ARTICLE IN PRESS
JID:FSS AID:7867 /FLA [m3SC+; v1.328; Prn:25/05/2020; 12:58] P.1 (1-17)
Available online at www.sciencedirect.com
ScienceDirect
Fuzzy Sets and Systems ••• (••••) •••–•••
FUZZY sets and systems
www.elsevier.com/locate/fss

A fuzzy model to enhance user profiles in *microblogging* sites using deep relations

Manuel Francisco *, Juan L. Castro

Department of Computer Science and Artificial Intelligence, University of Granada, Spain

Received 14 March 2019; received in revised form 15 January 2020; accepted 14 May 2020

Abstract

Social Networking Sites (SNS) have entailed a revolution for society. They have given a say to everyone regardless of their status and this has been translated into loads of data. The task of *profiling users* constitutes a way to learn from this data in order to show users only the content that is relevant to them. Several recommendation system techniques have been used to address this problem, being mainly based on what the user explicitly says about themselves, on what the user publishes in the SNS and on the similarities between users. However, in social media context, it is also possible to use relations between users. Considering basic relations like *follower* or *followee* to extract information from them may result in noise, since they do not imply that users share interest or even ideas. In this work, we present a fuzzy framework to enrich user profiles with complex properties in order to have an even better representation of them. We use basic relations defined by SNSs to complete the information available in user profiles with topics of interest and ideas towards them and to define deep relations that will enable new ways of analysis. We use these deep relations to create clusters of similar users that, ultimately, will allow the expansion of properties from known users to the rest of the cluster. We tested our proposal with a dataset of Tweets in Spanish related to a political event. Our experiments prove the potential that this approach has for a lot of applications in *microblogging* context.

Keywords: User profiling; Fuzzy relation; Fuzzy modelling; Incomplete information; Online behaviour; Social network

1. Introduction

Social Networking Sites (SNS) have revolved our lives as they have given a say to masses. It has been proven that they have the capacity to mobilise hundreds of people from all around the world as long as they seek a noble, common cause [1].

SNSs gained a lot of adepts in the last ten years [2], although in the last few of them, tendency has shifted in favour of anonymous social media [3]. It was expected that, by 2020, data on the Internet would have growth to 40000

* Corresponding author.
E-mail addresses: francisco@decsai.ugr.es (M. Francisco), castro@decsai.ugr.es (J.L. Castro).

https://doi.org/10.1016/j.fss.2020.05.006
0165-0114/

Please cite this article in press as: M. Francisco, J.L. Castro, A fuzzy model to enhance user profiles in *microblogging* sites using deep relations, Fuzzy Sets Syst. (2020), https://doi.org/10.1016/j.fss.2020.05.006

Table 1
$R^{m\times n}$ matrix representation of a user profile, for a total of m users and n attributes. In this table, each row stands for the preferences of a user meanwhile each column represents in which grade the users *matches* the attribute P.

	P_1	P_2	P_3	$\dots$	P_n
u_1	v_{11}	v_{12}	v_{13}	$\dots$	v_{1n}
u_2	v_{21}	v_{22}	v_{23}	$\dots$	v_{2n}
u_3	v_{31}	v_{32}	v_{33}	$\dots$	v_{3n}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
u_m	v_{m1}	v_{m2}	v_{m3}	$\dots$	v_{mn}

exabytes, reaching 175 zettabytes by 2025 [4,5]; most of this data is unstructured and unlabelled but, thanks to social media mechanics, it is possible to address with these issues in order to extract knowledge from *chaos*.

One step towards *learning* from SNSs is *profiling users*. This task consists in, given a set of users and topics of interest, gathering the properties that identify one user among the rest (within a certain degree of accuracy), and it is particularly useful for customising content in an always-growing information flow [6]. Profiling users and their interests are not trivial tasks. Traditional SNSs have, normally, three main features: a user profile, a list of connected profiles and the possibility to interact among other users and their lists [7]. There are several ways to define a user profile, but the vast majority of them rely on what users share about themselves and the content they create.

There are advanced aspects that are not being included in current analysis of social networks, such as the topics of interest of the users and their opinions about these. These concepts are inherently imprecise (the interest towards a topic will be up to a certain degree and the opinions may vary in different aspects regarding the same topic) and thus they might be modelled using fuzzy relations and sets [8,9]. In order to include those aspects, it is possible to analyse the content of the messages they share, but it may not be enough in real scenarios, because of (1) errors made by automatic analysis tools and machine learning techniques; (2) lack of a sufficient number of useful messages; and (3) users may avoid expressing their sincere opinions, especially in controversial topics where their friends or society may judge them [10].

However, we are not neutral individuals forging our personal perspective from the world but a product of society, and because of this, we are not owners of our ideas, at least to some extent. Our behaviour is a reflection of the world we live in, and SNSs behave like some sort of mirror from reality. Even though users try not to manifest *deep* beliefs, they interact with other users that may not have this restraint, and these interactions may disclose the actual profile of the user. An in-depth analysis of relations and interactions between users (and messages) will allow us to tackle with those issues (e.g. [11]).

For example, it is possible that a user declares that *"videogames are for children"* but, in reality, they follow accounts that keep them posted about novelties in this world, which might imply that they are interested but they do not want the world to know.

The purpose of this work is multiple: (1) we want to enhance user profiles in *microblogging* sites by adding new attributes, specifically, topics of interest and ideas about these, using fuzzy techniques; (2) we want to obtain high-level relations between users, as sharing interest and/or ideas; and (3) we want to refine profile attributes by the use of deep analysis of social networking sites. Thereby, we present in this document a fuzzy framework to attain these targets.

1.1. Problem statement

A user profile can be defined as a set of triplets (u, P, v), being u an user, P an item, property or attribute relative to the user u and v the value that the attribute P takes for the user u.

It is frequent that these triplets are stored as rating matrices $R^{m\times n}$ (see Table 1), where m is the number of users and n the number of items, in order to efficiently apply algebraic operations [12]. For each user u, the values v regarding each P will be either provided by the user or estimated through their interactions and content generated within the system.

Machine learning techniques are usually employed to infer attribute values, hereby leading to interpretability problems in most cases: state-of-the-art methods behave as *black boxes* which implies relying purely in the training data. Depending on the applications of profiling users, this can entail an issue, especially if we are using user profiles in a way that will cause a social impact [13].

The rest of this document is organised as follows. In the next section, we provide a review of related work in the field. Section 3 describes our proposal to extend user profiles with topics of interest and ideas towards them. Section 4 introduce the concept of deep relation, that will be used in section 5 to define social groups. In section 6, we show that social groups can be used to expand properties through the network of users. Section 7 shows experimental results for an application of our proposal. Finally, sections 8 and 9 show our conclusions and future work, respectively.

2. Related work

In this section, we present a review of literature related to our work.

User profiling has been a task for Recommendation Systems for many years now. These engines give recommendations based on the preferences a user (or a similar one) shows towards specific products. In order to perform this task, they try to define a profile for each user as well as relations between them.

Usually, a profile is defined as *the set of attributes (independent or not) that characterise a user and allows distinguishing them among others* [14,15]. Mainly, it is possible to define a profile by using previous knowledge from the user (*content based, CB*) or by relations with the rest of them (*collaborative filtering, CF*) [16]. The former method presents good accuracy as their recommendations are computed from explicit actions of the user, but it has a problem not easy to address regarding serendipity; the latter can solve this problem as it gives recommendations based on the similarity cluster, but it normally tends to very sparse feature matrices, especially at the beginning (*cold start*). Since there is not an optimal choice, it is common to use hybrid systems in which recommendations are based both on user previous actions and relations among them.

In the context of a SNS, content-based approaches work from what the user publishes in social media (e.g. text, photos, videos, audio, links, etc.) and what is explicitly expressed in them (e.g. we can understand from the text *"Lebron saved the game with a triple in the last second"* that the author is interested in basketball). Content-based systems have special relevance in fields like event recommendation, since events are time dependant and, normally, they have no attendance records [17]; also, they are being used to match user accounts across different SNSs in order to avoid and/or enrich profiles taking advantages of different privacy settings and published content [18]; most commonly, they are used in order to compute properties from users based on what they share (e.g. Sarna and Bhatia [19] apply content-based approaches to calculate the credibility of a given user).

On the other hand, collaborative filtering systems compute similarity measures between users and give recommendations based on interests of other users. In the domain of a SNS, this can be understood as common interests of the communities, leading to another problem regarding how to detect and model communities. CF methods are being used for detecting trust [20,21] and identifying malicious users (e.g. [22]), among others, since these applications rely on similarities between users.

As we can come to understand from the previous paragraphs, these techniques are not only used to recommend products (books, films, insurances, etc., e.g. [23–27]) but also people (friends or even a date, e.g. [28–31]). This is especially relevant in social networks where engines suggest who to follow based on your connections and the connections of your friends. Until now, traditional SNSs show bias towards friending as many people as possible in order to be popular or even an *influencer*, and at first, we can think that the task of recommending people is easy: we have more chances to know a friend of a friend that a random user, so at first glance, we can recommend friends of friends. Although, in social media, we also follow people we have not ever met just because they are famous or because our interests are alike (famous within a field). Once again, the logical approach consists in using a mixed technique that can suggest friends from close people (low distance between them) and from similarity clusters (same interests) [32].

Whereas suggesting people may be a complex but achievable task, it can be blurrier to detect user opinions. Predicting a user's rating of a product can be accomplished by using *CB* and *CF* methods together, but ideas are a bit tricky. We have a lot of connections (people who follow us and people we follow) in social media that do not necessarily imply sharing interests or opinions with them: they can be our neighbours, friends from school or college, family or even journalists, writers or politicians. Befriending them does not mean that we share their interests (we can be just interested in their lives) or ideas (we can follow them because we share topics but not opinion. This happens quite often with politicians and journalists). All in all, we can conclude that there is a lot of noise in these explicit relations (follower/following) for a recommendation system to work acceptably [33].

Microblogging social networks such as *Twitter* present peculiar mechanics that we can use to take advantage in the analysis process, specifically, actions like *retweet* (copy), reply, mark as favourite and naming another user. However, due to the shortness of the texts, traditional techniques of information retrieval (tag extraction, topic detection, and so on) generate very sparse matrices that stand in the way of recommendation systems [34,35].

In order to address this issue, we propose a fuzzy model that is able to propagate features through deep relations, diminishing the sparsity with low noise due to tailored relations, to enrich the profile of the users with properties of their environment that may reveal hidden features of the users.

3. Extending profiles

In this section, we address the process of extending the properties of user profiles.

Usually, properties on profiles can be defined by users themselves, they can be calculated (number of interactions per day, number of followers, mean length of messages...), they can be opinions from other users, and/or they can be defined by several other ways.

We want to enrich user profiles with *complex* properties in order to have an even better representation of each user. We consider that a property is *complex* when one of the following criteria applies:

a) If they require analysis in order to be extracted
b) If they are inherently imprecise (fuzzy)

As we already stated, the profile of a user goes beyond of what is explicitly shown. Our profile is influenced by other people in our social neighbourhood. What we say, how we say it and to whom we say it can be used to discover attributes buried deep into our personality and interests. This is part of what we have called *deep profile* of a user (non-superficial, possibly hidden at first sight), and we use complex properties in order to depict it precisely.

Throughout this document, we will consider that we have a dataset M of messages written by a set of users U, and a set of topics T that we are interested in studying.

Our start point will be the basic relations that come as a consequence of social networks mechanics, and we are going to describe them in this section. Mechanics in *microblogging* sites are determined by interactions between users, between users and messages and/or between messages.

$$\forall u, v \in U, follows(u, v) = \begin{cases} true & \text{if } u \text{ is subscribed to } v \text{ updates} \\ false & \text{in any other case} \end{cases} \tag{1}$$

$$\forall u \in U, \forall m \in M, author(u, m) = \begin{cases} true & \text{if } u \text{ is the author of } m \\ false & \text{in any other case} \end{cases} \tag{2}$$

$$\forall u \in U, \forall m \in M, favourite(u, m) = \begin{cases} true & \text{if } u \text{ likes the message } m \\ false & \text{in any other case} \end{cases} \tag{3}$$

$$\forall m \in M, \forall u \in U, mention(m, u) = \begin{cases} true & \text{if } m \text{ names user } u \\ false & \text{in any other case} \end{cases} \tag{4}$$

$$\forall m, n \in M, copy(m, n) = \begin{cases} true & \text{if } m \text{ is a verbatim copy of } n \\ false & \text{in any other case} \end{cases} \tag{5}$$

$$\forall m, n \in M, reply(m, n) = \begin{cases} true & \text{if } m \text{ is an answer to } n \\ false & \text{in any other case} \end{cases} \tag{6}$$

3.1. Analysing message content

Analysing message content can determine non-explicit features of the text, particularly if we combine it with the previous concepts (relations). In order to do this, we are going to use tools to measure message properties. We introduce the requirements we require for these tools in this section.

We want to add to user profiles the topics they are interest in and the ideas they have with respect to those topics. There are a number of algorithms that can be applied in order to extract topics (e.g. [36–38]) and sentiment (e.g. [39–41]) from messages. We will choose an algorithm for extracting topics and another one for analysing sentiment. The chosen algorithm should not affect the results of the rest of the paper, as long as they satisfy the requirements that we are going to present below.

Thus, we will have two functions:

- $sentiment(m)$ stands for the output of the chosen algorithm for sentiment analysis when passing the message m.
- $topics(m)$ will yield the topics that the message m is referring to, using the selected method for topic detection.

In order for the model to be consistent, we are going to consider the following premises:

1. For any given message m, if n is a copy of m, then they have the same sentiment score.

$$\forall m, n \in M, copy(n, m) \Rightarrow sentiment(m) = sentiment(n) \tag{7}$$

2. For any given message m, let n be a copy of m. Then, they need to refer to the same topics.

$$\forall m, n \in M, copy(n, m) \Rightarrow topics(m) = topics(n) \tag{8}$$

3. For any given message m, if n is a response to m, then they need to share at least one topic.

$$\forall m, n \in M, reply(m, n) \Rightarrow topics(m) \cap topics(n) \neq \emptyset \tag{9}$$

3.2. Adding interest to profiles

Our personality may be reflected in our actions. The news we read, the hobbies we have and the jokes we like are a useful factor in order to determine our profile. In a SNS, an approximation to this kind of information can be made by considering the topics of interest of users. They tend to follow others with similar interests and, ultimately, they will interact (copy, reply, mark as favourite...) with the content they are interested in. The purpose of this section is to present our proposal to include *topics of interest* to user profiles.

Let us remember that we have restricted the topics to a set of them that we are interested in studying (T). Then, we can consider a subset of T that will stand for any topic that the user u is interested in, even remotely:

$$\begin{aligned} Topics_u = \{t : t \in T \wedge [& \\ & \exists m \in M : (author(u, m) \wedge t \in topics(m)) \\ & \vee \exists m, n \in M : (author(u, m) \wedge copy(m, n) \wedge t \in topics(n)) \\ & \vee \exists m, n \in M : (author(u, m) \wedge reply(m, n) \wedge t \in topics(n))]\} \end{aligned} \tag{10}$$

Equation (10) is a crisp set that includes all the topics that:

a) the user has explicitly mentioned in their messages
b) the user has mentioned by republishing someone's publication
c) the user has referred to by replying to someone's publication.

It is noticeable that when a user replies to someone's content it does not necessarily imply that they are referring to the same topic than the original message. However, microblogging messages are short texts and it is not unreasonable to think that the noise would be minimum. It is preferable to cope with this noise than to exclude the information that *replies* may offer us. This could vary depending on the context (for example, if the topics are prone to interactions of users with unexpected behaviour, like *Internet trolls*), and it will enable a very substantial field for further research.

The interest that any given users has on these topics is not absolute. On the contrary, they will be more interested in some topics rather than others. This makes *interest* an imprecise concept, since users will be interested in some topic up to a certain degree. We modelled it as a fuzzy subset of T. Let us consider the following function:

Please cite this article in press as: M. Francisco, J.L. Castro, A fuzzy model to enhance user profiles in *microblogging* sites using deep relations, Fuzzy Sets Syst. (2020), https://doi.org/10.1016/j.fss.2020.05.006

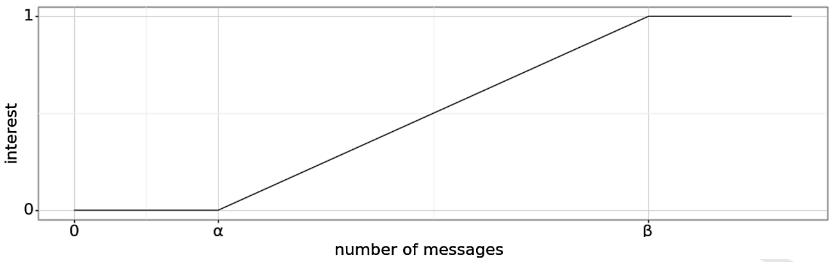

Fig. 1. Behaviour of the membership function (α, β)-$interest$.

$$interest(u, t) = |\{m : m \in M \wedge author(u, m) \wedge \big[t \in topics(m) \vee \\ \vee \exists n \in M : (reply(m, n) \wedge t \in topics(n))\big]\}| \quad (11)$$

that yields the number of times that the user u has written about the topic t, including also the number of times that he/she has replied to another user's message regarding topic t. Notice that copies are already included in *authored* messages (since they are republished messages). Consecutively, we can consider a normalised interest function

$$(\alpha, \beta)\text{-}interest(u, t) = \begin{cases} 0, & \text{if} \quad interest(u, t) <= \alpha \\ \frac{x-\alpha}{\beta-\alpha}, & \text{if} \quad \alpha < interest(u, t) < \beta \\ 1, & \text{if} \quad \beta <= interest(u, t) \end{cases} \quad (12)$$

where α and β are context-dependant parameters (Fig. 1).

Thereupon, it is possible to define the topic-of-interest profile of a user as follows:

Definition 3.1 *(Topics-of-interest profile).* The topic-of-interest profile of a user u, $Profile_T(u)$, is the fuzzy set $(T, (\alpha, \beta)\text{-}interest)$

$$Profile_T(u) = \{(t, (\alpha, \beta)\text{-}interest(u, t)) : t \in T\} \quad (13)$$

In this way, $Profile_T(u)$ stands for the degree in which the user is interested in each topic of T. $Profile_T(u)$ can be translated into a rating matrix $R^{|U|\times|T|}$, where the value of each cell will be the result of the mapping (α, β)-$interest$, and thus defining the profile of a user as stated in section 1.1.

3.3. Adding ideas to profiles

In several use cases, it will be useful to know the opinion of a user regarding a certain topic. For example, a given user may have interacted with content referring to *alcoholic beverages*. Thus, he/she will have a degree of interest towards them. However, if the interaction was in order to manifest their dislike, the user should have "negative" interest towards *alcoholic beverages*. Attending only to the Topic-of-Interest Profile, a system could recommend the user some beers, but that would not be relevant at all for the user. In this section, we propose adding the *actual thinking* of the user regarding the topics of interest.

Once again, modelling the ideas that a user may have regarding some topic is an imprecise concept. Users may have positive or negative opinions, but it is difficult to represent how much positive or negative their idea is. In fact, opinions may vary through time. They could even have contradictory opinions on the same matter whatsoever. Let us offer a few examples:

- *Jane considers herself a fervent patriot. She really loves her country, its traditions and its people. She is also against any kind of violence. She despises all kinds of wars and human suffering. She has* tweet'ed *a lot of times manifesting these opinions. Recently, another country has declared war to hers in order to take control of a region full of mineral resources. She finds herself in a contradiction. She does not like wars, but she cannot do anything to avoid it. In fact, she may be required to fight against the enemy in order to protect her life and their beloved ones. She has values, but she needs to betray them in order to survive. She has also* tweet'ed *about this.*

Please cite this article in press as: M. Francisco, J.L. Castro, A fuzzy model to enhance user profiles in *microblogging* sites using deep relations, Fuzzy Sets Syst. (2020), https://doi.org/10.1016/j.fss.2020.05.006

Table 2
Example of the rating matrix for a fictional user that likes beer (but despises alcohol-free ones), that is passionate about football (even though he/she thinks that it is opium for society) and that has controversial thoughts regarding climate change.

Topic	HP	P	Z	N	HN
beer	0.3	0.7	0.2	0	0.3
football	0.6	0.5	0	0.1	0.4
climate change	0.3	0.5	0.7	0.5	0.2

- *Mark likes psychoactive drugs. He has always enjoyed the feeling of losing control and letting his mind dream with alternate realities. He has published a few posts in Facebook with his stories under the influence, with reference to the drugs he had ingested. He has been taking drugs occasionally for the last 10 years now, even though he completed his medical degree 2 years ago. When he studied the effects that these harmful chemicals have on his health, he stopped publishing content praising drugs, since he understood that it is not good for public health. He even deleted all his posts regarding to this topic. However, he still* likes *a few posts every once in a while from the communities he used to participate in.*

In the first example, we could say that *Jane* has a *positive* interest in topics regarding her country. She has also a *negative* idea about war. When she encounters these controversial thoughts about fighting for her life and her country and keeping the peace, she betrayed her values. That could be interpreted as *positive patriotism* and *negative war* or, alternatively, as *positive patriotism* and *neutral war* (overall thinking). However, if we compare *Jane* to other people that will give their lives just because they refuse to fight, then *Jane* should be considered as *neutral war* or, at least, *less negative than* the others.

In the second example, *Mark* has deleted all his posts regarding drugs, so it is not possible anymore to classify his ideas as *positive* towards the topic *drugs*. For any profiling algorithm, he will be marked as *neutral drugs*. However, if we consider his *likes* with communities that revolve around drugs, we should mark him at least with *a bit positive* towards *drugs*.

For all these reasons, we modelled the ideas that a user may have as a fuzzy concept. Let us consider the following linguistic labels $L = \{HP, P, Z, N, HN\}$ (very positive, positive, neutral, negative, very negative) that measures the sentiment of messages. We represent the ideas of a user u for a specific topic t with a sentiment L by the following function.

$$sentiment(u,t,l) = |\{m : m \in M \wedge author(u,m) \wedge \wedge t \in topics(m) \wedge sentiment(m) \text{ is } l\}| \quad (14)$$

Analogously to interest, we are going to define a fuzzy set that relates users, topics and sentiment towards them. Considering the next function:

$$(\alpha,\beta)\text{-}sentiment(u,t,l) = \begin{cases} 0, & \text{if} \quad sentiment(u,t,l) <= \alpha \\ \frac{x-\alpha}{\beta-\alpha}, & \text{if} \quad \alpha < sentiment(u,t,l) < \beta \\ 1, & \text{if} \quad \beta <= sentiment(u,t,l) \end{cases} \quad (15)$$

Definition 3.2 *(Idea profile).* The Idea Profile $Profile_I(u)$ of a user u stands for the degree in which, given a topic $t \in T$ and a sentiment $l \in L$, the user agrees with the pair (t,l).

$$Profile_I(u) = \{(t, (\alpha,\beta)\text{-}sentiment(u,t,HP), (\alpha,\beta)\text{-}sentiment(u,t,P), (\alpha,\beta)\text{-}sentiment(u,t,Z), (\alpha,\beta)\text{-}sentiment(u,t,N), (\alpha,\beta)\text{-}sentiment(u,t,HN)) : t \in T\} \quad (16)$$

$Profile_I(u)$ can be translated into a rating matrix $R^{|U|\times|T|\times|L|}$, where the value of each cell will be the result of the mapping (α,β)-$sentiment$, analogously to the interest matrix, and thus defining the ideas of a user towards a topic (Table 2).

Please cite this article in press as: M. Francisco, J.L. Castro, A fuzzy model to enhance user profiles in *microblogging* sites using deep relations, Fuzzy Sets Syst. (2020), https://doi.org/10.1016/j.fss.2020.05.006

4. Deep relations

In the previous section, we presented our proposal to extend profiles with complex properties (*Topic-of-Interest* and *Idea Profile*). We used an algorithm of topic detection and another one for sentiment analysis of the messages, jointly with basic relations between users.

In this section, we are going to use both complex properties in order to define relations far more complex than the basic ones (or directly deduced from them) already presented.

These relations will be determined by two measures (*shareinterest* and $shareidea$) used to measure these complex aspects:

1. The degree in which two users share topics of interest.
2. The degree in which two users share the same idea regarding the topics of interest.

Normally, these relations go unnoticed in standard analysis of SNS datasets, especially in the cases were users themselves try to hide their opinion (e.g. in controversial topics), hence the reason to name them *deep relations* (they require a *deep analysis* and sometimes they may stay hidden). The interest of these relations is obvious, especially in order to be included in collaborative filtering approaches.

The rest of these sections is organised as follows. First, we introduce three metrics based of basic relations that will be used to calculate *deep relations*; then we present $shareinterest$ relation; and finally, we explain its analogous concept, $shareidea$.

4.1. Some metrics based on basic relations

We present in this section some metrics that will be used to define *shareinterest* and *shareidea* measures.

From the analysis of masses instead of individuals, it is possible to obtain fuzzy relations (between users) that are helpful in order to discover the non-explicit (or even hidden) properties of user profiles. Even if we do not know the content of the messages, we can argue that two users share interests up to a certain degree if they *like* or *reply* to the same messages.

Definition 4.1 *(Co-copies).* Given two users u and v, we define $cocopies(u, v)$ as the number of times that both users have *retweeted* the same message:

$$cocopies(u, v) = |\{x : x \in M \wedge \exists m, n \in M : \big[author(u, m) \wedge author(v, n) \wedge \wedge copy(m, x) \wedge copy(n, x)\big]\}| \quad (17)$$

Definition 4.2 *(Co-replies).* Given two users u and v, we define $coreplies(u, v)$ as the number of times that both users have replied to the same message:

$$coreplies(u, v) = |\{x : x \in M \wedge \exists m, n \in M : \big[author(u, m) \wedge author(v, n) \wedge \wedge reply(m, x) \wedge reply(n, x)\big]\}| \quad (18)$$

Definition 4.3 *(Co-favourites).* Given two users u and v, we define $cofavourites(u, v)$ as the number of times that both users have marked as favourite the same message:

$$cofavourites(u, v) = |\{x : x \in M \wedge favourite(u, x) \wedge favourite(v, x)\}| \quad (19)$$

4.2. Shareinterest

In this subsection, we introduce measures for the degree in which two users share topics of interest. We will propose two different measures, with the first one based on the *Topic-of-Interest Profile*. The intersection of the *Topic-of-Interest Profile* of two users will be a fuzzy subset of T, that represents the topics of interest shared by those two users. Thus, we can use the cardinality of these fuzzy subsets to measure the degree in which they share topics.

Please cite this article in press as: M. Francisco, J.L. Castro, A fuzzy model to enhance user profiles in *microblogging* sites using deep relations, Fuzzy Sets Syst. (2020), https://doi.org/10.1016/j.fss.2020.05.006

Definition 4.4 *(Shareinterest).* Given two users u and v, we define *shareinterest*(u, v) as addition of the common interest between them for each topic in T:

$$shareinterest(u,v) = \sum_{t \in Topics_u} \min\{(\alpha,\beta)\text{-}interest(u,t), (\alpha,\beta)\text{-}interest(v,t)\} \tag{20}$$

Despite the potential of this approach, it is complex to compute. It depends on (1) the *Topic-of-Interest Profile* of each user, (2) on the calculus of $topic(m)$ for each message on the dataset M and (3) on the calculus of $(\alpha, \beta) -$ *interest*(u, t) for each user and topic.

For this reason, we present here a second proposal to model this relation that will not depend on the algorithm that we used for $topic(m)$. It will also be less complex to compute. This proposal, that we will call *heuristic*, will be based on the relations defined in the previous section:

- *co-copies* and *co-favourites* are an obvious choice since they would directly share the topics.
- *co-replies* would be a good choice taking into consideration that the typical behaviour is to respond to publications that are interest to you, so you would not change the topic
- Lastly, we need to consider interactions between both users as if they were *co-replies*, *co-copies* and *co-favourites*, since similarities are as good between two authors than with a third person (if user u publishes a message and v marks it as favourite, they would share that message's topic).

Definition 4.5 *(Heuristic shareinterest).* Given two users u and v, we define the heuristic $H_{shareinterest}(u, v)$ as:

$$\begin{aligned} &H_{shareinterest}(u,v) = \gamma \cdot coreplies(u,v) + cocopies(u,v) + \\ &+ cofauvorites(u,v) + |\{m : m \in M \wedge author(u,m) \wedge \exists n \in M : \big[author(v,n) \wedge \\ &\wedge (copy(m,n) \vee reply(m,n) \vee favourite(u,n))\big]\}| \end{aligned} \tag{21}$$

For two users u and v replying to a third one, restriction from equation (9) states that u needs to share a topic with the original author just as v, but it does not mean that it is the same topic between u and v. However, taking into account that *tweets* are really short texts, it will not be unreasonable to think that they will be referring to the same one. To cope with this assumption, we can establish an empirical grade γ in which this actually happens, although it will be context-dependant.

4.3. Shareidea

In this subsection, we introduce measures for the degree in which two users share ideas towards topics of interest. We will propose two different measures, with the first one based on the *Idea Profile*.

Definition 4.6 *(Shareidea).* For two given users u and v, we define $shareidea(u, v)$ as:

$$shareidea(u,v) = \sum_{t \in Topics_u} \sum_{l \in L} \min\{(\alpha,\beta)\text{-}sentiment(u,t,l), (\alpha,\beta)\text{-}sentiment(v,t,l)\} \tag{22}$$

$shareidea(u, v)$ is also complex to compute. It depends (1) on the *Idea Profile* of each user, (2) on the calculus of $sentiment(m)$ for each message in the dataset M and (3) on the calculus of $(\alpha, \beta) -$ *sentiment*(u, t, l) for each user and topic.

Analogously to the case of *shareinterest*, we present a second proposal to model the relation, so it will not be influenced by the chosen algorithm for $sentiment(m)$ and in order to reduce the complexity. This proposal, that we will call *heuristic*, will be based on relations defined in the previous section. In this case, we will only contemplate *co-copies*, *co-favourites* and copies/favourites between them, since *replies* cannot considered to be in the same line of thinking (users can answer in order to manifest their agreement/disagreement and there is no *a priori* way of distinguishing between them).

Please cite this article in press as: M. Francisco, J.L. Castro, A fuzzy model to enhance user profiles in *microblogging* sites using deep relations, Fuzzy Sets Syst. (2020), https://doi.org/10.1016/j.fss.2020.05.006

Definition 4.7 *(Heuristic shareidea).* Given two users u and v, we define the heuristic $H_{shareinterest}(u, v)$ as:

$$H_{shareidea}(u,v) = cocopies(u,v) + cofauvorites(u,v) + |\{m : m \in M \wedge \\ \wedge\, author(u,m) \wedge \exists n \in M : \big[author(v,n) \wedge (copy(m,n) \vee favourite(u,n))\big]\}| \quad (23)$$

5. Social groups

In a social networking site, a Social Group is defined as a set of users that are related between them in a certain manner. Normally, basic relations are used to define them (explicit social groups). Since we introduced new relations that are more complex, it makes sense to use these to obtain not-so-obvious social groups.

In the previous section, we do not have relations *per se*, but tools to measure the degree in which two users are related. However, for a given measure μ between two users, we can normalise that metric in the interval [0, 1] to obtain the underlying fuzzy relation:

$$R_\mu(u,v) = norm(\mu(u,v)) \quad (24)$$

where

$$norm(x) = \frac{x - x_{min}}{x_{max} - x_{min}}$$

being x_{min} and x_{max} the minimum and maximum values of the variable x.

In this manner, we have the fuzzy relations $R_{shareinterest}$ and $R_{H_{shareinterest}}$ that can be used to model topics of interest for u and v; and we have the fuzzy relations $R_{shareidea}$ and $R_{H_{shareidea}}$ that can be used to model ideas that users u and v share regarding topics of interest.

Given a fuzzy relation R between users, we can represent R as a weighted graph $G_R = (U, R)$, where the set of all users U are the vertices and the relation being the links between them.

Definition 5.1 *(Social group).* A social group regarding R is a connected component of G_R.

Definition 5.2 *(Social group of a user).* Given a user $u \in U$, the social group of u regarding R is the maximal connected component of G_R with u as a member.

Arguably, the simplest relation R that we can use would be one of the obvious relations (*follower*, *friends*, etc.). Nonetheless, *deep relations* like the ones we described above can be used as well to give a different perspective on the links between users.

We can establish a protocol to translate each deep relation into a possible link in the following way: let $R(u, v)$ be the number of messages of u related in some manner with user v.

Definition 5.3 *(R-neighbours).* For any user u, we can define their neighbours with regard to relation R as

$$Neighbour(R,u) = \{v : R(u,v) > 0, \forall v \in U\} \quad (25)$$

$$Neighbour^t(R,u) = \{v : R(u,v) > t, \forall v \in U\} \quad (26)$$

Let us illustrate how will be denoted a social group of interest towards a specific topic. Consider a scenario where we want to study different opinions for the topic *climate change*. We detected that NASA is a relevant actor in discussions of this topic but, currently, the account of the agency in Twitter has 30.3 million followers. Most of them would not even be interested in *climate change* itself, so it does not make sense to analyse all these users' opinions. Fortunately, we can define a metric CC such that

$$CC(u,v) = |\{n : n \in M \wedge author(v,n) \wedge \\ \wedge\, \exists m \in M : \big[author(u,m) : reply(n,m) \wedge \\ \wedge\, \text{"climate change"} \in topics(n)\big]\}| \quad (27)$$

Please cite this article in press as: M. Francisco, J.L. Castro, A fuzzy model to enhance user profiles in *microblogging* sites using deep relations, Fuzzy Sets Syst. (2020), https://doi.org/10.1016/j.fss.2020.05.006

which stands for the number of messages of v that reply to one of u's messages that specifically talks about *climate change*. This metric will yield the relation R_{CC}.

Now, we can define a graph $G_{CC}(U, Neighbour(R_{CC}, \text{"@nasa"}))$ that would specifically represent users interested in NASA's *tweets* about *climate change*, regardless of these users being actual followers of the agency's account.

6. Improving profile by propagation of properties through social groups

Building profiles through content analysis may not be enough. There are a number of cases in which profile attributes may differ from real ones (e.g. users that do not explicitly manifest their opinion or users who use irony to do it). Social Groups may be used to alter the value of these properties. In this section, we introduce the concept of *Social Property* as a way of creating profile attributes that contemplate both content analysis and Social Groups.

As the old adage says, *you will be judged by the company you keep*. Despite prejudice, it is not unreasonable to think that users will interact with groups and communities related to them in a certain manner, from family and friends to completely unknown persons with common interests and hobbies. In the previous sections, we enriched profiles with the concepts of *Topic-of-Interest* and *Idea Profiles*, but those are based on what users publish (or republish). These are good approaches but they may not be sufficient on their own, since the analysis did not include relations with the environment. For example, not all users who like *fine arts* are artists. There is no need for them to create content related to arts, but that does not mean that it is not among their passions. However, it is more likely that they follow and interact with accounts of artists.

The benefit of defining Social Groups by particular relations is that we are assuring the connected users share characteristics (in the previous example, they share interest in the topic *climate change*). We can take advantage of this situation to infer properties of the users.

Definition 6.1 *(Direct property).* Being Q a property on messages. $Q(m)$ stands for the *degree in which message m verifies Q*. We define Q_{direct} of a user u as

$$Q_{direct}(u) = \sum_{m \in authored_u} Q(m) \tag{28}$$

where $authored_u = \{m : m \in M \wedge author(u, m)\}$.

This would be a *content-based* approach towards profiling users. The downside of this method is that they need to explicitly show the property in at least one of their messages. Consider a user that has written a message regarding *climate change*. It will not be possible to extract an opinion but as we said earlier, it is possible to infer it from your neighbours. In the previous example, we defined a graph G_{CC} that would represent users interested in *climate change*. Picking a random user v related in some way to NASA by the relation R_{CC}, it is likely that we find another user w related to v (but not to NASA) who is also *climate aware*. If w is also related to other users in G_{CC}, it would not be unreasonable to assert that w is interested in *climate change* to some extent.

Definition 6.2 *(Social property).* Considering Q_{direct} as a property on users. We can define a social property $Q_{extended}$ as

$$Q^{(i)}_{extended}(u) = Q^{(i-1)}_{extended}(u) + \varphi \sum_{x \in Neighbourhood(u)} Q^{(i-1)}_{extended}(x) \tag{29}$$

where i stands for the i-th iteration of the algorithm. Initial values (when $i = 0$) would be:

$$\forall u \in U, Q^{0}_{extended}(u) = Q_{direct}(u) \tag{30}$$

The main advantage of this approach is that it can model people bias. We are not only what we post but who we are, and our actions offline affects to how users of SNS interact with us. Following the same line of exemplification, imagine a *tweet* from Donald Trump that says "*Climate change is a very important issue*". Any model will probably say that the user is worried about the topic *climate change*. However, any person from all around the world would know that this message is ironic, since Donald Trump has expressed his disagreement with scientific community on

Please cite this article in press as: M. Francisco, J.L. Castro, A fuzzy model to enhance user profiles in *microblogging* sites using deep relations, Fuzzy Sets Syst. (2020), https://doi.org/10.1016/j.fss.2020.05.006

this specific topic for many years. If Barack Obama would issue the same statement, again, any sentiment model would describe the same concern about climate, and it will not shock us.

Let us elaborate this example to illustrate the behaviour of our proposal. Suppose that we collect a number of messages from 50 users, where a few of them are related to climate change.

- Trump has only one tweet related to climate change that reads: *Climate change is a very important issue*.
- Obama had already published the exact same tweet regarding climate change: *Climate change is a very important issue*. It is important to notice that none of them is a copy of the other.
- For any sentiment algorithm, the polarity of both tweets will be the same, since they are a verbatim copy. Thus, both Obama and Trump will be marked with a positive polarity towards climate change.
- There are 20 users that share ideas with Trump. The analysis of their messages has reported that their polarity is negative towards climate change.
- There are 15 users that share ideas with Obama. The analysis of their messages has reported that their polarity is positive towards climate change.
- The rest of the users are not connected to those 35.
- We apply our proposal for property expansion. The polarity of Trump's neighbourhood will be negative and it will decrease the calculated polarity for Trump. The polarity of Obama's neighbourhood will be positive, hereby reinforcing his polarity.
- After expansion, the polarity of Trump will be negative towards climate change meanwhile Obama's will still be positive.

7. Experiments and discussion

There are a number of potential applications of our proposals, one of them being the assistance in the process of labelling datasets. In this section, we present our results when applying the proposed methodology to tag and profile users in SNSs, particularly in microblogging sites such as *Twitter*.

We wanted to prove that we can assist the tagging process of user profiles through the expansion of properties using deep relations. We believe that we can achieve reasonable accuracy in the expansion with a low number of well-tagged users. Labelling tweets and users requires that an expert reads the content and, following guidelines, assigns one or more labels. This is a tedious process that could be simpler if the system asks the expert to tag specific users in key points (or even random ones) and expand the knowledge to the rest of them.

In order to evaluate our proposal, a dataset of labelled users and/or messages was required. Most Twitter datasets are built using the standard (or streaming) API from the microblogging site. Raw data coming from the official API is extremely complex, so dataset creators normally reduce the amount of data keeping basic information, like messages, entities and, in some cases, the original author. Unfortunately, they get rid of some columns that are necessary to evaluate our proposal (e.g. the field that states if the message is a *retweet* or not and, in such case, the reference to the original message). To the best of our knowledge, there is no dataset available in the literature that would offer us sufficient tagged data to validate our framework while keeping relations between users/messages.

For this reason, we created our own dataset from almost two thousand tweets in Spanish collected during a political event. It is important to notice that the number of instances is not enough to validate our proposal. The task of tagging dataset is extremely time consuming and we do not have enough resources to put together a big dataset. This is a proof of concept and results should be taken with caution since further experimentation and analysis are required.

We collected more than 1900 Spanish tweets (from more than 850 users) containing a hashtag of a political event after Spain National Election during April 2019 to ensure they were related to the same topic. The results of these elections were not sufficient for any political party to form a government. In the course of the event, they were discussing if there should be a coalition government or not. We labelled messages in three classes,

1. Positive, if they were in favour of a coalition government.
2. Negative, if they were against a coalition government.
3. Neutral, if the tweet was neither in favour nor against.

Please cite this article in press as: M. Francisco, J.L. Castro, A fuzzy model to enhance user profiles in *microblogging* sites using deep relations, Fuzzy Sets Syst. (2020), https://doi.org/10.1016/j.fss.2020.05.006

Fig. 2. At the left, graph of users using a basic *retweet relation*. At the right, same users are plotted using *co-copies*. There are remarkable differences in the distribution and centrality of the users.

Table 3
Distribution of message instances over the three classes in our dataset.

Class	No. instances
Positive	563
Neutral	185
Negative	110

Number of instances per class can be seen in Table 3. After that, we aggregated tweets belonging to the same author in order to decide on the polarity of the user. We used this as a direct property (see eq. (28)) and we use *co-copies* (see eq. (17)) to build social groups. Fig. 2 shows differences in distribution of user graph when using *retweets* (copies) and *co-copies*. There are noteworthy differences when using *co-copies*: the graph is not directed anymore, allowing that the knowledge can be transferred in both directions; relations no longer revolve around popular users; it is possible to find key users that connect clusters of the graph. Overall, the graph is not centred in popular users but on the groups that are relevant to the analysis of the topic in question.

Once we tagged the dataset, we randomly selected a certain percentage of users (hereafter *known users*) and we removed the label for the rest of them. Then, we expanded the labels from *known users* to the rest of them using a *cross validation* scheme to ensure consistency of the results.

Fig. 3 shows results of the expansion process. After 3 iterations, the accuracy stabilises and even decreases if we keep iterating. This suggests that there is an optimal number of iterations that needs to be determined. Results show that, with only 20% of known users (approximately 40), the proposed methodology is able to classify the rest of the users with an accuracy of almost 0.7 in 3 iterations (see Table 4).

However, from the second graph in Fig. 3, it is noticeable that the errors are mainly related to neutral uses. There is a significant decrease in the performance of our algorithm when classifying neutral users: it tends to tag them as either positive or negative (see Table 5). Despite this being true, we need to remember that we only tagged messages. Users' polarity was computed through the aggregation of the polarity of the tweets. For those users that we only have one tweet collected, if the message was said to be neutral, then we would have the user labelled as neutral. This is not enough to say that users were actually neutral, since there may be tweets manifesting a clear opinion on the topic that we did not collect.

We wondered how many of those users were, in fact, neutral. Due to the high cost of examining the full timeline for all users, we did a second sampling to check how many users were correctly classified by our algorithm in spite of the lack of information in their messages. By examining the full *timeline* of the sample, our experts confirmed that 51 out of 86 users were incorrectly classified as neutral (see Table 6). Everything seems to indicate that the proposed methodology is able to amend the errors of content-based approaches through expansion of the attributes.

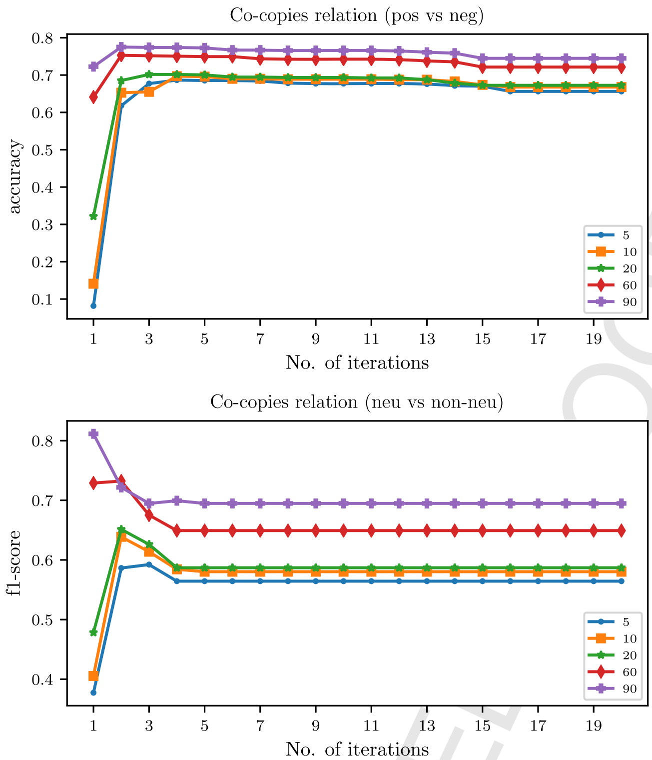

Fig. 3. Behaviour of accuracy and $f1$-score with different number of *known users*. We restricted the knowledge of the polarity of the users to those known users in order to see if the expansion algorithm is capable of estimating the polarity of the rest of them. First graph shows good performance when distinguishing between positive and neutral users. Second graph shows a decrease in this performance when referring to neutral users. However, this could be a problem related to the dataset (see discussion).

Table 4
Goodness of the positive vs. negative classification using a specific percentage of known users.

Known users	Accuracy		$f1$-score		Precision		Recall	
	mean	std	mean	std	mean	std	mean	std
5%	0.6402	0.1325	0.9844	0.0089	0.9694	0.0172	1.0000	0.0000
10%	0.6524	0.1211	0.9751	0.0306	0.9680	0.0171	0.9846	0.0576
20%	0.6683	0.0823	0.9827	0.0090	0.9661	0.0173	1.0000	0.0000
30%	0.7096	0.0536	0.9831	0.0093	0.9669	0.0179	1.0000	0.0000
40%	0.7140	0.0531	0.9836	0.0086	0.9678	0.0166	1.0000	0.0000
50%	0.7299	0.0240	0.9820	0.0090	0.9676	0.0170	0.9972	0.0126
60%	0.7326	0.0245	0.9824	0.0091	0.9682	0.0177	0.9974	0.0118
70%	0.7425	0.0176	0.9837	0.0087	0.9681	0.0168	1.0000	0.0000
80%	0.7503	0.0166	0.9829	0.0088	0.9689	0.0173	0.9975	0.0110
90%	0.7584	0.0141	0.9823	0.0086	0.9678	0.0170	0.9976	0.0107

8. Conclusions

In this article, we presented a fuzzy model to complete, amend and/or enrich user profiles in social networking sites, especially in those with the same interaction mechanisms as *microblogging* sites such as Twitter.

We designed a fuzzy framework that relies on the actual meaning of the actions we take on SNSs to add complex attributes to user profiles (interest and ideas towards specific topics). Copying or replying explicitly declare interest to what the user is saying, thus avoiding the noise from common relations like *following*. We also established heuristic mappings to determine the degree in which two users share interest or opinion in some topic, as a measure of similarity that can be applicable to define clusters.

Please cite this article in press as: M. Francisco, J.L. Castro, A fuzzy model to enhance user profiles in *microblogging* sites using deep relations, Fuzzy Sets Syst. (2020), https://doi.org/10.1016/j.fss.2020.05.006

Table 5
Goodness of neutral vs. non-neutral classification using a specific percentage of known users.

Known users	Accuracy		$f1$-score		Precision		Recall	
	mean	std	mean	std	mean	std	mean	std
5%	0.7871	0.1325	0.5575	0.0089	0.5512	0.0172	0.5897	0.0000
10%	0.8045	0.1211	0.5763	0.0306	0.5861	0.0171	0.5900	0.0576
20%	0.8174	0.0823	0.5866	0.0090	0.6063	0.0173	0.5873	0.0000
30%	0.8536	0.0536	0.6261	0.0093	0.7299	0.0179	0.5641	0.0000
40%	0.8544	0.0531	0.6244	0.0086	0.7428	0.0166	0.5511	0.0000
50%	0.8730	0.0240	0.6517	0.0090	0.8032	0.0170	0.5541	0.0126
60%	0.8761	0.0245	0.6586	0.0091	0.8169	0.0177	0.5586	0.0118
70%	0.8830	0.0176	0.6693	0.0087	0.8651	0.0168	0.5516	0.0000
80%	0.8930	0.0166	0.6896	0.0088	0.9187	0.0173	0.5568	0.0110
90%	0.9002	0.0141	0.7021	0.0086	0.9836	0.0170	0.5486	0.0107

Table 6
Statistics regarding the classification of neutral users. A thorough analysis of a second sample proved that our proposal correctly amended the polarity of 51 out of 86 users.

Errors that were, indeed, correctly classified		
a priori error	86	
false error (corroborated by experts second analysis)	51	**59.3%**
true error	35	40.7%

We established a methodology to determine social groups and expand specific attributes from users to their social group and vice versa, based on the premise that we are influenced by our environment. Altering user profiles through the use of social properties may highlight users that are trying to show an image of themselves that do not correspond to what their environment actually thinks of them. Alternative, it can also be used to reinforce or amend specific attributes of the user profile, closing the gap between the computed value of a model and the real value.

We tested our proposal with a small dataset in order to prove the interest that our fuzzy model has. We obtained good results when expanding the knowledge from well-labelled users to others connected using *co-copies*. Despite that further experimentation is required, our tests show that the method works well when predicting the actual polarity of *a priori* neutral users.

9. Future work

Although the intent of this document is to define a theoretical framework to create deep user profiles, we are currently in the process of developing a tool that put this scheme into practice. There are also a few parameters whose values require both theoretical and experimental research, since they are critic for the behaviour of the algorithm. This is the case of α and β for equations (12) and (15), γ for equation (21) and φ for the expansion of social properties (eq. (29)). We aim to tackle with these in particular domains, since they are context-dependant parameters. Finally, in order to validate our proposal, we need to create a dataset of labelled messages and include relations between users and messages. We plan to make this dataset available for the scientific community.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work has been financially supported by the Spanish Ministry of Economy and Competitiveness (MINECO) and European Social Fund (ESF), project FFI2016-79748-R. Manuel Francisco Aparicio was supported by the FPI

2017 predoctoral programme, from the Spanish Ministry of Economy and Competitiveness (MINECO), grant number BES-2017-081202.

References

[1] T. Lewis, How Social Media Mobilizes Society, 2013.
[2] K. Wagner, Facebook Passes 1 Billion Monthly Mobile Users, 2014.
[3] N. Gerhart, M. Koohikamali, Social network migration and anonymity expectations: what anonymous social network apps offer, Comput. Hum. Behav. 95 (2019) 101–113.
[4] J. Gantz, D. Reinsel, THE DIGITAL UNIVERSE IN 2020: Big Data, bigger digital shadows, and biggest growth in the Far East, in: IDC iView, 2020, pp. 1–16.
[5] T. Coughlin, 175 zettabytes by 2025, 2018.
[6] A. Singh, A. Sharma, A multi-agent framework for context-aware dynamic user profiling for web personalization, in: M.N. Hoda, N. Chauhan, S.M.K. Quadri, P.R. Srivastava (Eds.), Software Engineering, Advances in Intelligent Systems and Computing, Springer, Singapore, 2019, pp. 1–16.
[7] D.M. Boyd, N.B. Ellison, Social network sites: definition, history, and scholarship, J. Comput.-Mediat. Commun. 13 (2007) 210–230.
[8] H. Zhang, D. Liu, Fuzzy Modeling and Fuzzy Control, Control Engineering, Birkhäuser, Basel, 2006.
[9] M. Francisco, J.L. Castro, Extending clusters of social network users with deep relations, in: Proceedings of the 11th Conference of the European Society for Fuzzy Logic and Technology, EUSFLAT 2019, 2019, p. 52.
[10] The Associated Press, Whispers, secrets and lies? Anonymity apps rise, 2014.
[11] D. Chen, Q. Zhang, G. Chen, C. Fan, Q. Gao, Forum user profiling by incorporating user behavior and social network connections, in: J. Xiao, Z.-H. Mao, T. Suzumura, L.-J. Zhang (Eds.), Cognitive Computing – ICCC 2018, in: Lecture Notes in Computer Science, Springer International Publishing, 2018, pp. 30–42.
[12] B. Purkaystha, T. Datta, M.S. Islam, Marium-E-Jannat, Rating prediction for recommendation: constructing user profiles and item characteristics using backpropagation, Appl. Soft Comput. 75 (2019) 310–322.
[13] FAT/ML, Principles for accountable algorithms and a social impact statement for algorithms, 2019.
[14] M. Postigo-Boix, J.L. Melús-Moreno, Generating demand functions for data plans from mobile network operators based on users' profiles, J. Netw. Syst. Manag. 26 (2018) 904–928.
[15] M. Wang, Q. Tan, X. Wang, J. Shi, De-anonymizing social networks user via profile similarity, in: 2018 IEEE Third International Conference on Data Science in Cyberspace, DSC, 2018, pp. 889–895.
[16] P. Sánchez, A. Bellogín, Building user profiles based on sequences for content and collaborative filtering, Inf. Process. Manag. 56 (2019) 192–211.
[17] Z. Wang, Y. Zhang, H. Chen, Z. Li, F. Xia, Deep user modeling for content-based event recommendation in event-based social networks, in: IEEE INFOCOM 2018 - IEEE Conference on Computer Communications, 2018, pp. 1304–1312.
[18] Y. Li, Z. Zhang, Y. Peng, H. Yin, Q. Xu, Matching user accounts based on user generated content across social networks, Future Gener. Comput. Syst. 83 (2019) 104–115.
[19] G. Sarna, M.P.S. Bhatia, Content based approach to find the credibility of user in social networks: an application of cyberbullying, Int. J. Mach. Learn. Cybern. 8 (2017) 677–689.
[20] K. Akilal, H. Slimani, M. Omar, A robust trust inference algorithm in weighted signed social networks based on collaborative filtering and agreement as a similarity metric, J. Netw. Comput. Appl. 126 (2019) 123–132.
[21] J. Wu, J. Chang, Q. Cao, C. Liang, A trust propagation and collaborative filtering based method for incomplete information in social network group decision making with type-2 linguistic trust, Comput. Ind. Eng. 127 (2019) 853–864.
[22] F. Elmendili, Y. El Bouzekri El Idrissi, H. Chaoui, Detecting malicious users in social network via collaborative filtering, in: Proceedings of the 2nd International Conference on Big Data, Cloud and Applications, BDCA'17, ACM, 2017, pp. 44:1–44:7, Event-place: Tetouan, Morocco.
[23] C. Sirikayon, P. Thusaranon, P. Pongtawevirat, A collaborative filtering based library book recommendation system, in: 2018 5th International Conference on Business and Industrial Research, ICBIR, 2018, pp. 106–109.
[24] I. Hariadi, D. Nurjanah, Hybrid attribute and personality based recommender system for book recommendation, in: 2017 International Conference on Data and Software Engineering, ICoDSE, 2017, pp. 1–5.
[25] M. Ilhami, Film recommendation systems using matrix factorization and collaborative filtering, in: 2014 International Conference on Information Technology Systems and Innovation, ICITSI, 2014, pp. 1–6.
[26] M. Qazi, G.M. Fung, K.J. Meissner, E.R. Fontes, An insurance recommendation system using Bayesian networks, in: Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys '17, ACM, New York, NY, USA, 2017, pp. 274–278.
[27] A.A. Mustafa, I. Budi, Recommendation system based on item and user similarity on restaurants directory online, in: 2018 6th International Conference on Information and Communication Technology, ICoICT, 2018, pp. 70–74.
[28] S. Kutty, L. Chen, R. Nayak, A people-to-people recommendation system using tensor space models, in: Proceedings of the 27th Annual ACM Symposium on Applied Computing, SAC '12, ACM, New York, NY, USA, 2012, pp. 187–192.
[29] M. Manca, L. Boratto, S. Carta, Behavioral data mining to produce novel and serendipitous friend recommendations in a social bookmarking system, Inf. Syst. Front. 20 (2018) 825–839.
[30] M.B.S. Edith, W. Yu, Friend recommendation system based on mobile data, in: 2018 International Conference on Engineering Simulation and Intelligent Control, ESAIC, 2018, pp. 326–329.
[31] T.R. Kacchi, A.V. Deorankar, Friend recommendation system based on lifestyles of users, in: 2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics, AEEICB, 2016, pp. 682–685.

[32] T. Yuan, J. Cheng, X. Zhang, Q. Liu, H. Lu, How friends affect user behaviors? An exploration of social relation analysis for recommendation, Knowl.-Based Syst. 88 (2015) 70–84.
[33] S. Aral, D. Walker, Identifying influential and susceptible members of social networks, Science 337 (2012) 337–341.
[34] H. Ma, M. Jia, X. Lin, F. Zhuang, Tag correlation and user social relation based microblog recommendation, in: 2016 International Joint Conference on Neural Networks, IJCNN, 2016, pp. 2424–2430.
[35] H.-T. Zheng, Z. Wang, W. Wang, A.K. Sangaiah, X. Xiao, C. Zhao, Learning-based topic detection using multiple features, Concurr. Comput. 30 (2018) e4444, WOS:000438339700001.
[36] X. Guo, Y. Xiang, Q. Chen, Z. Huang, Y. Hao, LDA-based online topic detection using tensor factorization, J. Inf. Sci. 39 (2013) 459–469.
[37] Sayyadi Hassan, Raschid Louiqa, A graph analytical approach for topic detection, in: ACM Transactions on Internet Technology, TOIT, 2013.
[38] A. Alsanad, Arabic topic detection using discriminative multi nominal naïve Bayes and frequency transforms, in: Proceedings of the 2018 International Conference on Signal Processing and Machine Learning, SPML '18, Association for Computing Machinery, New York, NY, USA, 2018, pp. 17–21.
[39] J.R. Alharbi, W.S. Alhalabi, Hybrid approach for sentiment analysis of Twitter posts using a dictionary-based approach and fuzzy logic methods: study case on cloud service providers, Int. J. Semantic Web Inf. Syst. 16 (2020) 116–145.
[40] J. Awwalu, A.A. Bakar, M.R. Yaakub, Hybrid n-gram model using naïve Bayes for classification of political sentiments on Twitter, Neural Comput. Appl. 31 (2019) 9207–9220.
[41] M. Arora, V. Kansal, Character level embedding with deep convolutional neural network for text normalization of unstructured data for Twitter sentiment analysis, Soc. Netw. Anal. Min. 9 (2020) 12.

Please cite this article in press as: M. Francisco, J.L. Castro, A fuzzy model to enhance user profiles in *microblogging* sites using deep relations, Fuzzy Sets Syst. (2020), https://doi.org/10.1016/j.fss.2020.05.006

Sponsor names

Do not correct this page. Please mark corrections to sponsor names and grant numbers in the main text.

Ministry of Economy and Competitiveness, *country*=Spain, *grants*=FFI2016-79748-R

European Social Fund, *country*=European Union, *grants*=

Ministry of Economy and Competitiveness, *country*=Spain, *grants*=BES-2017-081202

Please cite this article in press as: M. Francisco, J.L. Castro, A fuzzy model to enhance user profiles in *microblogging* sites using deep relations, Fuzzy Sets Syst. (2020), https://doi.org/10.1016/j.fss.2020.05.006