International Conference on "Emerging Trends in Computer Engineering, Science and Information Technology"- 2015
Special Issue of International Journal of Electronics, Communication & Soft Computing Science and Engineering, ISSN: 2277-9477

Neighborhood Integrated Matrix Factorization for Temporal Prediction

Trishna Jadhav **Prof.S.S.Banait**

Abstract —**Link Mining and Temporal Link Prediction is becoming prominent trend in recent years. Link mining deals with different data sources that generates link data and this link data provide scope for collaborative filtering tasks, which is a prime requirement in recommending systems and a significant role is played in predictive analytics. The proposed introduction of Neighborhood Integrated Matrix Factorization method will improve the accuracy of missing value predictions as preheuristic task of Neighborhood Similarity Computation which produces object profiles.**

Key Words — **Collaborative Filtering, Neighbourhood Integrated Matrix Factorization, Predictive Analytics., Temporal link prediction**

I. INTRODUCTION

Many objects and entities in the world are inter dependent, and inter linked to many other objects through a diverse sets of relationships: people have same friends, family and collogues scientific papers have same authors, venues, and references to other papers; web pages links to other web pages and have hierarchical structures; proteins have locations and functions, and interact with other proteins. The linked data is typically very noisy and incomplete [9]. The data in different analysis applications for example social networks, communication networks, web analysis, and collaborative filtering consists of relationships, this relationships can be considered as links, between objects [3]. For instance, two people may be linked to each other if they sends emails to each other or exchange phone calls. Link mining refers to data mining techniques that explicitly consider the links when building predictive or descriptive models of the linked data. Commonly addressed link mining tasks include object ranking, group detection, collective classification, link prediction and sub graph discovery imaging produces an image in which the linked data is typically very noisy and incomplete [9].

The data in different analysis applications for example social networks, communication networks, web analysis, and collaborative filtering consists of relationships, this relationships can be considered as links, between objects [3]. For instance, two people may be linked to each other if they sends emails to each other or exchange phone calls. Link mining refers to data mining techniques that explicitly consider the links when building predictive or descriptive models of the linked data. Commonly addressed link mining tasks include object ranking, group detection, collective classification, link prediction and sub graph discovery actually an exciting, and rapidly increasing area. There is yet not comprehensive framework that can support a combination of link mining tasks as needed for many real applications. Link prediction is a sub-field of social network analysis [3]. It is concerned with the problem of predicting the (future) existence of links amongst nodes in a social network. The link prediction problems is interesting in that it investigates the relationship between objects, while traditional data mining tasks focuses on objects themselves [5].Dynamic interactions over time add another dimension to challenge of mining and predicting link structure. This study focuses on basic link mining tasks and the problem of temporal link prediction. In this problems given link data for T time steps, can one predict the relations among data objects at time T+1,T+2,…..,T+k. where (k> 0), is a point of interest. Objective of proposed work is as follows: To design a learning machine system using state space model such as Kalman Filter, by applying neighborhood matrix factorization to capture temporal dynamics and latent factors in link data. This project aims to improve the prediction accuracy of such collaborative filtering system by minimizing error and reducing noise in link data.

Relevant Objectives:

1. To capture temporal dynamics in link data.
2. To find the latent factors/ features.
3. To model predictive function.
4. To predict next step links.

The remainder of paper is organized as follows: Section 2 discuss the related work done and its shortcoming. Section 3 describes motivation. An overview of the implemented scheme is given in section 4. Section 5 discusses the results of the system. Finally, we conclude the paper.

II. LITERATURE SURVEY

Conventional approaches deals with object classification, ranking, entity resolution tasks like Object related tasks Graph related tasks and Link related tasks. This tasks can be approached based on similarity score, topological pattern mining, and content filtering. Recent work contributed in this regard is based on probabilistic approaches as well as graph based feature learning.
Aim of this survey is to set the basis for constructing a hypothesis. So that proposed hypothesis is to model a predictive function. One can set the base line to understand the predictive function as follows: Suppose state of a system is captured using X which is to be mapped to produce output say Y, then hypothesis can be set as h : X -> Y Such that for any input x which belongs to range of X given to function say h(x) shall provide the estimated output y which belongs to range of Y. To approach above function, one need to consider the error and need to track those changes in some variable, so as to more the

iterations of function shall result in identifying minimum required number of such variables to control the correcting and updating steps of Kalman Filter. Here one optimization step such as maximum likelihood or Expectation maximization is to be introduced to learn those variables. The above requirement of identifying variables needed for machine learning is a basic functional requirement. Therefore suitable technique is using matrix factorization [5] to decompose a variable in factors and learn the strengths of these factors over evolving time. Expectation maximization is to be incorporated, and is the second functional requirement, to deal with capturing temporal dynamics [5]. This configuration shall be suitably traced using state space model like Kalman Filter[3]. Resulting output is modelled predictive function that will be an optimal state of system that is to be measured as root mean square error over number of iterations.

III. MOTIVATION

Link mining is a fairly new research area that lies at the intersection of link analysis, web mining, predictive analytics, relational learning and inductive logic programming, and graph mining. Link mining tasks aims at learning and inferring the information about domain available in structural and temporal patterns. Possible tasks include learning models of the domain as well as inferring information about the entities, relationships, their attributes, and the connectivity structure of the domain. One of the key aspect in such domain typically consists of classifying inferred entities over time. The data in different analysis applications such as social networks applications, communication networks, web analysis consists of relationships, which are considered as links, between objects. Exploiting their structural information is been a major issue of many mining applications. Links among objects are formed to represent certain event of occurrence, and mining of such occurrences reveals useful information for predictive analytics.

Dynamic interactions over time add another dimension to challenge of mining and predicting link existence. This study motivates and finds scope in capturing temporal information

Table 1: RMSE PLOT FOR Fig 2.

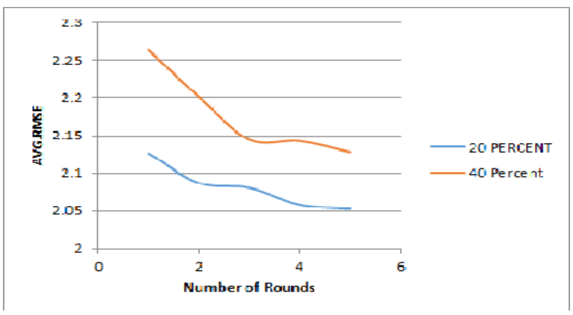

Fig.2 RMSE Measurement for 20% and 40% density

which can give significant contribution in solving temporal link prediction problems.

IV. IMPLEMENTATION DETAILS

A. PROPOSED SYSTEM

Function of each block is discussed below:

1. Load Dataset This module loads dataset obeying state space model so that to check effectiveness of proposed collaborative Kalman filter can be analyzed. Here generative dataset stores number of items, users, item factors, timeslots, observation ratio, and generates dataset obeying state model such that white noise is assumed to be added to dataset.
2. Kalman Filter This module is consist of functionality for following recursive operations of Predict- Correct phase further it undergoes through RTS smoother treatment for finding optimal state estimation: Predict phase 1.Project the state ahead 2.Project error covariance ahead.

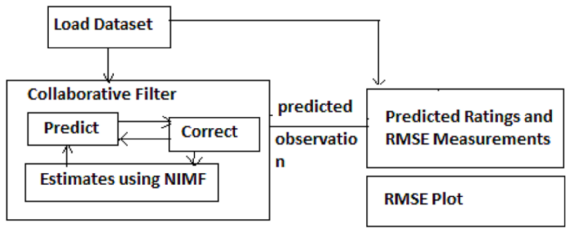

Fig.1 Block diagram of system

3. Update (correct) phase
   1.To compute Kalman Gain
   2.Update estimate with measurement.

3.Update covariance
1.To compute optimal estimates of states using NIMF
2.To compute predicted observations.
3.To compute missing value prediction

3.RMSE Measurement This module is conclusion of experiment and gives root mean square error measurements recorded at each iteration, objective of this measurement is to check for effective learning of machine and provide plot of the RMSE behavior.

Oratio=0.2	Number of Rounds				
	1	2	3	4	5
T=	2.1264	2.0866	2.0802	2.0579	2.0527
Oratio=0.4	1	2	3	4	5
T=	2.2633	2.2012	2.1444	2.1426	2.1281

V.

VI.

V. EXPERIMENTAL SETUP

Experiment is conducted to compare the accuracy of the proposed work with other collaborative filtering methods. Various parameters are identified to compare the accuracy of such as with respect to density of input rating matrix and dimensionality. The proposed work as compared with collaborative filtering method of state space approach gives the improvement in accuracy as shown in Fig 2.

A. Dataset

To evaluate the performance of system, following dataset is used Dataset 142 *4532 *64 time-aware Web service QoS dataset Real-world QoS evaluation results from 142 users on 4,532 Web services on 64 different time slots. This dataset includes[11][12] rtRate (480 MB): response-time values of 4,532 Web services when invoked by 142 service users in 64 time intervals.
The data format is as following:| Time Interval ID | Web Service ID | Service User
ID | Response-Time (s)
tpRate (571 MB): throughput values of 4,532 Web services when invoked by 142 service users in 64 time intervals[20].
The data format is as following:
| Time Interval ID | Web Service ID | Service User ID | Throughput (kbps)

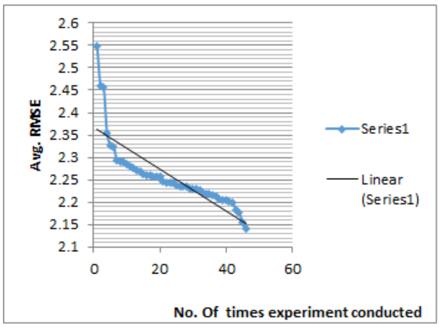

Fig.3.System Behaviour

B. Performance Measure

Performance of system is measured over Quality of service dataset. Performance of system is measured over Quality of service dataset. Following are discussions about rigorous testing conducted on available system, and data tables showing RMSE values for 10 iterations, observed over permutations of number of Items, Users, Factors, Timeslots and percentage of generated training dataset considered for experiment. Let:
M is number of Items (M > 0)
N is number of Users (N > 0)
K is number of factors (K > 0)
T is number of Timeslots considered for experiment out of T, T-1 is number of timeslots for which training is given
to proposed learning system and predicted data values are compared with true data values at Tth time slice. (T > 0) The system has gained a lot in terms of scalability. Time required to perform 5 iterations for data size of 142x4500x65 is 150 sec to 200 sec as compared to data size of 500x520x40
which is taking 25 to 30 minutes, results of [10]. The experiments were conducted with respect to different density of 20 percent and 40 percent respectively. It is observed that when density is less RMSE graph is smoother than the one with higher density matrix as shown in fig 2. Fig 3 shows the system behavior which is showing continuous improvement in system. Y axis represents no of times experiments conducted and X axis represents average RMSE values. continuous system improvement can be observed

VI. CONCLUSION

The presented work is motivated by temporal link prediction. This collaborative Kalman filtering approach and Neighborhood Integrated Matrix Factorization approach focus on capturing temporal dynamics so as to provide better accuracy in prediction tasks and missing value prediction task. A state space model called Kalman Filter captures temporal dynamics of time series data. Further optimal machine learning is derived using RTS smoother to predict optimal estimates of states. The presented work is validated on Quality of service dataset and shown significant improvements in reducing RMSE also tested for scalability issues and has shown remarkable performance in time requirement for execution.

REFERENCES

[1] Zibin Zheng,Hao Ma,Michael R.Lyu,"Collaborative Web Service QoS Prediction via Neighbourhood Integrated Matrix Factorization"."IEEE Trans. On services computing" Vol 6. No 3.July-September 2013
[2] Dunlavy D. M., Kolda, T. G., and Acar E. "Temporal link prediction using matrix and tensor factorizations." ACM Trans. Knowl. Discov. Data 5, 2, Article 10 (February 2011), 27 pages.
[3] "State Space Model for Link Mining".Kushal P.Birla,Prof.S.M.Kamlapur,IJETTCS.ISSN 2278-6856.
[4] Neural Networks and learning machines(3rd ed.).(2011) by simon Haykin,PHI,ISBN 978-81-203-4000-8
[5] Sun, J.Z.; Varshney, K.R.; Subbian, K, "Dynamic matrix factorization: A state space approach," Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference, vol., no., pp.1897-1900, 25-30 March 2012.
[6] Leskovec, J., J. Kleinberg and C. Faloutsos. 2007. Graph evolution: Densification and shrinking diameters. ACM Trans. on Knowledge Discovery from Data.
[7] J. Kleinberg. Authoritative sources in a hyperlinked environment. Journal of the ACM, 46(5):604632, 1999
[8] Q. Lu and L. Getoor. Link-based classi_cation. In In-ternational Conference on Machine Learning, 2003.

[9] D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. In International Conference on Information and Knowledge Management (CIKM), pg 556-559, 2003.

[10] "Link Mining and Temporal Link Prediction", Mr.Kushal Birla,Prof S.M,Kamalapur.

[11] Yilei Zhang, Zibin Zheng, and Michael R. Lyu, "WSPred: A Time-Aware Personalized QoS Prediction Framework for Web Services", in Proceedings of the 22th IEEE Symposium on Software Reliability Engineering (ISSRE 2011).

[12] (http://www.wsdream.net)

AUTHOR'S PROFILE

Ms.Trishna Jagannath Jadhav
Department of Computer Engineering,
KKWIEER,Nashik .
Email:trish.jgd@gmail.com

has completed Bachelor's degree of Computer Engineering in 2009 . She is working in college of Engineering as System Analyst since 2010 and currently pursuing M.E in Computer Engineering from K.K.Wagh Institute of Engineering and Education Research, Nashik 3. Her research area is Data mining,Link mining, and Machine Learning.

Prof Satish S. Banait completed his post-graduation from Government Engineering college Aurangabad,BAMU university Maharashtra. Presently he is working at K.K.Wagh college of engineering ,Nas hik,Maharashtra,India as a Assistant professor .He has presented papers at National and International conferences and also published paper in national and international journals on various aspect of the computer engineering and Data mining and Data Warehouse .His research of interest include Data Warehouse,Software testing and quality assurance ,Software design and architecture,parallel computing.