Belief Measure of Expertise for Experts Detection in Question Answering Communities: case study Stack Overflow

Dorra Attiaoui, Arnaud Martin, Boutheina Ben Yaghlane

HAL Id: hal-01568061
https://hal.science/hal-01568061

Submitted on 24 Aug 2017

International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, 6-8 September 2017, Marseille, France

Belief Measure of Expertise for Experts Detection in Question Answering Communities: case study Stack Overflow

Dorra Attiaoui[a,b], Arnaud Martin[b], Boutheina Ben Yaghlane[c]

[a]*LARODEC, ISG Tunis, University of Tunis, Le Bardo, Bouchoucha, Tunis, Tunisia*
[b]*DRUID, IRISA, University of Rennes 1, Rue E. Branly, 22300 Lannion, France*
[c]*LARODEC, University of Carthage, Carthage Prsidence, Tunis, Tunisia*

Abstract

Online Question Answering Communities (Q& A C) provide a valuable amount of information in several topics. The major challenge with Q& A C is the detection of the authoritative users. When manipulating real world data, we have to deal with imperfections and uncertainty that can occur. In this paper, we propose a belief measure of expertise allowing us to detect users with the highest degree of expertise based on their attributes. Experiments on a dataset from a large online Q&A Community prove that the proposed model can be used to improve the identification of most expert users.
© 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of KES International.

Keywords: Theory of belief functions; Expertise; Social Networks; Question Answering Communities; Classification

1. Introduction

With the increasing importance of social networks, identifying relevant users among other contributors represents a challenging task. Expert detection in social media and organizations has become important for several motivations. In [11], authors have enumerated some of them such as providing high quality content, getting valuable answers to specific questions, finding people with accurate skills in a given area, etc.

For the last decade, social networks and question answering communities have gained importance when seeking information. One of the most popular platforms is Stack Overflow (SO)[1]. It is the largest online community for pro-

E-mail address: dorra.attiaoui@univ-rennes1.fr

E-mail address: Arnaud.Martin@univ-rennes1.fr

E-mail address: boutheina.yaghlane@ihec.rnu.tn

[1] http://stackoverflow.com

grammers. Here, users can post questions, answer them, vote positively or negatively for both answers and questions in order to express their opinion in the quality of the posts. The reputation[2] is the summary of users' activity in the web site. It is earned by convincing other users that a contributor can provide helpful answers to technical questions. This shows the potential of a user to give important posts by sharing their knowledge.

Reputation actually reflects how a user is involved in the community and how other people see him. Most of the users within this platform aim to win as much reputation points as possible in order to obtain privileges like creating tags, moderating the forum etc. The easiest way to have a high reputation is by providing helpful answers. However, this measurement does not take into account the quality and the expertise. Indeed, if a question is considered as simple in a very popular topic, answers will be numerous and quick, creating a competitive spirit within the community. Therefore, if a question is seen as difficult in a less popular topic, contributors may not take the risk to post answers, due to their lack of knowledge or the risk of being evaluated negatively by the community and of losing some reputation points.

For[1], the authors proposed an analysis of Stack Overflow's reputation system. They focused on the contributors participation model. They considered the reputation as measurement of expertise. Any user with a reputation grater than 2400 point is an expert. However, their approach seems to be strict because it is only based on the value of the reputation gathered during users activity in the platform.

There is a difference between users' reputation and expertise or performance in the web site. The reputation is only based on other people's opinion while the expertise is only related to someone's ability to provide fair and well posted contributions. The main issue with the reputation is highlighted through the topic and other people's impressions in the contributor. The problem with the topics in stack overflow is some of them are very popular inducing a phenomenon of over votes, unlike unpopular topics that can be neglected because of the few number of persons interested in. While the expertise refers to the characteristics, skills, and knowledge that distinguish experts from novices and less experienced people, as described in[12], an expert is someone who is very skillful and well-informed in some special field.

In this paper we are interested in proposing a measure of expertise based on the theory of belief functions. Also called the mathematical theory of evidence, it is one of the most popular approaches for reasoning under uncertainty. It is a generalization of the probability and possibility theories. Managing and processing imperfect information, differs from manipulating certain data from real world applications, has become an important challenge. It has been used thanks to its ability to manage imperfect information. It is a strong formalism widely used in many research areas: image processing[15], clustering[7], etc. and more recently social networks[17].

The Belief Measure of Expertise (noted BME) will help us to identify the most influential users in the platform. This expertise measure is statistically defined and does not take into account any semantic information about questions or answers. Thus, the major challenge is to estimate users' expertise quantitatively without taking into account the reputation system defined by Stack Overflow.

This paper is organized as follow: in Section 2, we introduce some approaches for experts identification in question answering communities. In Section 3, we recall some basic concepts of the theory of belief functions . After this, we propose in Section 4 the formalism of the Belief Measure of Expertise. Finally, in Section 5, the proposed approach is used to measure the expertise of users in Stack Overflow, and users classification. We show the effectiveness of the results of the BME and compare the quality of the classification with other methods proposed for experts identification.

2. Related work

Detecting experts in online communities have been wildly investigated. We can distinguish two different methods: ranking based approaches and attribute based approaches as presented in[19].

On the one hand, the ranking based approaches intent to measure a score per user than select the top users[25]. The ExpertiseRank[27] which is an extension of PageRank[16] allowing to compute the expertise score of a users in a question answering community. Besides, the graphical features, this algorithm also includes a metric called "Z-Score" based

---

on both the number of answers and the number of questions asked by a given user. Their result suppose that a metric like "Z -Score" outperforms over complex graph based algorithm such as PageRank.

Another approach is proposed in[26] that is not founded on the reputation measure. They defined a metric called "Mean Expertise Contribution" that takes into account two indices: the debate generated by a question and the utility of the provided answers. The first index is related to the number of answers proposed for a given question. The second one is calculated according to the rating of an answer among all the answers provided.

On the other hand, attributes based approaches aim to identify a number of features for the users and then apply machine learning techniques in order to classify users. In[23], authors used a Gaussian Mixture Model (GMM) based clustering algorithm to identify clusters among expert users of question answering communities. They preferred a GMM based method because it overcomes traditional clustering methods such as K-Means. In[2], authors used a Beta Mixture Model (BMM) to identify authorities in online communities. They rated users in Yahoo!Answers based their activities. Recently in[19], authors extended the work presented in[3]. Their model based on a BMM evaluates the profiles of users to distinguish between authoritative and non-authoritative users based on voting mechanism. Their challenge is to locate important users by mining textual and meta data features. In[9], authors proposed an early detection of topical expertise based on the attributes of users and mainly the number of accepted answers. They also based their approach using textual, behavioral ad temporal characteristics of the users.

In this paper, we will demonstrate that attribute-based classification method can be considered as a solution to the problem of experts detection in question answering communities without taking into account posts provided by users.

In[14], the authors identified three levels of uncertainty in question answering communities. The first level is related to the extraction and integration of the data. The second one deals with information sources meaning the users of these platforms. The third level covers the uncertainty of the information itself. In our case, we are more interested in the evaluation of the sources and the part of uncertainty related to them. The main issue in these communities is that we are dealing with users that we do not usually have an apriori knowledge about them. We ignore everything about the sources' reliability, or expertise. In order to deal with type of uncertainty, we will use all the mathematical background provided by the theory of belief functions. This will help us to consider this problem with an uncertain point of view.

3. Theory of belief functions: background

We present in this section some notions related to the theory of belief functions, used in this paper. It has been developed by Dempster in his work on upper and lower probabilities[8]. Afterwards, it was formalized in a mathematical framework by Shafer in[18]. This theory is able to deal and represent imperfect (uncertain, imprecise and /or incomplete) information.

Let us consider a variable x taking values in a finite set $\Omega = \{\omega_1, \ldots, \omega_n\}$ called *the frame of discernment*.
A *basic belief assignment* (*bba*) is defined on the set of all subsets of Ω, named power set and noted 2^Ω. It affects a real value from $[0, 1]$ to every subset of 2^Ω reflecting sources amount of belief on this subset. A bba m verifies:

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

3.1. Particular belief functions

Mass function is the common representation of evidential knowledge. Basic belief masses are degrees of support justified by available evidences. This section recalls some particular mass functions.

- **Categorical mass function** is a normalized mass function which has a unique focal element X^*. This mass function is noted $m(X)$ and defined as follows:

$$m_{X^*}(X) = \begin{cases} 1 & \text{if } X = X^* \subseteq \Omega \\ 0 & \forall X \subseteq \Omega \text{ and } X \neq X^* \end{cases} \tag{2}$$

We distinguish two particular cases of categorical mass functions: the vacuous mass functions when $X^* = \Omega$ and the contradictory mass functions if $X^* = \emptyset$.

- **Vacuous mass functions** is particular categorical mass function focused on Ω. It means that a vacuous mass function is normalized and has a unique focal element which is Ω. This type of mass functions is defined as follows:

$$m_\Omega(X) = \begin{cases} 1 & \text{if } X = \Omega \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

A vacuous mass function emphasizes the case of total ignorance.

- **Simple support mass functions** is a special type that allow us to model both of the uncertainty and imprecision according the following equation:

$$\begin{cases} m(X) = 1 - \omega, \ X \subseteq \Omega \\ m(\Omega) = \omega \\ m(Y) = 0, \ Y \neq X \subseteq \Omega \end{cases} \tag{4}$$

where the mass on $m(\Omega)$ represents the ignorance.

3.2. Combination rules

In the belief function theory, Dempster in [8] proposed the first combination rule. It is defined for two bbas m_1, m_2, $\forall X \in 2^\Omega$ with $X \neq \emptyset$ by:

$$m_{DS}(X) = \frac{1}{1-k} \sum_{A \cap B = X} m_1(A)m_2(B), \tag{5}$$

where k is generally called the global conflict of the combination or its inconsistency, defined by $k = \sum_{A \cap B = \emptyset} m_1(A)m_2(B)$ and $1 - k$ is a normalization constant.

This combination rule is very useful, first because it decreases the vagueness and, then, it increases the belief of the observed focal elements.

3.3. Decision

Pignistic probability transformation was proposed in [22]. It transforms a *bba* "m" into a probability measure noted BetP, for all $X \in 2^\Omega$:

$$betP(X) = \sum_{Y \neq \emptyset} \frac{|X \cap Y|}{|Y|} \frac{m(Y)}{1 - m(\emptyset)}. \tag{6}$$

4. Users classification founded on belief functions

In this section, we highlight the difference between the reputation and expertise of a user in Stack Overflow. Note that in this paper we do not consider the reputation because it is a rough measurement based only on other people's opinion and not founded on both user's activity and the opinion of the community. For [12], authors consider the expertise as characteristics and skills that distinguish between experts and usual users.

4.1. Users description

We performed an ascendant hierarchical classification on Stack Overflow's data. This step allows us to regroup users according to their homogeneity within a same group and then estimate the number of potential classes.

We identified three different classes:

- **Experts (E)**: these users are very reliable and recognized by the community. They provide a considerable number of useful answers that are chosen as the best ones. They are very active in the platform and guarantee a high quality content.
- **Apprentices (A)**: these users may have some expertise in a given topic. They aim to increase their reputation. To do so, they post a lot of answers that are not always very useful. The quality of their posts is not guaranteed and their answers can be down-voted.
- **Occasionals (O)**: these users represent the major part of members of the platform. They do not have a lot of knowledge. They occur occasionally only when they need an answer to a specific question that have not been treated before.

4.2. Information modeling and experts characterization

Among the attributes that can describe a user in Stack Overflow, we select five different and important features characterizing users in this platform:

- **Number of votes related to answers** (AV_i): the sum of positive votes collected by posted questions and answers.
- **Number of votes related to questions** (QV_i): the sum of negative votes collected by posted questions and answers.
- **Time activity:** time of activity of users from their registration to their last connection.
- **Number of posted questions** ($NbQu_i$): number of questions posted in the dataset during the time activity of a user.
- **Number of posted answers** ($NbAn_i$): number of answers provided in the dataset during the time activity of a user.
- **Number of posted answers** ($NbAccAn_i$): number the answers chosen as the best answers.

We build the mass functions relative to some attributes presented previously in the frame of discernment Ω such as $\Omega = \{O, A, E\}$. Each equation describes a specific case. We obtain the mass functions as it follows:

- A high number of positive votes is represented by a mass function on the focal element "**Expert**" and the remainder is given to the ignorance. For a user i:

$$
\begin{aligned}
m_1^i (A \cup E) &= \alpha_1(1 - e^{-\gamma_1 NbUV_i}) \\
m_1^i(\Omega) &= \alpha_1 \, e^{-\gamma_1 NbUV_i}
\end{aligned}
\tag{7}
$$

- A high number of negative votes is represented by a mass function on the focal element "**Apprentice**" and the remainder is given to the ignorance. For a user i:

$$
\begin{aligned}
m_2^i (A) &= \alpha_2(1 - e^{-\gamma_2 NbDV_i}) \\
m_2^i(\Omega) &= e^{-\gamma_2 NbDV_i}
\end{aligned}
\tag{8}
$$

- A high number of posted questions is represented by a mass on the union of two classes "**Apprentice ∪ Expert**". Otherwise, when this value is low it is affected to the "**Occasional**" and the reminder to the ignorance. When a mass is on the union, this means that we can't decide which one of these classes is concerned by the mass. For a user i:

$$m_3^i (E \cup A) = \alpha_3 \left(1 - e^{-\gamma_3 NbQu_i}\right) \tag{9}$$
$$m_3^i(O) = \alpha_3 e^{-\gamma_3 NbQu_i}$$
$$m_3^i(\Omega) = 1 - \alpha_3$$

- A high number of answers is represented by a mass on the union of "**Apprentice ∪ Expert**" while on the opposite situation this goes to the "**Occasional**" and the reminder to the ignorance. For a user i:

$$m_4^i (E \cup A) = \alpha_4(1 - e^{-\gamma_4 NbAn_i}) \tag{10}$$
$$m_4^i(O) = \alpha_4 e^{-\gamma_4 NbAn_i}$$
$$m_4^i(\Omega) = 1 - \alpha_4$$

- A high number of accepted answers is represented by a mass on the focal element "**Expert**" and the reminder to the ignorance. For a user i:

$$m_5^i (E) = \alpha_5(1 - e^{-\gamma_5 NbAccAn_i}) \tag{11}$$
$$m_5^i(\Omega) = \alpha_5 \ e^{-\gamma_5 NbAccAn_i}$$

In the previous equations, we fix α_1, $\alpha_5 = 0.9$, $\alpha_2 = 1$, $\alpha_3 = 0.8$ and $\alpha_4 = 0.5$. The values are fixed after several experiments in order to have the best representation of each class of users. These values are used to represent the ignorance in every mass function as described in[6]. As the apprentices are modeled only one time as focal element in equation (9) unlike experts and occasionnals, we choose to affect the value of 1 to α_2. For γ after several experimentation, we decide to keep it as the maximum value of any attribute divided by 100.

We combine these mass functions using the Demspter's combination rule presented in equation (5). Next we apply the pignistic probability and classify the user into Expert, Apprentice or Occasional. The combination process allows us to estimate the actual belief expertise (noted BME) for each user during a period is expressed by the following equation:

$$BME_{t_1}(u_i) = {}^{\alpha^T} \oplus m_1^i \oplus m_2^i \oplus ... \oplus m_5^i \tag{12}$$

where α^T is the discounting operator related to the time activity of a user in the platform.

The value $\alpha_i^T = 1 - \frac{1}{nd_i}$, where nd is the number of days since the user first connected to the platform. The symbol $\oplus$ represents the operator of combination.

5. Illustrations

In this section, experiments on real data sets will be performed to show the effectiveness of the proposed measure of expertise. Results will be compared the Reputation system of Stack Overflow like presented in[1] and a Gaussian Mixture Model (GMM)[24].
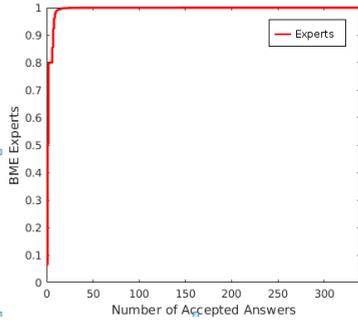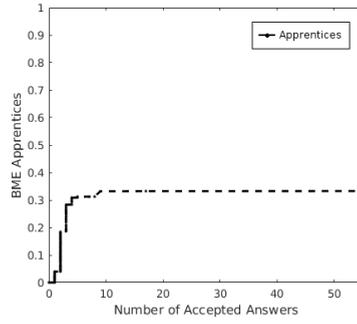
Fig. 1. BME Experts

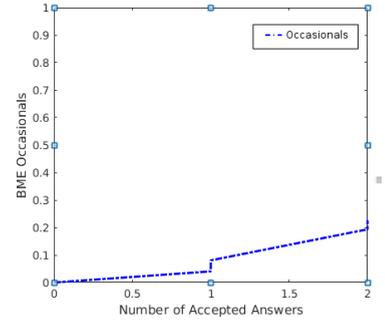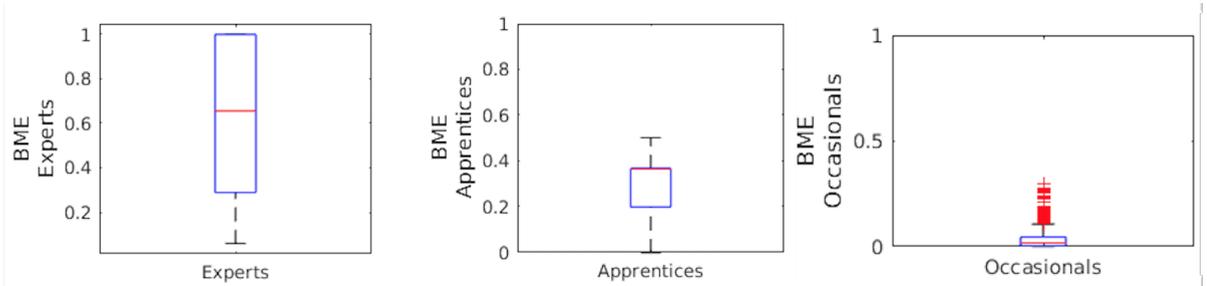

Fig. 2. BME Apprentices

Fig. 3. BME Occasionals

Fig. 4. Boxplots BME Experts, Apprentices, Occasionals

We use data provided by Stack Overflow from December 2013 to March 2015 [3]. The data set counts over 2 Million users, 2.5 Million answers and 1.7 Million questions. In Table 1 we present some statistics about the dataset provided by Stack Overflow and used for these experiments.

5.1. Belief detection of experts based on the BME

In this section, we show the results of experts detection based on the Belief Measure of Expertise. For every user, the BME takes a value in $[0, 1]$. It is the mass allocated to the focal element "Expert". When this value is close to 0 this means that the degree of expertise is weak. Otherwise, when it is near 1 we have a strong belief that this person is an expert. The BME is measured after the combination of the mass functions build from users' attributes and then reinforced by α^T related to the time of activity. After, we classify users according to their pignitic probability described in equation (6).

Figures 1, 2 and 3 show the evolution of the BME according to the number of accepted answers. This feature is considered as one of the most important index about a user's expertise. In figure 1, we witness that as the number of accepted answers of a user is high, their degree of expertise is increasing and reaching the value of 1. For the apprentices, the maximal value of their BME when considering the number of best answers given is low with a value of 0.3. The occasionals have the smallest BME due to their lack of knowledge and inactivity in the platform.

Figure 4 presents the box plot of the BME for every class. Here, we can see that the degree of expertise of "experts" is the highest. The median value of around 0.7 where some users reach the maximal value of 1. However, some users can be considered as experts even though their BME is lower (between 0.2 and 0.4). These users are experts but not as confirmed as the others due to their small time of activity. For the class of apprentices, their BME is smaller than the experts where the median is around 0.4. For the last class of occasionals, their BME is very close to zero. Though, some outliers occur in this class with a BME reaching a maximum value of 0.3.

The number of accepted answers is a very important index on estimating the expertise of a user.

---

[3] https://archive.org/download/stackexchange

Figures 1,2, 3 and 4 are related to each other. Thus, they reflect the value of the BME according to very important feature. As the number of accepted answers is big and especially for experts, the BME is high and can reach the value of 1. However, for Apprentices and Occasionals, as they don't have a lot of accepted answers, the value of their BME can be justified by the quality and the number of the questions they posted on the platform. Thus, their BME is small compared to the experts, especially for the occasionals as presented in Figure 4.

5.2. Indices of the classifications' quality

In this section, we present some indices used for the evaluation of the quality of the classification generated by the three approaches.

For C_i the class of index i between N_c different classes and n_i the number of elements of C_i We consider the following indices

- Davies-Bouldin (DB)[5]: this criterion treats every class individually. Its measures how similar a class is to the closest class. The best partition has to be minimizing the mean value calculated for every class.

$$DB = \frac{1}{n} \sum_{i=1}^{N_c} max\{\frac{I(C_i) + I(C_j)}{I(C_i, C_j)}\}$$ (13)

where $I(C_i)$ represents the mean of the distances between the objects of a class and its center. $I(C_i) + I(C_j)$ represents the distance between the centers of two classes.

- Calinski-Harabasz (CH)[4]: this criterion weighs the intra class variance by the number of classes. Its maximization is the optimal partition.

$$CH = \frac{SS_B}{SS_W} \times \frac{(N-C)}{(C-1)}$$ (14)

where SSB is the overall between-cluster variance, SSW is the overall within-cluster variance.

- Dunn[10]: The Dunn's index measures compactness (maximum distance in between data points of clusters) and clusters separation (minimum distance between clusters). The maximum value of the index represents the right partitioning. The goal is therefore to maximize the inter-cluster distance while minimizing the intra-cluster distance. Dunn's index is described according the following equation:

$$Dunn = min_i \left\{ min(\frac{d_{(C_i,C_j)}}{max(intra(C_i))}) \right\}$$ (15)

where $d_{(C_i,C_j)}$ is the distance between cluster C_i and C_j and $intra(C_i)$ the intra-cluster function of the cluster.

- Root-Square (RS): also called coefficient of determination. It measures the degree of difference between classes. When it is close to 0 this means that the predictive model is weak. Otherwise, when the RS near 1 we have a strong classifier. The best partition has to be close to 1. RS is expressed by the following equation:

$$RS = \frac{\sum_{i=1}^{N_c}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N_c}(y_i - \bar{y}_i)}$$ (16)

where y_i is the observed value, $\bar{y}$ as its mean, and $\hat{y}$ as the fitted value,
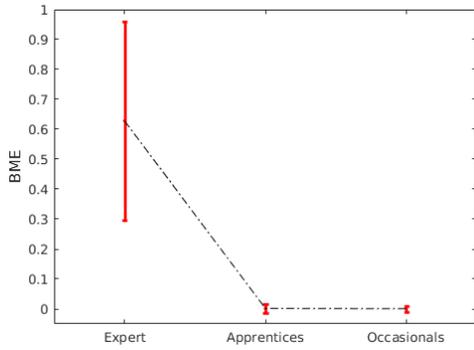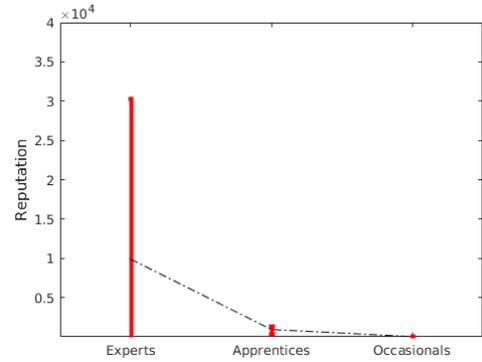
Fig. 5. Interval of confidence for BME

Fig. 6. Interval of confidence for Reputation

6. Classification results

In this section we present some classification indexes about the quality of the generated clusters.

Table 1. Statistics

	Mean	standard deviation
Reputation	264.98	$3.7189 * 10^3$
Number Accepted Answers	0.13	2.769
UpVotes	26.304	195.69
DownVotes	2.7463	108.896
Number of Answers	0.6647	9.1757
Number of Questions	0.3629	2.1626
Time Activity	556.5468	607.6323

Table 2. Indices of classification

	BME	Reputation	GMM
DB	0.397	$1.607 * 10^4$	1.023
CH	399.8	$9.2484 * 10^3$	$1.58 * 10^6$
Dunn	2.8	$9.29 * 10^{-5}$	0.117
RS	0.805	0.132	0.0534
Error RS	0.141	0.328	0.186

The results of the classification evaluation are presented in Table 2. Here, we compare the indices of our belief classifier with the reputation system of Stack Overflow and the Gaussian Mixture Model. For the reputation, we use the method described in[1], where experts have a reputation greater than 2400. We experiment the GMM with 3 clusters. The clustering output suggests that in fact there are three types of users in the platform.

The DB results prove that the proposed approach (BME) presents a better classification when studying the homogeneity of every class as the best partition has to be the smallest. However, the GMM presents a better partition when considering the intra-class variance for the Dunn criterion. Thus, the BME has better results than the reputation system. The Dunn criterion examines the distances between the clusters. The value of the reputation based method and the GMM both have small values (< 1) unlike our belief measure of expertise with a value of 2.8. The index of Dunn maximizes the inter-class distance while minimizing the intra-class distance presented in Table 2. This index has to be maximized. For the Root Square (RS) which reflects the degree of difference between classes, the BME has a value of 0.805. As the best partition must be close to 1, the belief expertise measure presents a better partition for the classification. This is confirmed with rate of Errors RS.

When we compare the results of the several indices of classification evaluation, the method based on the theory of belief functions outperforms the other approaches. The three classes generated by our belief model present a more stable classification and especially a better detection of experts.

Later, we calculate the confidence interval with 95% for every class generate by the BME and Reputation. Results are described in Figures 5 and 6. Confidence intervals consist of a range of values (interval) that estimates the unknown intra-class parameters. We are 95% confident that the true value of Belief Measure of Expertise for experts is between $[0.3, 1]$.

On Figure 6, we can see that the interval of experts' reputation is very large in $[2400, 35000]$. Thus a person whose reputation is equal to 2400 is an expert as an other contributor with 30000 reputation points. The fact that the reputation can not be enclosed, users may gain more and more points with no limitations, their expertise can not be similar to how the community member consider them based only on the reputation. As the values of the BME are in a limited interval of $[0, 1]$ and this measure takes into account both user's contributions and the time spent in the

platform, this allows us to have an overview on how does a user evolve in the community. The BME is a general measure of expertise that does not take into account topical issues because of the popularity of some of them. This belief degree allows us to detect general expert users in the platform of Stack Overflow.

7. Conclusion

In this paper we propose a general measure of expertise based on the theory of belief functions. First, we presented the limitations of the reputation measure proposed by Stack Overflow and several other approaches. The main issue is that none of the other methods takes into account data imperfection. With the theory of belief functions, we can manage these imperfections and deal with them. Thus, we propose a belief measure expertise that considers users attributes and the time they spend on the platform. This metric is based on the combination of users' features which allows us to have a global overview about how they behave. The performance of this measure is evaluated on real data from Stack Overflow. The results of the BME are then compared with the reputation measure of Stack Overflow and a GMM. As future work we will focus on analyzing time series of users' behavior in order to identify topical future potential experts.

References

1. Movshovitz-Attias, D., Movshovitz-Attias, Y., Steenkiste, P., Faloutsos, C.: Analysis of the reputation system and user contributions on a question answering website: Stackoverflow. In: Rokne, J.G., Faloutsos, C. (eds.) ASONAM. pp. 886893. ACM (2013)
2. Bouguessa, M. B. Dumoulin, and S. Wang .: Identifying authoritative actors in question-answering forums: the case of Yahoo! answers. In: KDD. pp. 866874 (2008).
3. Bouguessa, M., Romdhane, L.B.: Identifying authorities in online communities. ACMTIST 6(3), 30 (2015)
4. CaliÅski, T., Harabasz, J.: A dendrite method for cluster analysis. Communications in Statistics-Simulation and Computation 3, 1-27 (1974)
5. Davies, D.L., Bouldin, D.W.: A cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. 1(2), 224- 227 (1979)
6. Denoeux, T.: A k-nearest neighbor classification rule based on Dempster-Shafer theory.IEEE Trans. Systems, Man, and Cybernetics 25(5), 804-813 (1995)
7. Denoeux, T., Kanjanatarakul, O.: Evidential clustering: A review. In:IUKM. vol. 9978, pp. 24-35 (2016)
8. Dempster,A.P.: Upper and Lower probabilities induced by a multivalued mapping, Annals of Mathematical Statistics, vol. 38, (1967)
9. V Dijk, D., Tsagkias, M., de Rijke, M.: Early detection of topical expertise in community question answering. In: SIGIR. pp. 995998. (2015)
10. Dunn, J.: Well separated clusters and optimal fuzzy partitions. Journal of Cybernetics 4, 95-104 (1974)
11. Ehrlich, K., Shami, N.S.: Searching for expertise. In: CHI 08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems. pp. 1093–1096 ACM, Florence, Italy (2008).
12. Ericsson, K.A.: The Cambridge Handbook of Expertise and Expert Performance. Cambridge: Cambridge University Press (2006)
13. Guy, I., Avraham, U., Carmel, D., Ur, S., Jacovi, M., Ronen, I.: Mining Expertise and Interests from Social Media. In: WWW Brazil. pp. 515–526 , (2013).
14. Kasneci, Gjergji, GAEL, Jurgen Van, STERN, David, H. CoBayes: Bayesian knowledge corroboration with assessors of unknown areas of expertise. , WSDM, ACM, 9, p. 465–474, (2011).
15. Khaleghi, B., Khamis, A. M., Karray, F., Razavi, S.N.: Multisensor data fusion: A review of the state of the art. Information Fusion 14, (2013)
16. Page, L. Brin, S. Motwani, R. and Winograd T., The PageRank citation ranking: Bringing order to the Web, (1999)
17. Nguyen, V.D., Huynh, V.N.: Integrating with social network to enhance recommender system based-on Dempster-Shafer theory. In: CSoNet, vol. 9795, pp. 170-181. Springer (2016)
18. Shafer, G.: A mathematical theory of evidence, Princeton University Press (1976)
19. Sahu, T.P., Nagwani, N.K., Verma, S.: Multivariate beta mixture model for automatic identification of topical authoritative users in community question answering sites. IEEE Access 4,5343-5355 (2016)
20. Smets, P.: Constructing the pignistic probability function in a context of uncertainty, Uncertainty in Art Intelligence, vol. 5, pp. 29-39 (1990)
21. Smets, P.: The combination of evidence in the Transferable belief Model. IEEE Trans on Pattern Analysis and Machine Intelligence, vol. 12, pp. 447-458 (1990)
22. Smets, P., Kennes, R.: The transferable belief model, Art Intelligence. 66, pp. 191-234 (1994)
23. Pal, A., Harper, F. M., Konstan, J. A.: Exploring question selection bias to identify experts and potential experts in community question answering. ACM Trans. Inf. Syst. 30(2) (2012)
24. Pal.A, Chang S, and Konstan J: Evolution of Experts in Question Answering Communities. In ICWSM. The AAAI Press (2012).
25. Tang, X., Yang, C.C.: Ranking user influence in health care social media. ACM TIST 3(4), 73 (2012)
26. Yang, J., Tao, K., Bozzon, A., Houben, G.J.: Sparrows and owls: Characterization of expert behavior in Stack Overflow. In: UMAP. vol. 8538, pp. 266277. Springer (2014)
27. Zhang, J. Ackerman, M. S. and Adamic L., Expertise networks in online communities: Structure and algorithms, in Proc. 16th Int. Conf. World Wide Web, pp. 221230, (2007).