Causal Dependencies for Future Interest Prediction on Twitter

Negar Arabzadeh
Ryerson University
narabzad@ryerson.ca

Hossein Fani
University of New Brunswick
hfani@unb.ca

Fattane Zarrinkalam
Ryerson University
fzarrinkalam@ryerson.ca

Ahmed Navivala
Ryerson University
ahmed.navivala@ryerson.ca

Ebrahim Bagheri
Ryerson University
bagheri@ryerson.ca

ABSTRACT

The accurate prediction of users' future topics of interests on social networks can facilitate content recommendation and platform engagement. However, researchers have found that future interest prediction, especially on social networks such as Twitter, is quite challenging due to the rapid changes in community topics and evolution of user interactions. In this context, temporal collaborative filtering methods have already been used to perform user interest prediction, which benefit from similar user behavioral patterns over time to predict how a user's interests might evolve in the future. In this paper, we propose that instead of considering the whole user base within a collaborative filtering framework to predict user interests, it is possible to much more accurately predict such interests by only considering the behavioral patterns of the most influential users related to the user of interest. We model influence as a form of causal dependency between users. To this end, we employ the concept of *Granger causality* to identify causal dependencies. We show through extensive experimentation that the consideration of only one causally dependent user leads to much more accurate prediction of users' future interests in a host of measures including ranking and rating accuracy metrics.

ACM Reference Format:
Negar Arabzadeh, Hossein Fani, Fattane Zarrinkalam, Ahmed Navivala, and Ebrahim Bagheri. 2018. Causal Dependencies for Future Interest Prediction on Twitter. In *The 27th ACM International Conference on Information and Knowledge Management (CIKM '18), October 22–26, 2018, Torino, Italy.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3269206.3269312

1 INTRODUCTION

Users' topics of interest show the dynamic evolution of user behavior in online social networks such as Twitter, whose effective prediction can improve user experience and can increase advertising revenue, just to name a few. Common time series models, which use observations from past time intervals to predict the users' future topics of interest, have an independence assumption that users' behavior is considered to evolve independently from the other users. These methods overlook the explicit or implicit social interactions that are inherent to social networks. On the other hand, time-aware collaborative filtering approaches such as timesvd++ [7] and recurrent recommender networks (rrn) [10] propose a valuable step forward by integrating the individual and collective perspectives of the users in addition to their temporal evolution patterns (non-stationarity) under the traditional collaborative filtering framework. Successful as they are, these approaches, however, do not consider strict inter-user dependence (social influence) and only benefit from users' behavioral correlation to make predictions. In contrast, social-aware recommender systems such as trustsvd [5] and socialpmf [6] have already been proposed to address this issue but they in turn overlook temporal evolution.

In this paper, we propose to consider both the temporal evolution of users' interests as well as a stricter form of inter-user influence through the notion of *causal dependency*. We employ Granger causality to determine the degree of inter-user influence that can be used to identify which users play influential roles in the behavioral evolution of one or more other users. Based on Granger causality, we identify a causing user c to influence the affected user e if and when the past observations of c lead to a more accurate prediction of the behavior of e above and beyond the information contained in past observation of e alone.

To capture the influence between users, we build a topic preference time series representing users' interests towards topics over time. Then, we employ Granger causality [4], G-causality for short, to identify influence relations between users on a per topic basis. This leads to a weighted directed network of users, denoted by the *influence network*, in which the edges depict the influence direction of its adjacent users. Finally, we use the influence network to perform interest prediction. Specifically, given a topic of interest z and a user e, we find e's influential neighbour(s) from the influence network such as c and build a vector autoregression model (var) based on e and c's topic contribution time series to predict e's degree of interest toward topic z in the future.

We perform experiments based on a Twitter corpus and compare our work for future user interest prediction with several state of the art baselines. The objective of the experiments is to see whether our proposed notion of inter-user influence, which incorporates both the temporal evolution of user interest behavior captured through a times series representation and user causal dependencies determined based on G-causality, can show a stronger predictive power compared to other baselines in terms of both ranking and rating metrics. Summarily, we find that (1) the prediction of user's future interests on Twitter can be much more accurately predicted based on their influence network compared to other state of the art baselines especially compared to variations of temporally and

CIKM '18, October 22–26, 2018, Torino, Italy

ACM ISBN 978-1-4503-6014-2/18/10…$15.00
https://doi.org/10.1145/3269206.3269312

socially-aware collaborative filtering methods, and (2) within the influence network, the consideration of the most highly influential user is sufficient for the accurate prediction of users' future interests and the inclusion of additional influential users does not lead to any statistically significant improvement.

2 RELATED WORK

Most existing work perform user interest prediction based mainly on two information sources: users' posting behaviour over time [11], and underlying user interaction on the social network [5]. There are also work that employ both of these information sources in tandem to improve prediction performance [1, 7, 10].

Koren [7] proposed timesvd++ for modeling the non-stationarity of user preferences and topic popularity over time in the context of a collaborative filtering system. While timesvd++ uses the same characteristic vector as in svd++, it alters the base estimators for the user's degree of interest towards a topic, topic popularity as well as user's characteristic vector as a function of time. The authors have mentioned that this model interpolates past observations and as such might fall short in extrapolating future temporal dynamics. Wu et al. [10] have recently proposed a recurrent neural network model, namely recurrent recommender networks (rrn), to perform future behaviour prediction. Specifically, two long short-time memory (lstm) models have been used to learn the dynamics of users' degree of interest and items' popularity separately such that the user's degree of interest towards a topic in the future depends on the state representation of the user and the topic in previous time intervals. Contrary to timesvd++, rrn is nonparametric. While rrn complies with the precedence rule in causality, that is, rrn always relies on the previous states to estimate the next states in contrast to timesvd++, it does not explicitly model any inter-user influences such as causality or social interaction, e.g., friendship.

Similar to Wu et al.'s goal to predict users' interest in the future, Bio et al. [1] have extended the social probabilistic matrix factorization (socialpmf) model [6] to incorporate users' topics of interest through an exponential time decay function. Like rrn and unlike timesvd++, temporal-socialpmf follows the precedence rule of causality. Additionally, it respects the inter-user influence because its underlying socialpmf model enables the integration of a social network structure. This work is similar to ours in a sense that we both try to predict future user interests in social networks by taking the inter-user influence into account. However, instead of the social network structure, our work considers an influence network derived based on users' causal dependencies.

3 USER INTEREST PREDICTION

3.1 Problem Definition

Given a set of topics $\mathcal{Z}$ from Twitter within T time steps (e.g. days) extracted by a topic detection method (e.g. lda) and a set of users $\mathcal{U}$, we represent the time-aware topic preferences of each user $e \in \mathcal{U}$ towards each topic $z \in \mathcal{Z}$ over time steps $1 \leq t \leq \mathrm{T}$ as a time series $X_{ez} = [x_{ez,1} : x_{ez,\mathrm{T}}]$, namely *topic preference timeseries*, where $x_{ez,t} \in \Re^{[0,1]}$ indicates the preference by user e for topic z at time step t. The main objective of our work is to accurately predict $x_{ez,\mathrm{T}+1}; \forall z \in \mathcal{Z}, \forall e \in \mathcal{U}$.

3.2 Proposed Approach

Our approach consists of three pipelined phases: users' topic preference detection, users' influencers identification, and users' future topics of interest prediction. In the following, we describe the details of each step.

3.2.1 Topic Preference Detection. Our work relies on users' dynamic behavior towards a set of topics within a time period. To incorporate both users' topics of interest and temporality, for each user $e \in \mathcal{U}$, we model her inclination towards each topic $z \in \mathcal{Z}$ at each time step $1 \leq t \leq \mathrm{T}$ by a topic preference time series. To instantiate the topic preference time series $X_{ez} = [x_{ez,1} : x_{ez,\mathrm{T}}]$ for user e, we find i) a set of topics $\mathcal{Z}$ that have been observed within T time steps, and ii) e's degree of interest at time t towards each topic z, i.e., $x_{ez,t}$.

We derive the set of topics from the collection of users' posts using lda. To this end, we view all tweets authored by each user e at time step t as a single document $d_{e,t}$. Given the document corpus $\mathcal{D} = \{d_{e,t} | \forall e \in \mathcal{U}, 1 \leq t \leq \mathrm{T}\}$ and the number of topics $|\mathcal{Z}|$, lda distills $\mathcal{D}$ into two probability functions following a Dirichlet distribution: i) distribution of words in each topic ϕ_z, describing what each topic z is about, and ii) distribution of each topic z in each document $\theta^z_{d_{e,t}} \in \Re^{[0,1]}$, showing e's degree of interest toward z at time step t. Formally, $x_{ez,t} = \theta^z_{d_{e,t}}$.

3.2.2 Influencer Identification. We leverage the influence between users when delivering their topics of interest over time to predict users' topics of interest. The influence that one user might exert on the other is identified through Granger causality [4]. Granger causality has been perceived as a predictive notion of causality between time series [2] and as such in our case, it can be applied to the topic preference time series. In the bivariate case, user e, the effect, is said to be influenced (Granger caused) by another user c, the cause, with respect to topic z, if and only if, regressing on past values of both e and c's topic preference time series is statistically significantly more accurate than doing so with past values of e alone. Formally, let $X_{ez} = [x_{ez,1} : x_{ez,\mathrm{T}}]$ and $X_{cz} = [x_{cz,1} : x_{cz,\mathrm{T}}]$ be two stationary topic preference time series of user e and c with respect to topic z, the two regression models are:

$$\begin{aligned} \mathrm{H}_1 : x_{ez,\mathrm{T}} &= \sum_{l=1}^{\mathrm{L}} a_l x_{ez,t-l} + \sum_{l=1}^{\mathrm{L}} b_l x_{cz,t-l} + \epsilon_1 \\ \mathrm{H}_0 : x_{ez,\mathrm{T}} &= \sum_{l=1}^{\mathrm{L}} a_l x_{ez,t-l} + \epsilon_2 \end{aligned} \tag{1}$$

where L is the maximal time lag, a_l and b_l are the regression variable coefficients, and ϵ_1 and ϵ_2 are the residual terms, which are i.i.d according to a standard Gaussian $\mathcal{N}(0, \sigma^2)$. If H_1 is a significantly better model than H_0 (e.g., provides more precise predictions), we conclude that X_{cz} Granger causes X_{ez}; notationally $c \xrightarrow{\mathrm{G}}_z e$. Among other techniques, the significance level can be tested using the F statistic by the Granger-Sargent test [4], defined as follows:

$$F = \frac{(rss_{\epsilon_2} - rss_{\epsilon_1})/\mathrm{L}}{(rss_{\epsilon_1})/(\mathrm{T} - 2\mathrm{L})} \sim F(\mathrm{L}, \mathrm{T} - 2\mathrm{L}) \tag{2}$$

where rss_{ϵ_2} is the restricted residual sum of squares under H_0, rss_{ϵ_1} is the unrestricted residual sum of squares under H_1, T is the number of time steps, and F follows the F-distribution. We reject the null hypothesis that c does not Granger cause e if the above calculated F is greater than the critical value of the F-distribution for some desired false-rejection probability, e.g., 0.05.

Based on Granger causality between all users for all topics, i.e., $c \xrightarrow{G}_z e; \forall e, c \in \mathcal{U}, \forall z \in \mathcal{Z}$, it is possible to find causal dependency between pairs of users in order to identify influencers. The set of influencers for a user form its influence network.

3.2.3 Interest Prediction. We use estimated vector autoregression (var) model to do one-step-ahead prediction of users' topics of interest at time step T+1. Given a user e and a topic z, we build a var model whose variables are e's topic preference time series and her influencer network, identified by Granger causality, up to time step $t = \mathrm{T}$. Formally,

$$Y_{z,\mathrm{T}+1} = b + \sum_{l=1}^{\mathrm{L}} A_l Y_{z,t-l+1} + \epsilon \tag{3}$$

where $Y_{z,t}$ is a vector whose first element is equal to e's degree of interest toward topic z at time step t; notationally $Y^{(1)}_{z,t} = x_{ez,t}$. The other elements belong to e's influencers such as c, i.e., $Y^{(i)}_{z,t} = x_{cz,t}; i > 1$. Here, b is a vector of constants (intercepts), A_l is a time-invariant matrix of coefficients and ϵ is a vector of error terms. After model estimation (training) to learn b, A_l and ϵ, the predicted degree of interest for user e towards topic z at time step T+1, denoted as $\hat{x}_{ez,\mathrm{T}+1}$, will be $Y^{(1)}_{z,\mathrm{T}+1}$.

Overall, $|\mathcal{U}| \times (|\mathcal{Z}| \times |\mathcal{U}| + |\mathcal{Z}|); |\mathcal{U}| >> |\mathcal{Z}|$, var models should be trained for pairwise Granger causality tests and one time step ahead predictions. While the time complexity of our method is quadratic function of the number of users, its parallel implementation is able to reduce the complexity to linear complexity; $|\mathcal{U}|$ users in parallel with each other.

4 EVALUATION

4.1 Dataset and Setup

We adopted a dataset consists of 2,948,742 tweets authored by 135,731 unique users posted in Nov. and Dec. 2010. We sampled the dataset to obtain the active users who posted more than 100 tweets, resulting in a total of 2,458 users, as our user set $\mathcal{U}$ for evaluating our approach. Additionally, we collected the followership networks of the users using Twitter api. The whole two months time period is sampled on a daily basis, i.e., T+1 = 61 days. The settings in each step of our method are as follows:

Topic Preference Detection. We applied lda using Mallet[1] after removing stopwords. The number of topics for this dataset has already been investigated in [3] and as such set to 50. We created $X_{ez} = [x_{ez,1} : x_{ez,\mathrm{T}=60}]; \forall e \in \mathcal{U}$, up to day 60 as our observation to find the users' influencers and estimating the var model in order to predict e's degree of interest at the future day 61, i.e., $\hat{x}_{ez,\mathrm{T}+1=61}$.

Influencer Identification. Our users' topic preference time series satisfy the Granger causality assumption of stationarity as they passed different stationarity tests, namely the phillips-perron, Augmented Dickey-Fuller and KPSS. The significance level and the maximum number of lags were set to 0.05% and 2, respectively. Bayesian information criterion (bic) was used to find optimal lag.

Interest Prediction. We used the first 4 values of the users' topic preference time series as the presamples to initiate the var models estimation. Appropriate number of lags has been determined similar to the influencer identification step.

4.2 Baselines

We compare our work against the following baselines:

Scrs [9] is a non-temporal baseline. Based on this method, user's interests in the future are semantically similar to the ones a user has been interested in the past. The authors use linked open data to extract item features to compute the similarity of two items. To apply their approach in our context, we consider each topic of interest as an item and the constituent Wikipedia entities of a topic as its content. Given each topic is a distribution over Wikipedia entities, this method predicts $x_{ez,\mathrm{T}+1}; \forall z \in \mathcal{Z}, \forall e \in \mathcal{U}$ as follows:

$$\hat{x}_{ez,\mathrm{T}+1} = \frac{1}{\mathrm{T} \times |\mathcal{Z}|} \sum_{t=1}^{\mathrm{T}} \sum_{z' \in \mathcal{Z}} x_{ez',t} \times s_{zz'} \tag{4}$$

where s denotes the similarity of two topics calculated by the cosine similarity of their respective entity weight distribution vectors.

Zarrin et al. [11] is similar to the scrs approach as it follows a content-based approach. However, the authors first model high-level interests of a user over Wikipedia categories, denoted by C, then predict the user's future interests of based on her categories of interest. This method predicts $x_{ez,\mathrm{T}+1}; \forall z \in \mathcal{Z}, \forall e \in \mathcal{U}$ as follows:

$$\hat{x}_{ez,\mathrm{T}+1} = \sum_{a \in C} r_{za} \times i_{ua} \tag{5}$$

where r_{za} is the degree of relatedness of topic z to category a and i_{ua} denotes the degree of interest of user u to category a.

Timesvd++ [7] is the temporal extension to svd++. The implementation in librec[2] was used in our experiments. We performed a grid search over the bin size in {1,2,4,8,16,32,64} and factors size in {10,20,40,80} to select the best settings. Other settings were left to default value, i.e., learning rate=0.01 and regularization $\lambda = 0.1$.

Rrn [10][3] is a temporal collaborative filtering approach based on recurrent neural nets. We performed grid search over bin size in {1,2,4,8,16,32,64} and users and topics' dynamic states size in {10,20,40,80}. Other hyperparameters were set to default: single-layer LSTM with 40 hidden neurons and embeddings size of 40.

Temporal-socialpmf (tspmf) [1] is another temporal baseline which incorporates social context in terms of both collaboration and social connections into probabilistic matrix factorization (pmf) [6]. We implement this method using librec's socialpmf method. We set β, the kernel parameter, to 3 and θ, the weight parameter that indicates how important the whole previous time points are to the current one, to 0.2 in the exponential decay function.

Granger is our proposed approach in which the set of top influencers for each user with respect to a topic is considered in a bivariate var model. We performed experiments on an increasing number of influencers $k \in \{1, 2, 5, 10, 20\}$.

[1] `mallet.cs.umass.edu/topics.php`

[2] `www.librec.net`

[3] The implementation is kindly provided by its authors.

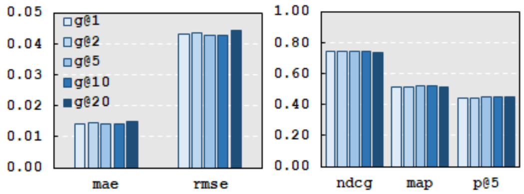

Figure 1: Performance with varying number of influencers.

4.3 Metrics

We evaluate the predictive power of the baselines using *rating* metrics including mean absolute error (mae) and root-mean-squared error (rmse). We also assess the baselines in terms of ranking metrics, ndcg, map, and p@5, to show that our proposed approach provides not only the most accurate predictions, but also the best rankings compared to the baselines.

4.4 Results

We first explored the impact of the size of the influencer network on the performance of our model and then compared our work to other baselines. We evaluated the performance of our approach for varying number of top-influencers $k \in \{1, 2, 5, 10, 20\}$ that were used in the influence network. As seen in Figure 1, the accuracy of our approach did not show statistically significant improvement or deterioration on any of the metrics for different number of influencers (g@k). This can be due to two factors: i) the var model structure which is used to incorporate the k influencers' topic preference time series selects the top (k=1) influencer's topic preference time series as its salient component for all baselines $k > 1$, and ii) pairwise (bivariate) Granger causality test could potentially lead to misleading influencers as mentioned by Ding et al. [8] for cases when more than one causes are considered. For instance, let c, m, and e be three users where $c \xrightarrow{G}_z m$ and $m \xrightarrow{G}_z e$. Pairwise Granger analysis would yield $c \xrightarrow{G}_z e$ and not be able to distinguish whether the causality between e and c is direct or mediated by m. As such, we conclude that only considering each user's top influencer is sufficient to accurately predict the user's future interests; therefore, without loss of generality, we compare our proposed approach with the baselines based on g@1.

The comparative results of the baselines in terms of prediction error and ranking metrics are shown in Figures 2 and 3, respectively, based on the average over all user-topic predictions at time step T+1. We report rrn and timesvd++ with the bin and factor size based on grid search, which we found to be 2 and 20, respectively.

As seen in Figure 2, our proposed approach outperforms other baselines in terms of prediction error and the difference in all cases is statistically significant based on a paired t-test at 0.05. Amongst the temporal methods based on collaborative filtering, while rrn outperforms timesvd++ which is in line with the results reported in [10], it is still weaker than content-based methods such as Zarrin et al. and scrs, and tspmf, as well as our proposed approach. We attribute the accuracy of our approach to the fact that it directly models and leverages the impact of user's top influencers on her future interests, an impact that is overlooked in all other baselines.

The baselines also fail to output better ranking compared to our proposed approach as shown in Figure 3 where the difference between our proposed approach and all baselines on the three metrics

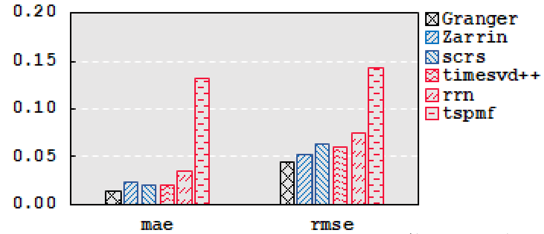

Figure 2: Comparison on rating metrics (lower is better).

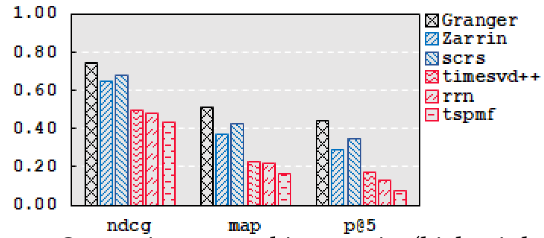

Figure 3: Comparison on ranking metrics (higher is better).

is statistically significant. A similar performance trend could be seen in ranking as in prediction error for all baselines. Our approach performs consistently better, the content-based baselines scrs and Zarrin are the runner-up, and last are the temporal collaborative filtering baselines timesvd++, rrn, and tspmf. There are, however, some exceptions. Rrn, which has a better prediction error shows weaker performance compared to timesvd++ in terms of ranking.

5 CONCLUDING REMARKS

We addressed the problem of predicting users' future interest on Twitter. We propose a method that considers (1) temporal evolution of users' interests through a time series representation and (2) the impact of causal dependencies between users by constructing an influence network based on the concept of Granger causality. We have shown that compared to several strong state of the art baselines, our method provides statistically significantly better performance on both ranking and rating metrics.

REFERENCES

[1] Hongyun Bao, Qiudan Li, Stephen Shaoyi Liao, Shuangyong Song, and Heng Gao. 2013. A new temporal and social PMF-based method to predict users' interests in micro-blogging. *Decision Support Systems* 55, 3 (2013), 698–709.
[2] Prasanta Bhattacharya and Rishabh Mehrotra. 2016. The Information Network: Exploiting Causal Dependencies in Online Information Seeking. In *CHIIR'16*. 223–232.
[3] Hossein Fani, Ebrahim Bagheri, and Weichang Du. 2017. Temporally Like-minded User Community Identification through Neural Embeddings. In *CIKM'17*. 577–586.
[4] C. W. J. Granger. 1969. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* 37, 3 (1969), 424–438.
[5] Guibing Guo, Jie Zhang, and Neil Yorke-Smith. 2015. TrustSVD: Collaborative Filtering with Both the Explicit and Implicit Influence of User Trust and of Item Ratings. In *AAAI'15*. 123–129.
[6] Mohsen Jamali and Martin Ester. 2010. A matrix factorization technique with trust propagation for recommendation in social networks. In *RecSys'10*. 135–142.
[7] Yehuda Koren. 2009. Collaborative filtering with temporal dynamics. In *SIGKDD'09*. 447–456.
[8] Ding Mingzhou, Chen Yonghong, and Bressler Steven L. 2006. *Granger Causality: Basic Theory and Application to Neuroscience.* Wiley-Blackwell, 437–460.
[9] Tommaso Di Noia, Roberto Mirizzi, Vito Claudio Ostuni, Davide Romito, and Markus Zanker. 2012. Linked open data to support content-based recommender systems. In *8th International Conference on Semantic Systems, 2012*. 1–8.
[10] Chao-Yuan Wu, Amr Ahmed, Alex Beutel, Alexander J. Smola, and How Jing. 2017. Recurrent Recommender Networks. In *WSDM'17*. 495–503.
[11] Fattane Zarrinkalam, Hossein Fani, Ebrahim Bagheri, and Mohsen Kahani. 2017. Predicting Users' Future Interests on Twitter. In *ECIR'17*. 464–476.