Indian Journal of Fundamental and Applied Life Sciences ISSN: 2231– 6345 (Online)
An Open Access, Online International Journal Available at www.cibtech.org/sp.ed/jls/2015/03/jls.htm
2015 Vol. 5 (S3), pp. 2443-2455/Moradi et al.
Research Article

THE IDENTIFICATION OF WEB NAVIGATION PATTERNS OF SOCIAL NETWORKS TO IMPROVE MARKETING OF VIRTUAL COMMUNITIES

***Hamid Saidi Moradi[1, 2], Fardin Akhlaghian Tab[3] and Fardin Ahmadi Zar[4]**
[1]*Department of Computer Engineering, Islamic Azad University of Sanandaj, Sanandaj Branch, Janbazan Boulevard, Sanandaj, Iran*
[2]*Department of Computer Engineering, Kurdistan Science and Research Branch, Islamic Azad University, Sanandaj, Iran*
[3]*Department of Computer Engineering, University of Kurdistan*
[4]*Department of Industrial Engineering, University of Kurdistan*
**Author for Correspondence*

ABSTRACT
After development and increasing expansion of social networks, new markets and new marketing methods became available in the online world. In this regard, competition among different producers and service providers raise the need for e-marketing and as a result, it is the type of marketing, the channels and persons by whom products are introduced which increase efficiency and business success. Today, the marketing managers of online social networks are concerned with identification and understanding of clients' behaviors due to absence of visual determination of sectors to which they belong, products which attract more clients as well as invisibility of clients. Therefore, more information on clients regarding their behaviors, interests and web-navigation patterns in social networks to provide them with better services along with information on usages of social networks for clients to develop of list of their favorite items is the objectives of present study. In other words, one can determine natural nodes based on navigation patterns of clients. The present study uses the associated information of social networks, clients' sequence of navigation, URLs of their most-visited pages, and web-mining algorithms to develop a model for prediction of sequence and frequency of clients' web-navigation. Through this model, the marketers can identify sectors (categories) of clients and targeted market(s) in association with demographic data. Availability of the information, assignment of new clients to existing clusters and observing marketing strategies contribute to offering favorite products to clients. Finally, the results of present study including diagram analysis of clusters, descriptive profiling of clusters, discrimination analysis of clusters, etc. can assist the companies to properly predict and analyze current and future conditions of clients along with improvement of marketing in virtual communities.

Keywords**:** *Social Network Websites, Web Navigation Patterns, Web Mining Algorithms, Social Network Marketing (SNM)*

INTRODUCTION
The Identification of Web Navigation Patterns of Social Networks to Improve Marketing of Virtual Communities
In the contemporary competitive e-business environment, there is nothing called "*Primitive Production*" and "*Monopolies*". The increasing expansion of internet and globalization led to establishment of thousands of firms manufacturing a distinctive product. Here, the question is "Who can survive victoriously in this chaotic environment?" The answer is "those who have comprehensive marketing measures". Marketing is the key of successful e-business without which the best products generate no profit and top-quality services find no client (Phillip *et al.*, 2005).
The marketers face problems in accessing clients through paper and TV advertisements. However, rapid developments of clients' population in the internet and Web.2 along with emphasis on interaction between two people have drawn the attention of market to themselves. In this regard, social networks are especially favored by the public and they attract an increasing number of users. Social network websites

have changed into new resources and platforms of marketing and this is because of their increasing popularity among users.

Therefore, marketing managers need to obtain more information of clients to offer proper services based on their interests and behaviors as denoted in patterns of web-navigation. It is also essential to extract and analyze usages of websites for clients and list of their favorite items. In other words, one can determine natural groups based on navigation patterns of clients. A marketing manager seeks to understand how clients use web pages so as to find the list of their attractive products. He/she also seeks to find natural groups in regard to navigation pattern of clients based on which targeted partitioning of markets can be planned and implemented.

In the present study, sequential clustering algorithm is used to analyze the sequence of clients' web navigation. These sequences are organized based on similarities of natural groups. This algorithm is designed and implemented in a way that it can analyze sets of records with sequential data and group them into homogeneous clusters based on the existing similarities of sequences. Through this algorithm, one can provide personal sales guidelines in association with profiles of each client. The sequential clustering algorithm is a combination of sequencing and clustering techniques.

The present study utilizes social network marketing (SNM) and web mining to develop a framework for identification of targeted markets in social networks. The utilized data include the information of clients of social networks in Iran based on their sequence and URL of different visited web pages. The development of model is followed by application of prediction functions to estimate the condition of sequential specifications. Then, the associated models were developed by web mining queries and combination of clustering algorithm with Markov chain. At last, the profile making of each cluster and discrimination analysis of clusters is discussed.

Review of Literature

Social Network Marketing (SNM)

The increasing significance of online social networks has had some visible influences upon companies. Marketing experts believe that social networks will generate significant changes in business of so-called traditional industries (Heidemann *et al.,* 2012). Social networks affect the client's behaviors from many perspectives such as information searching strategies, decision-making processes and decision to consume (Joo *et al.,* 2011). An online social network such as Facebook is an instance of a virtual community. Within these communities, information sharing is done in an easy and quick manner. So, a more significant role for social networks in effective marketing of social networks is expected. For instance, Casteleyn *et al.,* (2009) pointed out that data of social networks can be regarded as a "crystal ball" which shows the future intentions and acts of consumers. Jakson *et al.,* (2011) in a study called "*An Overview of Social Networks and Economic Applications,*" provide a summary of studies on social networks and their role in forming the economic behaviors and outputs. This study includes discussions on experimental and theoretical analyses of the role of social networks in different markets and transactions, learning domain and its expansion as well as online games. It also includes a background of features and measurements of social networks, models of network information, models of statistical analysis of social networks and social monitoring. Jothi *et al.,* (2011) in a paper called "*Analysis of Social Networking Sites*" did a case study on the impacts of communication strategies of individuals so as to discuss the products and services which create significant competition among different market brands. They found out the virtual social networks are full of potential clients that are young forces with a lot of time to use the internet and such networks.

Bernoff and Li (2008) found out that how OSNs can be used as new channels and ways of more effective and efficient marketing measures. Saravankumar and Suganthalakshmi (2012) presented a comprehensive discussion on popularity of social networks and marketing advantages of these media. They stated that social network marketing is a way to promote a website, brand or business through establishing communications with clients within social media channels. Kumar *et al.,* (2009) discussed social networks from different perspectives along with features which they provide for users. Qiao (2008) studied the positive and negative aspects of e-business in social networks and stated that social networks have significantly modified e-business. He also believe that the limitations of this area can be successfully resolved. Hartline *et al.,* (2008) examined different types of marketing strategies in social networks, smart implementations of sales strategies and theories on viral marketing.

Web Mining: Web mining is application of data extraction to discover and determine the latent knowledge of databases. So, web mining can be regarded as a significant tool for studying web recognition (Yu *et al*., 2009). There are numerous studies on this subject. For instance, Zhang and Segall (2008) studied the technology of web mining. Compton (2006) referred to the necessity of using modern technologies such as web mining. Sumathi and Sivanandam (2006) wrote a book on web mining and its applications which includes a chapter on using web mining for customer value and customer relationship management (CRM). Jian-Guo *et al.,* (2003) discussed the problems of web mining in e-business. Mobasher (2006) provided more detail on the issue that actual web mining consists of different resources and types of data such as reports of web server operations. Chen *et al.,* (2001) discussed the application of web mining to obtain web image. In their study, data mining of web pages to explain some images and posts constitutes one of the key ideas behind their discussions. Charen and Ross (2007) studied text mining in a retailer company which consisted of transformation of unstructured data to structured data. They enabled the integration of mined text through smart business tools. Goel (2006) introduced a comprehensive discussion of data mining in e-business domain and used Amazon Website as an instance. Ross (2007) discussed the creation of client-oriented datacenter and found out that lack of client-specific data, low quality of data and lack of trust in data are the main three setbacks in this regard. That study showed that individual and heterogeneous data cannot be used for data analysis. In regard to different analytical resources and objectives, one can categorize web mining into three types of web usage mining, web content mining, and web structure mining (Cooley *et al.,* 1997). Yu *et al.,* (2009) used web usage mining to analyze the way a website can be used and how orientation and direction behaviors of website users are. The main data resources for this technique are click stream data of servers (i.e. log files). However, client-side data (i.e. log files and cookies of client) can sometimes be used too. Web content mining is used for discovering and extraction of useful information from contents such as text, multimedia data and metadata. Web structure mining is used to analyze links and structure of a website and it applies the theory of graph to attain this objective. Recently, studies on combination of web mining techniques have been done to attain more optimal results. In the present study, web usage mining is applied.

MATERIALS AND METHODS

Methodology

The first step of present study was collection of data from social network websites and then, pre-processing them. The last step refers to modeling web mining technique. The framework of present study generally consists of 3 steps.

The first step is data collection which includes clients' data along with sequence and URL addresses of different visited pages of a social network. The second step is pre-processing data during which data are collected in a place and missing and irrelevant data are excluded. The third step is modelling in which the intended algorithms and essence of data are considered during identification of different data orders and latent data knowledge is codified in a distinctive format.

Data Collection

The used data includes clients' information along with their sequence and URL of different visited pages of digital media in social network. The targeted population consists of one of the most popular social network website in Iran "Cloob" (www.cloob.com).

There were two reasons behind selection of this statistical population. First, the industry of trading digital tools in Iran is highly competitive and there are different online stores. Second, the shared discussions and information on this industry and its products are significant and shared in social network websites. Our dataset was collected during May 2014. The samples were selected through simple random sampling and sample population consists of 16527 users.

Pre-Processing: The most significant step in the process of knowledge mining is data preparation. The objective of this step is to supply sufficient input for modelling step. In pre-processing step, the unprocessed data of all data sources are extracted and then, they undergo initial pre-processing in an independent stage. The output of data preparation step is pre-processed data from which modeling can be done.

In this step, the data distributed among different sources are collected in a site and a central database is created. The reason of collecting these data is lack of data concentration in a definite site. In addition, the data of different sections might be stored in different file formats.

The second step in data preparation step is pre-processing mined data. The most significant objective of this step is to resolve different potential problems of the data. In this stage, the excess data of clients are excluded and entry of missing and irrelevant data is prohibited. Table (1) and (2) are data sources of developing the intended model (Han and Kamber, 2006).

Table 1: Profile-Maker with Client Data

Customer Guide	Daily Online Time	Night Online Time	Browser Type	Comment Number	Chat Number	Geological Location
327	20	1	IE	0	0	Tehran
1965	68	2	MF	0	0	Kerman
266	73	9	IE	2	0	Isfahan
6955	31	43	MF	0	0	Tehran
2944	40	28	IE	0	0	Isfahan
7237	53	48	IE	0	0	Mashhad
429	75	7	IE	0	1	Mashhad
11038	26	33	GC	0	0	Tehran
1903	52	12	GC	1	0	Tehran

Table 2: Maintenance of Sequential Status of Clicked URLs for Each Client

Customer Guide	URL Category	Sequence ID
1965	Printer	1
1965	Mobile	2
1965	RAM	3
1965	Tablet	4
266	Laptops	1
266	Laptops	2
6955	Mobile	1
6955	Mobile	2
6955	Mobile	3

Modelling Step: In this step, the intended algorithms and essence of data are considered during identification of different data orders and storage of latent data knowledge in desired format. To model, the modelling methods should be properly known so as to use the right method in the right conditions.

Introduction of Applied Algorithm

Introduction of Sequential Clustering Algorithm: The sequential clustering algorithm is a combination of sequential and clustering techniques. This algorithm is designed and implemented in a way to analyze cases with sequential data and group the cases into homogeneous clusters based on similarity of sequences. Sequence is a sequel of discrete events so that their numbers in a sequence is limited and bounded.

For instance, if the purchasing order of a product is insignificant, the business problem for analysis of market basket is in domain of web mining with association rules but if the purchasing order of products is significant, the problem is associated with sequence of web mining (Ferreira *et al.*, 2008; Li and Cheung, 2012).

Principles and Concepts of Sequential Clustering Algorithm: This algorithm is obtained by combining two techniques of clustering and sequential analysis. The sequential analysis is done by Markov chain model. Markov chain is a sequel in which the next variable is defined by the present variable. In other words, Markov chain is a randomized process without memory. This means that the conditional probability distribution of the next state is dependent on present state and independent of previous states. Figure (1) shows an instance of Markov chain for sequence of DNA. This chain includes a set of states and a matrix of transition probabilities. The changes of system states are called “Transition” and the probability assigned to these change of states are called “Transition Probability”.

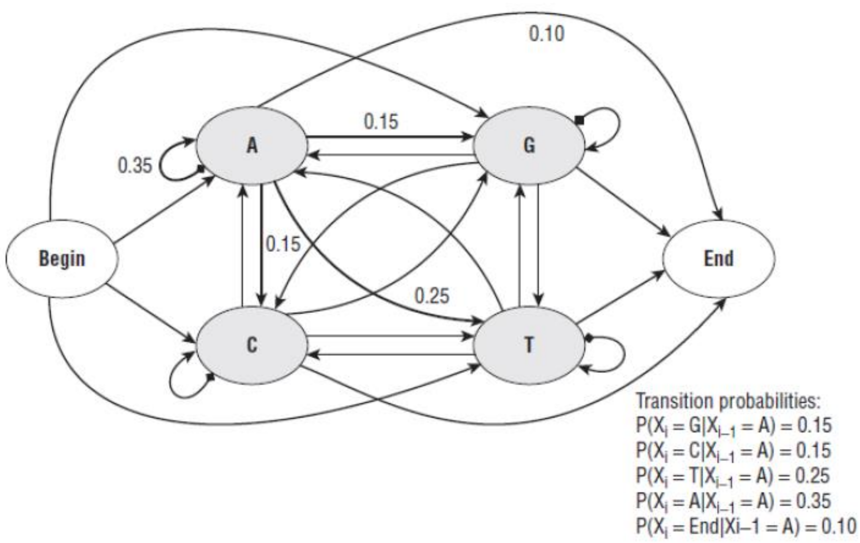


Figure 1: Markov Chain (SQL Server Books Online, 2012)

The expression of conditional probability $P(x_i=G|x_{i-1}=A)=0.15$states that the probability that the next state is G equals 0.15 if the present state is A (Online SQL Server Books, 2012).

Order of Markov Chain: One of the significant characteristics of Markov chain is order. This characteristic states that the conditional probability distribution of the next state depends upon the present system state. In Markov chain, the n order means that the probability of a state depends upon n previous states.

The most usual Markov chain is of first-order in which the probability of each Xi state depends upon Xi-1 state (i.e. previous state). Therefore, models with higher orders can be developed with n memory of previous state (Hahsler and Dunham, 2010).

The Markov chain of n orders for k states represents a first-order Markov chain for k^n state(s). For instance, second-order Markov chain for DNA model is represented as a first-order Markov chain with the following state.

{AA,AC,AG,AT,CA,CC,CG,CT,GA,GC,GG,GT,TA,TC,TG,TT}

The total number of states equals 4^2. Higher order of Markov model is associated with higher memory and time for processing. Based on the Markov chain, for L sequences of random variables x {x1, x2, x3,…,xL}, the probability of each sequence is calculated from the following equation.

P(x) = P(xL.xL-1,…,x1)

= P(xL|xL-1,…,x1) P (xL-1|xL-2,…,x1)…P(x1) (1)

In the first-order Markov chain, the probability of each Xi sequence only depends on Xi-1 which is calculable from the following equation.

P(x) = P(xL.xL-1,…,x1)

$= P(xL|xL\text{-}1)\ P(xL\text{-}1|xL\text{-}2)\ldots P(x2|x1)P(x1)$ (2)

State Transition Matrix: The Markov chain remembers the probability of transition in different states. Figure (2) shows first-order state transition matrix. Each box of the following table is associated with transition probability from one state to another. In this matrix, higher probabilities are coded with darker colors.

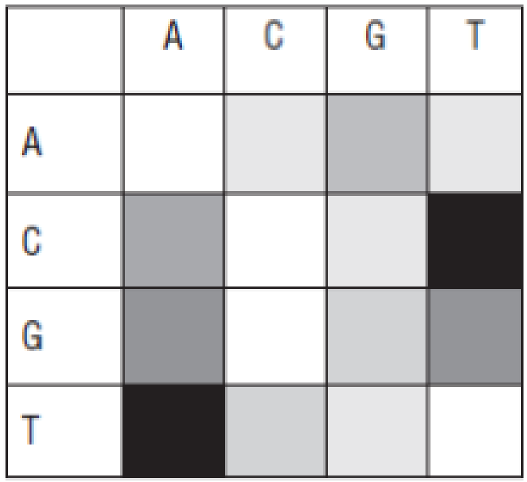


Figure 2: State Transition Matrix (SQL Server Books Online, 2008)

The state transition matrix for first-order Markov chain is a square M x M matrix in which M represents the number of states. In cases in which number of states is high, it is recommended to store the probabilities which exceed a predefined limit (SQL Server Books Online, 2008).

Clustering by Markov Chain: Learning in sequential clustering algorithm is based on its combination with Markov chain so that each mixed item is associated with a definite cluster. The data development method in this algorithm is done through the following steps.

- Each cluster is selected randomly by using inter-cluster probability distribution.
- Depending on the selected cluster, a sequence of Markov chain associated with the cluster is developed (each cluster is associated with different Markov chain).

Learning and processing model are done based on mixed model parameters. These parameters include mixed weights (distribution probability among clusters) and parameters of each Markov chain. Expectation-Maximization algorithm parameters are set based on maximization of probability of assigning data to clusters. The clustering algorithm processing is done through the following steps:

1- The model parameters are valued randomly.
2- Based on the present parameters of the model, each case is assigned to K cluster with definite probability (E steps).
3- The model is reevaluated based on the assigned weights to each case (M steps).
4- The model convergence is controlled and if the model does not converge, the step 2 is repeated.

In regard to the sequence item, the model parameters for each cluster have a state transition matrix in association with sequences. For sequence X, probability of placing and assigning it to cluster C can be calculated through the following equation.

$P(x|C) = P(xL|xL\text{-}1)P(xL\text{-}1|xL\text{-}2)\ldots P(x2|x1)P(x1)$ (3)

In the above equation, P(xj|xi) is the probability of transition from state *i* to state x in cluster C. Then, Bayes' rule is used to calculate the probabilities of membership of x in cluster C and P(C) represents the final probability of cluster C (i.e. weight of cluster C in total statistical population).

Using the Bayes' rule to calculate the probabilities of membership of C in cluster X, we have:

$P = (xi = G|xi\text{-}1 = A) = 0.15$

The data used in sequential clustering algorithm are stored in nested table. In the sequential nested table, a key sequence column is defined. This key column can contain any type of data as a sequential value (SQL Server Books Online, 2012; McLennan, 2009).

Development of Sequential Clustering Model

To develop the model, profiling table for customers was used which are denoted as "Case Table". The Customer-guid was used as the main key of the table and another transaction table (Click Path) was used as nested table with the following columns.

Customer Guid: As external key to table (Customer)

URL Category: Keeping the state of sequences and as a predictable column

Sequence ID: A sequential column in which the sequence of web clicks is stored.

The sequence in this model refers to a sequence of web clicks stored in *URLCATEGORY*. This sequence has a variable length because the behavior of customers in web pages and their length of stay at each page are different.

Cluster Prediction

The most significant stage in web mining processing is cluster prediction and it is often the final target of web mining projects. Model prediction stage uses the obtained patterns of trained model on a new set of data and predicts the values of predictable columns for each case. Prediction refers to discovering information regarding unknown cases through patterns developed from data histories (Hahsler and Dunham, 2010).

The sequential clustering algorithm supports prediction functions. To predict the cluster members, the function *Cluster ()* is used which returns a case for each member or defines the *Customer Guid* representing a cluster. In other words, it predicts the assignment of new members to clusters. Table (3) represents the results of query of predicting the number of cluster for each client.

Table 3: Predicting the Cluster ID for Each Client

Customer Guid	**$CLUSTER**
327	Cluster 8
1965	Cluster 5
266	Cluster 1
6955	Cluster 14
2944	Cluster 9
7237	Cluster 2
429	Cluster 13
11038	Cluster 6

Execution of Sequence Prediction

The sequence prediction algorithm also predicts the sequels of each sequence. In this regard, the table with sequential data is denoted as "Predictable". As a result, in web mining query the function "Predict Sequence" is used the output of which is a table. The resulting table includes a computed column called "$Sequence" which returns the sequence order. In this table, the next sequence is predicted for each client. In other words, if a client visits three pages with definite sequences which are introduced as input to the query, its sequence and URL title are predicted. Table (4) shows the results of executing prediction query for the next two steps in which a client visits the pages associated with Headphone, Mobile Cover, and Micro-SD Cable.

Table 4: Three-Stage Prediction after Visiting Pages by Client

Sequences	
$Sequence	**URL Category**
1	Mp3 Player
2	Mobile Cable

Using Histogram to Predict Sequences

During the prediction of sequences, the algorithm arranges all of the next steps based on their probabilities in the previous state and then, it returns the highest probability. For instance, if a user is in the webpage "Headphone", for the next step he might visit page "Mobile" with probability of 0.5 or go to another page such as "Tablet" with probability of 0.28. The *Predict Sequence* function always returns the highest probability of visiting the next page which in the present instance, it is page "Mobile".

Mining Unusual Sequence Clustering Patterns

A part of sequential algorithm can also be used for another interesting scenario. This algorithm can be used for identification of unusual sequential patterns and clients who cannot find their favorite webpage. It can be used for analysis of network sequences such as viruses. Using *Predict Case Likelihood* function, one can estimate the probability of closeness of a case as an unusual sequence. If the result is close to zero, the case is unusual (SQL Server Books Online, 2008; MacLennan *et al.*, 2009).

RESULTS AND DISCUSSION

Results

After developing the model and processing it, one can observe and analyze the content of the model. The content of the model consists of 5 discrete parts. These parts are diagram of clusters, profilers of each cluster, characteristics of each cluster, discrimination analysis of clusters, analysis of transition from one URL state to another or sequence transition.

Analysis of Clusters' Diagram

A cluster refers to a set of data objects which are similar to each other in some respects. The cluster analysis determines the number of clusters in data. A proper method of clustering defines clusters with high quality. This means that clustering is done in a way that reduces the similarity between clusters but increases the intra-cluster similarity. In other words, the cluster members are very similar to each other rather than to members of other clusters (Li and Cheung, 2012).

Figure (3) shows the diagram of clusters with their arrangement based on existing similarities. In this diagram, the arrangement of clusters is determined based on the connections. The similar clusters are those with same distribution probabilities which are paced close to each other in diagram. The clusters 3, 9 and 12 are the ones whose distribution probabilities are close to each other.

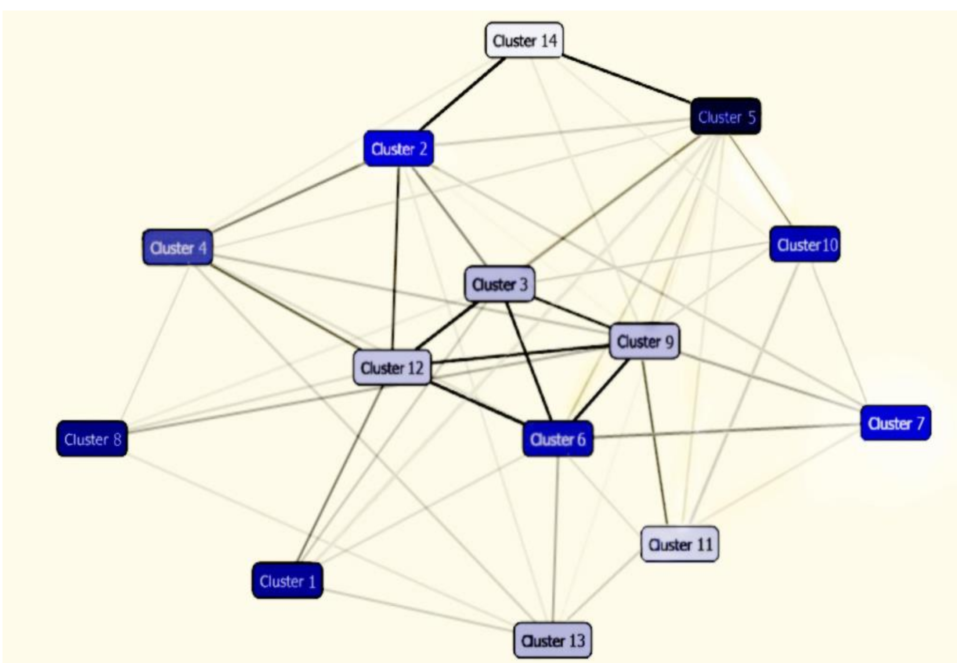


Figure 3: Graphic Diagram of Clusters' Arrangement

The darkness degree of color of each cluster represents its size (i.e. number of cases in each cluster). In this diagram, the cluster 5 has the highest and cluster 14 has the lowest distribution of statistical population. In this diagram, each cluster can be analyzed based on the characteristic which determines the sequence of URL Category and in regard to its states. For instance, the cluster 10 has the highest statistical probability (67%) in webpages the titles of which are Laptop.

Cluster Profilers: Sometimes, the objective of web mining is merely description of what is currently happening in a complicated database. The result of profiling increases our understanding of the individuals, products or processes which generate data in development stage. A good description of behavior is accompanied by a proper description of the time when suitable explanation is expected. Figure (4) represents the descriptive profiling of clusters.

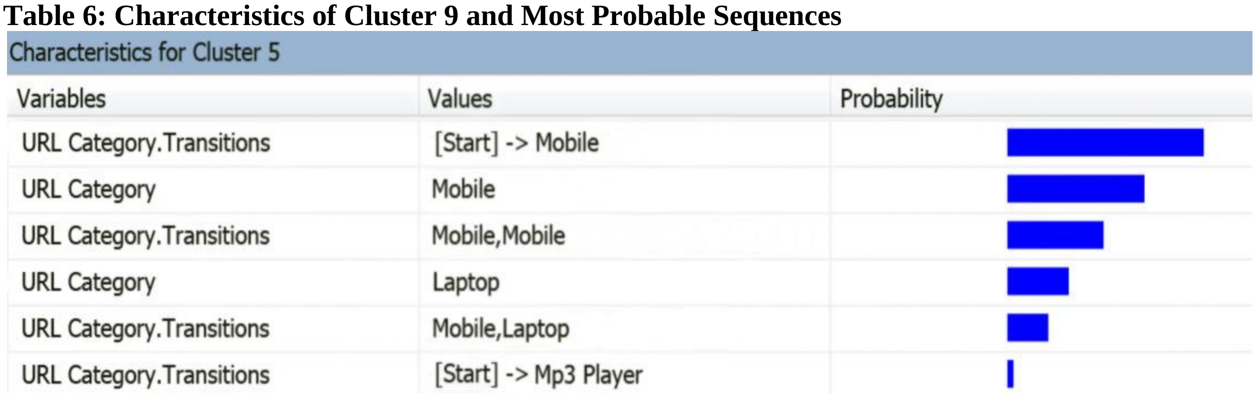
Figure 4: Descriptive Profiling of Clusters

Table 5: Probability Distribution of Clicks in Cluster 10

Mining Legend		
Color	Meaning	Distribution
	Mobiles	0.005
	Loptops	0.674
	Tablets	0.003
	Cameras	0.004
	Printers	0.003
	Monitors	0.004
	Desktop Comp...	0.002
	Scanners	0.002
	Loptop Bags	0.006
	Keyboards	0.002

Each cluster is associated with references of users and the pages they visited. In this analysis, each column represents a cluster and each row denotes a characteristic. The row *URL Category* represents the sequential characteristic and each box of the row consists of a histogram of sequences of each cluster. The second row denotes the distribution probability with discriminating URLs of each cluster. In the following figure, different characteristics are shown with a color and thickness of each colored band represents the percentage of that characteristic.
The above table shows the information of cluster 10. This information includes the sequence of clicked URLs in each cluster and probability distribution of each characteristic. For instance, one can determine the highest probability distribution of clicks per each cluster. Therefore, the list of provided products for each cluster and the sequence of clicked URLs for each case (i.e. client) can be observed. In this table, the probability distribution of 9.674 is shown for Laptops webpage of cluster 10. So, one can conclude that most of the probability distribution of the cases are in the "Laptops" webpage of cluster 10. For other clusters, the same method is used.
Characteristics of Clusters: In this section, the characteristics of each cluster are extracted. Each row represents the frequency (probability) of each characteristic and its value in the cluster. Each sequence (including initial and final events) consists of a discrete value for sequence characteristic. The list of values is stored based on frequency of observations. For instance, cluster (6) is the most probable characteristic of [Start]-> Monitors.
This means that most of those visiting the website start their web navigation with seeing the pages on monitors. Table (6) represents the characteristics of cluster (5) as a graphic histogram of highest probability of visiting the website along with sequence of transitions (i.e. clicked URLs). For instance, one could see that cluster 5 has the highest probability of visits of individuals to mobile webpage and this shows that most of the people start this web navigation cluster with watching mobile pages. Table (7) includes the values of probabilities.

Table 6: Characteristics of Cluster 9 and Most Probable Sequences



Characteristics for Cluster 5

Variables	Values	Probability
URL Category.Transitions	[Start] -> Mobile	
URL Category	Mobile	
URL Category.Transitions	Mobile,Mobile	
URL Category	Laptop	
URL Category.Transitions	Mobile,Laptop	
URL Category.Transitions	[Start] -> Mp3 Player	

Table 7: Most Probable Clicked Sequences at Cluster 5

Variables	Values	Probability
URL Category. Transitions	[Start] -> Monitors	**%15.712**
URL Category	Keyboards	**%7.431**
URL Category	Desktop Computers	**%6.543**

Discrimination Analysis: Discrimination analysis is used to compare the characteristics of the two selected clusters. Table (8) shows the discrimination analysis of cluster (2) and (6). There is a significant difference between web-navigation patterns of clients of the two clusters so that in cluster (2), the clients end their web-navigation in section of "Scanner" while in cluster "6" the clients end their web-navigation in the section of "Speakers". In cluster (6), the clients are interested in visiting pages on "External Hard Drives" and "Mouse" while in cluster (2), they are interest in web pages on "Flash" and "Leans".

Table 8: Discrimination Analysis of Clusters based on Sequence of Clicks

Discrimination scores for Cluster 6 and Cluster 2

Variables	Values	Favors Cluster 6	Favors Cluster 2
URL Category.Transitions	Scanner-> [End]		
URL Category.Transitions	Cameras-> Printers		
URL Category.Transitions	Extrenal Hard Drives-> Mouse		
URL Category.Transitions	Leans-> Flash		
URL Category.Transitions	Camera Bags-> Tripod		
URL Category.Transitions	Speakers-> [End]		
URL Category.Transitions	Mobile Covers-> Hands faree		
URL Category.Transitions	Monitors-> Webcam		

Analysis of States of Sequences: This analysis is done to examine the sequence navigation pattern of each cluster. Each node shows the state of a sequence and the arrows represent the transition from one state to another among the sequences. These arrows have definite weight and direction. The weights represent the probability of transition from one state to another state of sequences. As shown in figure (5), the clients in cluster (6) are more interested in keyboards, monitor and desktop computers as their dark colors show their statistical density. There is a link from section of "Monitors" to that of "Keyboards". Of the clients visiting the webpage of keyboards, 50 percent also visit the pages on "Desktop Computers".

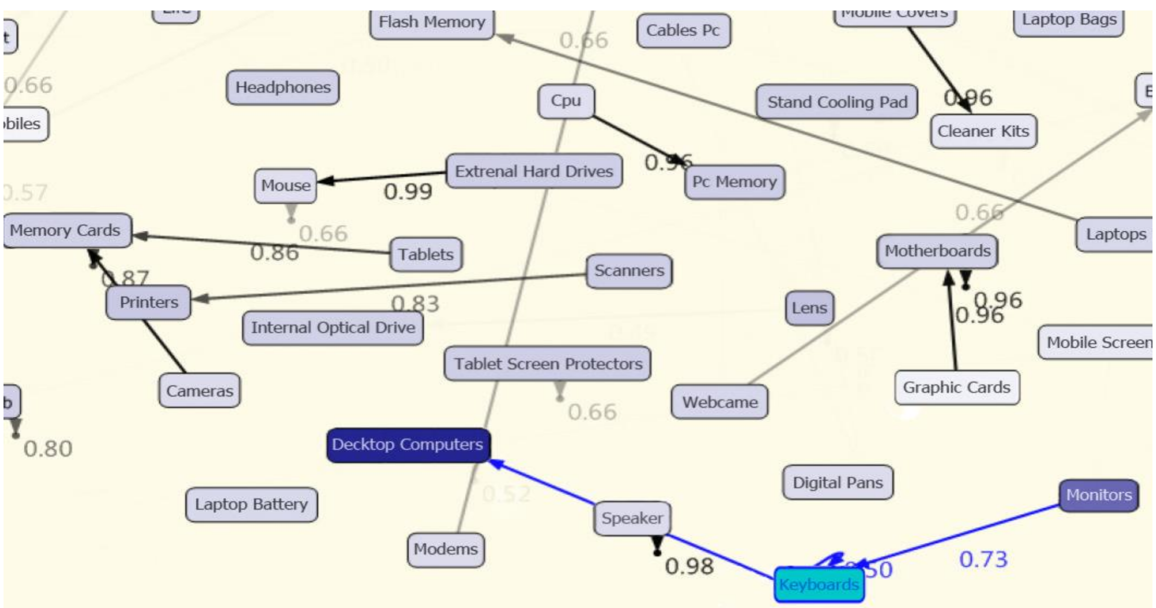


Figure 5: Analysis of State Transition of Sequences

Conclusion

The expansion of internet in the society has been accompanied by increasing influence of social networks among the public and higher number users. Therefore, the issue of business and its associated procedures in social networks draw more attention to themselves. The competition among producers and service providers led to development of distinctive types of e-marketing. To survive in existing competitive market, the companies need to know the current behaviors of clients and predict their future behaviors. Precise understanding and prediction of clients' behaviors enable the companies to keep their clients, improve their marketing methods and develop their relationships with clients. In the present study, the segmentation of clients of online stores within social networks was done through combination of their demographic characteristics and their sequence and patterns of web navigation were determined through sequence clustering algorithm. After developing the model, different analysis were done the results of which are discussed in the following.

- Cluster Diagram Analysis: Initial segmentation of clients and visual determination of similarities of each cluster.
- Descriptive Profiling Analysis of Each Cluster: Extraction of characteristics of each cluster with the list of provided products based on sequence of visits to webpages.
- Discrimination Analysis: Determination of characteristics of clusters with maximum distance.
- State Transition Analysis: Extraction of sequential pattern of visits to webpages based on sequences of state transition, percentage of visiting each webpage
- Prediction Querying: Determination of probability of assigning new clients to new clusters and provision of personal guidelines of sales and marketing based on extracted descriptive profilers for each cluster, listing favorite products based on marketing pattern and strategy

A general review of results of present study provides one with a lot of information which opens the way for new studies and analyses. The results of present study help the companies to analyze the present and future conditions of clients and improve marketing in virtual communities.

REFERENCES

Bernoff J and Li C (2008). Harnessing the power of the oh-so-social web. *MIT Sloan Management Review* **49**(3) 36–42.

Casteleyn J, Mottart J and Kris Rutten (2009). How to use Facebook in your market research. *International Journal of Market Research* **51**(4) 439-447.

Charen S and Ross D (2007). Mining text in a retail enterprise, *B-Eye Network*, August 21. Available: www.b-eye-network.com/view/5810.

Chen Z, Wenyin L, Zhang F, Li M and Hang H (2001). Web mining for web image retrieval. *Journal of the American Society of Information Science* **52**(10) 831-9, Available: www.citeseer.ist.psu.edu/chen01web.html.

Compton J (2006). What your customer data should be telling you... and how to be a better listener. *1to1 Magazine*, Available: www.intelligentresults.com/news/articleArchive/articleIndex/1to1article_dec06.html

Cooley R, Mobasher B and Srivastave J (1997). Web mining: Information and pattern discovery, World Wide Web. *Proceedings of the 9th IEEE International Conference on Tool with Artificial Intelligence, Newport Beach, CA* 558–567.

Ferreira D, Zacarias M, Malheiros M and Ferreira P (2008). Approaching process mining with sequence clustering: Experiments and findings. In: *Business Process Management.*

Goel S (2006). Data mining in e-business. *Impact Solutions*, Waltham, MA, Available: www.impact-sol.com/datamining.html.

Hahsler M and Dunham MH (2010). Extensible Markov model for data stream clustering. *Journal of Statistical Software*.

Han J and Kamber M (2006). *Data Mining: Concepts and Techniques*, 2nd edition (Morgan Kaufmann Publishers) ISBN 10: 1-55860-901-6.

Hartline J, Mirrokni VS and Sundararajan M (2008). Optimal marketing strategies over social networks, Refereed Track: *Internet Monetization-Recommendation & Security*.

Heidemann J, Klier M and Probst F (2012). Online social networks: A survey of a global phenomenon, *Computer Networks*, Available: http://dx.doi.org/10.1016/j.comnet.08.009.

Jakson MO (2011). An overview of social networks and economic applications. In: *Handbook of Social Economics*, edited by Benhabib J, Bisin A and Jackson MO (Elsevier Press).

Jian-Guo L, Zheng-Hong H and Wei-Ping W (2003). Web mining for electronic business application. *Proceedings of the Fourth International Conference on Parallel and Distributed Computing, Applications* *and Technologies, PDCA Tapos, August 27-29* 872-6, Available: www.ieeexplore.ieee.org/iel5/8749/27719/01236436.pdf.

Joo YH, Kim Y and Yang SJ (2011). Valuing customers for social network services. *Journal of Business Research*, doi:10.1016/j.jbusres.2011.06.030.

Jothi PS, Neelamalar M and Shakthi PS (2011). Analysis of social networking sites: A study on effective communication strategy in developing brand communication. *Journal of Media and Communication Studies* **3**(7) 234-242.

Kotler Ph, Armstrong G, Saunders J and Wong V (2001). *Principles of Marketing*, third edition (Prentice Hall, UK).

Kumar SD, Karkee P and Majumder K (2009). Managing social networking in web based society: A Need-based approach of present generation online users. *7th International CALIBER*.

Li X and Cheung JCK (2012). Sequence clustering and labeling for unsupervised query intent discovery. *WSDM '12 Proceedings of the Fifth ACM International Conference on Web Search and Data Mining* 383-392.

MacLennan J, Tang Z and Crivat B (2009). *Data Mining with Microsoft SQL Server 2008* ch10 (Wiley Publishing, Inc).

Mobasher B (2006). Web data mining. In: *Web Data Mining: Exploring Hyperlinks, Contents and Usage Data*, Chapter 12, edited by Liu B (Springer, Berlin).

Qiao Y (2008). Social networks and E-commerce. *TKK-110.5190 Seminar on Internetworking*.

Ross D (2007). Building a customer-centric data warehouse. *Business Intelligence Network*, April 3, Available: www.b-eye-network.com/view/4222.

Saravanakumar M and SuganthaLakshmi T (2012). Social media marketing. *Life Science Journal*.

SQL Server 2008 Books Online. (2012). Exploring the sequence clustering model, Microsoft Corporation.

SQL Server 2012 Books Online (2012). Mining Model Content for Sequence Clustering Models (Analysis Services - Data Mining). Available: http://technet.microsoft.com/en-us/library/cc645747.

Sumathi S and Sivanandam SN (2006). Data mining in customer value and customer relationship management, introduction to data mining and its applications. *Studies in Computational Intelligence* **29** 321-86.

Yu G, Xia Ch and Guo X (2009). Research on web data mining and its application in electronic commerce. *International Conference on Web Intelligence and Intelligent Agent Technology, IEEE*.

Zhang Q and Segall RS (2008). Web mining: a survey of current research, techniques, and software. *International Journal of Information Technology and Decision Making* **7**(4).