Temporal Absence in Recommendations: a survey of Temporal Patterns in Netflix Prize Data

Mulang' Isaiah Onando[1], Waweru Mwangi[2]

[1] Computing Department, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

[2] Computing Department, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Abstract

Research on evaluating recommender systems shows that algorithms in this area are still deficient in prediction accuracy but recent works prove that modeling with temporal dynamics improves the degree of recommendation accuracy. Recommendations are invariably based on similarities of users and/or items in the user-item matrix of a system, user profiles, and rating information which presumes the presence of users or items in the matrix. In Social Media the matrix is takes a different form that may include a user-user or user-attributes. There is limited work focused on the temporal absence as an indicator of preference or concept drift: and hence a factor for inclusion in the recommender algorithms and models either to improve accuracy or to enhance user interaction in social based recommendations. This paper defines temporal absence in the context of recommender systems and verifies, through examination of the Netflix Prize data, the extent of temporal absence and the significance of such information in future research and improvement of recommendation algorithms.

Keywords: Temporal, Absence, Collaborative Filtering, CF, Prediction, Accuracy Recommendation, Recommender, RS, IR, IS.

1. Introduction

In this paper we present the necessary shift in paradigm from the normal view of recommender systems that bases them inherently upon full time presence of users and items to the irrevocable reality that occasionally or fully, a user could be absent from the system, whether passively: in situations where s/he actually logs in but does nothing or does other things apart from accepting the recommendations or rating them, or completely in situations where they could be dead or out of touch. These are just probable courses of temporal absence but recommender algorithms should take into considerations that one of the major reasons for such absence is disinterest or dissatisfactions from the recommendations, especially in systems in which rating, or ecommerce is the only activity of concern.

This paper explores the need to concentrate on finding a way to model temporal absence through mathematical modeling techniques in linear algebra, correlations and regression statistical models and to use the same within recommendation algorithms. An exploration of algorithms in the area of collaborative filtering more specifically in Matrix factorization and Aspect modeling is initially carried out and is followed by an analysis of the Netflix prize data to verify existence of user absence. Time dynamic models with a bias towards absence rather than presence forms a concentration point.

2. Theoretical Background

Matrix Factorization (Bell, Koren, & Volinsky, 2009) are techniques derived from the mathematical fields of numerical linear algebra and matrix manipulation, which has come to be accepted among the best state-of-the-art techniques. Due to its wide expanse of application domains and disciplines such as information retrieval and machine learning, recommender systems stay a relevant and active area of research with most work currently drifting towards evaluation frameworks and user centric approaches[4], [8].

There are early attempts to include time based variations in usage by a customer as apparent from work by Steve Hanneke and Eric Xing who propose a family of statistical models for social network evolution over time, based on an extension of Exponential Random Graph Models. Yehuda Koren (2009) [10] adapts an approach to model the temporal dynamics along the whole recommender systems time period; this is an approach that would allow a separation of lasting and meaningful factors from instantaneously decaying factors.

Recommendations should be able to recommend exactly what the user wants or provide a close near miss in predicting the same. When a user's requirements change, his/her preferences and activities changes and hence his behavior will change accordingly; a recommender system should be able to track such changes intuitively so as to alter recommendation to suite the new preferences, Absence from activity in the system either passively or completely can be a good indicator of shifting interest of a user. Some users are inherently dissimilar from their peers who would appear similar given their profiles or otherwise rating information, a recommendation algorithm should be able to closely monitor a user's reaction to recommendations and conclusively determine whether they are dissimilar rather than similar; if a user takes a considerable amount of time before s/he reacts to a recommendation then such absence could point to an inherent dissimilarity of the user and hence a call to change operational mode of the system.

Complete absence from the system can be tracked for a reasonable time, especially in social networking recommender systems, to be able to recommend to the user's friends a possible danger or mishap to their friends. If used in this manner, Social recommendation systems become not only a tool to increase interaction between socially present users but a social life assistant tool.

Regardless, current recommender systems algorithms have not yet fully embraced the aspect of temporal absence as a factor. As Yehuda Koren (2009) notes; "a mere decay of older instances or usage of multiple separate models loses too much signal, thus degrading prediction accuracy of recommender systems". Further, modeling time based use patterns of a user and time variations of preference to items can intelligently separate transient factors from lasting ones the only twist being that even time dynamics models have still not incorporated temporal absence. If well modeled, temporal absence can point out aspects like shifting customer preferences, absence due to natural phenomena like deaths in the cases of Social networks, dissimilarities of customers rather than similarity portrayed by peers and many more.

On another view point, such modeling of temporal absence can immeasurably reduce the runtime complexity of the algorithms that is if the temporal absence is done a priory hence used as a factor in dimensionality reduction. Also such algorithms as proposed by Yehuda Koren (2009) can easily lend themselves to model temporal absence and increase prediction accuracy.

2.1 Matrix Factorization

For a detailed discussion of a model based on these techniques refer to Alexandros Karatzoglou and Markus Weimer, 2010 [1]. As noted by Melville and Shindhwani (2010) [7], Matrix factorization fall under the famous wider category of algorithms referred to as latent factor models. In these models, two column vectors that intrinsically stand for some underlying features are employed to represent users and items within the system. Since Recommender systems rely upon the determination of users' probable rating on items; the recommended values are obtained by carrying out inner products of the vectors in question in relation to a defined error optimization function.

Currently perceived as the most successful in recommendation, Matrix factorization (MF) is comparable to other approaches like the Principal Component Analysis (PCA) and Singular Value Decomposition (SVD) as noted by (Mark Graus, 2011) [6]. In this sense, MF operates by breaking down the larger user-item matrix, always sparsely populated or mostly zero valued, into two matrices and subsequently, as earlier mentioned, obtains the inner products of these derived matrices in an attempt to reproduce the original matrix. Matrix factorization entirely relies on the perception that there are some hidden factors that influence a user's rating on an item. As an example, in social media, a user may refrain from commenting on another user's post on which s/he is expected to comment (probably she has been mentioned therewithal) because that post contains strong political sentiments or perhaps the friend has recently changed their religious views; several of these factors may result in the user totally avoiding social media activity (a state which will later become known as Temporal Absence). On the other hand, in the most common research field of ecommerce, a good example would be when a user gives higher ratings to products manufactured by the same company or they rate a given lip balm because a famous actress uses it (if you think of Lupita Ngong'os lip balm sales URL1 [12], URL2 [13]).

The whole combination of these factors becomes known as the feature set and it is therefore the fundamental function of the MF algorithm to learn the features associated with the user and the items. According to Mark Graus, 2011[6]; the features associated with a user should match with the features associated with an item pair on which a certain rating prediction is to be applied. From our social media example it is important to note that the user-item matrix may not necessarily be an equivalent of the strict ecommerce scenario but may include social media recommender systems that deal with user-user, usermessages or even user-time matrices, likewise, the ratings in these other domains would have different meaning from the traditional number valued ratings. In trying to discover the different features, we also make the assumption that the number of features would be smaller than the number of users and the number of items because it is impossible to have a user associated with a unique feature (www.quuxlabs.com, 2012) [3].

2.2 User-Item Matrix and Prediction Model

In Matrix Factorization methods for Collaborative Filtering, the known data is interpreted as a sparse matrix $\boldsymbol{M} \in \mathbb{R}^{n \times m}$. In this case of a recommender algorithm, it tries to describe the user – item matrix $\boldsymbol{M}$ as a $\boldsymbol{n} \times \boldsymbol{m}$ matrix (with $\boldsymbol{n}$ the number of users and $\boldsymbol{m}$ the number of items in the catalogue [6]. The cells in $\boldsymbol{M}$ denoted as $\boldsymbol{M_{i,j}}$ contain the rating of item $\boldsymbol{j}$ by the corresponding user $\boldsymbol{i}$, if such a rating is known. The predicted rating $\boldsymbol{F_{i,j}}$ of item $\boldsymbol{j}$ by user $\boldsymbol{i}$ is modeled as a linear combination of item factors $\boldsymbol{Q_{j*}} \in \mathbb{R}^{\boldsymbol{k}}$ and user factors $\boldsymbol{U_{i*}} \in \mathbb{R}^{\boldsymbol{k}}$**,** given the assumption that we would like to discover K latent features:

$$\boldsymbol{F_{i,j}} = \boldsymbol{U_{i*}}, \boldsymbol{Q_{*j}}$$

Where $\boldsymbol{U_{i*}}$ is the factor vector for user $\boldsymbol{i}$ and $\boldsymbol{Q_{*j}}$ the factor vector for item $\boldsymbol{j}$. Let $\boldsymbol{U} \in \mathbb{R}^{n \times k}$ denote the matrix of all user factor vectors and $\boldsymbol{Q} \in \mathbb{R}^{m \times k}$ the matrix of all item factor vectors. We can then express this prediction rule as a matrix product:

$$\boldsymbol{M} = \boldsymbol{UQ'}$$

That is, each row of $\boldsymbol{U}$ would represent the strength of the associations between a user and the features. Similarly, each row of $\boldsymbol{Q}$ would represent the strength of the associations between an item and the features. As expressed by Istv´an Pil´aszy (2009), if we consider the matrices as linear transformations, the approximation can be interpreted as follows: matrix $\boldsymbol{Q}$ is a transformation from $\mathbb{R}^{\boldsymbol{M}}$ into $\mathbb{R}^{\boldsymbol{K}}$, and matrix $\boldsymbol{U}$ is a transformation from $\mathbb{R}^{\boldsymbol{K}}$ into $\mathbb{R}^{\boldsymbol{N}}$, and $\boldsymbol{M}$ is a transformation from $\mathbb{R}^{\boldsymbol{M}}$ into $\mathbb{R}^{\boldsymbol{K}}$. Typically, K≪N and K≪M, therefore the intermediate K-dimensional vector space acts as a bottleneck in the approximation of $\boldsymbol{M}$ as a sequence of two linear transformations.

That is to say that the number of parameters to describe $\boldsymbol{M}$ can be reduced from $|\boldsymbol{M}|$ to NK + MK, if K is small. However, there are cases when NK + MK is larger than $|\boldsymbol{M}|$, for example, when K is greater than the average number of ratings per user.

A good learning algorithm should be able to handle these cases as well, since even for relatively small values of K, there will be many users who have less than K ratings. Note that $\boldsymbol{U}$ and $\boldsymbol{Q}$ typically contain real numbers, even when $\boldsymbol{M}$ contains only integers (Istv´an Pil´aszy, 2009).

On approaches and methods on how the actual error function is derived and algorithms used in optimizing the values including how these predicted ratings are calculated together with related issues of overfitting and Regularization refer to [2][6][7][10] and [11].

Illustrations or pictures: All halftone illustrations or pictures can be black and white and/or colored. Supply the best quality illustrations or pictures possible.

2.1 Footnotes

Footnotes should be typed in singled-line spacing at the bottom of the page and column where it is cited. Footnotes should be rare.

3. Temporal Absence

According to [10] modeling the way user and item characteristics change over time in a factorization model will help distill longer term trends from noisy patterns while in an item-item neighborhood model; the more fundamental relations among items can be revealed by learning how influence between two items rated by a user decays over time. In both of the scenarios, inclusion of temporal dynamics proved very useful in improving quality of predictions, more than various algorithmic enhancements, temporal absence can be a basic addition that would be used to indicate these decaying characteristics of time dynamics based models. Luheng et al., (2011) [5], proposed a unified framework for actively acquiring item and attribute feedbacks based on the random walk with restart (RWR) model that incorporates User-Item Edges, User-Attribute Edges and Item-Attribute Edges. From [5], the User-Attributes and Item-Attribute edges can easily incorporate time as an attribute and hence during dormancy or absence, the modeling can take a whole different dimension.

3.1 Perceived Benefits

Incorporating temporal absence in different algorithms would induce probable improvements as follows:

i) If the evaluation score is known a prior, it can help prune the user-item matrix to a manageable size by eliminating users who have been absent with longevity or items that have long since stopped being rated hence improve the runtime complexity of an algorithm.

ii) In social network recommender systems, for users who have long stopped using the system, if the algorithms could valuate such absence, they could give valuable recommendation to their friend on possible illness or demise.

iii) Passive Temporal absence of activities on recommendations given by a system could be modeled to provide a need to possibly change the mode of operation of the algorithms e.g. shifting from similarities to dissimilarity.

3.2 Research Focus and Target

The general objective of this research is to establish the significance of giving weight to temporal absence and use patterns of recommender systems users and incorporating the same into the recommender systems algorithms to track drifting user preferences for better recommendation results.

1. To provide a stepwise definition of temporal absence in the context of Recommender systems and algorithms.
2. To identify existing temporal dynamics model that lends itself for temporal absence modelling.

4. Methodology

Considering research on recommender systems have consistently employed only datasets prepared by other researchers or companies in this paper we evaluate freely available online datasets from Recommender systems that have been collected over time e.g. (as employed by Rashid et al., 2005) “there is a publicly available dataset from www.grouplens.org. The dataset is a fraction of the usage data drawn from MovieLens (www.movielens.org), a CF-based online movie recommendation system. It contains 6,040 users, 3,593 movies, and about one million ratings on a 5-star scale. Each user has rated at least 20 movies in the dataset”.

4.1 Netflix Dataset

About a decade ago, Netflix, the online DVD rental company announced a contest that would bring together many researchers in Machine Learning and Information Retrievals and several other Mathematical modeling based sub disciplines. The aim of this contest was to help improve the recommendation accuracy of the company’s system. “To enable this, the company released a training set of more than 100 million ratings spanning about 500,000 anonymous customers and their ratings on more than 17,000 movies, each movie being rated on a scale of 1 to 5 stars” [10] .

From the results of the competition that concluded in 2009 with a grand prize of $1 million prize given to the first team that surpassed the 10 percent increase; research on recommendation algorithms exploded with a dedicated conference being set in place by the ACM. Lessons learnt from this contest also include the fact that temporal dynamics incorporated within recommendation algorithm makes their performance better as observed from the winning paper [10]

This dataset has been employed by nearly all research from that period to date since it is the largest existing dataset known since the history of recommender systems. In addition to its size and availability, the dataset is 4 tuple with each entry consisting of {user id, item id, rating, and time}, the time attribute is normally absent in other datasets like the movielens dataset. This follows that it is the best dataset for modeling temporal aspects of the systems and algorithms, though no longer available due to legal issues, this research has acquired this dataset and has used it for the foregoing analysis.

4.2 Temporal Absence Connotations in Netflix Data

To study the temporal absence from the Netflix data; given that this is a very large dataset of over three (3) million ratings spanning a period of over six (6) years, I divide this time into time bins, a concept discussed by [10] that creates smaller time windows from the larger period. I then get the unique customer ids within each time bin and produce a graph of the same. This will provide a view of presence patterns within the system. The presence graph is shown in figure 5.

4.2.1 Binning

We used a bin size of 10 days resulting into two hundred and twenty five (225) time bins for this study but the 225th bin has only two important days so it’s discarded leaving 224 effective time bins; table1 bellow indicates the binning periods sample for the first twenty (20) time bins. It should be noted that the last time bin period is 2005-12-19 – 2005-12-28.

IJCSI International Journal of Computer Science Issues, Vol. 11, Issue 4, No 2, July 2014
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

#BIN	PERIOD	#BIN	PERIOD
bin1	1999-11-11 - 1999-11-20	bin11	2000-02-19 - 2000-02-28
bin2	1999-11-21 - 1999-11-30	bin12	2000-02-29 - 2000-03-09
bin3	1999-12-01 - 1999-12-10	bin13	2000-03-10 - 2000-03-19
bin4	1999-12-11 - 1999-12-20	bin14	2000-03-20 - 2000-03-29
bin5	1999-12-21 - 1999-12-30	bin15	2000-03-30 - 2000-04-08
bin6	1999-12-31 - 2000-01-09	bin16	2000-04-09 - 2000-04-18
bin7	2000-01-10 - 2000-01-19	bin17	2000-04-19 - 2000-04-28
bin8	2000-01-20 - 2000-01-29	bin18	2000-04-29 - 2000-05-08
bin9	2000-01-30 - 2000-02-08	bin19	2000-05-09 - 2000-05-18
bin10	2000-02-09 - 2000-02-18	bin20	2000-05-19 - 2000-05-28

Fig. 1 First 20 10-day bin periods

5. Results

5.1 Obtaining Absence

Since the Netflix data does not come with the customers' registration information; we are only provided with the movies and rating information, yet our interest for this study is more inclined to the customers/users of the system, it becomes a challenge to know exactly how many customers were not present from the system. To circumvent this problem we use the customer id fields to get the lower bound number of customers present in the system. I.E. since the customer id is auto incremented from 1, we can confidently tell that if for bin X, the highest customer id is a given figure Y, and then the least number of customers present within the system at this bin time is Y. Subsequently, for any successive bin, if there is no higher customer id than Y, then Y remains the least number of customers within the system for that bin, otherwise we update Y with the new highest customer id from the bin. $P_{LB,Y} = \{\max(customer_{id}, Bin_{1...Y})\}$

From the yellow circle points in the windowed comparisons bellow, we note that the number of customers in the system change with time; in most cases within this data, the highest number of rating users remains constant but at some points it shifts as indicated by the bins between bin3 to bin10 as opposed to the overall graphs that hid such fine details. It is therefore prudent to assume that the overall numbers in the system will constantly be changing hence the number of absent uses of the ratings system is not a mere reflection of the absence based on some fixed total number of users.

We therefore define Temporal Absence as the complete or partial lack of activity from the system by a user over a transitive period of time long enough to suggest a drift in preference or otherwise to affect the natural order of his/her sub conscience. It should be noted though that absence in this context does not necessarily mean the opposite of presence and by this we mean that a user can be passive in the system yet actually present or on the other hand completely stay out.

If the user is actually present in the system then his/her absence can be validated by lack of their usual activity e.g. for a rating system, a user who is expected to rate recommended item yet they do it not, to the system, this is a partial absence and could suggest patterns about their interests in the recommendations. Temporal absence, can otherwise mean partial or complete lack of activity by a user from the system or lack of activity on an item within the system for a measurable duration of time long enough to indicate probable change of behavior or interest.

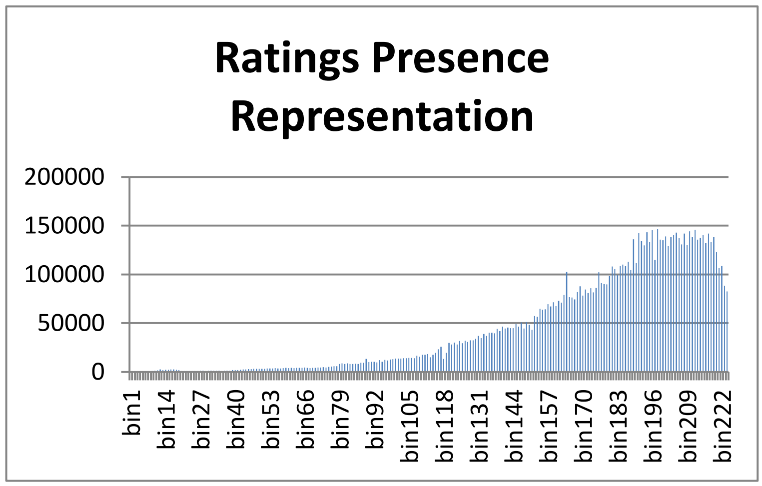

Fig. 2 Bin ratings presence graph

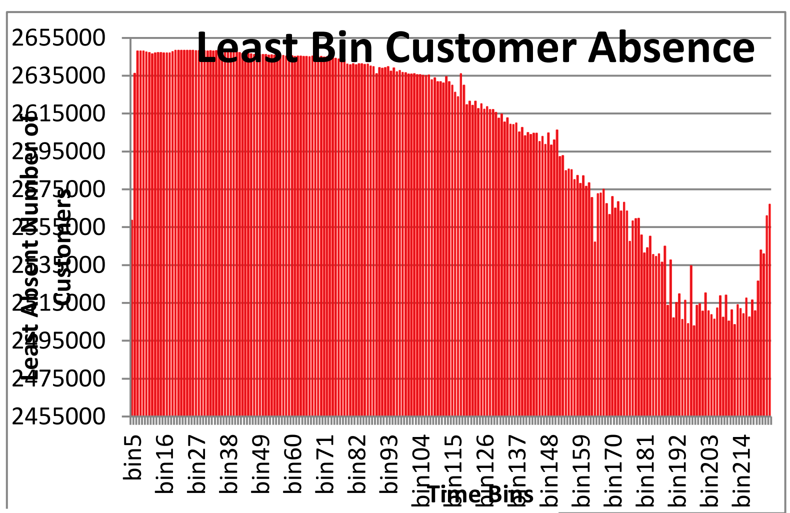

Fig. 3 Least Bin Customer Absence graph

IJCSI International Journal of Computer Science Issues, Vol. 11, Issue 4, No 2, July 2014
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

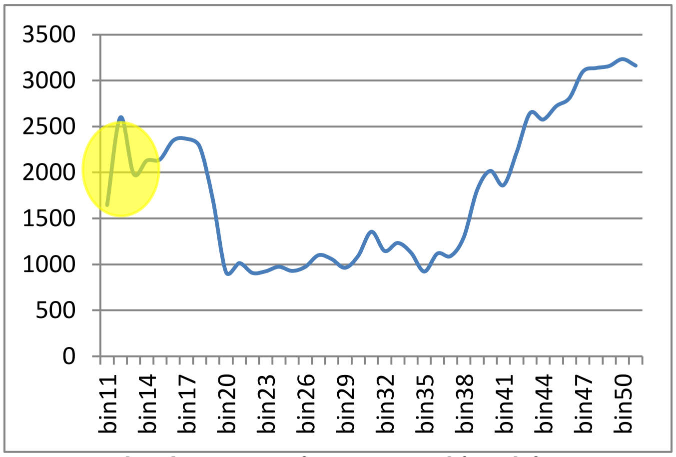

Fig. 4 Windowed comparison of present to reveal finer shifts in customer numbers

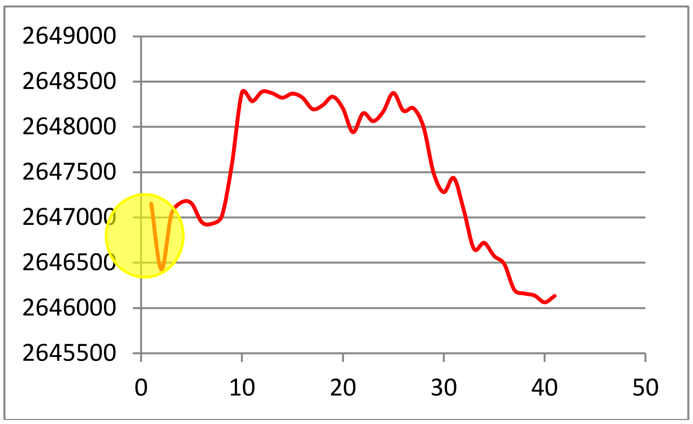

Fig. 5 Windowed comparison of absence/present to reveal finer shifts in customer numbers

6. Discussion

From the results of the analysis above, it is clear that the numbers that are absent from the system at any given bin period way exceeds the numbers present. Given that the actually present individuals in the system or those that actually do the ratings at any given instance of time change from bin to bin, it is prudent to determine which particular users would be present in a bin so that during the algorithmic run, we have a highly reduced number of users to work with. Such is the modeling that is required to be able to be achieved to attain efficiency and accuracy from temporal absence connotations in the data. To achieve this, the following steps would be taken:

Determination of each user's absence pattern from the rating system; here we note that the given user might actually be present in the system or otherwise be signed in to the system but not actively involved in system ratings or for the Netflix case, purchase of movies. To achieve this, a regression analysis algorithm could be used; the end result of which is to have a separation of which users are temporally absent from the ratings system and during which time bins. The results of step one would then be fed into Matrix Factorization algorithm such as the SVD++ algorithm [10][11] in a bid to predict future absence patterns. This would mean that the information fed to the algorithms would have to be reconstructed into a suitable low rank matrix.

There after we could use the outcomes as a discriminative criterion for selecting which particular user will be required for a given instance run of the ratings prediction algorithm. Since the number of users actually present in the system within any bin period is actually way less than the actual population of the user – item matrix, this will be very cost effective in terms of execution time.

The last step of the improved algorithm would be to determine whether a user who has been absent from the system should be considered temporarily absent or is s/he within their absence patterns, is considered temporally absent then the mode of operation of the algorithm is altered to encompass the absence factor but if otherwise found to be within their absence patterns, their ratings predictions would be computed as usual. These steps form the four major steps of a temporal absence model of recommendation algorithms, any developer of such algorithm must seek to address the given steps during design.

In other systems like Social media systems, we observe that determining a user's absence patterns and eventually being able to predict future absence can improve interaction of users of the system and other aspects such as the people you may know recommendations

7. Conclusions

The Netflix dataset is a widely used dataset for recommender systems research, this paper provides an introduction to relatively new dimension to modelling recommender systems by the proof of existence of temporal absence connotations within recommendation data. With this, we propose a paradigm shift from an all presence oriented recommendation to an absence inclusion approach to recommendations. In furthering this research, the big question is majorly on how to build a temporal absence model and how to incorporate the same within the existing or otherwise new algorithms for recommendations with an eye to improve both efficiency and accuracy of recommendations. Such models thus built from has the potential of forming a backbone for several application

IJCSI International Journal of Computer Science Issues, Vol. 11, Issue 4, No 2, July 2014
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

domains such as E – commerce as previously mentioned, other areas of Application for Temporal Absence Models is in social media, for enhancing presence and expectance of peers i.e. the system should be able to suggest to the user the time usage patterns of their friend and peers so as to schedule online meeting times and chats appropriately. Security in devices can also be enhanced by use of temporal usage patterns of the users and recommending steps of actions in case the absence patterns are abnormal.

References

[1] Alexandros Karatzoglou and Markus Weimer, "Quantile Matrix Factorization for Collaborative Filtering", 2010

[2] Bell R., Volinsky C. "Matrix factorization techniques for recommender systems". AT&T Labs—Research, August 2009.

[3] Http://www.quuxlabs.com/blog/2010/09/matrix-factorization-a-simple-tutorial-and-implementation-in-python/., 12th, July 2012.

[4] Li Chen and Pearl Pu, "Evaluating recommender systems from the user's perspective: survey of the state of the art". Springer Science +Business Media B.V. March, 2012.

[5] Luheng He, Nathan N. Liu, and Qiang Yang, "Active Dual Collaborative Filtering with Both Item and Attribute Feedback", 2011

[6] Mark Graus., "Understanding the Latent Features of Matrix Factorization Algorithms in Movie Recommender Systems", Eindhoven, March 25th, 2011.

[7] Melville P., Sindhwani V., "Recommender Systems". Encyclopedia of Machine Learning, Springer-Verlag Berlin Heidelberg, 2010.

[8] Paolo Cremonesi, Franca Garzoto and RobertoTurrin, "User Effort vs. Accuracy in Rating-based Elicitation"

[9] Rashid A.M., Karypis G., Riedl J., "Influence in Ratings-Based Recommender Systems: An Algorithm-Independent Approach". 2005.

[10] Yehuda Koren, 2009; "Collaborative Filtering with Temporal Dynamics". KDD'09, June 28–July 1, 2009, Paris, France. Copyright 2009 ACM 978-1-60558-495-9/09/06

[11] Yehuda Koren. "Factorization Meets the Neighborhood: a Multifaceted Collaborative Filtering Model". KDD'08, August 24–27, 2008, Las Vegas, Nevada, USA

[12] URL1; "Lupita Nyong'o's Oscars Lip Balm Sells Out. http://www.hollywoodreporter.com/news/lupita-nyongos-oscar-lip-balm-686506, Accessed: 30/May/2014.

[13] URL2; "Lupita Nyong'o's Clarins lip balm from the Oscars sells out, becomes a social media sensation". http://www.nydailynews.com/life-style/oscars-2014-lupita-nyong-lip-balm-viral-article-1.1710752#ixzz33KKl03c9. Accessed: 30/May/2014.

Mr. Isaiah Mulang Holds a First class honors degree in Computer Technology, 2009 from JKUAT and has recently submitted his MSc. Thesis from where this paper has been deduced; has been a Teaching assistance at the Department of Computing (SCIT) in JKUAT sin 2010t; a student member of the ACM; Research interest include Recommender Systems, Information Retrieval, Machine Learning and Big Data.

Prof. Ronald Waweru Mwangi Currently associate professor at Jomo Kenyatta University of Agriculture and Technology is the immediate former director of School of Computing and Informatics. Is a known researcher in the field of mathematics and Computer Science and has supervised many PhD. And Maters Students within his ten years as director of SCIT.