Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 6, June 2021: 1571-1578

Analysis of Effectiveness of the Browsing Behavior of Web User using Sequence Weight Knowledge

Sheetal Sahu [a*], Rajendra Gupta [b], Amit Dutta [c]

[a*,b,c] Rabindranath Tagore University, Raisen, Madhya Pradesh (M.P.), India. All India Council for Technical Education (AICTE), New Delhi, India.

***Corresponding author:** sheetalsahu.kanu@gmail.com, . rajendragupta1@yahoo.com amitdutta07@gmail.com

Abstract

A number of web users worldwide regularly follow the web to share their information, converse various topics, share, stay connected and obtain information. As a result, enormous amounts of data are generated by the web users and then one could employ this data to obtain useful predictions of certain web user related behavior. There are many platforms where web users could communicate and exchange information. These platforms comprises of social networks, and digital communication networks. In this research paper, the effectiveness of the browsing behavior of web user is analysed using sequence weight learning. The outcome shows that the proposed personalized recommendation framework using the Competence Scoring process for sequence weight learning able to achieve significantly higher accuracy. The investigation result of web uses also shows the personalized framework predicts future items reliably and could be employed to automatically recommended next-items to indented web users.

Keywords: Web user behavior, Sequence Weight Learning, Sequential Pattern Mining

1. Introduction

User behavior analysis over web makes user information view and searches the rules of user behavior from massive web user behavior data. It helps enterprises better understand the web users' preferences, developing the value of web users and ultimately bringing more benefits to enterprises. In the development of the technological data, the multi-dimensional mass storage data acquisition and storage technologies has become maturing the structured and un-structured web user behavioral data which has a large increase, making web user behavior assumption and application research that is important.

The data mining for sequential data, having lots of research areas for the constructive information gathering from unknown, employ and reliable samples related to sequential database. Now a day the website growth increases and the complexity for web users to browse effectively. Therefore the web form of data in different format has been growing rapidly in previous years. This bulk quantity of data has produced in highly volume of hyperlinked documents

which contain text, audio, video etc. which is not possible to store in memory [1,2].

In the present scenario, the websites are having drastically growth in size with new process for processing of customer requirement. The online applications are having lots of data about material and supplier. Since the website data in the size rapidly growing and produces the resulting network of information with lacks of structure. Due to this reason, the website web users are getting now and again mismanaged and lost in that

related is information overload that continues for expansion of data. The detection of most related information is needs to be fetching by involving the new search with indexing of the content in the websites. The collection of some meaningful knowledge is most useful that is also available on the web form. As per the concern of individual web users' needs and its interests in the product selection by personalizing and provide relative information and its services [3-5].

2. Web User Behaviour On Web

A numeral of web users worldwide regularly follows the web to share their information, converse various topics, share, stay connected and obtain information. As a result, enormous amounts of data are generated by the web users and then one could employ this data to fabricate useful predictions of certain web user related behavior. These predictions could be employed in a variety of domains, including products marketing, finance, social dynamics, public health, and politics. The consequences of this are that an increasing number of researchers have been attracted to do study on this subject [6].

Following points discusses the short overview of web access platforms:

A. Social Networks

The social networks are the well-liked platforms for socialising, interacting and sharing information on the web. Societal Network platform consist of Twitter and

Facebook; while Question & Answers (Q & A) forum includes Quora and Stacks Exchange. In addition, digital newspapers, such as the Daily Newspaper and The weekly Newspaper allow web users to send their comments and interact with other web users.

Following are the common features of these social platforms,

- the existence of mechanisms to interact and propagate information
- being there a social structure

In earlier days, two most successful and the largest social networks are Facebook and Twitter:

- Facebook: In this platform, or social network, the people could make `friend' to every other and communicate with every other. A mutual abstract model for representing the 'Facebook' assemblage is a social graph where people are nodes and their relationships are edges of the graph.
- Twitter: In this social media platform, the people

`follow' every other and be follow by others/ or follow somebody.

B.. Digital Communication Networks

The Internet services have provided people the ability to communicate worldwide through e-mail and instant messengers. Top furthermost brands like Google, Microsoft and Yahoo are the type of such email services. The Facebook messenger, Whats-App and Kiki are examples of instant messengers.

3. Models Of Web User Behaviour On The Web

The web user behavior on the web could be considered as a procedure of communication and interaction amongst web users on a web. This is one of the modest models of communication is responded in the earlier work. This model consists of a transmitter, a message, a channel in which the communication travels, noise or interference, and a receiver [7,8,12].

The models of web user behavior are split into two large groups which are-

(1) A Dynamic models

(2) Graph-based models

The dynamic model is based upon the control assumption and system assumption whereas graph-based models employ graph assumption and social networks assumption.

3.1. Dynamic Model

A dynamic model characterizes the behavior of an object on time. Generally such models are measured as a set of states ordered in a sequence. In case of web user behavior on web, such objects are peoples and the

dynamic models represent their behavior. Every internal cerebral state could be written as a single dynamic process:

$\dot{x}_k = f_k(x_k, t) + \xi(t)$

$y_k = h_k(x_k, t) + \nu(t)$

where, the function fk models the dynamic evolution of state vector xk at a time denoted with k. Both ξ and ν are white noise processes with known spectral density matrices. The explanation denoted by yk is a function hk of the state vector xk.

In the study of sequential data mining, the concept of dynamic model has been functional for various domains. Such system consists of input, an output and state variables that are generally associated with differential equations.

3.2. Graph based Model

The graph based model presumes that the graph is employed for modeling web user behavior. Typically, the nodes of such a graph are connected with people and the edges connecting the nodes are related with some sort of communication or connection among the people. The modeling of web user behavior as a graph allows capturing structural properties of the network formed by the group.

A graph-based model of a social media Twitter network have been proposed and then built. In the research study, the researchers proposed a graph namely

$tt = (U,E)$

where the nodes U linked with people and the edges E is linked with the relationships of following web users or web users being followed by other web users.

The Time-Varying Graphs (TVGs) have been established to describe a wide range of dynamic networks. The nodes of a Time-Varying Graphs are denoted as a set of entities U and edges are a set of relation denoted with E in among these entities. In addition, an alphabet L accounts for any possessions of a relation. This could be denoted as,

$E \subseteq U \times U \times L$

The meaning of labels L is a domain specific and left opens. It is having a distinct label L and the set E permit multiple relations among entities.

The relations amongst entities are defined on a time span T

$\in T$ called the life-time of the system. The temporal

domain denoted with T is N for discrete time system for continuous time systems. The dynamics of the system could be described as a Time-Varying Graphs, in which

$tt = (U ; E; T ; \rho; \zeta)$,

where $\rho : E \times T \rightarrow \{0, 1\}$, called occurrence function, indicates whether a given edge is available at a given time. Whereas $\zeta : E \times T \rightarrow T$, called Latency function that indicates the time it takes to cross a given edge if starting at a given period (the latency of an edge could vary with time).

The sequence $ST(tt) = sort(\cup\{ST(e) : e \in E\})$, which is called characteristic dates of tt, correspond to the sequence of dates when appearance or disappearance of an edge occur in the system. As such events could be capable of evolution of the graph tt. The evolution of tt is described as the sequence of graphs

$Stt = tt1, tt2, .., ttn-1, ttn$

where tti corresponds to static snap-shot of a graph at time t

= i. In general case, tti f= tti+1.

Time-Varying Graphs concept is a useful mathematical ideal for capturing temporal properties of a societal network sites.

4..Extracting Features For Prediction

Prediction of web user behavior on the web involves building predictive models of web user behavior that depends on historical data. In this paper, an algorithm is employed for building web user behavior model. The

most important requirement for such models is being able to produce a truthful prediction [13-15]. The process of identification of features for predicting web user behavior is mentioned below :

4.1 Identifying Features

There is a great number of probable ways to construct and compute features. These features are essential for producing a perfect prediction and some of them are employing. In this Sub-section of research study, a set of features for Twitter trends prediction is analysed. The investigator employed three types of features which are -

- Content
- Node
- Structure

The content features are extracted from the data; for example, the number of persons. The node features are associated with the information of web users. The structure features are related to the metrics on the topological structure of the network.

The features related to Structural features, Usage of

network and Profile features the author extracted for building a predictive model. These features are :

- Structural features (in which number of edges, internal density of social network usual total degree formed by a community)
- Usage of network (like chat days, audio-video days)
- Profile features (includes number of countries, cities, gender, age of web users)

All over again, the structural features are high significant for the prediction. A model for predicting how web user connected in on-line location-based societal networks was proposed by the researcher [9,10]. The researcher classified the features into social features pointed for friends-of- friends, place features computed for place-friends and global features. The place features include, the amount of check-ins, the numeral and the fraction of common places among two web users.

4.2 Complexity of Feature Extraction

To identify the features that considerably inspiration the realization of a sequential sample task is one side of furthermost sequential sample mining work. The complexity of calculating such features could be analysed using Big ‘O’ notation. To do this, several transformations of feature space model are introduced [16]. These transformations are as following:

- Standardisation:

The features could signify comparable objects then be measured in different units. For instance, two dissimilar features represent the same dimension but the first one is in seconds and the second one is in minutes. The standardisation prevents this from happening.

- Aggregation:

Aggregation involves several features aggregation.

- Normalisation:

Employed to eliminate the requirement of a feature on the size of the feature

- Non-linear expansions :

In case of the problem is very hard and complex, the available features do not enough to derive respectable results in the form of accuracy of prediction;

It is observed that some transformations transform the dimensionality of a problem while others do not. For example, Standardisation & Normalisation doesn’t change the dimensionality of a difficulty while Aggregation decreases the problem dimensionality.

On the basis of above, choosing a process for feature selection task depends on the -

- Total number of features
- Size of data

- Algorithm computational complexity

5. Sequence Weight Knowledge For Next Items Recommendation

The personalization of weights is initially assigned to all sequences in the Sequence Database as an off-line operation using the aforementioned learning progression. The author define the personalized count and support to obtain into account the learned sequence the weights so that the web user-specific sequence knowledge is efficiently exploited by the proposed framework to personalize the sequential mining:

It is necessary to check the sample count. In this regard, the every sequence si Sequence Database has a learning weight w(si,sq) with respect to web user sequence sq, the count of sample x in sequence database is:

$$Count\ (x, s_q) = \sum_{s_t} w(s_i, s_q)$$

The above equation sums the weights of all sequences that contain the sample x, such that a higher count(x, sq) means a higher support for sample x in the sequence database.

Now next task is to count the sample support. The support of x in Sequence Database is defined as:

$$Support\ (x, s_q) = \frac{Count\ (x, s_q)}{\sum_{s_t} w(s_i, s_q)}$$

where the denominator is the entire sequence the weight in the sequence database.

The above support definition satisfies the monotonically- decreasing property: given samples A and B, if B A, then support(B) support(A),

as B must be part of all sequences containing A. Therefore, it is readily appropriate to any of the conventional Apriori- inspired sequential sample mining algorithms to mine the high-support samples. An an example of such sequential sample mining algorithms is Consequence weight learning, the admired sample-growth approach in which a sequence database is recursively predictable into a set of smaller databases, and sequential samples are grown in every projected database by exploring merely locally frequent fragments [17].

In the sequential pattern mining, samples with personalized support not less than the specified minimum support δ are productivity as the web user-specific frequent samples Fq:

Fq = {x|support(x, sq) ≥ δ, δ > 0}

In the above equation, the author employed sequence the weight knowledge to eliminate all the web user-irrelevant records in Sequence Database which are not related to the target web user (since the author will not include x in Fq if support(x, sq) = 0). As such, the author could significantly progress the mining efficiency. The frequent sample set Fq is then employed in our target task to recommend next- items to the target web user [18], where the majority recent items of the object web user are considered further valuable for predicting the next-items.

Algorithm 1. Personalized Sequential Mining-based Recommendation

INPUT: The target web user u_q's sequence, s_q, and the database Sequence Database of other web user sequences
BEGIN
// sequence the weight learning step
for all sequence $s_i \in SEQUENCE\ DATABASE$ **do**
Compute $b_{i,q}(t)$ // backward-compatibility
$f_{i,q}(t)$ // forward-extensibility
$c_{i,q}(t)$ // Competence Scoring
end for
// the Sequence Database with sequence the weights personalized to web user
Apply a sequential mining algorithm to get personalized frequent samples, F_q to compute the personalized support
// next-items recommendation step (target task)
Employ the algorithm to compute the personalized support of every candidate next-item in frequent samples F_q and recommend the items with highest support to web user u_q
END

6. Experimental Evaluation

The author presents a two part evaluation of proposed personalized sequential mining-based next-items suggestion framework. In the first part, the author evaluates the effectiveness of learning sequence with respect to the sequential weight mining. In the second part, the author evaluates the framework's efficiency in terms of prediction accuracy of the next-items recommendation.

6.1. Evaluating the Framework Efficacy

We have chosen MSNBC dataset to assess the effectiveness of the framework as it has to a great extent more sequences. The learning of web user-specific (personalized) sequence the weights ought to decrease the time taken for sample mining, because by eliminating all the web user-irrelevant sequences (i.e. the weights are

equal to 0) in Sequence Database. We merely need to handle a much smaller personalized hypothesis space consisting of sequences that is extra relevant to the target web users [19,20]. On the divergent, without exploiting personalized sequence the weights, sequential sample mining algorithm will be performed inefficiently since it has to go through all the transactions in Sequence Database. To test a dataset an analytical tool MATLAB 2016a is employed over the dataset.

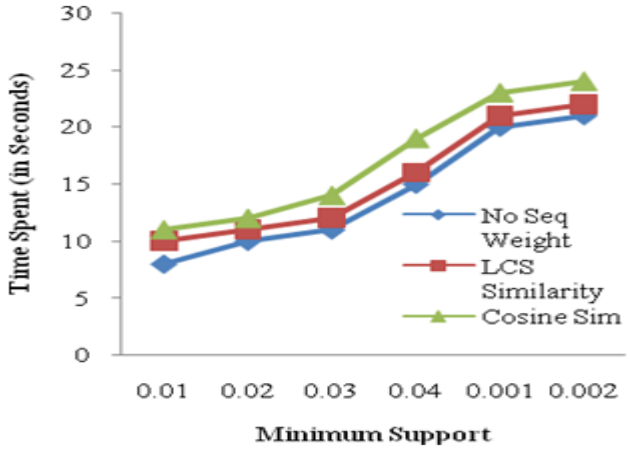

Fig a

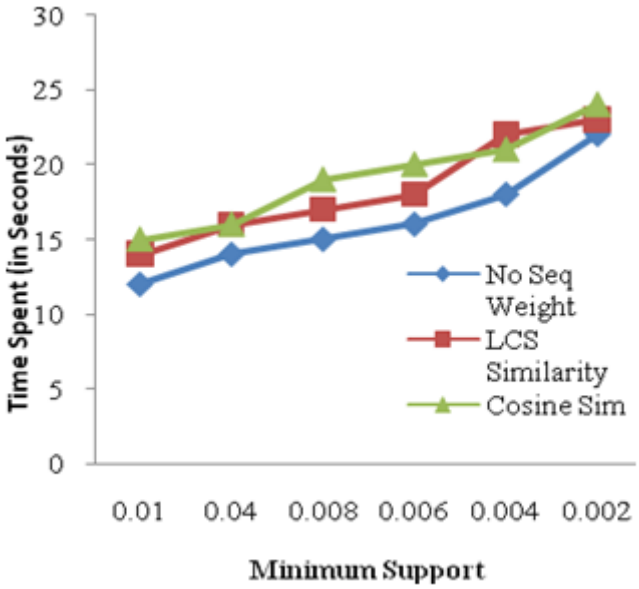

Fig b

Figure (b) : Improvement in time taken as a result of Personalized Learning Framework (database size from 50000-100000 sequences

Table 1 : Performance of algorithms with % improvement over and above sequential weight mining with no sequence weight learning

Figure 1(a) shows the time taken to mine sequential weight for dissimilar minimum support values when there is no personalized sequence of weight learning, versus with sequence the weight learning using the related process. Indeed, the framework considerably condensed the time for the sequential sample mining, particularly with our Competence Scoring the weighting procedure. Figure 1(b) compares the total time taken for both learning sequence of the weights and running the sequential weight mining algorithm. Taking into account the sequence in weight learning instance, the web user-independent learning process is applied. This completely ignores the target web users and computes sequence the weights based on their items' recognition, added significantly to the time taken, making it fewer striking compared to sequential sample mining without sequence the weight learning.

6.2. Accuracy of the Next-Items Recommendation

For different processes for next-items recommendation using MSNBC and book-loan datasets has compared. We need to do experiment with sequences having sufficient items (at least ten items in our experiments). For every sequence, we release the primary five items (known portion) and hold-out the remaining items because the ground- truth for evaluation. The author allow every process to recommend up to ten items with the highest support values and evaluate results in terms of recall, precision and F1-measure, all of which are standard evaluation metrics for recommendation systems [21,22]. As there are test sequences for which certain processes don't give any recommendation, we also measure the applicability of every process in terms of the percentage of test cases for which recommendations are given. The author set the min-sup for MSNBC and book- loan as high as possible to 0.1 and 0.05, respectively, so that the majority of the less frequent samples are effectively filtered off by the sequential sample mining, while maintaining a minimum applicability of around 40% for the recommendation.

(a) Performance on MSNBC dataset

The results for the MSNBC and book-loan datasets are presented in Table 1. The experimental results clearly demonstrate that the process for learning of sequence the

weights which are more related to the target web users could indeed yield significantly more accurate next-items recommendations. In particular, sequence knowledge learning using our proposed recommendation Competence Scoring produced the greatest improvement in performance among the competing process; it provide recommendations for almost 95 percentage of the test cases (a vast improvement in applicability), and considerably increased the recall, precision, F1-measure (i.e., proportion of test cases where the top-1 recommended item is correct). Particularly, for the MSNBC dataset, 70.6% of the test cases contained the first recommended item when Competence Scoring was employed, compared to just around 45 percentage for the competing process. Similarly for the book-loan dataset, our proposed Competence Scoring process was able to outperform all the competing procedure in terms of the various evaluation metrics.

7.Conclusions

The approach of sequential mining for web user behaviour allows one to extract semantic groups of features linked with web users, their relationships, messages, temporal, and structural features. The proposed approach for feature extraction shows that using structural and temporal features is advantageous for the accuracy of a forecasting but extracting such features could significantly enhance the time of a feature extraction. As a consequence, a guideline for selecting and constructing an efficient feature set based on the classification is proposed.

The proposed approach for feature mining shows that using temporal features is advantageous for the accurateness of a prediction but extracting such features could increase the time of a feature extraction. The research results thus show that our personalized framework predicts future items reliably and could be employed to automatically recommend next-items to indented web users.

References

[1] Smith, M. A. and Fiore, A. T. (2016), "Visualization components for persistent conversations". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '01, pages 136–143, New York, NY, USA. ACM.

[2] Sontag, E. D. (2017). "Mathematical Control Theory: Deterministic Finite Dimensional Systems", Springer- Verlag New York, Inc., New York, NY, USA.

[3] Spiro, E., Irvine, C., Du.Bois, C., and Buts, C. (2016). “Waiting for author " modeling waiting times in

[4] information propagation”, In *2012 Neural Information Processing Systems (NIPS) workshop of social networks and social media confe ence*, volume 12, pages 1–8, Montreal, Canada.

[5] Stahl, F. and Bramer, M. (2016). Scaling up classification rule induction through parallel processing. *The Knowledge Engineering Review*, 28:451–478.

[6] Steurer, M. and Trattner, C. (2016). "Predicting interactions in online social networks: An experiment in Second Life", In *Proceedings of the 4th International Workshop on Modeling Social Media*, MSM ’13, pages 1–8, New York, NY, USA. ACM.

[7] Strrle, H. and Fish, A. (2013). “Towards an operationalization of the physics of notations for the analysis of visual languages”, In *Model-Driven Engineering Languages and Systems*, Volume 817 of Lecture Notes in Computer Science, pages 104–120. Springer

[8] Su, Jo. and Zhang, H. (2006), “A fast decision tree learning algorithm”, In Proceedings of the 21st National Conference on Artificial Intelligence - *Volume 1*, AAI’06, pages 501–505, Boston, Maschemploytts. AAAI Press.

[9] Tange, J., Musolei, M., Mascolo, C., and Latora, V. (2016). “Temporal distance metrics for social network analysis”, In *Proceedings of the 2nd ACM Workshop on Online Social Networks*, WOSN ’09, pages 31–36, New York, NY, USA. ACM.

[10] Teevan, J., Morris, M.R., and Panvich, K. (2011). “Factors affecting response quantity and speed for questions asked via social network status messages”, In Adamic, L. A., Baeza-Yates, R. A., and Counts, S., editors, *ICWSM*. The AAAI Press.

[11] Teinemaa, I., Leotjeva, A., Duma, M., and Kikas, R. (2015). “Community-based prediction of activity change in skype”. In 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)", pages 731–748, Beijing, China.

[12] Toivonen, R. (2010). “Social Networks: Modeling Structure and Dynamics”. A Technical Report on published in Advances in Computing of Sequential Data.

[13] Vartak, M., Huong, S., Siddiqui, T., Madden, S., and Paramswaran, A. (2015), “Towards visualization Recommendation System”, In *Workshop on Data Systems for Interactive Analytics (DSIA)*, pages 01–06, Chicago, USA.

[14] Venolia, G. D. and Neustaedter, C. (2013). “Understanding sequence and reply relationships within email conversations: An mixed-model visualization”. Technical Report MSR-TR-2002-102, Microsoft Research.

[15] Vi´egas, F.B., Smitha, M. (2014). “Newsgroup crowds and researcher lines: Visualizing the activity of individuals in Conversational cyberspaces", In *Proceedings of the the 37th Annual Hawaii International Conference on System Sciences (HICSS’04) - Track 4 - Volume 4*, HICSS ’04, pages 1– 10, Washington, DC, USA. IEEE Computer Society