Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media

Using Co-Following for Personalized Out-of-Context Twitter Friend Recommendation

Ingmar Weber and **Venkata Rama Kiran Garimella**
Qatar Computing Research Institute
iweber@qf.org.qa, gvrkirann@gmail.com

Abstract

We present two demos that give personalized "out-of-context" recommendations of Twitter users to follow. By out-of-context we mean that a user wants to receive recommendation on, say, musicians to follow even though the user's tweets' contents and social links have no connection to the "context" of music. In this setting, where a user has never expressed interest in the context of music, many existing methods fail.

Our approach exploits co-following information and hidden correlations where, say, a user's political preference might actually provide clues about their likely music preference. For example, a user u might be recommended a particular music band b because u also follows a set of politicians P, and other users who follow members of P tend to follow b, rather than an alternative b'.

We implement this framework in two very distinct settings: one for recommending musicians and one for recommending political parties in Tunisia. Our framework is simple and similar to Amazon's "users who bought X also bought Y" and can be used not only for explainable out-of-context recommendations but also for social studies on, say, which music is "closest" to users of a particular political affiliation. It also helps to introduce and to "link" a user to an unknown domain, say, politics in Tunisia.

Our two web-based demos are publicly accessible at http://scd1.qcri.org/twitter/musicians/ (for recommending musicians) and http://scd1.qcri.org/twitter/tunisia/ (for recommending Tunisian parties).

Introduction

More and more users publicly express their interests in various topics on social networking sites such as Twitter. To keep users engaged and to help them find more relevant content, both the sites themselves as well as third party solutions try to suggest users to follow. The best-performing methods for this task all make use of the user's immediate social network and look for *triadic closure*: if a user u is already following a, b and c and all of a, b and c are also following d then d is a good friend suggestion for u as such an edge (u, d) closes several triangles (Golder and Yardi 2010; Brzozowski and Romero 2011). These methods perform well due to how social circles work, where the "circles" are more like "cliques" (in a graph theoretic sense), and due to homophily where my friends would most likely follow the same kind of politicians as me. But such approaches work not so well when a sports fan who is only connected to other sports fans is asking for music recommendations or when a U.S. technology junkie wants to know which political parties in Tunisia are most similar to him. Similarly, content-based approaches would still require a user to mention the topic of interest at least once (Hannon, McCarthy, and Smyth 2011).

We call such recommendations "out of context" recommendations and we present a simple framework that uses collaborative filtering to make such recommendations. For a Twitter user following all of the Los Angeles Lakers players, who is asking for a music recommendation, we would try to find musicians that are followed by Twitter users who also follow Los Angeles Lakers players.

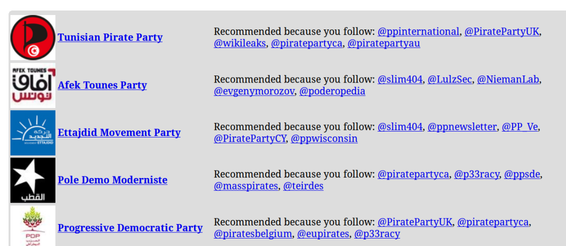

Figure 1: Screenshot of *http://scd1.qcri.org/twitter/tunisia/* showing recommendations of political parties to follow for Twitter user *@ChatoX*.

As our demos make "links" used for the recommendation explicit in an attempt to explain the recommendations, they can also be used for studies in computational social science. For example, Figure 1 shows the Tunisia party recommendations for the user @ChatoX, a user who supports open source software and is politically left-of-center. The recommendations shown are appropriate, even though the querying user does however not follow a single user from Tunisia.

How it Works

Our demos solely use Twitter follower network information and no textual tweet content, trading a loss in information for a more general, language-agnostic setting. Our method requires as input a set of *potential recommendation targets*, also called *seed accounts*, centered around a common topic. For each element in the set we then obtain all incoming following links, i.e., all users that follow it. If this set is very large[1] then for further processing we use a uniform random sample of followers of each seed account. For each of the followers selected we then obtain up to 5,000 of their outgoing following links, i.e., friends they follow.[2] This "followers of x also follow y" information is then turned into a single vector summary as follows.

First, for each follower of a seed user x the lists of their friends are treated as high dimensional vectors, with each user being followed corresponding to one dimension. The dimensions corresponding to any seed user x_i are ignored to ensure *out-of-context* recommendations and the remaining dimensions are IDF-weighted with less weight assigned for following globally popular accounts. The vector obtained is then normalized in 2-norm. For all x's followers, these normalized vectors are added up and re-normalized in 2-norm. This vector s_x then serves as a summary of the interests of x's followers.

At query time, for a user u we get up to 5,000 of u's friends. This list is turned into a high dimensional vector and normalized in 2-norm to obtain s_u.[3] For each of the seed users x_i we then compute the cosine similarity between s_u and s_{x_i} and rank the x_i correspondingly.

In practice, we slightly deviate from this process and, e.g., prune users that are only followed by a single user following the seed account of interest.

Data Sets

Tunisian Political Parties. We collected follower information for 14 Twitter accounts representing political parties in Tunisia. These Twitter accounts were identified using http://www.partistunisie.com/, http://en.wikipedia.org/wiki/List_of_political_parties_in_Tunisia and with the help of domain experts.[4] Two accounts, @UGTT_TN and @NahdhaTunisie, had more than 10k followers and for these we considerered a uniform random sample of size 10k.

Musicians on Twitter. We compiled a list of musicians on Twitter by combining the lists at bit.ly/1gl73Iv, on.mash.to/1fQjSjk, bit.ly/1lpFTVt and bit.ly/1kb19yc. Musicians with fewer than 10k followers were ignored and we manually removed some false list entries, such as music producers (not musicians) and relocated accounts. In the end 295 musicians were retained and for each of these the feature vectors were constructed as described above.

[1] @JustinBieber alone has more than 50 million followers.

[2] The limit of 5,000 friends is the maximum that can be obtained in a single call to the Twitter REST API. However, as users require more than 4,000 followers to be able to follow more than 5,000 users, "normal" users are well below this limit.

[3] Though this normalization does not change the ranking of the recommendations, it ensures that the cosine similarity values fall within $[0.0, 1.0]$.

[4] We thank Yasmine Ryan (Al Jazeera) and Muzammil Hussain (University of Michigan) for compiling this list.

Example Applications to Computational Social Science

As an example from the music domain, the first two musicians recommended for @Kaka (the soccer star) are (i) Sandy Leah (@SandyLeah), a Brazillian pop artist, and (ii) Ivete Sangalo (@ivetesangalo), who plays Latin pop/Samba-reggae in general. This makes sense because Kaka is also from Brazil. Again the results are very different for a basketball star, say Shaquille O'Neal @SHAQ, with rapper T.I (@Tip) and R&B artist Mary J. Blige (@maryjblige) ranked highest.

Just as music and sports are linked, also food preferences and music are linked. For example, @EatingWell who follows lots of accounts on healthy eating, is recommended (i) @DaveJMatthews (@davejmatthews) and (ii) Jennifer Lopez (@jlo), whereas @rdublife (head of social media at McDonals) is recommended (i) Neko Case (@NekoCase) and (ii) Wayne Coyne (@waynecoyne) (both related to Indie/Alternate rock).

Of course, not all recommendations are "correct" but even the explanations, given by the online demo of the incorrect ones are often worth exploring as they show that unexpected accounts are "closer than expected".

Conclusions

We presented two demos that construct out-of-context recommendations for Twitter users for two very distinct domains: Tunisian politics and music.

Although our approach works beyond a user's immediate ties, it still requires the querying user to have expressed some interests in the form of following Twitter users. It is impossible for our system to bridge the contextual gap between a user u and potential recommendation targets t unless there is at least a single 3-hop bridge of the form $u \rightarrow x \leftarrow u' \rightarrow t$, where $u \rightarrow x$ indicates that u follows x. Anecdotally, from trying many examples, we remark that the reach of these 3-hop bridges is impressively far due to the globalized nature of Twitter. Given the folklore "six degrees of separation" the broad reach we observed might not be surprising after all.

In the future, we want to use our out-of-context recommendations to build "bridges" from a user to a new topic. For example, someone without knowledge about Tunisia might be more inclined to read news articles about upcoming Tunisian elections if hidden links can be pointed out to them such as "users like you who follow @wikileaks and @LulzSec are interested in the Tunisian Pirate Party." (see Figure 1).

References

Brzozowski, M. J., and Romero, D. M. 2011. Who should i follow? recommending people in directed social networks. In *ICWSM*.

Golder, S. A., and Yardi, S. 2010. Structural predictors of tie formation in twitter: Transitivity and mutuality. In *SocialCom/PASSAT*, 88–95.

Hannon, J.; McCarthy, K.; and Smyth, B. 2011. Finding useful users on twitter: Twittomender the followee recommender. In *ECIR*, 784–787.