An Adaptive Technique for Weighting Multiple Factors in Followee Recommendation Algorithms

Antonela Tommasel and Daniela Godoy
ISISTAN Research Institute, CONICET-UNCPBA, Tandil, Buenos Aires, Argentina

Abstract

The accurate suggestion of interesting friends arises as a crucial issue in recommendation systems. This work argues that the criteria for recommending friends (or followees) needs to be adapted and combined according to each user's preferences. A technique is proposed for adapting such criteria to the characteristics of previously selected followees. Experimental evaluation showed that the technique improved the precision of static weighting strategies. Results highlighted the importance of adapting to changes in user preferences over time.

1 Introduction

Online social networks have an important place in the life of millions of users who actively use them for finding new friends. The decision to start following other users simultaneously attends to several reasons, which might differ for each individual user. Thus, understanding how users select followees emerges as a key design factor of strategies to personalise recommendations. Interestingly, most followee selection approaches are only based on equally important and independent factors, disregarding how users' interests can affect the followee selection. This work argues that followee recommendation criteria needs to be personalised according to users' preferences. A technique is proposed for adapting such criteria to each user considering the characteristics of previously selected followees.

2 Related Work

Several approaches have proposed to suggest interesting users in social networks based on a unique and independent factor [Golder and Yardi, 2010; Hannon *et al.*, 2010]. Approaches that combine several factors assume that they are equally important to each user by averaging or multiplying them [Armentano *et al.*, 2011]. Closely related to this work, Agarwal and Bharadwaj [2013], and Garcia and Amatriain [2010] personalised factors' weights. However, unlike the technique proposed by this work, changes over time in user preferences were not considered for adapting the weights.

3 An Adaptive Technique for Personalising Followee Recommendation

The technique suggests a list of interesting followees by optimally combining different recommendation factors. The combination is particular to each user as it is based on his/her preferences reflected on previously selected followees.

Computing Factor Weights. The overall similarity between users u and v ($Similarity(u,v)$) can be defined as a linear combination of the similarity for each followee recommendation factor ($sim_i(u,v)$) and its corresponding weights (α_i) as follows: $\sum_{i=1}^{n} \alpha_i * sim_i(u,v)$. As recommendation systems aim to find the most similar potential followees, factors' weights (α_i) should accurately capture user preferences. Thus, they are defined by considering the characteristics of the previously selected followees. Followees are assumed to be chosen by a determined factor if their similarity with the target user for such factor is higher than a pre-defined threshold. The preference of users regarding each factor is computed as the percentage of followees for whom the similarity is higher than the threshold. Then, percentages are used as the similarity weights that will be further updated according to user preferences.

Updating Factor Weights. The computed weights are used for assessing the similarity between each potential followee and the target user in the recommendation process. The target user is presented with the set of most similar potential followees. For each accepted followee, i.e. each potential followee the target user has accepted or manifested interest in, weights are updated to reflect the new interests of the target user.

Ranking Recommended Followees. In standard similarity-based algorithms, as all recommended candidates are similar to the target user, they are likely to be similar to each other. Thus, such algorithms will never uncover certain items, which although less similar to the target user, are also important [Hurley and Zhang, 2011]. Consequently, it would be desirable to include novel or diverse items in the recommended list. Novelty could be introduced to similarity-based algorithms aiming at balancing both, the relevance of candidate followees (i.e. its similarity to the target user) and the diversity of recommendations. Novelty can be measured in terms of the degree to which is unusual regarding the target user normal interests (i.e. the previously selected followees). It can be computed as $\frac{\sum_{i \in followees(u)} abs(Similarity(u,i) - Similarity(u,pf))}{|followees(u)|}$, where u represents the target user, pf represents the potential followee, $followees(u)$ represents the previously selected users of u and $Similarity$ is the overall similarity. If previously selected followees are similar to the target user, and the new potential followee is dissimilar to the target user, he/she will also be dissimilar to previously selected followees. The higher the absolute differences, the higher the dissimilarity, and thus the novelty introduced. Consequently, the novelty of a potential followee can be assessed without computing the actual dissimilarity between the potential followee and each previously selected followee.

Finally, the potential followees are ranked by considering the linear combination of relevance and novelty. The weight of the novelty is computed as the percentage of the previously selected followees for whom the novelty score was higher than a pre-defined threshold. Similarly, the weight of the relevance is computed as the percentage of the previously selected followees for whom novelty was lower than the threshold. Both weights are updated as previously described.

4 Experimental Evaluation

This section presents the experimental evaluation performed to assess the effectiveness of the proposed technique.

Factors for Followee Recommendation. Although the presented technique could be applied to any arbitrary number of recommending factors, this work focuses in the two main followee recommendation factors: topology and content.

Topology. Most link prediction algorithms are based on network topology. The number of common followees is one of the most common metrics applied to *Twitter* network. It can be defined as $\frac{|\Gamma_{out}(x) \cap \Gamma_{out}(y)|}{\Gamma_{out}(x) \cup \Gamma_{out}(y)}$, where x and y are nodes (i.e. users), k_x is the degree of node x, and $\Gamma(x)$, $\Gamma_{out}(x)$ and $\Gamma_{in}(x)$ are the set of neighbours, followees and followers of x , respectively.

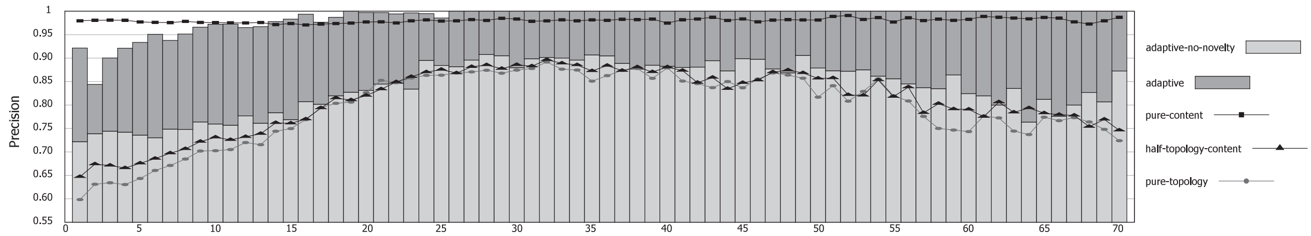

Figure 1: Comparison of Precision Results

Content. The interest of a user can be characterised by profiles based not only on the information they create and publish (*publishing profile*), but also on the information they consume (*reading profile*), for example the retweets. The *publishing profile* of user u_j is built by considering all of the user tweets ($tweets(u_j)$), which can be defined as: $pub-profile(u_j)=tweets(u_j)$. The *reading profile* of a user u_j can be defined as: $read-profile_{RT}(u_j)=tweets_{RT}(u_k)\,\forall k \in followees(u_j)$. The similarity between the *reading profile* of a user and the *publishing profile* of their potential followees is assessed using the cosine similarity.

Experimental Settings. To evaluate the performance of the proposed technique, potential followees were ranked and the top-*N* users were selected. For each user, their actual followees and a equal proportion of randomly selected non-followees were added to the pool of potential followees to be recommended. To simulate the actual behaviour of target users over time, actual followees were added to the pool of potential followees in the same order in which the user started following them.

The proposed technique (*adaptive*) was compared against three static baselines: *pure-topology*, *pure-content* and *half-topology-content*. Additionally, *adaptive* was compared to a version that ignores the novelty factor: *adaptive-no-novelty*.

The quality of recommendations was evaluated by selecting a ranked sub-set of the potential followees and computing the overall precision immediately after the weights were updated. As there is no explicit feedback from target users available, the evaluation assumes that items that were not originally part of the followee set are uninteresting for the user. This assumption might not be completely accurate as recommended users might not be selected simply because the user was unaware of them. As a result, precision might be underestimated.

The pool of potential followees comprised 20 users, out of which 10 were recommended to the user. Factors' weights were updated after 10 accepted recommendations. Initially, the technique assumes that no followee has been selected. Thus, all factors are assigned equal weights. The minimum similarity threshold was set to 0.7 for the content-based factor, and to 0.2 for the topology factor. The novelty threshold was set to 0.05.

Dataset. The dataset was obtained by crawling a set of 3,453 target users listing the language account as English, and having at least 10 followees and 10 published tweets. All user information was retrieved by means of the TwitterAPI[1].

Experimental Results. Figure 1 shows the evolution of the average recommendation precision for the first 70 weights updates performed. As regards the baselines, the best results were achieved when considering the *pure-content* alternative, which achieved a precision higher than 0.95, with differences up to a 58% regarding the worst baseline (*pure-topology*). These results indicated that although the majority of the followee relations were content driven, there were also followee relations that were not found with a pure content oriented strategy. Topology-based results further highlighted the fact that the majority of the followee relations are content driven.

[1] `https://api.twitter.com`

Regarding the proposed technique, the *adaptive-no-novelty* achieved the worst results. As a result, although the combination of weights is adapted to each user, it is not sufficient for further improving results. Also, it can be inferred that although users have a particular preference for a certain type of followees, they also select followees that do not exactly match such preferences. Consequently, the search and ranking of users should not be only guided by similarity, but also by novelty. Adding novelty (*adaptive*) improved the best baseline. As the figure shows, the *adaptive* alternative was able to achieve an optimal precision after 26 weights updates. These results evidenced the importance of recommending both similar and novel followees. Finally, it is also shown the precision stability once the preferences of users are learned and adapted.

Regarding the differences between the weights predicted by the technique, and the real preferences of the target users), the absolute differences were below 0.1 for the 76% of target users, highlighting the usefulness of the proposed technique not only for adequately capturing users' interests, but also for adapting to the changes in user preferences over time.

In summary, precision of recommendations can be improved when considering an adaptive technique for defining the weights of the recommendation factors. Results emphasised the importance of adapting the relevance or weights of the factors to changes in user preferences over time, and also considering diversity in followee recommendations.

5 Conclusions

This work proposed a technique for adapting the followee selection criteria to the decisions of each particular user regarding the characteristics of his/her previously selected followees. Experimental evaluation showed that the proposed technique helped to improve precision results regarding static weighting strategies. Furthermore, results highlighted the importance of adapting to the changes of the user preferences over time.

References

[Agarwal and Bharadwaj, 2013] V. Agarwal and K. K. Bharadwaj. A collaborative filtering framework for friends recommendation in social networks based on interaction intensity and adaptive user similarity. *Social Netw. Analys. Mining*, 3(3):359–379, 2013.

[Armentano *et al.*, 2011] M. Armentano, D. Godoy, and A. Amandi. A topology-based approach for followees recommendation in Twitter. In *Proceedings of the ITWP at 22nd IJCAI*, pages 22–29, 2011.

[Garcia and Amatriain, 2010] R. Garcia and X. Amatriain. Weighted content based methods for recommending connections in online social networks. In *Proceedings of the 2nd RSWeb*, pages 68–71, Barcelona, Spain, 2010.

[Golder and Yardi, 2010] S. A. Golder and S. Yardi. Structural predictors of tie formation in twitter: Transitivity and mutuality. In Ahmed K. Elmagarmid and Divyakant Agrawal, editors, *SocialCom/PASSAT*, pages 88–95. IEEE Computer Society, 2010.

[Hannon *et al.*, 2010] J. Hannon, M. Bennett, and B. Smyth. Recommending Twitter users to follow using content and collaborative filtering approaches. In *Proceedings of the 4th ACM Conference RecSys*, pages 199–206, 2010.

[Hurley and Zhang, 2011] N. Hurley and M. Zhang. Novelty and diversity in top-n recommendation – analysis and evaluation. *ACM Trans. Internet Technol.*, 10(4):14:1–14:30, March 2011.