USER INTEREST DETECTION IN SOCIAL MEDIA USING DYNAMIC LINK PREDICTION

by

Chaitra R Hosmani, B.Eng., Visvesvaraya Technological University, 2008

A Major Research Paper
presented to Ryerson University
in partial fulfillment of the requirements for the degree of

Master of Science
in the Program of
Data Science and Analytics

Toronto, Ontario, Canada, 2019

© Chaitra R Hosmani 2019

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A MAJOR RESEARCH PAPER (MRP)

I hereby declare that I am the sole author of this Major Research Paper. This is a true copy of the MRP, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this MRP to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this MRP by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my MRP may be made electronically available to the public.

Chaitra R Hosmani

USER INTEREST DETECTION IN SOCIAL MEDIA USING DYNAMIC LINK PREDICTION

Chaitra R Hosmani

Master of Science 2019

Data Science and Analytics

Ryerson University

ABSTRACT

Social media provides a platform for users to interact freely and share their opinions and ideas. Several researches have been conducted to predict user interests in social media. Because of the dynamic nature of social media, user interests change over time. In this paper, given a set of emerging topics and user's interest profile over these emerging topics we are interested to predict the user interest profile for the future. We conducted this experiment on twitter data captured for 2 months from 1 November 2011 to 1 January 2012. We will be using temporal latent space to infer characteristics of users and then predict user's future interests over these given topics. We will evaluate the results with different ranking metrics like MAP and nDCG. We will also compare our results with the results of Zhu et al. [4] temporal latent space which uses the same methodology but on a different dataset.

Key words:

Social Media, User interest prediction, Matrix factorization, Temporal network

TABLE OF CONTENTS

LIST OF EQUATIONS

LIST OF FIGURES

I. Structured Abstract

1. Background

Several researchers have experimented with different approaches to accurately predict future interests for users in social media. This ability to predict future interests has a variety of applications such as recommender system, news feeds, suggesting contact. The popularity of Twitter as a social media network and ease of access to datasets has made many researchers employ twitter data for their analysis. We will also be using Twitter dataset for our project for its rich content and ease of access.

2. Aim

In this paper we are interested to predict user's interests in future given historical interests. i.e. given graph $(G_1, G_2, \dots, G_t)$ our objective is to find G_{t+1}. By exploiting Twitter data from 1 November 2010 to 1 January 2011 we will build a temporal latent space which is a representation of each user in an unobserved latent space. A low rank dynamic temporal latent space is built with non-negative matrix factorization and will be used to infer user interest in future.

3. Methodology

i. Data extraction: A Graph is represented as G = (V, E) with V as set of nodes and E set of edges connecting the nodes. A graph is formed with the given dataset which depicts users, active topics and user's interest over these topics. Multiple snapshots of graph are created as $(G_1, G_2, \dots G_t)$.

ii. Building temporal latent space: Each interaction graph (G) is mapped into a temporal latent space (Z) with k dimensions. This low rank latent space is inferred with the assumption that dimension of latent space is much smaller than the number of nodes and these nodes change the positions gradually in their latent space. We use fast BCGD (Block Coordinated Gradient Descent algorithm) which is a modified version of standard BCGD for the construction of latent space as $Z = (Z_1, \dots Z_t)$

iii. Link prediction: The latent space Z_{t+1} is approximated by all historic latent space $(Z_1, \dots Z_t)$ and user's future prediction G_{t+1} is inferred from Z_{t+1}

4. Results

The classification accuracy using AUC score is in the range of 0.38 to 0.48, which means model has no separation between relevant and non-relevant topics for users.

5. Conclusions

Even though temporal latent space modeling is a powerful technique which has been successful in predicting links in different networks, it was not successful in predicting user's interests on Twitter dataset. User interest in Twitter might be changing rapidly and abruptly influenced by real-world events. The assumption of temporal smoothness may not hold true for this dataset.

II. Introduction

People use social media to communicate their ideas and opinions. Social media also reflects current world events faster than other media. Researchers are interested in finding valuable insights from users' posts on social media. Analysing social media data has proved its usefulness in multiple applications. Recommending a product, evaluating sentiments on a product, suggesting a contact, predicting evolution of topic, detecting anomalous activities, news feeds are few applications of social media analysis.

Twitter is a popular social media platform enables users to post messages in the form of tweets. User can reply to tweets in the form of retweets. The unique characteristics of Twitter is that user gets to express her opinions within limited 140 characters.

In this paper we will discuss our approach to predict user's future interests in Twitter based on Zhu et al. temporal latent space method. Our entire section is divided as below. In section III, we will review the previous work done in regard to extracting topics from social media and detecting future interests of users in social media and also conduct exploratory analysis on our dataset. In section IV, we will elaborate the methodology to obtain future interests and describe the type of experiments and evaluation methods used. In section V, we will present the results and analysis and lastly in section VI, we arrive at conclusions and future work.

III. Background and Literature Review

Literature Review

Several researches have been conducted on social media analysis with different objectives and approaches. We will begin with unique challenges of using Twitter dataset and explore previous work on extracting topics, finding user's interest over these topics and evolution of user's interest over time.

1. Extracting topics and Interest prediction

Since the tweets are short, messy and unstructured it's a non-trivial problem to extract topics from the tweets. Several methods have been experimented to extract the prevalent topics in the social media. The text in the tweets, mentions, URLs and hashtags giveaway clues about the topic that user is interested in. The naive method is to extract topics is 'Bag of words approach' used by Chen et al. [17] and Shin et al. [18]. The topic is extracted by using the terms used in a tweet. However, this approach may not lead to accurate prediction of topics as tweets suffers from polysemy and synonymy. Moreover, the topics generated in this way are high level, generic and may not capture emerging trends in social media.

Next approach for topic formation would be 'Mixture of Model' to create topics. The most commonly used algorithm LDA (Latent Dirichlet Algorithm) creates soft clustering of documents on different topics. All the tweets of a user are aggregated into a document and LDA is applied over these documents [20].

Another approach to extract topic would be 'Bag of Concepts'. Kapanipathi et al. [19] has identified the topics as set of concepts from external knowledge base. Each tweet is annotated with concepts from external knowledge base like DBPedia. A concept graph is formed which represents ephemeral semantic correlation between the concepts during that period. Group of concepts which have high internal cohesion will form a topic. Topics generated in this manner are more granular and can capture emerging trends in social media. Interest prediction is based on intuition that when user has tweeted more on a specific topic, more the user is interested in that topic. User inclination towards these extracted topics is calculated by applying the distribution of user's tweet over these topics.

2. Implicit interest prediction

Many social media users are not active users, i.e. they might consume a lot of content but may not explicitly contribute to the tweets. Users' contribution to tweets follows the power law distribution. Finding the implicit interest of users will increase the accuracy of recommender system or news feeds and boosts users' engagement in social network.

For implicit interest prediction, Zarrinkalam et al. [2] used graph-based link predictions over multiple heterogeneous graphs. The three heterogeneous graphs are created as following. (i) User's followership graph which is unweighted and directed (ii) User's explicit interest graph which is weighted and undirected graph representing user's inclination towards the topics and (iii) Topic relatedness graph which is weighted and undirected. The topic graph representing relationship between topics is formed by either (a) Semantic similarity (b) Collaborative relatedness or (c) Hybrid approach. The implicit user's interests are estimated by applying various link prediction methods like SimRank, Common neighbors and Jaccard's coefficient.

Researchers have identified extracting the information about topics in user's tweet has an important clue in detecting user's implicit interest. If the user has explicitly tweeted about a topic in the past, user might be also interested in semantically similar topics in future. Based on this intuition Trikha et al. [11] proposed frequency pattern mining for detecting implicit user's interests. The topics are extracted by annotating the tweets with Wikipedia concepts and applying LDA over collection of these concepts. The frequent topics are mined using FP-growth algorithm for it's better performance over other algorithms like Apriori and Eclat. Interest profile for user is constructed using a transaction database. For each user we will search if any explicit topic of interest is present in the frequently mined set, if yes then other topics in the set are added to user’s interest profile. The resulting interest profile will contain both user’s explicit and implicit interests.

3. Extracting future interest prediction

Our next literature review explores research work on finding future interest prediction. Methods like Collaborative filtering, frequency pattern mining can not be applied as they suffer from cold start problems. Zarrinkalam et al. [3] predicted future interest of users over set of unobserved topics. This research assumes that even though user’s interest changes over time, they still revolve around broader related category. User’s interest at different snapshots are mapped into a Wikipedia category structure. By generalising the category structure and using the temporal evolution of user’s interests Zarrinkalam et al. was able to infer (i) the future topics and (ii) user interests over these future topics.

Another approach is to infer future link prediction is by latent space modelling of network. Given a sequence of graphs from 1 to t with each graph representing a snapshot of links in the network, the aim is to predict links for t+1 time. The static latent space would not capture dynamic changes in the network. In paper [4] author proposes temporal latent space model for dynamic link prediction. The interaction graph is represented $G = (G_1, \dots, G_t)$ where G_1 is the graph at time 1 and G_t is the graph at time t. These graphs are mapped into k dimensional latent space $Z = (Z_1, \dots, Z_t)$ where each dimension represents unobserved characteristics of each node. Two nodes which are closer in the observed graph are also closer in temporal latent space. Z_{t+1} is approximated from $Z = (Z_1, \dots, Z_t)$ and G_{t+1} is inferred from Z_{t+1}. Our work uses the same design and methodology as this paper.

Time Series models assume overlook the dependencies of users with other users in the network. This has overcome in recurrent recommender network(rrn) combines individual and collective information of the network. Arabzadeh et al. [10] modeled influence between users along with temporal changes of network and changes in user’s interest over topics. Author states that Granger causality which determines the degree of influence a user can have from another user which helps to predict user’s interests. The experiment identifies a causing user ‘c’ who influences affected user ‘e’ such that by observing past predictions of ‘c’ will have more accurate predictions for ‘e’ than using just the past predictions of ‘e’.

In our study we are leveraging the data used in Arabzadeh et al. [10] research on casual dependencies where we have the Twitter dataset extracted from 1 November 2010 to 1 January 2011. We also have the prevalent topics extracted using LDA topic modelling and distribution of user’s interests over these topics from this existing research dataset. We will use the design and methodology used in Zhu et al. [4] dynamic temporal latent space which is successfully applied on other networks like Facebook, YouTube, DBLP (computer science bibliography website).

Exploratory Analysis

The dataset consists of 2,948,741 tweets from 130,168 unique users. Users who have tweeted more than 100 times are considered as active users and these users are used for training and evaluation purposes. From our dataset we obtained 2458 unique users for our evaluation criteria. The number of topics is set as 100. Each topic is a collection of multiple concepts from Wikipedia. The user's interest over these topics are created by giving weights based on their explicit tweets.

1. Top 20 words for a given topic using LDA

Below table shows the top 10 concepts for a single topic. The value column shows the importance of this concept in forming the specific topic. Based on some of the concepts listed such as "Qatar", "Happiness", "FIFA World Cup" we can infer this topic as "the announcement made on 2 December 2010, about Qatar hosting the 2022 FIFA World Cup".

Id	Topic Id	Concept Id	Value	Concept Name
22974	560	8238258	0.046553218	Qatar
22976	560	169409	0.043675078	Happiness
22982	560	11370	0.042669126	FIFA World Cup
22968	560	11175	0.035264202	Political freedom
22964	560	11049	0.035208316	FIFA
22970	560	25391	0.032525778	Russia
22966	560	5417956	0.027076871	RT (TV network)
22965	560	10568	0.023248666	Association football
22987	560	13327177	0.019364574	2018 FIFA World Cup
22975	560	17742072	0.019336631	2022 FIFA World Cup
22990	560	9904	0.014726018	England national football team
22991	560	9316	0.012658228	England
22998	560	36493580	0.012071423	Russia national football team
22972	560	32817	0.01195965	Vladimir Putin
22981	560	36639348	0.01151256	Grammy Award
22952	560	38714	0.011456674	World
22967	560	62028	0.008997681	CNN
22989	560	22844966	0.008494705	England 2018 FIFA World Cup bid
22954	560	3434750	0.008103501	United States

22979	560	867877	0.007488753	Share (finance)

2. Examples of Concepts

Below table shows extracted summary of a few concepts. This summary is extracted from DBPedia and each concept is uniquely identified with an Id. A group of concepts will together form a unique topic.

Id	Summary	Concept Name
31653	The First Amendment (Amendment I) to the United States Constitution is part of the Bill of Rights. The amendment prohibits the making of any law respecting an establishment of religion, impeding the free exercise of religion, abridging the freedom of speech, infringing on the freedom of the press, interfering with the right to peaceably assemble or prohibiting the petitioning for a governmental redress of grievances.	First Amendment to the United States Constitution
39862	An entrepreneur () is an enterprising individual who builds capital through risk and/or initiative. The term was originally a loanword from French and was first defined by the Irish-French economist Richard Cantillon. Entrepreneur in English is a term applied to a person who is willing to help launch a new venture or enterprise and accept full responsibility for the outcome. Jean-Baptiste Say, a French economist, is believed to have coined the word entrepreneur" in the 19th century - he defined an entrepreneur as "one who undertakes an enterprise	especially a contractor
90451	Amazon.com, Inc. () is an American multinational electronic commerce company with headquarters in Seattle, Washington, United States. It is the world's largest online retailer. The company also produces consumer electronics - notably the Amazon Kindle e-book reader - and is a major provider of cloud computing services.	Amazon.com
2471540	Hacking means finding out weaknesses in a computer or computer network, though the term can also refer to someone with an advanced understanding of computers and computer networks. Hackers may be motivated by a multitude of reasons, such as profit, protest, or challenge. The subculture that has evolved around hackers is often referred to as the computer underground, but it is now an open community.	Hacker (computer security)

3369375	Cyberwarfare refers to politically motivated hacking to conduct sabotage and espionage. It is a form of information warfare sometimes seen as analogous to conventional warfare although this analogy is for both its accuracy and its political motivation.	Cyberwarfare
8877168	WikiLeaks is an international, online, self-described not-for-profit organisation publishing submissions of private, secret, and classified media from anonymous news sources, news leaks, and whistleblowers. Its website, launched in 2006 under the Sunshine Press organisation, claimed a database of more than 1.2 million documents within a year of its launch. Julian Assange, an Australian Internet activist, is generally described as its founder, editor-in-chief, and director. Kristinn Hrafnsson is the only other publicly known acknowledged associate of WikiLeaks as of 2011. Hrafnsson is also a member of the company Sunshine Press Productions along with Assange, Ingi Ragnar Ingason and Gavin MacFadyen.	WikiLeaks
36322844	PayPal is a global e-commerce business allowing payments and money transfers to be made through the Internet. Online money transfers serve as electronic alternatives to paying with traditional paper methods, such as checks and money orders.	PayPal

IV. Methodology

1. Problem Statement

Let there be a graph $G(V, E)$ where V is the number of nodes and $E \subseteq (V \times V)$ being the edges between any two nodes. This will be a bipartite graph connecting users and topics with weight as user's interests over these topics. Let $G_\tau = (V_\tau, E_\tau)$ is the snapshot recorded as time τ where V_τ is the set of nodes and $E_\tau \subseteq V_\tau \times V_\tau$ is the set of interactions. Let $\triangle G_\tau = (\triangle V_\tau, \triangle E_\tau)$ represents the change in network representing changes in vertices and edges at time τ. Individual nodes are represented as u and v and t and τ represents the timestamps.

[4]A dynamic network G is a sequence of network snapshots within a time interval and evolving over time $G = (G_1, \ldots, G_t)$. Given graphs with different snapshots our aim is to predict future graph G_{t+1}.

2. Temporal latent space

Let Z_τ be the graph representing the $V\tau$ in temporal latent space. Z_τ is a k dimensional space where k is much smaller than number of nodes V. [4] For each individual u at time τ, we use a row vector $Z_\tau(u)$ to denote its temporal latent space representation and a scalar Z (u, c) to denote its position in cth dimension.

Based on Zhu et al. [4] temporal latent space, following assumptions are made.

i. **Temporal smoothness:** The nodes in latent space change their positions gradually over the time.

ii. **Network embedding:** The distance between two nodes in interaction network and distance between two nodes in latent space are correlated i.e. if two nodes are closer in interaction network, they are also closer in temporal network space.

iii. **Latent Homophily:** Nodes which are closer to each other interact more frequently than the nodes which are far away.

Predicting G_{t+1} is implemented as follows.

- First, creating the temporal latent space $(Z_1, \ldots, Z_t)$ from $(G_1, \ldots, G_t)$ and approximating Z_{t+1} from $(Z_1, \ldots, Z_t)$
- Second, inferring G_{t+1} from Z_{t+1}

3. Creating temporal latent space

[4] Given a dynamic social network $G = (G_1, \ldots, G_t)$ we aim to find a k-dimension latent space representation at each timestamp Z_τ that minimizes the quadratic loss with temporal regularization

[4] Equation 1

$$arg \min_{z1,\ldots zt} \sum_{\tau=1}^{t} \|G_\tau - Z_\tau Z_\tau^T\|_F^2 + \lambda \sum_{\tau=1}^{t} \sum_{u} (1 - Z_\tau(u) Z_{\tau-1}(u)^T) \qquad Equation\ 1$$

Subject to $\forall\, u, \tau, Z_\tau \geq 0, Z_\tau(u) Z_\tau(u)^T = 1$

[4] λ is regularization parameter and the term $(1 - Z_\tau(u) Z_{\tau-1}(u)^T)$ penalizes node for suddenly changing its latent position. The value of term $(1 - Z_\tau(u)\ Z_{\tau-1}(u)^T)$ is higher when value of Z_τ and $Z_{\tau-1}$ are closer to each other.

The resulting latent space Z_τ will always be greater than zero as model imposes non-negative constraint. This helps to represent each node in latent space as an additive information about a node. Each dimension in latent space Z_τ is representing some unknown characteristics of the node like city, hobby, age etc.

4. Link Prediction

[4] Given that we have inferred $(Z_1, \ldots, Z_t)$ by optimizing Equation 1, our goal is to predict the adjacency matrix G_{t+1} at the next timestamp t + 1. The expectation value of G_{t+1} can be expressed as below.

[4] Equation 2

$$Y_{t+1} = E[G_{t+1} | Z_1, \ldots Z_t] \qquad Equation\ 2$$

As described by Zhu et al., By assuming that the temporal dynamics of latent positions is Markovian and satisfies $E[Z_{t+1}|Z_1, \dots, Z_t] = Z_t$ and $cov[Z_{t+1}^T \,|Z_1, \dots, Z_t] = D_t$ (a diagonal matrix), as well as $Z_{t+1}Z_{t+1}^T$ as an unbiased estimate of G_{t+1} we form

[4] **Equation 3**

$$Y_{t+1} = E[G_{t+1}|Z_1, \dots Z_t]$$

$$Y_{t+1} = E\ [E\ [G_{t+1}|Z_1, \dots Z_t, Z_{t+1}] \mid Z_t, \dots , Z_t]$$

$$Y_{t+1} = E[Z_{t+1}\, Z_{t+1}^T \,\big|\, Z_1, \dots Z_t]$$

$$Y_{t+1} = cov[\, Z_{t+1}^T \,\big|\, Z_1, \dots Z_t]$$

$$+ E[Z_{t+1} \mid Z_1, \dots Z_t]\, E[Z_{t+1} \mid Z_1, \dots Z_t]^{\,T}$$

$$Y_{t+1} = D_t + Z_t Z_t^T \qquad Equation\ 3$$

The resulting matrix in our scenario will be symmetric as we are using undirected graph. We can ignore the diagonal entries of the adjacency matrix as we are not interested to find values of self loops. Hence leaving $D_t$diagonal matrix, we use only $Z_t\ Z\ {}_t^T$ to predict G_{t+1}.

To efficiently and effectively infer latent space Zhu et al. [4] proposes Block Coordinate Gradient Descent (BCGD) algorithm as means to solve Equation 1. We will touch upon standard BCGD algorithm which is still computationally expensive and later cover local BCGD algorithm which one of the variations in fast BCGD algorithms.

5. Standard BCGD Algorithm

Below section describes details of how BCGD algorithm works. The Equation 1 can be decomposed as linked and non-linked part.

[4] **Equation 4**

$$arg \min_{z1,\dots,zt} \sum_{\tau=1}^{t} \sum_{u,v \epsilon E_\tau} (G_\tau(u,v) - Z_\tau(u) Z_\tau(v)^T\,)^2$$

$$+ \sum_{\tau=1}^{t} \sum_{u,v \notin E_\tau} (Z_\tau(u)\, Z_\tau(v)^T)^2 \ +\ \lambda \sum_{\tau=1}^{t} \sum_{u} (1 - Z_\tau(u)\, Z_{\tau-1}(u)^T) \qquad Equation\ 4$$

Subject to $\forall\, u, \tau, Z_\tau \geq 0, Z_\tau(u) Z_\tau(u)^T = 1$

[4] Unfortunately, the decomposed objective function is still fourth order polynomial non-convex. Hence block coordinated gradient descent algorithm is used to solve the Equation 4. [4] We update $Z_\tau(u)$ for each

node u at time τ by fixing both latent positions $Z_\tau(v)$ of all other nodes v at time τ as well as all the temporal latent positions other than at time τ

Now the optimization problem will be, $arg \min_{Z_\tau(u)>0} J(Z_\tau(u))$

[4] **Equation 5**

$$J\big(Z_\tau(u)\big) = \sum_{v \epsilon N(u)} (G_\tau(u,v) - Z_\tau(u) Z_\tau(v)^T)^2 + \sum_{v \notin N(u)} (Z_\tau(u)\, Z_\tau(v)^T)^2 + \lambda\,(1 - Z_{\tau+1}(u)\, Z_\tau(u)^T) + \lambda\,(1 - Z_\tau(u)\, Z_{\tau-1}(u)^T) \quad Equation\ 5$$

[4] In the following, we use the projected gradient descent algorithm to find the approximation solution with a non-negativity constraint. With the gradient descent optimization algorithm, for each node u at timepoint, we could iteratively update $Z_\tau(u)$ in each iteration r + 1 with the following rule:

[4] **Equation 6**

$$Z_\tau^{(r+1)}(u) = Z_\tau^{(r)}(u) - \eta\, \nabla\, Z_\tau(u)\ \ J\left(Z_\tau^{(r)}(u)\right) \quad Equation\ 6$$

η is the step size.

Theoretically even though we can calculate adjacency matrix using above formula, the exact algorithm is computational expensive in terms of both time and space for large scale data. The above equation requires all the graphs from$(G_1, \dots, G_t)$ to update $(Z_1, \dots, Z_t)$ jointly and cyclically. from Hence Zhu et al. [4] proposes multiple variations block coordinate gradient descent algorithm such as local BCGD and incremental BCGD. In this paper, we will be discussing only on Local BCGD.

6. Local BCGD algorithm

In local BCGD algorithm, the local objective function to compute $Z(t)$ is as follows

[4] **Equation 7**

$$arg \min_{Z_\tau} \sum_{u,v \epsilon E_\tau} (G_\tau(u,v) - Z_\tau(u) Z_\tau(v)^T)^2 + \sum_{u,v \epsilon E_\tau} (Z_\tau(u) Z_\tau(v)^T)^2 + \sum_{u \epsilon V_\tau} (1 - Z_\tau(u)\, Z_{\tau-1}(u)^T) \quad Equation\ 7$$

[4] Using the same BCGD approach, we iteratively update the latent position $Z_\tau(u)$ of each node u by fixing the latent positions of all the other nodes. This leads to the following update rules for $Z_\tau(u)$ in the (r + 1) th iteration:

[4] Equation 8

$$Z_{\tau}^{(r+1)}(u) = \max\left((1+2\alpha) Z_{\tau}^{(r)}(u) + \alpha\lambda Z_{\tau-1}(u) + 2\alpha \sum_{v \in N(u)} G_{\tau}(u,v) Z_{\tau}^{(r)}(v) - 2\alpha\, Z_{\tau}^{(r)}(u) Z_{\tau}^{(r)^T} Z_{\tau}^{(r)}, 0 \right) \qquad Equation\ 8$$

Where α is $\frac{(a_{(r+1)} + a_r - 1)}{a_{r+1}\, L}$ is the Lipschitz constant and $a_r = \begin{cases} 1\ if\ r = 0 \\ \frac{1+\sqrt{4a_{r-1}^2+1}}{2}\ if\ r > 0 \end{cases}$

Algorithm: The local BCGD algorithm for sequentially inferring temporal latent space
Input: Graphs $\{G_1, \ldots, G_t\}$ and latent space dimension k
Output: Y_{t+1} and latent space $\{Z_1, \ldots Z_t\}$
1: Nonnegative initial guess for Z_1
2: for each τ from 1 to t
3: Initial Z_τ based on $Z_{\tau-1}$
4: repeat
5: for each u in graph G_τ
6: update Z_τ (u) by Eq. 7 and normalize it
7: until Z_τ converges.
8: **return** $\boldsymbol{Y_{t+1} = \Phi([Z_1, \ldots, Z_t])\ and\ \{Z_1, \ldots, Z_t\}}$

Figure 1:[4] Local BCGD algorithm

Algorithm for local BCGD is as shown in Figure 1. In Global BCGD algorithm for each iteration we jointly and cyclically updated all the temporal latent space $(Z_1, \ldots, Z_t)$ is updated for each iteration from all network graphs $(G_1, \ldots, G_t)$. This is expensive in terms of computation and storage. In Local BCGD algorithm, Z_t is computed in one iteration showed in step 4-7. i.e. Z_t is updated using G_t and Z_{t+1}. Figure 2 depicts the iterations that happen while using the local BCGD algorithm.

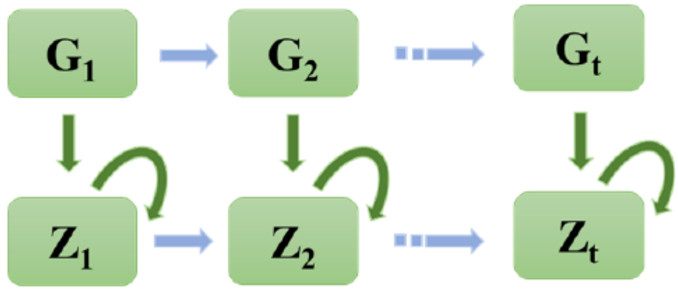


Figure 2:Local BCGD algorithm updates

7. Experiments

In this section, we describe our experiments in terms of the dataset, setup and experimentation methodologies. The experiments are conducted on an Intel(R) Core™ i5-7440HQ CPU @ 2.80GHz 16GB RAM.

A. Dataset

As mentioned in exploratory data analysis, we are using 2,948,741 tweets 130,168 unique users. all the tweets of the users that are published in each day are extracted into single document and applied LDA over all documents to find the prevalent topics. The number of topics is set to 100. We have the data on user's distribution over these topics available from previous research dataset.

To create graph G, we use all the users and topics are set of nodes and user's interests over the topics as edges. We obtain a bipartite graph for depicting the degree of interests for all users. With 100 topics and 2458 users we have total 2558 unique nodes in our interaction network. We extract $G = (G_1, G_2, \dots G_{60}, G_{61})$ from the data set of which the graphs from $(G_1, \dots, G_{60})$ will act as training set and G_{61} acts as ground truth.

B. Evaluation metrics

We will first calculate predictive accuracy measures like RMSE (Root Mean Squared Error), MSE (Mean Squared Error) and MAE (Mean Absolute Error) to see how the algorithm performs for different parameters. Scores for RMSE, MSE and MAE is calculated based on predicted weights of all 100 topics. We will also evaluate with classification accuracy metrics such as MAP (Mean Average Precision), nDCG (Normalized Discounted Cumulative Gain). We will evaluate top 5 topic predictions as MAP@5, nDCG@5 i.e. we will compare top 5 predicted topics against top 5 ground truth topics. To compare our results with dataset used in Zhu et al. [4] paper we will compute the AUC (Area Under the Curve) using ROC (Receiver Operating Characteristic).

C. Parameter variation

This set of experiments is aimed at finding the optimal parameters so that it can be run for large scale of data.

The effect of dimensionality

Higher the dimension in temporal latent space Z leads to a better prediction results but also utilizes higher online computational cost. In our experiments we will run the algorithm with several values for k to find the optimum value for k.

The effect of regularization parameter

Zhu et al. [4] mentions effects of extreme low or extreme high value for regularization parameter λ. Extreme low value for λ implies absence of smoothness and extreme high value for λ implies strong temporal smoothness. We will evaluate the results for our dataset with different values of λ.

The effect of number of snapshots

We would also experiment to see if number of snapshots has any effect on prediction accuracy. We would like to evaluate how the results are varied when only a few historic snapshots are used versus when 60 historic snapshots are used.

V. Results and Discussions

A. The effect of dimension

As we are transforming the graphs from interaction network into temporal latent space with k dimensions, we would run the experiment by varying the dimension(k) and keeping all other parameters constant. Table 1 shows the results for different dimensions. The evaluation scores are the average values obtained for multiple repeated experiments. Scores for RMSE, MSE and MAE is calculated based on predicted weights of all 100 topics whereas scores for MAP and nDCG is calculated based on top 5 predicted topics.

Other parameters: λ =0.01, number of iterations=300, snapshots=60, algorithm = BCGD local

Table 1:Effect of dimension(k)

Dimensions	RMSE	MSE	MAE	MAP_at_5	nDCG_at_5
20	0.018572162	0.000344925	0.001740318	0.023129485	0.050354497
30	0.017221816	0.000296591	0.001646962	0.031002935	0.061784485
40	**0.016561177**	**0.000274273**	**0.001566978**	**0.031381279**	**0.065710032**
50	0.016975636	0.000288172	0.001625035	0.029874429	0.062461921
60	0.017660142	0.000311881	0.001700776	0.023783431	0.050276011
70	0.016832092	0.000283319	0.001638459	0.028176778	0.062825266

80	0.016977893	0.000288249	0.001667697	0.029294912	0.061752792
90	0.016977893	0.000288249	0.001698545	0.024372146	0.052031578
100	0.017033246	0.000290131	0.001616579	0.030228311	0.063113361

As we can see in the Figure 3, the error rate is minimum when dimension is 40. Even though accuracy increased by increasing the dimension from 20 to 40, higher dimension not necessarily increased the prediction accuracy. Figure 4 also showed MAP and nDCG values resonated with RMSE score. Highest values for MAP and nDCG were obtained for dimension 40. One reason for stagnant prediction accuracy might be that higher the dimension, the model starts accumulating more noise from the dataset. We will use the dimension as 40 for all our experiment for efficiency and effectiveness.

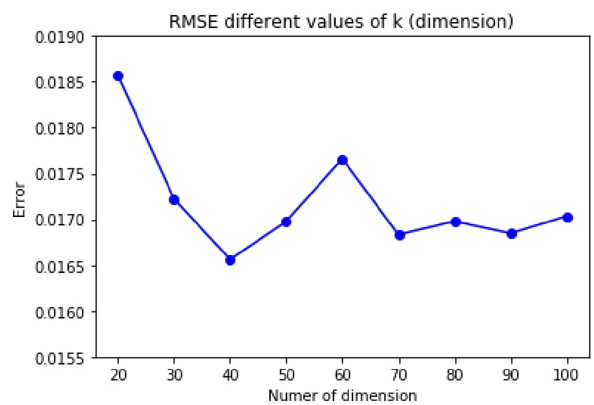


Figure 3:Effect of dimension(k) on RMSE score (lower is better)

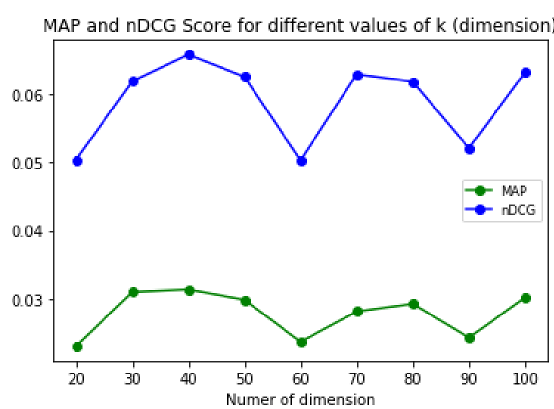


Figure 4:Effect of dimension(k) on MAP and nDCG score (higher is better)

B. The effect of temporal smoothness

By keeping all other variables constant, we will study the effect of regularization parameter λ. We varied λ in terms of logarithmic scale and reported the prediction errors. Extreme low value for λ implies absence of smoothness and extreme high value for λ implies strong temporal smoothness. We will evaluate the results for our dataset with different values of λ. Table 2 lists the results for different λ.

Other parameters: dimension(k) =40, number of iterations=300, snapshots=60, algorithm = BCGD local

Table 2:Effect of regularization parameter

lambda(λ)	RMSE	MSE	MAE	MAP_at_5	nDCG_at_5
10	0.017023721	0.000289807	0.001629813	0.029964123	0.061705991
1	0.017333653	0.000300456	0.001689898	0.031099152	0.068857336
0.1	0.018277531	0.000334068	0.001715139	0.025370189	0.052840276
0.01	**0.018572162**	**0.000344925**	**0.001740318**	**0.031002935**	**0.061784485**
0.001	0.018078044	0.000326816	0.001704653	0.025986628	0.057921322
0.0001	0.016955966	0.000287505	0.001562893	0.02958578	0.063535233
0.00001	0.016535055	0.000273408	0.001500801	0.030342466	0.063753028

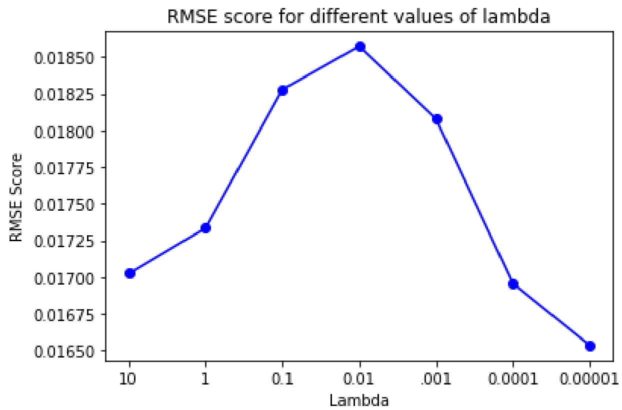


Figure 5:RMSE Score for different λ on RMSE (lower is better)

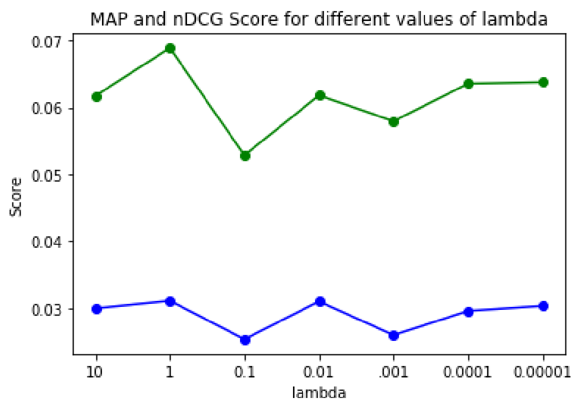


Figure 6:RMSE Score for different λ on MAP and nDCG Score (higher is better)

Figure 5 is plotted with RMSE score for each value of λ. We can see the highest error rate is when λ =0.01. Interestingly in Figure 6 we have got the best results in terms of MAP and nDCG score when λ =0.01. This might be because RMSE is on all 100 topic predictions and MAP and nDCG value is for top 5 topic predictions. Since we are interested in limited top predictions we will keep the lambda value as 0.01

C. Effect of number of historic days

We conducted our experiments with different number of snapshots. We are evaluating how many numbers of snapshot i.e. historic days data is best to predict the future interests. Table 3 shows the next day's user's interest based on previous number of historic snapshots.

Other parameters: dimension(k) =40, λ=0.01, number of iterations=300, algorithm = BCGD local

Table 3:Effect of number of snapshots

# of snapshots	experiments	MAP_at_5	nDCG_at_5
1 day	4	0.049333333	0.090844797
5 days	**4**	**0.044222222**	**0.100984763**
10 days	4	0.043555556	0.085326051
20 days	4	0.042927667	0.08152317
30 days	4	0.0376897	0.06721233
60 days	4	0.027915967	0.05923327

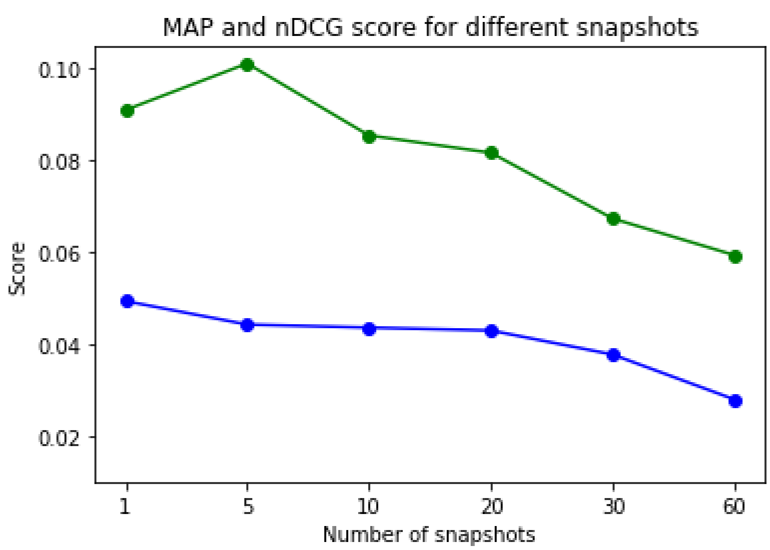


Figure 7:MAP and nDCG score for different historic snapshots (higher is better)

We can see in Figure 7, the prediction results for 5 historic snapshots were comparable with the prediction results for higher number of snapshots. We can infer that larger snapshot may not necessarily mean better predictive power. Perhaps this network may be rapidly evolving, and user's interests are highly influenced by current trends in the world.

D. Comparison with other networks

Zhu et al. [4] experimented with real temporal data collected from Koblenz Large Network such as Facebook, DBLP (computer science bibliography website), YouTube. Each network had varying set of nodes and edges representing the interaction graph. Zhu et al. [4] evaluated the results from three aspects, parameters, efficiency and accuracy of link predictions in terms of AUC(ROC) score. We will compare the prediction results of these network with results from our dataset. To perform the comparison, we experimented with all three types of BCGD algorithm and with the same parameters used in Zhu et al. experiment.

Parameters: λ=0.01, number of iterations=300, dimensions k=20, snapshots=60

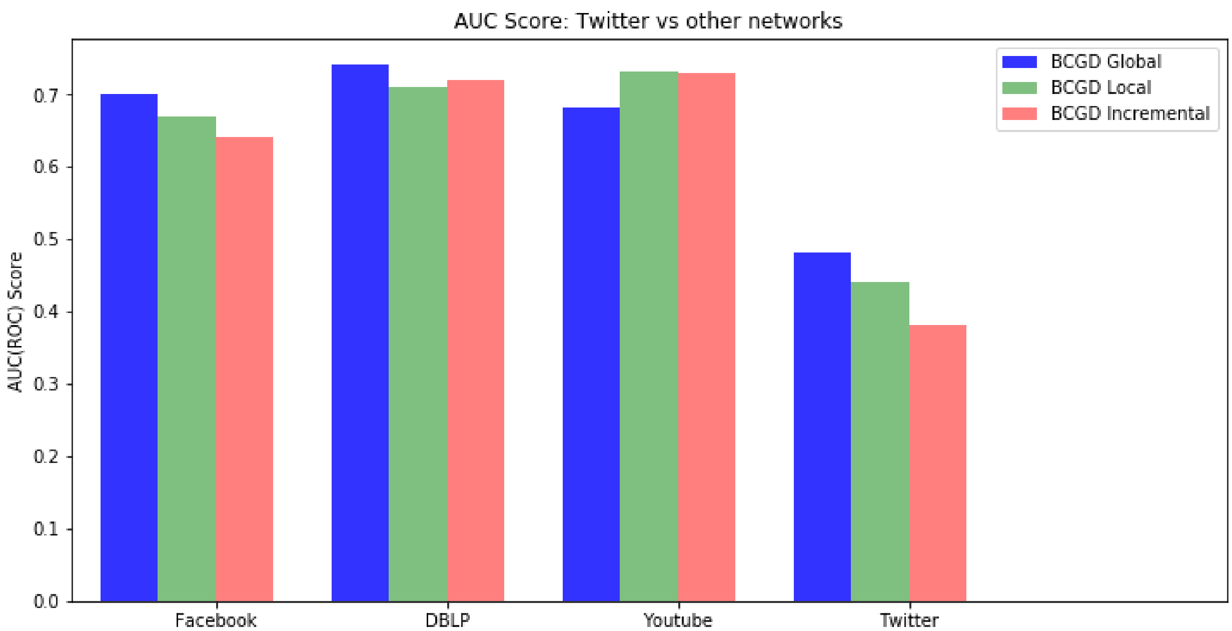


Figure 8:AUC Score: Twitter vs other networks (Higher is better)

Figure 8 shows that our results score least among the results of all different network. Since the AUC value is less than 0.5, the resulting prediction has no separation between relevant and non-relevant topics for users, it is as good as a random guess. This implies that even though temporal latent space modeling is a powerful non-negative matrix factorization technique, it was not successful in predicting the user's interests on Twitter dataset.

As we evaluate the root cause for this, we would like to highlight some of the factors that might have contributed to this result. One of the assumptions in dynamic temporal latent space is that even though the interaction network changes continuously they only evolve smoothly. This is a strong assumption as the users on Twitter might be highly influenced by external events and hence, temporal smoothness assumption may not hold true.

Another perspective is that we have only the information about the relationship between users and topics. We don't have information like user followership information or relationship between topics in our current dataset. More information about the user would have been supportive for better prediction accuracy.

VI. Conclusions and Future Work

In this paper our aim was to predict user's future interest for Twitter dataset using temporal latent space which has been successfully applied for different networks. We evaluated the results with various predictive accuracy metrics and classification accuracy metrics. The results were not successful in predicting the user's future interests for our dataset. We evaluated different factors that might have contributed to this result. We conclude that user interest in Twitter might be changing rapidly and abruptly, hence the assumption of temporal smoothness may not hold true for this dataset.

In our future work we would like to add more information about users while modelling temporal latent space such as a directed graph to represent user's followership information, a directed graph to represent the relationship between topics to improve results accuracy.

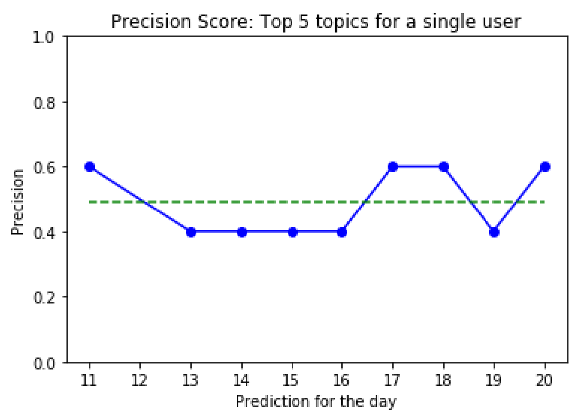
Appendix

A. Detailed code is available at the following location in GitHub:

https://github.com/chaitrahosmani/User-interest-detection-in-social-media

B. Prediction for single user

To start with we conducted our experiment with prediction of topics for a single user. Prediction for next day is generated from previous 10 historic days. finding top topics for a user for day11 using graphs from day 1 to 10, for day 12 using graphs from day 2 to 11 and so on. Below graph showing the average precision as 0.5 i.e. out of every 20 topics predicted for user, 10 are relevant topics.



References

1) Semantics-enabled User Interest Detection from Twitter, Fattane Zarrinkalam, Hossein Fani, Ebrahim Bagheri, Mohsen Kahani and Weichang Du
2) Inferring Implicit Topical Interests on Twitter, Fattane Zarrinkalam, Hossein Fani, Ebrahim Bagheri and Mohsen Kahani
3) Predicting Users' Future Interests on Twitter, Fattane Zarrinkalam, Hossein Fani, Ebrahim Bagheri, and Mohsen Kahani
4) Scalable Temporal Latent Space Inference for Link Prediction in Dynamic Social Networks, Linhong Zhu, Dong Guo, Junming Yin, Greg Ver Steeg, Aram Galstyan
5) P. Ferragina and U. Scaiella, "Fast and accurate annotation of short texts with Wikipedia pages," IEEE Software, vol. 29, no. 1, pp. 70–75, 2012. (tweet annotating)
6) Entity Recommendations Using Hierarchical Knowledge Bases, Siva Kumar Cheekula, Pavan Kapanipathi, Derek Doran, Prateek Jain and Amit Sheth
7) Inferring User Interests in the Twitter Social Network, Parantapa Bhattacharya, Muhammad Bilal Zafar, Niloy Ganguly, Saptarshi Ghosh, Krishna P. Gummadi
8) User interest prediction over future unobserved topics on social networks, Fattane Zarrinkalam, Mohsen Kahani, Ebrahim Bagheri
9) Hossein Fani, Ebrahim Bagheri, and Weichang Du. 2017. Temporally Like-minded User Community Identification through Neural Embeddings. In CIKM'17. 577–586.
10) Causal Dependencies for Future Interest Prediction on Twitter Negar Arabzadeh, Hossein Fani, Fattane Zarrinkalam, Ahmed Navivala, Ebrahim Bagheri
11) Topic-Association Mining for User Interest Detection, Anil Kumar Trikha, Fattane Zarrinkalam(B), and Ebrahim Bagheri
12) N. Rajani, K. McArdle, and J. Baldridge, "Extracting topics based on authors, recipients and content in microblogs," in 37th International ACM SIGIR Conference on Research and development in Information Retrieval, 2014, pp. 1171–1174.
13) Chao-Yuan Wu, Amr Ahmed, Alex Beutel, Alexander J. Smola, and How Jing. 2017. Recurrent Recommender Networks. In WSDM'17. 495–503.
14) Y. Koren. Collaborative filtering with temporal dynamics. In Knowledge discovery and data mining KDD, pages 447–456, 2009.
15) [39] S. A. Vavasis, "On the complexity of nonnegative matrix factorization," J. on Optimization, vol. 20, no. 3, pp. 1364–1377, 2009.
16) Setting Goals and Choosing Metrics for Recommender System Evaluations, Gunnar Schröder, Maik Thiele, Wolfgang Lehner,

17) J. Chen, R. Narin, L. Nelson, M. Bernstein, and E. Chi, “Short and tweet: experiments on recommending content from information streams,” in 28th international conference on Human factors in Computing Systems (CHI ’10), 2010, pp. 1185–1194.
18) C. R. Y. Shin and J. Park, “Automatic extraction of persistent topics from social text streams,” World Wide Web, vol. 17, no. 6, pp. 1395–1420, 2013.
19) P. Kapanipathi, F. Orlandi, A. Sheth, and A. Passant, “Personalized filtering of the twitter stream,” in SPIM Workshop at ISWC 2011, 2011 pp. 6–13.
20) J. Weng, E. Lim, J. J, and Q. He, “Twitterrank: finding topic-sensitive influential twitterers,” in 3rd ACM international conference on Web search and data mining (WSDM ’10), 2010, pp. 61–270.