Reasoning about Weighted Semantic User Profiles through Collective Confidence Analysis: A Fuzzy Evaluation

Nima Dokoohaki, Mihhail Matskin

Abstract User profiles are vastly utilized to alleviate the increasing problem of so called information overload. Many important issues of Semantic Web like trust, privacy, matching and ranking have a certain degree of vagueness and involve truth degrees that one requires to present and reason about. In this ground, profiles tend to be useful and allow incorporation of these uncertain attributes in the form of weights into profiled materials. In order to interpret and reason about these uncertain values, we have constructed a fuzzy confidence model, through which these values could be collectively analyzed and interpreted as collective experience confidence of users. We analyze this model within a scenario, comprising weighted user profiles of a semantically enabled cultural heritage knowledge platform. Initial simulation results have shown the benefits of our mechanism for alleviating problem of sparse and empty profiles.

Key words: Confidence, Fuzzy inference, Semantic user profiles, Personalization, Reasoning, Uncertainty evaluation.

1 Introduction

Increasing overload of information scattered across heterogeneous information ecosystems and landscapes, has increased the importance of user profiling. Profiling is seen as a facilitator and enabler for personalization. Personalization is a method-

---

Nima Dokoohaki
Department of Electronics, Computer and Software Systems, School of Information and Communications Technology, Department of Information and Computer Science, Royal Institute of Technology (KTH), Stockholm, Sweden, e-mail: nimad@kth.se

Mihhail Matskin
Department of Information and Computer Science, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, e-mail: misha@kth.se

ology used for filtering the information on user behalf. As a result, profiles are increasingly implemented and utilized to allow intelligent information systems to disseminate selected and filtered information to individual or group sets of users, based on gathered personal information, stored in their respective profiles. Reasoning about uncertain knowledge is increasingly important. There has been a strong emphasis on the problem of reasoning in the face of uncertainty in Semantic Web [1]. Fuzzy Logic[2] has become an important focus area to Semantic Web research community [3]. While strong attention has been given to present fuzzy ontological concepts[3], and reason about them [4], still how to process and infer the uncertain degrees and truth ranges, is of interest in many fields. Within the profiling domain certain concepts such as trust, privacy and ranking carry vague and uncertain semantics. While ontological fuzzy languages can be used to present these concepts, analyzing the fuzzy degrees of each of these notions, as well as processing them is of our interest. We have proposed a profile format [5], through which we consider trust, privacy and rank as weights to items the user has visited and they are stored in profiled records. We record values for each of these three weights, creating a multiweighted profile structure. We have used RDF as the language for presentation of profiled information. Since profile is used to reflect both interests of the user and store previous experiences of the user, we create a hybrid notion of user profile. Confidence is defined as the state of being certain. As certainty of an experience is affected by situation-dependent measures of usage, we can consider these weights as parameters affecting usage confidence. As a result, we can take each of weight-triple values and process them to model and evaluate the confidence of user during profiled experience. To this end we take a fuzzy approach, through which we process each three-weight values of profiled records and infer confidence values. We demonstrate our model in the context of the SMARTMUSEUM scenario [6], a physical exhibition of art, in which users interact with a personalized ubiquitous knowledge platform that uses profiling for providing users with their information services of their choice. The organization of the paper is as follows: following background study in section 2, our framework is presented in section 3, a simulation of our framework is presented in 4, while we conclude in section 5 and present a future work in section 6.

2 Background

User profiling has its roots in human studies. A user profile is defined as gathering of raw personal material about the user, according to Koch [7]. User profiles gather and present cognitive skills, abilities, preferences and interaction histories with the system [8]. According to Gauch et al.[8], User profiling is either knowledge-based or behavior-based. Knowledge-based approaches construct static models of users and match users to the closest model. Behavior-based methods consider the behavior as a modeling base, commonly by utilizing machine-learning techniques [9] to discover useful patterns in the behavior. Behavioral gathering and logging is used in order to obtain the data necessary to detect and extract usage patterns, according to Kobsa [10]. Personalization systems are based on user profiles, according to Gauch et al, [8]. A category of personalization techniques is based on cognitive patterns (such as interests, preferences, likes, dislikes, and goals) a user has. These methods are known as filtering and recommendation techniques [10]. They filter resources based on features (mostly metadata) extracted and gathered from a resource or according to ratings (generally weights) of a user of similar profile, according to Weibelzahl [11]. Ontologies, at the heart of Semantic Web technologies, are used to formalize domain concepts which allow describing constraints for generation or selection of resource contents belonging the domain the user is keen towards, as well as being used to formalize the user model or profile ontology that helps making decision which resources to be adapted (for instance, shown or not shown) to the user. Ontologies along with reasoning create formalization that boosts personalization decision making mechanisms, according to Dolog et al, [12] [13]. Ontological user profiles are becoming widely adopted. For instance, within the domain of digital cultural heritage, CHIP project is definitely a significant stake holder. Considerable amount of research attention has been paid to semantically formalizing the user domain [14], as well as personalization of information retrieval. Hybrid ontological user models are consumed to learn, gather, store and use personal user data, according to which semantically-enriched art works are recommended to, during both on-line and on-site visit to exhibition. We have considered utilizing hybrid user models [5], which incorporate a semantic presentation of personal information about users as well as incorporating notions of trust, privacy and ranking for items the user has interest towards in the form of weight-descriptors. Fuzzy logics have been considered as a means for mining, learning and improving user profiles [15] [16]. Fuzzy notions of trust [17] [18] [19] [20], privacy [21] and ranking [22]have been proposed. In the context of, e-commerce Multi-agent settings, a fuzzy framework for evaluating and inferring trustworthiness values of opinion of agents has been proposed, by Schmidt et al, [19]. Agents state their evaluations about a particular (trustee) agent, agent being evaluated, with respect to agent initiating the transaction (truster). We have adopted and utilized the framework to our problem. At the same time we have adopted the privacy approach, proposed by Zhang et al., [21] to privacy and ranking, while for trust evaluation, we have taken approach proposed by Schmidt et al,[19]. In addition, uncertain notions of confidence modeling have been proposed [20]. In the context of PGP key-chaining, Kohlas et al. [20], proposes a naive approach to confidence evaluation based on uncertain evidence. Considered as a close notion to trust and belief, confidence is modeled, as an important element in the fuzzy inference mechanism.

3 Fuzzy Confidence Framework

In this section we present our approach for modeling and evaluating overall confidence from the three-weight descriptors of trust, privacy and rank, assigned to semantic user profiles. We refer to the inferred resulting values as overall confidence of the users. Before we present the process first, we describe the presentation format for the profiled records containing values of trust, privacy and rank, and describe the motivation for using them. We limit the application of the profiles to the scenario that we are eager to apply our framework towards.

3.1 Presenting Profiled Weight Descriptors

In addition to interest capturing (known as a traditional approach in profiling), we assign extra weights for capturing trust, privacy and rank to user and customer profiles. As an example, in the SMARTMUSEUM [6] case, the three-weight descriptors (privacy, trust and rank) are gathered in form of sensor data, given directly by users from mobile devices which they carry during exhibition, or are unobtrusively gathered without their consensus from environmental sensors such as RFID tags, GPS location services and Wireless networks. These weights represent the perception of users with respect to their experience, in our case exhibition and visit from art and cultural artifacts [5] [6]. Rank or score presents the amount of interest a user has with respect to his/her visit. Privacy presents the secrecy of users with respect to the disclosure of their personal information. Trust describes the self-assurance of the experience of users. As a result, user has the ability to tell the system how much his/her current experience is secret and cool. Main motivation for processing these weights to profiled items is that using these extra weights we can alter and perhaps improve the behavior of the system. If services provided by platform can be seen as system's behavior sets, perhaps these weights could alter system's behaviors to an extent that system provides better and improved services. Raw values of privacy, trust and ranking are gathered during the exhibition visit from the interactive interfaces implemented on smart handheld devices. Software interface depicts the three values in form of a scale which user can change in the preferences section. During the visit, experience data (mainly items visited and weigh values assigned) are retrieved from the handheld devices by SMARTMUSEUM servers and are stored in user profiles. The structure of the profiles [9] has been specified flexible and generic enough to accommodate ontological (RDF triplets) data about visited artifacts, context of the visit and weight descriptors. As an example, following profiled slice:

```
<http://www.smartmuseum.eu/ns/context/weather#rainy,
visited,St.JeromeWriting,atDate 20081210,0.8,0.6,0.5>
```

Conveys the following semantics;

In a rainy weather (context), at a certain date (20081210), anonymous user (subject) visited (predicate) Saint Jerome writing (object) artwork and liked it very much (rank value = .80) and user trusts moderately his/her own experience (trust value= .60) and has average secrecy (privacy value= .50).

3.2 Fuzzy Confidence Modeling Process

In order to evaluate the overall confidence of a user, we extract the weight values (privacy, trust and rank) from user profiles described previously, and we process them accordingly. The process is made of two main phases; *pre-processing* and *postprocessing*. The following steps are taken in order to evaluate the overall confidence of the user: In the pre-processing phase, first step involves application of weighting methodologies to raw values. This step includes fuzzification of each of the weightdescriptors. We take different approaches per each weight-descriptor, depending on the usage and semantics of each of weight-descriptors. Second step involves defining membership functions where per each fuzzy-weight input we define a membership function which translates the linguistic fuzzy rules and axioms into fuzzy numbers and values, as members of fuzzy sets. The final step involves application of fuzzy rules where fuzzy rules, which are defined and embedded in the fuzzy rule-base, are applied to fuzzy and weighted sets. In the post-processing phase, first step involves feeding input values to membership functions, where fuzzy sets are created as a result of this process. Second step involves application of defuzzification methodology to fuzzy sets. In the following sections, we describe each step in more details.

3.2.1 Fuzzification Phase

In this phase, each weight value will be taken separately and converted into fuzzy values which are fed into fuzzy inference engine afterwards.

Weight Fuzzification (Secrecy, Self-Reliance and Opinion Weighting)

As stated, privacy in our model gives the users this ability to specify the secrecy of their experience. This allows system to treat their personal data according to their choice. Within a similar fashion, an uncertain privacy model is introduced by Zhang et al. [21], in which privacy is defined as a role that allows the user to manage personal information disclosure to persons and technologies, with respect to their privacy preferences. This motivates us to adopt this approach to our own problem. With respect to self-reliance, trust in our approach allows the user to describe if they can rely on their own experience. To model such form of uncertain trust model, we have adopted approach proposed by Schmidt et al.,[19]. The agent-oriented approach undertaken allows modeling trust as a weighted factor. In the case of rating, we take the same approach for privacy [21], with major difference that the importance (sensitivity) of certain item has direct relationship with resulting weighted rating and most importantly, the raw rating has direct effect to weighted resulting rating value. Meaning that, for instance the more important an item is the higher the rating of a user is. As the focus of this work is on confidence values, we advise the reader to refer to [19][21] for detailed description of formulas used for calculation of weighted values.

Defining and applying membership functions and fuzzy rules

Before the fuzzy values are fed to inference engine, membership functions should be formed and values need to be grouped according the degree of membership of each input parameter. Existing membership functions for fuzzy values comprise of Exponential, Sigmoid, Trapezoid, Gaussian, and bell-shaped [3]. We take the simple approach of Triangular shape [2][3] with our three fuzzified sets (for trust, privacy and rank). Fuzzy rules [2], allow the combination and specification of the output model from the inference engine. We utilize the fuzzy rules in our approach to characterize the confidence output model. We would like to describe the degree of user's confidence with respect to the scores he/she has assigned as weights for trust, privacy and rank. An example for a rule in our confidence model could be:

```
If the (Fuzzy) Trust Value is High AND
the (Fuzzy) Ranking Value is High AND
the (Fuzzy) Privacy Value is High Then
Confidence is High.
```

In this case AND operator narrows down the output result of the rule, as it represents a conjunction of membership functions. Finally, membership functions translate the fuzzy rules into fuzzy numbers. Fuzzy numbers are then used to give input to the fuzzy expert system.

3.2.2 Defuzzification phase

For defuzzification methodology, several approaches exist. Existing approaches include center of gravity method, the center of area method, the mean of maxima method, first of maximum method, the last of maximum method [2], bisector of area method, or the root-sum-square method. Root-sum-square is considered here as the main method. Other methods could be considered and evaluated, consequently but for the sake of simplicity we only consider this approach in this paper. In the defuzzification phase, the calculated membership function results are taken, grouped according to fuzzy rules, raised by power of two, and summed following the consequent side of each asserted fuzzy rule. Let *FR*, be a fuzzy rule. Following the defuzzification approach, formula (1), is used to defuzzify the values:

$$\overline{FR_m} = \sqrt{\sum (FR_m)^2} \tag{1}$$

Let m = ['-' , '0' , '+'], which represents the labels used to group the rule sets. This allows us to distinguish between rules, describing negative (low), neutral and positive (high) confidence outcomes. Now that we have defined the outcomes, we can apply weights and scale the weighted output. For instance, if positive confidence weights are more in favor of our approach then more weight can be given to positive confidence rather than negative or neutral. As a matter of fact, W_0 represents neutral weight, W_+ represents positive weight, W_- represents the negative weight.

Adopted from Schmidt et al.,[19], formula (2) allows us to create a weighted overall confidence output:

$$C(U_x) = \frac{\overline{FR}_{-}W_{-} + \overline{FR}_0 W_0 + \overline{FR}_{+}W_{+}}{\overline{FR}_{-} + \overline{FR}_0 + \overline{FR}_{+}} \tag{2}$$

Where C represents the evaluated Confidence and U_x represents the user being evaluated. We can derive a Collective Confidence Factor (CCF), where we consider the confidence degree of other users with respect to the same information item and we calculate and process the confidence value of a user with respect to an item, bearing in mind overall derived confidence. Let us define View as the inferred confidence of the user with respect to a certain item. If we consider a user's self-view as internal, we can define an Internal View, while other users' views can be seen as External Views. By taking into account this assumption we can assign weights to average confidence of others and to a user's confidence, altogether. Adapted from OTV (Overall Trustworthiness Value), proposed by Schmidt et al.,[19], we formulate CCF using formula (3):

$$CCF(U_x) = W_{IView} C(U_x) + W_{EView} \frac{\sum_{l}^{i=0} \left(\frac{\overline{FR}_{-}W_{-} + \overline{FR}_0 W_0 + \overline{FR}_{+}W_{+}}{\overline{FR}_{-} + \overline{FR}_0 + \overline{FR}_{+}} \right)}{l} \tag{3}$$

Where CCF is Collective Confidence Factor, U_x is user being evaluated,$C(U_x)$ is the confidence of user with respect to the item being viewed, W_{IView} represents internal view weight while W_{EView} represents external view weight, and l represents total number of users, for whom we have considered their confidence values. At this stage we can scale the resulting values on a specific confidence scale and expand the range of resulting CCF values.

4 SMARTMUSEUM Simulation

In an experimental evaluation, taking into account a SMARTMUSEUM setting, we simulated 100 weighted user profiles. Weight values are intentionally sparse; weights assigned to profile slices contain blank values in order to reflect the real-world scenarios where users dont provide much input data into system, or sensors are faulty. We considered artifacts of two physical museums as items being experienced by exhibitors. In order to apply our model and demonstrate it in the context of our laid out scenario, we follow the steps, described previously in section 3.

In the first step we fuzzify all input raw values for three weight-descriptors at hand. We consider three main qualifiers for preferred outcome. Fig. 1 depicts an excerpt of weight values for raw and fuzzified trust values of 10 simulated users. Crisp trust values (left) and weighted trust values (right) for 10 users. Simulated raw values are intentionally sparse, as depicted with broken lines in left diagram.We have considered *high* confidence, *neutral* confidence and *low* confidence:

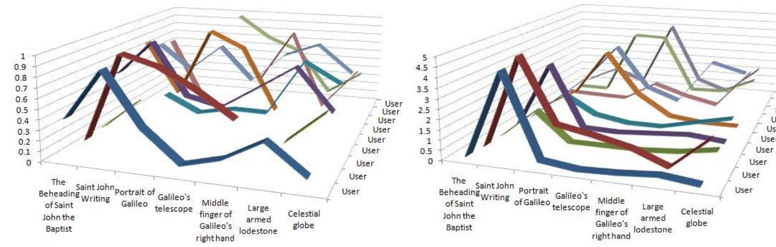

Fig. 1 Linear presentation of crisp trust values (left) and weighted trust values (right)

```
If Trust,Privacy&Rank all High, Confidence High.
If Trust,Privacy&Rank all Average, Confidence Neutral.
If Trust,Privacy&Rank all Low, Confidence Low.
```

Now that fuzzy sets are formed we apply defuzzification methodology. This allows us to filter negative results, in the case negative values were available. In our scenario all input values are positive. As fuzzy sets are grouped based on the preferred output, we can scale the output and gain more flexibility using weights. We can define weights for each type of output. Weight degree is taken from the range of [0,1]. We have given the maximum weight to positive values, while neutral values are considered more important than low values. Now we're able to evaluate the confidence.

Fig. 2. depicts the resulting confidence evaluation for user/item in our scenario. Now that confidence values are derived, we can infer CCF for the users we have evaluated so far. More flexibility can be gained through giving weights to internal and external Views of user information that we have processed. Since we don't have any preference over difference of processed user's view (W_{IView} internal view) and other user's views (W_{EView} external view), then we assign equal weight to both views. Confidence values and Collective Confidence Factors are depicted in Fig. 2. Results were generated for 100 user profiles. For Confidence Evaluation weight set, W=[$W_+ = 1, W_0 = 0.05, W_- = 0.5$] while for Collective Confidence Factor weight set, W= [$W_{IView} = 0.5, W_{EView} = 0.5$]. Horizontal axis represents users, while vertical axis plots confidence degree distribution. We tried to generate values that represent real-world user inserted values, in many cases either one or two, or in one or two exceptions, all three values were kept empty. This reflects the sparsity problem of training data for profiling (in general personalization) services such as recommendation, or matchmaking. Such problem hinders the performance of personalization services by creating infamous problem of cold-start. By comparing input (raw) values with resulting confidence degrees, we realize that results are not uniform and that is justifiable with respect to different preferences or interests of users. In certain cases values have improved, while in many cases values haven't changed. We observed that empty values in many situations haven't changed and this is mainly because of the naive rules considered, where we weigh positive and neutral outcomes higher than low outcomes. Simple approach could be considered to address

Fig. 2 Stacked linear presentation of (top) confidence and (bottom) collective confidence

empty or zero values, by using an offset for trust fuzzification. Although in comparison between pure confidence values and collective confidence factors, we realize that considering collective opinions while evaluating the confidence of an individual user over a certain item, could give more improved results. As seen in Fig. 2., CCF values are more uniformly distributed over diagram in comparison to pure confidence values. The uniformness in distribution of values in CCF comes from quantification of others confidence while calculating one's confidence. The other reason can be seen as the flexibility given by incorporating further weight for Views with respect to user being evaluated or collective Views of other users. As a result we can use CCF values instead of classic, pure confidence values for boosting personalization services. All and all, we have managed to replace all empty values with a single value (although zero) and at least sparsity is alleviated with respect to that.

5 Conclusion and Future Work

We have introduced a fuzzy approach to modeling and analyzing confidence based on weights assigned to profiled information of users stored in semantic profiles. Based on our approach weights can be processed through a fuzzy reasoner and create a weighted outcome based on factors affecting the context of calculation. We have tested our approach with simulation data from a real-world scenario, where exhibitors of visual art experience personalized services of distributed knowledge-platforms. We have introduced a classic and a collective notion of confidence where values could be used to improve quality of adaptive personalized services or allow us to detect similar individual or group behavioral patterns. As a future work, we will use resulting confidence degrees to improve personalization services provided by the knowledgeplatform, such as recommendation, matchmaking, and etc. We would like to also see how collective notion can be used to enable group-based services such as group recommendations.

Acknowledgements This work has been done within the FP7-216923 EU IST funded SMART-MUSEUM project. The overall objective of the project is to develop a platform for innovative services enhancing on-site personalized access to digital cultural heritage through adaptive and privacy preserving user profiling.

References

1. Z. Huang, F. van Harmelen, A. T. Teije: Reasoning with Inconsistent Ontologies, International Joint Conference on Artificial Intelligence, vol 19, pages 454-459. ISBN ISSN 1045-0823. Lawrence Erlbaum associates ltd, USA,(2005).
2. G. J. Klir, B. Yuan: Fuzzy sets, Fuzzy Logic, and Fuzzy Systems: selected papers by Lotfi A. Zadeh, World Scientific Publishing Co., Inc. River Edge, NJ, USA, (1996) .
3. E. Sanchez: Fuzzy Logic and the Semantic Web. Elsevier. ISBN 0444519483 9780444519481,(2006).
4. Y. Li, B. Xu, J. Lu, D. Kang: Reasoning with Fuzzy Ontologies. Contexts and Ontologies: Theory, Practice and Applications, (2006).
5. N. Dokoohaki, M. Matskin: Personalizing human interaction through hybrid ontological profiling: Cultural heritage case study, in 1st Workshop on Semantic Web Applications and Human Aspects, (SWAHA08), M. Ronchetti, Ed., In conjunction with Asian Semantic Web Conference 2008. AIT epress, pp. 133-140, (2008).
6. EU IST FP7 SMARTMUSEUM project: http://www.smartmuseum.eu, last accessed June (2009).
7. N. Koch: Software Engineering for Adaptive Hypermedia Systems. Munich, Germany: PhD thesis, Ludwig-Maximilians University, (2005).

8. S. Gauch, M. Speretta, A. Chandramouli, A. Micarelli: User Profiles for Personalized Information Access. NUMB 4321, pages 54-89 . Lecture notes in computer science, Springer-Verlag, (2007).
9. E. Bloedorn, G. Mason, I. Mani, T. R. MacMillan: Machine Learning of User Profiles: Representational Issues. Proceedings of the national conference on artificial intelligence, number 13, pages 433-438. John Wiley and Sons ltd, (1996).
10. A. Kobsa: User Modeling: Recent work, prospects and Hazards, Auth. Schneider-Hufschmidt, m.,(1993).
11. S. Weibelzahl: Evaluation of Adaptive Systems: PhD thesis. University of Trier, (2003).
12. P. Dolog, W. Nejdl: Challenges and Benefits of the Semantic Web for User Modelling - Proceedings of Adaptive Hypermedia, (2003).
13. P. Dolog, W. Nejdl: Semantic Web Technologies for the Adaptive Web. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.): The Adaptive Web: Methods and Strategies of Web Personalization, Lecture Notes in Computer Science, Vol. 4321. Springer-Verlag, Berlin, (2007).
14. Y. Wang, L. Aroyo, N. Stash, L. Rutledge: Interactive User Modeling for Personalized Access to Museum Collections: The Rijksmuseum Case Study. In Proc. 11th User Modeling Conference, Greece, (2007).
15. O. Nasraoui, H. Frigui, R. Krishnapuram, A. Joshi: Extracting Web User Profiles Using Relational Competitive Fuzzy Clustering. international journal on artificial intelligence tools, vol 9; PART 4, pages 509-526, ISSN 0218-2130, (2000).
16. M. J. Martin-Bautista, D. H. Kraft, M. A. Vila, J. Chen, J. Cruz: User profiles and fuzzy logic for web retrieval issues. soft computing, vol 6; numb 5, pages 365-372. springer verlag. Issn1432-7643, (2002).
17. S. Nefti, F. Meziane, K. Kasiran: A Fuzzy Trust Model for E-Commerce. E-Commerce Technology,CEC (2005).
18. R. Falcone, G. Pezzulo, C. A. Castelfranchi: Fuzzy Approach to a Belief-Based Trust Computation ISSU 2631, pages 73-86. Springer-Verlag; ISSN 0302-9743, (2003).
19. S. Schmidt, R. Steele, T. S. Dillon, E. Chang: Fuzzy trust evaluation and credibility development in multi-agent systems- Journal of Applied Soft Computing Journal, Elsevier, (2007).
20. R. Kohlas, U. Maurer: Confidence Valuation in a Public-Key Infrastructure Based on Uncertain Evidence, Springer-Verlag, ISSU 1751, pages 93-112, ISSN 0302-9743, (2000).
21. Q. Zhang, Y. Qi, J. Zhao, D. Hou, Y. Niu: Fuzzy privacy decision for context-aware access personal information- Wuhan University Journal of Natural Sciences, Springer Verlag, (2007).
22. P.A. Raj, D.N. Kumar: Ranking alternatives with fuzzy weights using maximizing set and minimizing set- Fuzzy Sets and Systems, (1999).