ISSN 2007-9737

A probabilistic Profile Reliability Approach to Improve the Richness of User's Interests Based on Social Information

Leila Ghorbel[1], Corinne Amel Zayani[1], Ikram Amous[1], Imen Boutouria[2]

[1] Sfax University,
MIRACL-ISIMS,
Tunisia

[2] Sfax University,
Faculty of Sciences of Sfax,
Laboratory of Probability and Statistics,
Tunisia

leila.ghorbel@gmail.com, {corinne.zayani, imen.boutouria}@fss.rnu.tn, ikram.amous@isecs.rnu.tn

Abstract. The user requires personalized information that depends on his/her current profile information (needs, interests, preferences, etc.). Interests are the most important information of the user's profile. There are many sources which may provide beneficial information to model the user's interests, such as social neighbors' profiles. However, the latter may provide conflicting interests (erroneous, duplicated, out of date, ambiguous, etc.) and consequently they can be considered as non-reliable sources. In this paper, we propose a probabilistic approach to handle conflicts by detecting reliable profiles so as to improve the richness of user's interests. Our approach is different from those that are previously proposed as it takes into account the organizational aspects of interests in terms of their evolutionary aspect (freshness and popularity) as well as their semantic relationships. Our experiment was conducted in a social-learning context in order to take the case of the improvement of the learner's profile content based on his/her social network. The experiment led to satisfactory results.

Keywords. User's profile, interests, reliability, conflict resolution, probability.

1 Introduction

For years, taking into account the user's profile information, particularly the user's interests, has been significant in different systems (adaptive [2], recommender [27], etc.), in terms of returning adapted results to the user according to his/her information.

There are many sources which may stand as beneficial information to improve the richness of the user's interests such as the social behavior (tagging behavior), the social network (neighbors) or the distributed profiles (e.g. his/her profiles in e-learning systems). For example, in an e-learning system, the user's (learner's) profile may contain incomplete interests. Therefore, no learner adaptation can be fulfilled. As a consequence, there is a strong need to improve the learner's profile content based on different sources. On the one hand, this improvement can be performed based on the learner's profiles existing in other e-learning systems [18, 9]. On the other hand, it can be performed based on the learner's social behavior as a rich source of information [19, 7].

However, the user may be influenced in a negative way by his/her neighbors. In fact, the neighbors' profiles may contain conflicting interests (erroneous, duplicated, out-of-date, ambiguous, etc.) and consequently they can be considered as non-reliable sources. In fact, the spread of conflicting interests over the profiles enable to provide irrelevant results to the user. For this reason, conflicting interests should be resolved.

Conflict resolution is still a critical problem in different fields such as health-care, crowd-sourcing and information extraction.

In literature, two main methods [16] for conflict resolution are distinguished, which are majority voting and source reliability. These methods are based on non-organized information in terms of their evolutionary aspect as well as their semantic relationships. In fact, the user's profile as a source is organized either by deleting the out-of-date interests (by using the machine learning techniques [33]) or keeping some interests relative to a specific period of time (by using the notion of temperature which includes the freshness and popularity of interests [17, 19, 31]) without taking into account the semantic relationship between interests.

In this paper, we propose a probabilistic approach to detect reliable neighbors' profiles of a user resting on organized profiles. Each profile is organized by taking into account the evolutionary aspect of interests as well as their semantic relationships.
This approach was validated, in social-learning context, based on learners who are members of Moodle e-learning system and Delicious social network. Results show that our probabilistic approach based on organized profiles improves the results of conflict resolution by detecting the reliability weights of profiles.

In the remaining part of this paper, some existing studies about profile reliability and conflict resolution are presented. Afterwards, our approach is established before it is evaluated. Finally, the closing section concludes the manuscript and offers some prospects for further works.

2 Related Works

Generally, improving the richness of a profile information (interests, preferences, etc.), based on other profiles, depends on the reliability of the sources (profiles) where the information is stored [26, 14]. This information should be evacuated from conflicts. These latter mean that information in different sources may be: i) semantically similar (having different values) and ii) up-to-date in a source and out-of date in another one. For this reason, conflict resolution approaches appear to detect the reliability of sources and trustworthy information (data value) in these sources [16].Conflict resolution is performed through two methods which are the majority voting and source reliability [16].

The majority voting method consists in merging in the corresponding source information with the highest number of occurrences existing into the other sources (profiles). The major shortcoming of this method is that it assumes that all sources providing information are equally reliable [16]. As a matter of fact, the second method emerged in order to estimate the source reliability degrees and infer true information. The sources providing true information will be assigned higher reliability, and information supported by reliable sources will be regarded as true information.

Source reliability method has emerged as a powerful tool to resolve conflicts. In literature, we find three main categories of the latter method [16] that distinguish between reliable and non-reliable sources by inferring their reliability degrees and derive true information [5, 20, 15, 35, 13].

In the first category, the source weight reliability is calculated iteratively until convergence. Then, the true information can be inferred through weighted aggregation, such as weighted voting in [20, 8, 5, 13]. The second category is based on an optimization formulation which is based on a distance function that measures the distance between a source and the identified truth [15]. The third category is based on a probabilistic graphical model [12] as a relay between the source weight and the identified truth [34, 35, 23].

Authors in [16] claimed that the first category is easier to understand and interpret, while the optimization and the probabilistic graphical model based categories (the second and the third categories) are interpretable but complex and need more explanation. These categories of source reliability method adopt some characteristics that are summarized in three aspects: input information, source reliability and output.

The input aspect describes the pre-processing of the input information which can be duplicated. The freshness value stands for a crucial criterion to solve this problem [22]. Moreover, the input

information can be structured [5, 20, 4, 26] or unstructured [30, 3], categorical [8, 11] or continuous [20].

The source reliability aspect describes the used assumptions the most popular of which are related to the source dependency. Some studies assume that sources are independent as they do not copy information from one another [28, 8, 20, 15, 29].

Other studies assume that sources are dependent [5, 24, 6, 10, 14]. The authors adjust the weight of each source based on the copying relationship between sources. In fact, a source can copy information from a non-reliable source (direct copying) and this information can be copied to another source (co-copying, transitive copying).

As for the output aspect, these approaches use either the labeling technique [20] which assigns a label (true or false) to each information or the scoring technique which assigns a score to each information in the form of a probability [5].

A deeper look in literature reveals that most of popular approaches rely on the probabilistic source reliability method leading to accurate results. However, these approaches do not take into account the semantic relationship between information which can be structured [5, 20, 4]. In fact, information can be semantically dependent from one source to another, which affects the reliability results. In addition, some approaches have ignored information evolution over time in each source [20, 6, 15, 4, 10], which improves the result of source reliability in [29, 30, 11, 26, 14].

In fact, the user's profile as a source contains different values, especially those of interests which are semantically dependent on each other in the same period of time. The user's interests can be characterized by their temperature values [17] including their freshness and popularity values that change over time.

Thus, we need to know how far these interests are up-to-date (freshness) and popular and decide whether some of them will improve the richness of the profile content. As a consequence, the organization of the user's profile by taking into account the semantic relationship between interests and their temperature values is required.

Moreover, we notice that the majority of approaches that are based on source dependency assumption assign scores to sources [29, 6, 10, 14]. These approaches are motivated to choose the scoring technique because sources have a certain probability (score) to be reliable. However, with the labeling techniques [20, 26] this information is lost.

In this paper, we propose a probabilistic profile reliability approach to improve the richness of user's interests based on his/her social information that are stored in the profiles of his/her friends or neighbors.

The originality of this approach resides in the fact that it detects reliable neighbors' profiles of a user by applying some aspects of the source reliability method based on organized profiles. A user's profile is organized by generating a semantic and hierarchical structure of the user's interests based on our previous study [32]. In this study, we have generated a semantic and hierarchical interest structure based on the K-means machine learning algorithm by using two features: i) the temperature of interests (freshness and popularity), which deals with the evolutionary aspect of the interests over time and ii) the semantic relationship between interests, which deals with the interest dependency, duplication and so on. With the semantic and hierarchical structure, the processing of interests becomes easier and more meaningful.

3 Proposed Approach

In this section, the principle of the proposed approach is described. Afterwards, the proposed probabilistic approach is detailed.

3.1 Principle of the Proposed Approach

The proposed approach attempts to detect the reliability weights of neighbors' profiles in order to extract interests for user's profile improvement. In fact, a neighbor's profile may be non-reliable as it may provide conflicting interests, namely false interests.

We consider $\mathscr{P}$ a set of neighbors' profiles relative to a user: $\mathscr{P} = \{p_1, p_2, ..., p_p\}$. Each profile contains a set of interests each of which is represented by

three elements: a word (name), its freshness and its popularity.

The proposed approach takes into account the following aspects (cf. Section 2):

— Input information: most of the proposed approach used only structured information for conflict resolution [5, 20, 4]. However, we consider in the proposed approach the temperature values (freshness and popularity) and semantic similarity measures in order to represent interests in a semantic and hierarchical structure based on the hierarchization process that is proposed in our previous work[32].

— Profile dependency: assuming that profiles are dependent. In fact, a profile can contain a copy of false interests coming from a non-reliable profile. The dependency between two profiles is computed based on our probabilistic approach which takes into account the semantic and hierarchical structure of interests of each profile.

— Output: assigning a weight (score) to each interest in the form of probability. The weight is calculated on the basis of the profile dependency result and the temperature values. Therefore, the weights of interests are aggregated to compute the reliability weight of each profile.

The proposed probabilistic approach takes into account three assumptions that we have inspired from the proposed approaches [6, 24, 10, 14]. We have adapted these assumptions for our approach as follows:

Assumption 1 : A profile is reliable for a user only if it contains his/her interest true values. In fact, a profile may contain two interest types: i) local (true) interests, which are not provided (e.g. interests related to the keywords of the visited resources) and ii) provided (copied) interests from other profiles. A copied interest may be considered false because a false value can be spread through copying, which reduces the reliability weight. However, a copied interest may be considered true with the increase of its temperature values (freshness and popularity) or if there are other interests that are semantically close to this copied interest and have better temperature values. Figure 1 shows an extract of a semantic and hierarchical structure of user's interests. These latter are local (eg. computer, learning, etc.) or provided (eg. web, science, etc.)

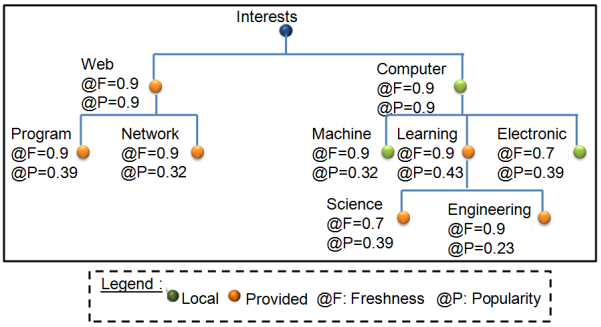

Fig. 1. An extract of the semantic and hierarchical structure of user's interests

Assumption 2 : the profiles are dependent through the copying relationship. In other terms, the dependence of p_1 on p_2 is relative to the number of false interests that are copied from p_2 to p_1.

Assumption 3 : the dependence relationship between profiles is acyclic (i.e. the dependence of p_1 on p_2 is different from the dependence of p_2 on p_1)

3.2 Probabilistic Profile Dependency

The dependency of p_1 on p_2 is the probability $P(p_1 \rightarrow p_2)$ of having a number of false interests in p_1. Denoting $\mathscr{P}1 = \{I1_1, I1_2, ..., I1_{n1}\}$ and $\mathscr{P}2 = \{I2_1, I2_2, ..., I2_{n2}\}$ are the sets of n_1 and n_2 interests respectively in p_1 and p_2. An interest in p_1 is false if one of these conditions is satisfied:

— It figures with the same name in p_2, it is provided (copied) in p_1 and local in p_2 or it is provided in p_1 and p_2 and its freshness value in p_1 is lower than that in p_2.

— Its parent in the hierarchy is false.

Thus, the probability $P(p_1 \rightarrow p_2)$ varies according to the number of false interests in p_1.

As a matter of fact, a discrete random variable X is characterized by a set of values k∈{0, 1, 2, ..., n_1} and by a mathematical expression of the probability of these values. This expression is called the probability distribution of the X random variable [12].

This X variable satisfies the required conditions [12] to follow a Binomial distribution $\mathscr{B}$ with parameters n_1 and p. Probability p is the discrete probability distribution of the amount of success in a sequence of n_1 independent experiments. The amount of success expresses the number of having k false interests in p_1. For example, p(X=5) is the probability of having 5 false interests (successes) in p_1. In order to compute the probability of success p, a tree of probability is constructed as illustrated in figure 2. The probability symbols that figure in the tree are described in table 1.

The illustrated probabilities depend on the probability of existence of an interest I in $\mathscr{P}2$ denoted p(E). The mathematical expression of this probability is also determined through a discrete random variable, which is characterized by a set of values $\alpha \in \{0, 1, 2, ..., n_1\}$. These values constitute the number of interests existing in $\mathscr{P}1$ and $\mathscr{P}2$. Thus, the equality of p(E) is as follows:

$$p(E) = \binom{\alpha}{n_1} * (\frac{1}{2})^{n_1}.$$

Based on probability tree, probability p represents the intersection of different probabilities. It is detailed in the following equality:

$$\begin{aligned} p =& 10 * (1 * \frac{1}{2} * \frac{1}{2} * \frac{1}{2} * \frac{1}{4} * \binom{\alpha}{n_1} * (\frac{1}{2})^{n_1}) \\ &+ 1 * \frac{1}{2} * \frac{1}{2} * (1 - \binom{\alpha}{n_1} * (\frac{1}{2})^{n_1}) \\ =& \frac{3}{2^{n_1+4}}\binom{\alpha}{n_1} + \frac{1}{4}. \end{aligned}$$

As a consequence, the Binomial distribution $\mathscr{B}$ is defined by:

Table 1. Probabilities description

Probability symbols	Description
E	I exists in p_2
$\bar{E}$	I does not exist in p_2
A	I is provided in p_1 and is local in p_2
B	I is provided in p_1 and p_2
C	I is local in p_1 and is provided in p_2
D	I is local in p_1 and p_2
H	I has a parent in p_1
$\bar{H}$	I is the parent in p_1
R	The parent of I in the hierarchy of p_1 has a false value
$\bar{R}$	The parent of I in the hierarchy of p_1 has a true value
F1	The freshness value of I in p_1 is lower than its freshness value in p_2
$\bar{F}1$	The freshness value of I in p_1 is higher than its freshness value in p_2
F	I has a false value
T	I has a true value

X $\sim \mathscr{B}(n_1, \frac{3}{2^{n_1+4}}\binom{\alpha}{n_1} + \frac{1}{4})$ and

$$\begin{aligned} p(X = k) = \binom{k}{n_1} * (\frac{3}{2^{n_1+4}}\binom{\alpha}{n_1} + \frac{1}{4})^k * \\ (\frac{3}{4} - \frac{3}{2^{n_1+4}}\binom{\alpha}{n_1})^{n_1-k}. \end{aligned}$$

The profile p_1 depends on p_2 in cases where the number of false interests in p_1 exceeds half of the overall interests in i) p_2 if $n_1 > n_2$ or ii) p_1 if $n_1 < n_2$. Thus, the probability of dependency of p_1 on p_2 is the sum of the probabilities for having more than $\frac{min(n_1,n_2)}{2}$ false interests until n_1 ($p(\frac{min(n_1,n_2)}{2} <$

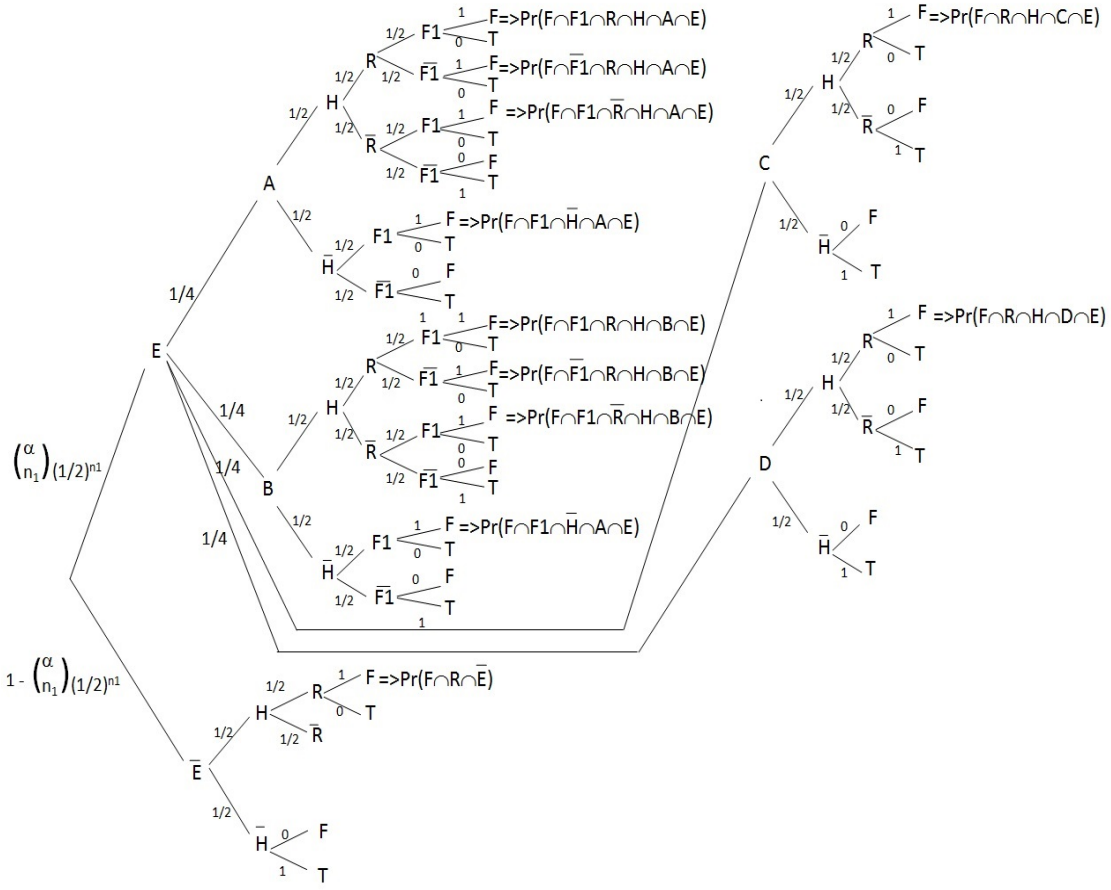

Fig. 2. Probability tree

$X \leq n_1$)). This probability is defined in equation 1:

$$p(\frac{min(n_1,n_2)}{2} < X \leqslant n_1) = p(p_1 \to p_2) = \sum_{k=E(\frac{min(n_1,n_2)}{2})+1}^{n_1} p(X=k), \quad (1)$$

when n_1 exceeds 30, the computation of $pr(\frac{min(n_1,n_2)}{2} < X \leqslant n_1)$ is more complex [12].

In this case, an approximation of this Binomial distribution by a normal distribution should be achieved. This approximation is detailed below:

$$X \sim \mathscr{B}(n_1,p) \approx X \sim \mathscr{N}(n_1 p, \sqrt{n_1 p(1-p)}) \approx X \sim \mathscr{N}(n_1(\frac{3}{2^{n_1+4}}\binom{\alpha}{n_1}+\frac{1}{4}) \sqrt{n_1(\frac{3}{2^{n_1+4}}\binom{\alpha}{n_1}+\frac{1}{4})} * \sqrt{(\frac{3}{4}-\frac{3}{2^{n_1+4}}\binom{\alpha}{n_1})}).$$

The X random variable is transformed into T which follows the centered reduced normal distribution $\mathscr{N}(0,1)$. This transformation is portrayed in the

following equality:

$$T = \frac{X - n_1 p}{\sqrt{n_1 p(1-p)}},$$

$$p(\frac{min(n_1,n_2)}{2} \leqslant Y \leqslant n_1) \approx p(\frac{\frac{min(n_1,n_2)}{2} - n_1 * p}{\sqrt{n_1 p(1-p)}} < T \leqslant \frac{n_1 - n_1 * p}{\sqrt{n_1 p(1-p)}}).$$

Consequently, the probability of dependency of p_1 on p_2 defined in equation 1 is approximated in equation 2:

$$p(p_1 \rightarrow p_2) = \Phi(\frac{n_1 - n_1 * p}{\sqrt{n_1 p(1-p)}}) - \Phi(\frac{\frac{min(n_1,n_2)}{2} - n_1 * p}{\sqrt{n_1 p(1-p)}}). \quad (2)$$

3.3 Profile Reliability Weight

In order to compute the profile reliability weight, an algorithm (cf. algorithm 1) is proposed. This algorithm browses the semantic and hierarchical structure of each profile $p_i \in \mathscr{P}$. Therefore, it assigns a weight to each interest I in p_i by taking into account the temperature values and the result of the probabilistic dependency of p_i on the other profiles in $\mathscr{P}$ (cf. equation 3). Thus, the weight of each interest is calculated as the sum of the products of these two main values relative to each profile $\in \mathscr{P}$. The first value represents the product of the freshness and popularity of the interest in p_j.

The second value is the probability of non-dependency (1-$p(p_i \rightarrow p_j)$) which is proposed by source reliability approaches that assume the dependency between sources [29, 6, 10, 14]. Through this probability, the weight of dependency is subtracted from the first value (product of the freshness and popularity).

Therefore, the weight of the interest decreases which implies also the decrease of the reliability weight:

$$\text{Weight}(I[p_i]) = \frac{\sum_{j=0}^{j=\text{SizeOf(p)}} \text{Temperature}(I[p_j]) * (1\text{-p}(p_i \rightarrow p_j))}{\text{SizeOf(p)}}. \quad (3)$$

Finally, the algorithm aggregates the weights of interests in p_i in order to compute its reliability weight (cf. equation 4). This weight is the quotient of the sum of its interest weights by the total number of interests in p_i (P[i]):

$$\text{Reliability}(p_i) = \frac{\sum_{n=1}^{n_i} \text{Weight}(I_n[p_i])}{n_i}. \quad (4)$$

Algorithm 1 *Profile reliability*

Require: $P\ arraylist\ of\ profiles\ related\ to\ a\ user$: $tree$
Ensure: $reliability\_weight\ arraylist\ of\ size\ of\ P$: $Real$
1: **for** $(i = 0, i < P.length, i++)$ **do**
2: $reliability\_weight[i] = reliability(i, P)$;
3: **end for**
4: **Function** $reliability(i, P)$
5: **begin**
6: **for each** $Interest\ in\ P[i]$ **do**
7: $update\ its\ weight\ based\ on\ equation\ 3$;
8: $update\ the\ reliability\ weight\ of\ p[i]\ based\ on\ equation\ 4$;
9: **end for**
10: **End Reliability**
11: **End**

end

4 Evaluation

In this section, datasets and metrics used for the evaluation are described. Then, the obtained results are presented.

4.1 Datasets and Metrics

Our experiment was conducted in a social-learning context in order to take the case of the improvement of the learner's profile content based on his/her social network. For this reason, we are based on learners who are at the same time members of the e-learning system Moodle[1] and the social network Delicious[2]. Moodle contains the learner's interests which are explicitly provided

[1] https://Moodle.org
[2] https://del.icio.us/

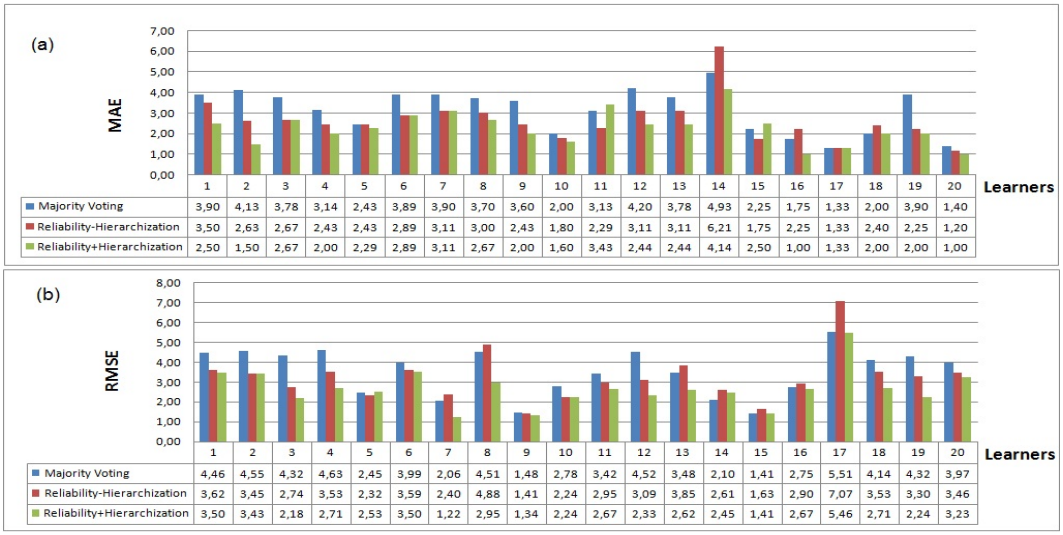

Fig. 3. Profile reliability evaluations with MAE and RMSE

by the learner or implicitly based on the visited and learned courses or lessons belonging to various domains. Delicious dataset contains social networking, bookmarking, and tagging information. Besides, it provides information about the user's friend relationships and the tagging behaviors ($< user, tag, resource >$). Tagging behavior is provided according to time (day, month, year, hour, minute and second). A tag reflects a user's interest, which is described in terms of its name, freshness and popularity. It can be related to an educational resource and can improve the richness the user's interests in Moodle.

As far as the evaluation is concerned, we extracted from Delicious the profiles of friends (explicit neighbors of each user) of a set of learners in Moodle. We applied the proposed approach in order to detect the profile reliability weights so as to improve the richness of interests of each learner in Moodle. Afterwards, we assessed our probabilistic approach to detect the reliability weight for each profile of a learner's friend.

Table 2 presents some characteristics of our dataset.

In order to check the accuracy of the detected profile reliability weights, we calculated, firstly, the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) [1, 25].

Table 2. Description of the dataset

Description	Number
Learners	100
Generated hierarchies	545
Average of neighbors (profiles) per learner	6
Average of interests in a profile	79

To calculate these metrics, each learner in Moodle was provided with the profile of his/her friends in Delicious for manual ranking based on their reliability. Then, these profiles were ranked based on their reliability weights. Therefore, MAE (cf. equation 5) was computed with the deviation between predicted rank (p_i) and manual rank (r_i) which constitutes the sound truth of a profile:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|p_i - r_i|\,. \tag{5}$$

RMSE (cf. equation 6) is similar to MAE. What differs is that much more emphasis is put on larger deviation. Smaller MAE or RMSE indicates better accuracy:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - r_i)^2}. \tag{6}$$

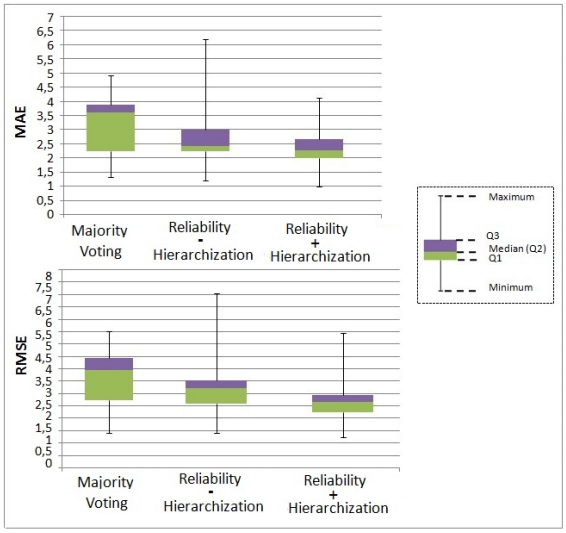

Fig. 4. Boxplots of MAE and RMSE values for: (i) majority voting, (ii) reliability without hierarchization, and (iii) reliability with hierarchization

4.2 Evaluation of the Profile Reliability Result

Through this evaluation, we attempt to validate the impact of our probabilistic approach based on the semantic and hierarchical structure of interests and the temperature values for profile reliability weight detection.

For this reason, the potential and superiority of our approach is checked compared to some existing approaches. For each user, 3 values of MAE and RMSE are computed. At a first stage, the accuracy of the majority voting approach is evaluated, which assigns the highest weight to the profiles having the highest number of interests that figure in the majority of other profiles. This approach does not require computing the dependency weight between profiles.

At the second stage, the accuracy of reliability results is assessed based on the dependency aspect without taking into account the semantic and hierarchical structure of interests (hierarchization). Finally, the accuracy of reliability results is assessed based on our probabilistic profile reliability approach.

Figure 3 demonstrates the evaluation results relative to a sample of 20 learners. The results illustrated in figure 3.a show a clear improvement of MAE in our approach for all learners. We notice that the results from the majority voting (3.16) as well as those from dependency-based approach without hierarchization of interests (2.64) are not very satisfactory. One possible reason to account for these results is that interests are regarded as separate (there is no relationship between interests either in semantic or in temperature). However, considering the semantic and hierarchical interest structure, the results have largely improved and remain always the best with the lowest MAE average value (2.28).

Figure 3.b illustrates RMSE results which confirm that our approach remains the best with the lowest RMSE average value (2.67). This value indicates that there is a little deviation between the detected ranks and the manual ranks compared to other approaches in which an RMSE average values of 3.54 and 3.23 is recorded. For more clarification about the results relative to all learners (100), we represent, in figure 4, the MAE and RMSE values in box plots.

Each box plot summarize the MAE or RMSE values, for each approach of comparison, through visualizing five values which consists of the minimum value, first quartile (Q1), median (Q2), third quartile (Q3), and maximum value. The rectangle exhibits all the values situated between Q1 and Q3, which are 25 % of the values situated below Q1, and 25% of the values situated above Q3. Thus, the inter-quartile range corresponds to 50% of the values situated in the central part.

Figure 4 shows that the median of MAE (RMSE) values of our approach is lower than the medians of the majority voting and profile reliability approach that do not take into account the hierarchization of interests. The inter-quartile ranges are reasonably similar, though the overall range of the MAE (RMSE) values of our approach is lower than that in the other approaches (as shown by the distances between the minimum and the maximum values for each box plot). The overall conclusion is that the proposed approach does vary with MAE (RMSE) values, with the other approaches having, on average, larger MAE (RMSE) values.

Having proved the effectiveness and feasibility of our approach compared to others, we attempt to

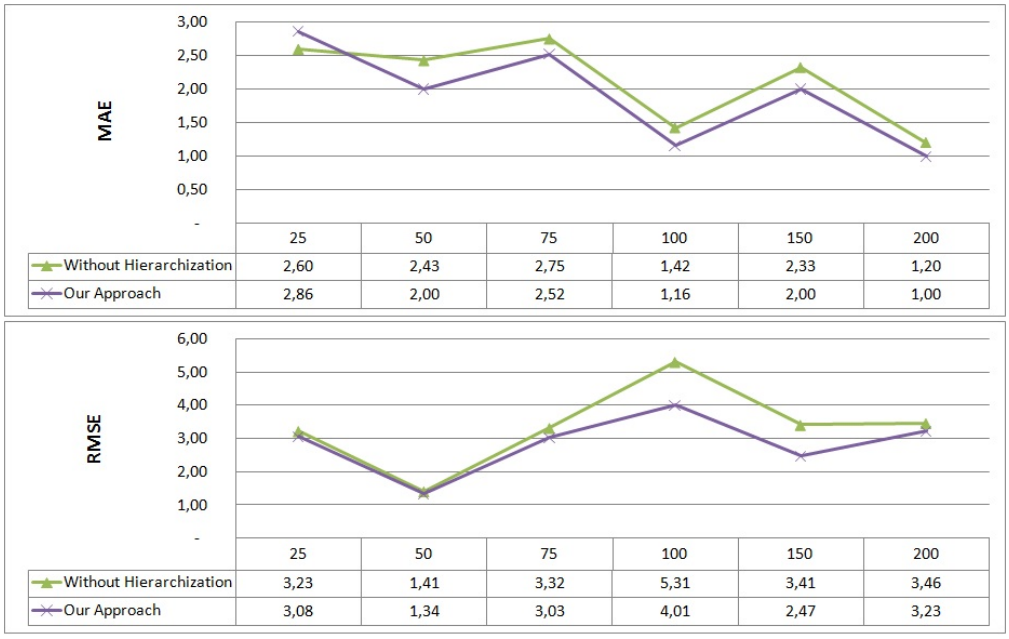

Fig. 5. Accuracy of the proposed approaches by the variation of the number of interests

evaluate the impact of the number of interests on the accuracy of the reliability results.

Figure 5 illustrates the result of the average of the reliability weights of a set of profiles having a number of interests varying between 25 and 200. This result is further evidence proving the efficiency of our approach regardless of the evolving number of interests.

5 Conclusion

In this paper, we have detailed our probabilistic profile reliability approach which leads to improve the richness of user's interests based on the profiles of his/her neighbors in social networks. This approach attempts to resolve conflicts which may occur in the neighbors' profiles based on a semantic and hierarchical structure of each neighbor's interests. It consists in detecting the profile reliability weights which rely on the probability of dependency between profiles.

We have validated the proposed approach, in a social/learning context, based on learners/users in Moodle and Delicious. Results show the effectiveness of our approach and prove that it can improve the learner's interests by resolving conflicts. However, these results should be improved by the variation of the reliability weight computation way.

For example, we can fix a threshold relative to the level of the semantic and hierarchical structure of interests and aggregate the weights of interests that exist in the higher levels compared to this threshold. This way shed more light on the effectiveness of the use of the semantic and hierarchical structure of user's interests within a profile reliability approach.

In future works, we aspire to apply our approach based on other social networks so as to improve the user's interests from different profiles. Therefore, a great number of profiles which may disturb the user's satisfaction need to be reconsidered for conflict resolution. In order to respond to the user's satisfaction, some approaches have inserted to the concept of intrusion detection [21] which is meant to detect and react to the presence of unauthorized users of a network. Likewise, we tend to detect

unauthorized users (profiles) as a first step before applying the proposed approach.

References

1. **Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013).** Recommender systems survey. *Knowledge-based systems*, Vol. 46, pp. 109–132.
2. **Brusilovsky, P. & Millán, E. (2007).** User models for adaptive hypermedia and adaptive educational systems. *The adaptive web*, Springer-Verlag, pp. 3–53.
3. **Derczynski, L., Maynard, D., Rizzo, G., van Erp, M., Gorrell, G., Troncy, R., Petrak, J., & Bontcheva, K. (2015).** Analysis of named entity recognition and linking for tweets. *Information Processing & Management*, Vol. 51, No. 2, pp. 32–49.
4. **Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., Strohmann, T., Sun, S., & Zhang, W. (2014).** Knowledge vault: A web-scale approach to probabilistic knowledge fusion. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp. 601–610.
5. **Dong, X. L., Berti-Equille, L., & Srivastava, D. (2009).** Integrating conflicting data: the role of source dependence. *Proceedings of the VLDB Endowment*, Vol. 2, No. 1, pp. 550–561.
6. **Dong, X. L., Berti-Equille, L., & Srivastava, D. (2013).** Data fusion: resolving conflicts from multiple sources. In *Handbook of Data Quality*. Springer, pp. 293–318.
7. **Farnadi, G., Tang, J., De Cock, M., & Moens, M.-F. (2018).** User profiling through deep multimodal fusion. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, ACM, pp. 171–179.
8. **Galland, A., Abiteboul, S., Marian, A., & Senellart, P. (2010).** Corroborating information from disagreeing views. *Proceedings of the third ACM international conference on Web search and data mining*, ACM, pp. 131–140.
9. **Ghorbel, L., Zayani, C. A., Amous, I., & Sèdes, F. (2018).** Semi-supervised algorithm with knowledge-based features for learner's profiles interoperability. *International Journal of Technology Enhanced Learning*, Vol. 10, No. 1-2, pp. 137–159.
10. **Huang, C. & Wang, D. (2016).** Topic-aware social sensing with arbitrary source dependency graphs. *Proceedings of the 15th International Conference on Information Processing in Sensor Networks*, IEEE Press, pp. 7.
11. **Huang, C., Wang, D., & Mann, B. (2017).** Towards social-aware interesting place finding in social sensing applications. *Knowledge-Based Systems*, Vol. 123, pp. 31–40.
12. **Johnson, N. L., Kemp, A. W., & Kotz, S. (2005).** *Univariate Discrete Distributions, Set*, volume 444. John Wiley & Sons.
13. **Li, F., Dong, X. L., Langen, A., & Li, Y. (2017).** Discovering multiple truths with a hybrid model. *arXiv preprint arXiv:1705.04915*.
14. **Li, F., Lee, M. L., & Hsu, W. (2017).** Profiling entities over time in the presence of unreliable sources. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 29, No. 7, pp. 1522–1535.
15. **Li, Q., Li, Y., Gao, J., Zhao, B., Fan, W., & Han, J. (2014).** Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation. *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, ACM, pp. 1187–1198.
16. **Li, Y., Gao, J., Meng, C., Li, Q., Su, L., Zhao, B., Fan, W., & Han, J. (2016).** A survey on truth discovery. *Acm Sigkdd Explorations Newsletter*, Vol. 17, No. 2, pp. 1–16.
17. **Manzat, A.-M., Grigoras, R., & Sèdes, F. (2010).** Towards a user-aware enrichment of multimedia metadata. *Workshop on Semantic Multimedia Database Technologies*.
18. **Martinez, M. L., González-Mendoza, M., & Valle, I. D. D. (2014).** Enrichment of learner profile with ubiquitous user model interoperability. *Computacion y Sistemas*, Vol. 18, pp. 359–374.
19. **Mezghani, M., Péninou, A., Zayani, C. A., Amous, I., & Sèdes, F. (2017).** Producing relevant interests from social networks by mining users' tagging behaviour: A first step towards adapting social information. *Data Knowl. Eng.*, Vol. 108, pp. 15–29.
20. **Pasternack, J. & Roth, D. (2010).** Knowing what to believe (when you already know something). *Proceedings of the 23rd International Conference on Computational Linguistics*, Association for Computational Linguistics, pp. 877–885.
21. **Peng, J., Choo, K.-K. R., & Ashman, H. (2016).** User profiling in intrusion detection: A review.

Journal of Network and Computer Applications, Vol. 72, pp. 14–27.

22. **Rekatsinas, T., Dong, X. L., & Srivastava, D. (2014).** Characterizing and selecting fresh data sources. *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, ACM, pp. 919–930.
23. **Rekatsinas, T., Joglekar, M., Garcia-Molina, H., Parameswaran, A., & Ré, C. (2017).** Slimfast: Guaranteed results for data fusion and source reliability. *Proceedings of the 2017 ACM International Conference on Management of Data*, ACM, pp. 1399–1414.
24. **Sarma, A. D., Dong, X. L., & Halevy, A. (2011).** Data integration with dependent sources. *Proceedings of the 14th International Conference on Extending Database Technology*, ACM, pp. 401–412.
25. **Tarus, J. K., Niu, Z., & Yousif, A. (2017).** A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining. *Future Generation Computer Systems*, Vol. 72, pp. 37–48.
26. **Varma, S., Sameer, N., & Chowdary, C. R. (2017).** Relic: Entity profiling by using random forest and trustworthiness of a source-technical report. *arXiv preprint arXiv:1702.00921*.
27. **Villanueva, D., Lagares, M., Gómez, J. M., & González, I. (2018).** Resys: Towards a rule-based recommender system based on semantic reasoning. *Computación y Sistemas*, Vol. 22, No. 3.
28. **Yin, X., Han, J., & Philip, S. Y. (2008).** Truth discovery with multiple conflicting information providers on the web. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 20, No. 6, pp. 796–808.
29. **Yin, X. & Tan, W. (2011).** Semi-supervised truth discovery. *Proceedings of the 20th international conference on World wide web*, ACM, pp. 217–226.
30. **Yu, D., Huang, H., Cassidy, T., Ji, H., Wang, C., Zhi, S., Han, J., Voss, C. R., & Magdon-Ismail, M. (2014).** The wisdom of minority: Unsupervised slot filling validation based on multi-dimensional truth-finding. *COLING*, pp. 1567–1578.
31. **Zarrinkalam, F., Kahani, M., & Bagheri, E. (2018).** Mining user interests over active topics on social networks. *Information Processing & Management*, Vol. 54, No. 2, pp. 339–357.
32. **Zayani, C. A., Ghorbel, L., Amous, I., Mezghani, M., Péninou, A., & Sèdes, F. (2017).** Semantic-based reconstruction of user's interests in distributed systems. *Computación y Sistemas*, Vol. 21, No. 3.
33. **Zghal, R., Ghorbel, L., Zayani, A., & Amous, I. (2013).** An adaptive method for user profile learning. *ADBIS: Advances in Databases and Information Systems*, pp. 126–134.
34. **Zhao, B., Rubinstein, B. I., Gemmell, J., & Han, J. (2012).** A Bayesian approach to discovering truth from conflicting sources for data integration. *Proceedings of the VLDB Endowment*, Vol. 5, No. 6, pp. 550–561.
35. **Zhao, Z., Cheng, J., & Ng, W. (2014).** Truth discovery in data streams: A single-pass probabilistic approach. *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, ACM, pp. 1589–1598.

Article received on 08/01/2019; accepted on 24/06/2019. Corresponding author is Leila Ghorbel.