Decision Support Systems, Vol. 29, No. 4*, 2000, pp. 1-9.

[14] Raykar, V., Yu, S., Zhao, L., Jerebko, A., Florin, C., Valadez, G., Bogoni, L., Moy, L., Supervised learning from multiple experts: Whom to trust when everyone lies a bit. In *Proc. 26th ICML 2009*, 2009, pp. 889-896. ACM (2009)

[15] Raykar, V., Yu, S., Zhao, L., Valadez, G., Florin, C., Bogoni, L., Moy, L., Learning from crowds. *JMLR, Vol. 11, No. 2, 2010*, pp. 1297-1322.

[16] Shafer, G., A mathematical theory of evidence, *Princeton University Press*, Princeton, 1975.

[17] Schubert, J., Conflict management in dempster- theory by sequential discounting using the degree of falsity. In Int. *Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems IPMU 2008*, Madrid, 2008, pp. 298-305.

[18] Smets, P., Analyzing the combination of conflicting belief functions. *Information Fusion, Vol. 8, No. 3*, 2006, pp. 387-412.

[19] Smyth, P., Fayyad, U., Burl, M., Perona, P., Baldi, P., Inferring ground truth from subjective labelling of venus images. *In Proc. 9th NIPS 1995*, 1995, pp. 1085-1092.

[20] Spiegelhalter, D., Stovin, P., An analysis of repeated biopsies following cardiac transplantation. *Statistics in Medicine, Vol. 2, No. 1*, 1983, pp. 33-40.

[21] Stencl, M., Stastny, J. Neural network learning algorithms comparison on numerical prediction of real data. *In MENDEL 2010, 16th International Conference on Soft Computing. Brno University of Technology*, 2010, p. 280-285.

[22] Tubaishat, M., Madria, S., Sensor networks: an overview. Potentials. *IEEE, Vol. 22, No. 2*, 2003, pp. 20-23.

[23] Whitehill, J., Ruvolo, P., Wu, T., Bergsma, J., Movellan, J., Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In *Proc. 23rd NIPS 2009, vol. 22*, pp. 2035-2043.

[24] Wu, O., Hu,W., Gao, J.: Learning to rank under multiple annotators. In: *22nd IJCAI, 2011*, pp. 1161–1168.

[25] Yan, Y., Rosales, R., Fung, G., Dy, J.: Active learning from crowds. In: *Proc. 28th ICML*, 2011, pp. 1161–1168.

[26] Yan, Y., Rosales, R., Fung, G., Schmidt, M., Hermosillo, G., Bogoni, L., Moy, L., Dy, J., Malvern, P., Modeling annotator expertise: Learning when everybody knows a bit of something. *Journal of Machine Learning Research, Vol. 9*, 2010, pp. 932-939.