Position Papers of the of the 17th Conference on Computer Science and Intelligence Systems pp. 65–68
DOI: 10.15439/2022F280
ISSN 2300-5963 ACSIS, Vol. 31

Evaluation without Ground Truth: a Comparative Study on Preference Mining Techniques in Twitter Social Network

Fabiola S. F. Pereira
Faculty of Computing
Federal University of Uberlândia
Uberlândia, Minas Gerais, Brazil
Email: fabiola.pereira@ufu.br

***Abstract*—In social media research the lack of ground truth for evaluation is a recurrent problem. We study the preference mining task in Twitter network which suffers from this lack of ground truth problem. We implement three different methods from literature, considering a common preference domain of news and carry a comparative study among them. Our preliminary findings show that is possible to combine methods in order to avoid unfeasible user surveying baselines and enable the evaluation of techniques. In the future, our target is to completely eliminate ground truth sets and evaluate based on correlation and causality techniques.**

I. Introduction

IN SOCIAL media research evaluation without ground truth is a pressing need [1]. Social networks are characterized by a huge volume of unstructured data, which can reveal from spatiotemporal and causal patterns to users sentiments. The problem is that mining such patterns is challenging due to the lack of reference values previously established. For example, according to [1] researchers are interested in discover when and where certain user activity is likely to occur – when the user is going to search for restaurant reviews? Where the user will be in the evening? Without surveying this user, however, the gap between prediction algorithms and reality can be deep.

We investigate the user preference mining problem in social networks [2]. Specifically, we look for patterns that describe preferences of a given social network user. For instance, considering a domain of news, we want to discover what are the user u preferred themes through her interactions in her social network. Thus, we can discover that u prefers to read news about *politics* over *sports* news, for example. This task fits exactly in the above discussed problem where we do not have ground truth preference values neither is feasible to manually survey each user about her preferences in the whole social network.

To tackle this problem, we conduct a comparative study among three different preference mining methods in Twitter dataset. The goal is to analyze the behavior of independent models when mining preferences and then overlap results. This resultant intersection set of mined preferences can perform as ground truth. Our contributions are two-fold. (i) Bring a set of different preference mining methods to the same context of Twitter and (ii) compare and overlap results building a trustworthy set of preferences to fill the gap caused by the lack of a ground truth.

II. The User Preference Mining Problem

User preference is a specific type of opinion that establishes an order relation between two objects. For example, when a user says: "I prefer sports than economy", we clearly identify her preference to sports themes over economy ones. These preference order relations (or preferences, for short) respect the irreflexive and transitive properties. We denote by $o_1 \succ_u o_2$ the preference of user u by the object o_1 over o_2 for $o_1, o_2 \in D$, where D is a preference domain.

Methods for learning and predicting preferences in an automatic way are among the recent research topics in disciplines such as machine learning, knowledge discovery and recommender systems. Approaches relevant to the area range from approximating of an as effective as possible question-answer process (preference elicitation) to collaborative filtering where customer preferences are estimated from the preferences of other customers [2]. In fact, problems of preference learning can be formalized within various settings, depending on the underlying type of preference model or the type of information provided as an input to the learning system. We explore the role of user preferences in recommender systems.

In a general way, to build an effective preference prediction system the following process is executed (Figure 1): first elicit patterns from feedback, which can be explicit (e.g. rating movies) or implicit (e.g. social data, visual perception, clicks, logs). The preference mining task consists in deriving a model from feedback able to infer a preference order between two given objects. This model is often referred to as prediction model. In some proposals, any preference mining task is used, and there is just a user profiling module that seeks to represent preferences through feature vectors or tensors. In the end, given some items and a target user u, the goal is to predict a preference order or a ranking (a special case of total orders of a set of alternatives) of these items according to u's preferences.

Many approaches have been used the term *user preferences* for different purposes. In recommender systems, this term refers to user profiling, i.e., the way that users' tastes are *represented*, generally by means of a feature vector or a tensor. In general Artificial Intelligence (AI) research, this *user preferences* term refers to the preference *order* over objects or ranking inferred by a preference model. In our work we refer to *user preferences* as well as in AI research line: the *preference order* induced over objects.

Problem definition. Given a social network $\mathcal{N}$, a user u and a preference domain D, return a set P of order relations $\succ_u$ over D that describes u's preferences.

III. The Twitter Social Network

Through Twitter Streaming APIs[1], during the course of 95 days, we collected tweets related to Brazilian news. All tweets, retweets and quoted-status[2] containing some mention to the Brazilian newspaper Folha de São Paulo, whose Twitter user is *@folha*, were considered. In all, we collected 1,771,435 tweets and 292,310 distinct users in a time span of tweets posting times from Aug 7 2016 to Nov 9 2016.

Based on such tweets, we applied Latent Dirichlet Allocation (LDA) [3] to extract a set of topics to represent user's preferences. We got a total of 50 topics, that then were manually grouped in 7 more general topics. In the end, the preference domain is $D = \{politics, international, sports, entertainment, security, economy, others\}$. The same crawling and topic extraction strategies were used in [4], but for a short dataset (3 weeks time span).

IV. Mining User Preferences from Twitter

We compare two literature methods for preference mining [5], [6] based on implicit feedback and the baseline method given by explicit feedback. Despite proposed in literature, neither of them have been evaluated.

Favorites (FV). This is the method based on explicit feedback. Intuitively, we can build a preference ranking based on the number of favorites (likes) a user gives over a topic in preference domain. Though apparently straightforward, this method still face to the lack of ground truth problem as users not necessarily assign as favorite their preferred topics. Sometimes, favorite can be a strategy to store some important post which not necessarily is preferred.

Follower Network (FN). This method proposed in [5] is based on Twitter following relationship. The intuition is that if a user u_1 follows some personalities or celebrities, then u_1 prefers topics related to what those personalities represent. For instance, if u_1 follows some representative celebrity user u_2 from fashion world, and u_1 does not follow u_3 which is a religious leader, then u_1 prefers *fashion* topics over *religion* topics. Bringing to our context, we match each item in the preference domain D with specific Twitter users referring to publishing channels from Folha de S. Paulo. These users compose the set A. The match is defined by the function $f : A \rightarrow D$. For instance, topic $t = sports$ is assigned to user $u =$*@folhaesporte*, $t =$*politics* is $u =$*@folhapoder* and so on. Then, we built a follower/following network considering all users in our dataset and extract preferences of a given user u according to the following steps : (1) if u does not follow any $v \in A$ then $others \in D$ is preferred by u over all $o \in D - \{others\}$. Else, (2) for each $v \in A$ followed by u, add $t \succ o$ in P_{FN}, for $t, o \in D$ and $f(v) = t$. Remark that in this strategy we just have two levels of preferences: the most preferred objects and the others.

Topic Distribution (TD). This method was proposed in [6]. In this comparative study we slightly adapted it as our goal is not recommendation, just user profiling and preference extraction. The preference mining strategy is based on the number of tweets/retweets about some topic $t \in D$. The most tweeted/retweeted topic by u is the preferred one over the second most tweeted/retweeted which is preferred over the third one and so on. As example, if u posts three times about *politics* and two times about *sports*, then we can establish a preference order between *politics* and *sports* ($politics \succ sports$ and *sports* is preferred over the remaining topics $o \in D$). Here preference order levels can be deeper than in FN method.

In face of three different methods FV, FN and TD, and a common preference domain D, our proposal in this article lies at the combination of these methods in order to obtain a final set P_{GR} containing trusty preference relations for a given user, which can supply the lack of ground truth.

V. Preliminary Findings

Our goal in these preliminary experiments is to observe the preference set mined for each method for a given user. We seek to answer *how the methods FV, FN and TD overlap in their results?* We define the agreement score S_u for a given user u as

$$S_u = \frac{|P_{M_1} \cap P_{M_2} \cap ... \cap P_{M_n}|}{|P_{M_1} \cup P_{M_2} \cup ... \cup P_{M_n}|} \tag{1}$$

where P_{M_i} is the resultant preference set mined by method M_i and $n > 1$ is the total number of methods being combined. The higher S_u the higher the agreement among methods and thus, more reliable is the resultant preference ground truth set $P_{GT} = |P_{M_1} \cap P_{M_2} \cap ... \cap P_{M_n}|$.

Considering user u_1 (id=279635698), the mined preference sets are described below.

$P_{FV} = \{politics \succ_{u_1} international, politics \succ_{u_1} sports, politics \succ_{u_1} entertainment, politics \succ_{u_1} security, politics \succ_{u_1} economy, politics \succ_{u_1} others\}$

[1]https://dev.twitter.com/streaming/

[2]Quoted-status are retweets with comments

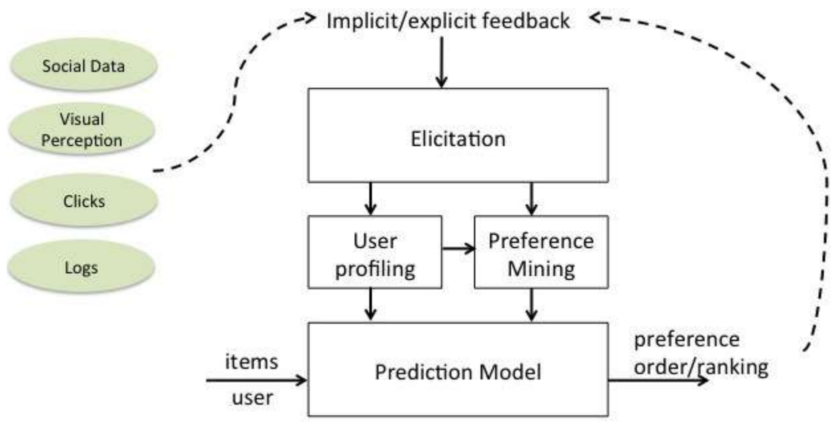

Fig. 1. Schema of a traditional preference prediction system.

$P_{FN} = \{politics \succ_{u_1} international, politics \succ_{u_1} entertainment, politics \succ_{u_1} security, politics \succ_{u_1} economy, politics \succ_{u_1} others, sports \succ_{u_1} international, sports \succ_{u_1} entertainment, sports \succ_{u_1} security, sports \succ_{u_1} economy, sports \succ_{u_1} others\}$

$P_{TD} = \{politics \succ_{u_1} international, politics \succ_{u_1} sports, politics \succ_{u_1} entertainment, politics \succ_{u_1} security, politics \succ_{u_1} economy, politics \succ_{u_1} others\}$

The corresponding agreement scores for u_1 are in Figure 2. Each $C_{M_1,\ldots,M_n}$ corresponds to the methods combination run.

Combination	S_{u_1}
$C_{FV,FN}$	0.454
$C_{FV,TD}$	1.0
$C_{FN,TD}$	0.454
$C_{FV,FN,TD}$	0.454

Fig. 2. Agreement scores among methods FV, FN and TD for user u_1

Combination	S_{avg}
$C_{FV,FN}$	0.061
$C_{FV,TD}$	0.429
$C_{FN,TD}$	0.066
$C_{FV,FN,TD}$	0.03

Fig. 3. Agreement scores among methods FV, FN and TD averaged for all users.

Figure 3 summarizes our results so far. The agreement score S_{avg} is the average of scores of all users in our dataset. Notice that the best score is for combination $C_{FV,TD}$. Also, FN is the worst performance, penalizing the full combination $C_{FV,FN,TD}$ score.

Discussions. There are other social network preference mining methods in literature not embraced in this study [7], [8]. The challenge is in applying methods proposed for very specific and diversified contexts in the same preference domain study. The technique from [7], for example, could not be applied in our news preference domain due to the lack of comparative sentences in tweets. In [8] the approach is ranking preference learning and the preferences are extracted from labels indicating fan page's political view in Facebook. A problem not tackled yet relies on consistency issues in resultant preference set P_{GT}. Given the transitive property of a preference relation $\succ$ over a domain D, P_{GT} is consistent if there is not any inferred preference $o \succ o \in P_{GT}$ for $o \in D$.

VI. Final Remarks

We have raised the discussion about evaluating without ground truth in social media research. In this context, we are studying the problem of preference mining in Twitter network. Three different existent methods have been implemented considering a common preference domain of news categories (sports, politics etc). In order to supply the lack of ground truth in our problem, we have proposed a combination strategy of the resultant set of preferences of each method to generate a final trustworthy set of user preferences. Our next steps will be study more methods [7], [8], [9] and social networks (Last.fm, Instagram, TikTok), and organize them according to their main features. Our final goal is to propose a framework able to extract preferences based on correlation and causality patterns, to eliminate the need of ground truth sets of preference relations. We expect that our method generalizes over other sources of data, for instance IoT domain [10] and web media [11].

References

[1] R. Zafarani and H. Liu, "Evaluation without ground truth in social media research," *Com. ACM*, vol. 58, no. 6, pp. 54–60, 2015.
[2] J. Furnkranz and E. Hullermeier, *Preference Learning*. Springer, New York, 2010.
[3] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[4] F. S. F. Pereira, S. de Amo, and J. Gama, "Detecting events in evolving social networks through node centrality analysis," *Large-scale Learning from Data Streams in Evolving Environments with ECML/PKDD*, 2016.

[5] ——, "On using temporal networks to analyze user preferences dynamics," in *Discovery Science: 19th International Conference, DS 2016, Bari, Italy, 2016.*, 2016.

[6] X. Liu, "Modeling users' dynamic preference for personalized recommendation," in *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI'15)*, 2015, pp. 1785–1791.

[7] F. S. F. Pereira and S. de Amo, "Mineracao de preferencias do usuario em textos de redes sociais usando sentencas comparativas," in *Symposium on Knowledge Discovery, Mining and Learning (KDMiLe)*, 2015, pp. 94–97.

[8] M. A. Abbasi, J. Tang, and H. Liu, "Scalable learning of users' preferences using networked data," in *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, ser. HT '14. New York, NY, USA: ACM, 2014, pp. 4–12.

[9] H. Al-Jarrah, M. Al-Asa'd, S. A. Al-Zboon, S. K. Tawalbeh, M. M. Hammad, and M. AL-Smadi, "Resolving conflict of interests and recommending expert reviewers for academic publications using linked open data," in *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 2019, pp. 91–98.

[10] T. Elsaleh, S. Enshaeifar, R. Rezvani, S. T. Acton, V. Janeiko, and M. Bermudez-Edo, "Iot-stream: A lightweight ontology for internet of things data streams and its use with data analytics and event detection services," *Sensors*, vol. 20, no. 4, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/4/953

[11] T. R. Tangherlini, S. Shahsavari, B. Shahbazi, E. Ebrahimzadeh, and V. Roychowdhury, "An automated pipeline for the discovery of conspiracy and conspiracy theory narrative frameworks: Bridgegate, pizzagate and storytelling on the web," *PLOS ONE*, vol. 15, no. 6, pp. 1–39, 06 2020. [Online]. Available: https://doi.org/10.1371/journal.pone.0233879