

Personalized Marketing in Facebook using Hidden Markov Model

Gayana Fernando
Management and Science University
Malaysia

Md Gapar Md Johar
Management and Science University
Malaysia

ABSTRACT

In the current dynamic business world, each business is compelled to keep up with the state-of-the-art technologies to survive in the ever competitive market. With the dawn of the digital era, marketers were opened up to a brand new path of marketing strategies that allows them to not only manage their existing consumers but also to capture the world wide attention of potential consumers at a trivial cost. Social networks are now pronounced to be the next generation of consumer-centric interactions. This is proven further by the data released by popular social networks like Facebook regarding their consumer interactions. Big data analysis is another innovative technology concept used by businesses to recognize consumer buying patterns. Big data analysis has made path to web data mining. With the growth of social networks its profile holders have become the creators and distributors of business data. Due to this, social networks hold terabytes of raw data. With these massive advantages in line, web mining techniques and theory requires to be researched thoroughly to uncover the potential patterns and user access behaviors of these large amounts of online data. The major dilemma faced by researches in this field is the dynamic behavior of online data. To add to this dilemma the data is also quite unstructured. With these difficulties, the need to discover an efficient approach to mine unorganized and dynamic data in social networks is of utmost importance. This research aims to fill this research gap by introducing a framework that can be used to mine profile user data extracted from Facebook and use the knowledge in personalized marketing. This research paper explains framework build based on Hidden Markov Model to predict the next movement to handle the dynamic behavior of web data.

Keywords

Social Networks, Web Data Mining, Personalized Marketing, Hidden Markov Model

1. INTRODUCTION

Social Networks (SNS) are defined as “web-based services that allow individuals to construct a public or semi-public profile within a bounded system articulate a list of other users with whom they share a connection, and view and traverse their list of connections and those made by others within the system” [1]. SNS facilitates the virtual collaboration among the members. Word of mouth marketing is a powerful and cost effective marketing technique where business use from early days. Since social networks create a communication hub for the people to share thoughts, opinions and ideas, businesses can use this media in word of mouth marketing. Social media is pronounced as the next generation of consumer-centric interactions. Social network sites like Facebook claims that over one billion people like and comment an average of 3.2 billion times every day.

Businesses are now using social networking sites as opportunities for globalization. Internet-based networks such as

these encourage businesses to take advantage of marketing solutions and other services the social media has to offer.

Facebook web site was introduced in February 2004, by Mark Zuckerberg along with his collage mates in Harvard. The initial objective of the website was to enhance collaboration among Harvard students, later it has been expanded to be used in worldwide [2]. The users should register in Facebook site in order to create their own profiles. They are allowed to add other members to their profiles as “Friends”. Facebook allows the user to set privacy settings, so user can maintain friend lists with different privacy options. Basically user would be able to control “who would do what” in his profile. Facebook business allows any business to make use of the site in advertising and promotion. It allows the businessmen to create a Facebook Page. So the company can have one to one discussions with the customers via messaging. The customers can like the page, read the posts and share them with their friends. The page can be used to target large groups of people globally and frequently.

Personalization is defined as the customization of some or all the elements of the marketing mix to an individual level. It is intimately connected with the idea of interactive marketing. The personalization is automated by the marketer on behalf of the customer, as opposed to customization that a customer requests on her own behalf [3].

There is large number of data content available on World Wide Web. Analyzing and mining this data new knowledge can be produced. The problem is online data is unstructured and changes within seconds. Using traditional web crawler methods to extract the data will not be efficient. The research introduces a new approach based on Hidden Markov model in order to extract and mine online data. This specifically focus on social network sites, but the concept it commence can be used in any other public network. The framework proposed would be able to filter and extract public information from social network profile pages. A procedure is being introduced to process dynamic and unstructured web data and mine to identify patterns and new knowledge. This knowledge can be used by business parties to persist profitable and efficient marketing strategies in the company. The researcher has tested the system prototype with Facebook.com

The research paper highlights the previous work on the specific domain, the methodology followed and the framework of the proposed system.

2. LITERATURE REVIEW

Since the usage of social network sites, blogs, knowledge sharing sites and email systems are being widely used, Internet holds large amount of data.

Rennie et.al in 2011 mentioned that new data will come in great part from the social web. Already, Facebook users share more than 30 billion pieces of content web links, news stories, blog posts, photos each month. Twitter users generate more than 155

million tweets per day (up from 55 million one year ago). These companies are establishing themselves as platforms for data aggregation, granting other companies access to the results through APIs, or application programming interfaces. For example, Facebook Connect, this allows sites to access public data from Facebook user. The growth of these interconnections, of course, spawns still more online data, in a widening spiral [4].

H.Ma et.al states in “Mining Social Networks Using Heat Diffusion Processes for Marketing Candidates Selection” that millions of users participate in these social networks, and act as different roles. All of these social networks provide valuable information for decision making in marketing campaigns, especially in marketing of new products from start-up businesses [5]. Based on the Empirical analysis done by Tucker, it is stated that when the private information of the profile holders is given to public is beneficial in advertising supported media and advertisers on social network sites [6].

Many applications have been introduced to conduct social network analysis. Graph Characterization Toolkit [7] TweetHood [8] , Meerkat [9] , NetDriller [10] , HiTS/ISAC Social Network Analysis Tool [11] D-Dupe [12] and X-RIME, a cloud-based library for large scale social network analysis [13]. In today’s era of technology enhancements the practices of data mining and web content extraction is a significant and growing field. Centrifuge Analytics system is a big data discovery technology that provides the power and flexibility to connect, visualize and collaborate without complex data integration and costly services [15]. Cuttlefish is a network workbench application that visualizes the networks with some of the best known layout algorithms. It allows detailed visualizations of the network data, interactive manipulation of the layout, graph edition and process visualization as well as different input methods and outputs in tex using Tikz and PSTricks [16].

A hidden Markov model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. An HMM can be considered as the simplest dynamic Bayesian network [17].HMM is a popular technique used in bioinformatics and character recognition. Da Silva and Ferreira in 2009 have conducted a research on applying HMM for process mining. They have successfully used HMM with sequence clustering for log tracing [18]. There are many new algorithms have been created based on HMM. Zaki et.al introduced a variable order hidden Markov model with state durations combining pattern mining and data modeling named VOGUE. This algorithm has been tested on web usage mining, intrusion detection and as a spell checker [19].

Markov chains are mathematical algorithms that undergo transitions from one state to another, among a finite or countable number of possible states. It is a random process where the next state depends only on the current state and not on the sequence of events that preceded it [20]. Markov models can be used in web mining to predict users’ next action. The Social network can be mapped to a where nodes will be users previous visits. So based on the node information by using Markov models users next visit can be predicted [21]. The social networks can be mapped as a Markov chain Crnovrsanin et.al introduces a new procedure for social network discovery based on the Markov centrality of a node [22]. Peng et.al have used the Hidden Markov model along with SVM classifier to predict sports, weather and social activities on Twitter messages [23].

3. METHODOLOGY

This experimental research aims to mine web data in order to achieve personalization in social network marketing. An

empirical study has been carried out to analyze the current trends in web data mining and existing data mining frameworks. The major limitations of the current studies are that they do not tackle the dynamic behavior of the web data. A new system is designed and developed to cover the loopholes of the existing systems. Object oriented approach is used to design and system and followed the prototype methodology in developing.

The profile pages of Facebook site would be analyzed mined in order to identify the patterns which can be effectively used in social network marketing strategies. The data gathering will be done among the friends network so it will not arise private and ethical issues. A customize web crawler is designed to gather the data. An algorithm has to be developed to filter and gather important facts by using data classifications. Hidden Markov Model will be used in data mining process to handle the dynamic behavior of the web data.

According to the literature the traditional process of personalization is illustrated in figure 3. The customer, customer data, customer profile and marketing output can be taken as objects. Interactions, Processing (analyses, data mining, differentiation, segmentation, targeting), customization and delivery are operations [24].

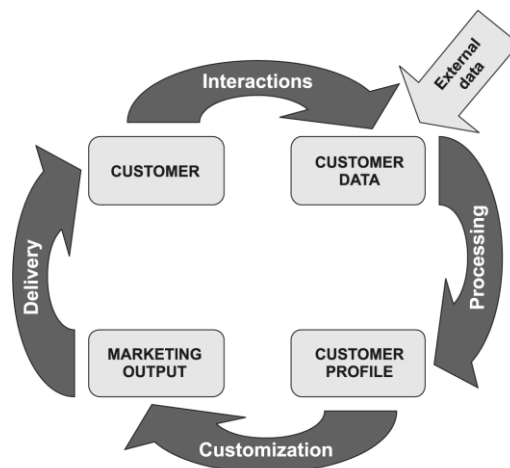


Fig. 3. Traditional Personalization Process [4]

The research uses a customized version of the above mentioned process with relevant to social network marketing. This research introduces the following framework to adopt traditional personalizing marketing techniques to modern social network marketing. New process is illustrated in figure 4.

Customers profile pages would be analyzed via social network analysis techniques. The information available on public would be collected and mined to analyze the patterns. The mining algorithm would be based on the HMM in order to tackle the dynamic aspect of the web data. Based on the results generated from the mining process, the customer profile would be identified. Those profiles can be used in promotional activities.

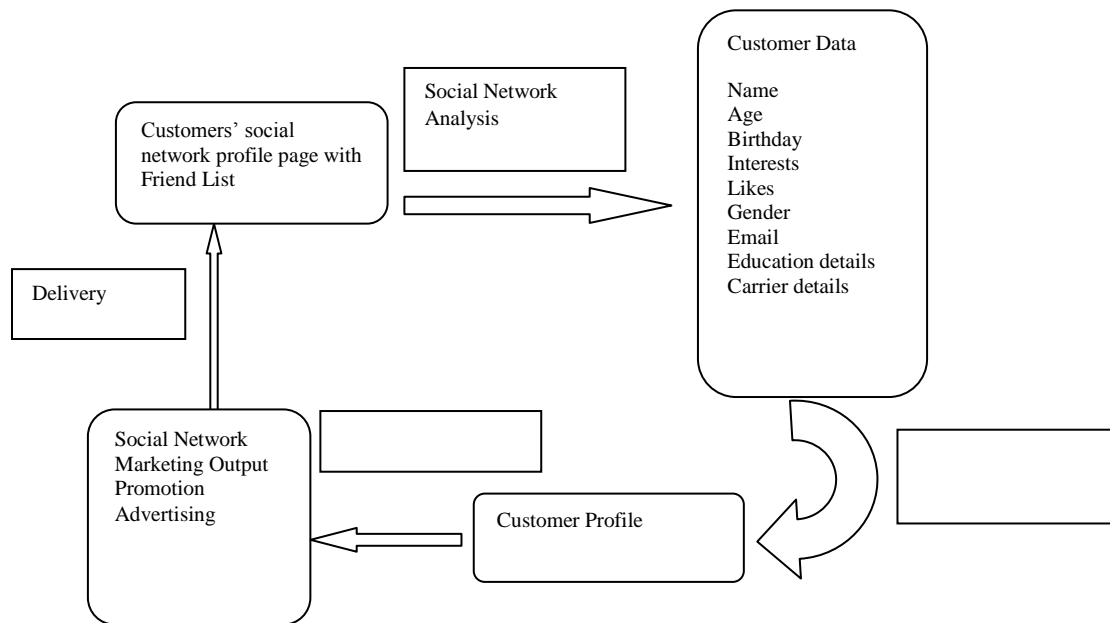


Fig 4. Modified Personalization Process

The system would consist of three major modules. The web crawler to extract the data from three social network sites, the data repository and the data mining component based on HMM model. The web crawler is the software component developed to extract the data from Facebook. The extracted details can be public information available in social network profile pages, for an example it can include birthdays, email, likes, interests, professional qualification etc. The crawler is capable of starting from a given profile page, and access the friends networks. Based on the access permission set by the profile holder, the data can be extracted. It will not violate privacy policies. These details

would be stored in a database. This process would be executed periodically. These data would be cleaned via a software script and would be inserted to the data warehouse. A data mining structure is created with a mining model, where sequence clustering based on HMM can be used as the algorithm. Since the system is developed in experimental level, researcher did not focus much on graphical user interfaces. The reporting service can be implemented to view the appropriate profile users for personalized marketing. Figure 5 illustrates the design of the proposed system

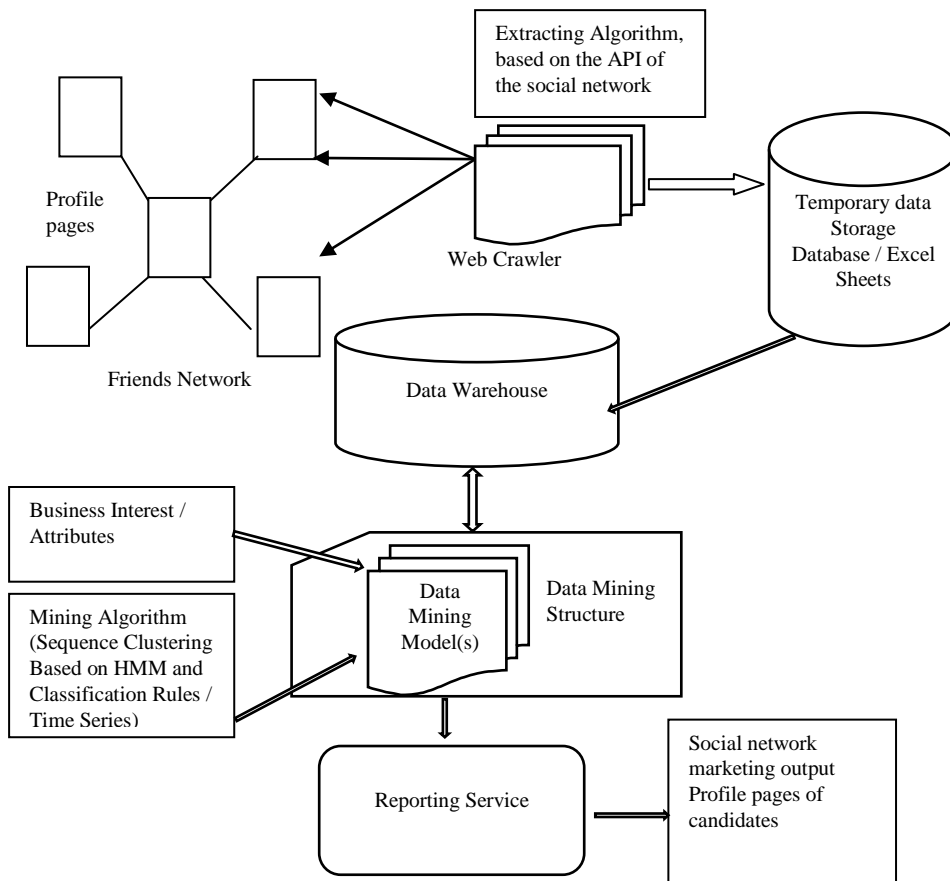


Fig. 5. System Design

4. RESULTS AND DISCUSSION

The research was carried out to find an optimal solution to handle the dynamic behavior of web data. According to the empirical study done on data mining techniques, it was found that statistical model like HMM can be used to handle temporal data. Sequence Clustering is based on Markov Chains that predict the future state of the attributes by looking at the current states. The proposed system adopts the clustering based on HMM to predict possible future states, the use classification rules to profile customer information.

First objective of the research was met by developing a web crawler algorithm that can be used to filter and extract public information from social network profile pages.

The pseudo code for the crawler algorithm would as follows :

```

Initialize the starting URL with profile owners URL
Execute authentication
    IF authentication false
        Exit
    IF authentication true
        Create a List l with URLs of friends profile pages
        Until l empty or page or time limit exhausted:
            Pop profile ID, i, from next index of A.
            If i authentication false
                Continue next index
            If i authentication true
                Read public data
            Parse A to obtain list of friends of friends N.
            Append N to the end of l
    
```

However the system is tested on prototype level only. The actual performance levels have not been tested. The target customer profiles generated by mining model can be used to customize the content of the advertisement or to figure out the date and the time it should be placed etc. The data was collected through the APIs of the network, so the data is somewhat structured. The statistical model HMM was used to predict all possible future states, so it might tackle the temporal behavior of the data. In order to use the HMM model in clustering there should be an attribute that has a sequence of occurring. So when analyzing the data collected via the profile pages, the sequence of liking pages, sequence of changing relationship status etc are identified. Each profile user may have a long sequence of liking pages but less sequence in changing relationship status. Figure 6 illustrates how HMM is mapped with page liked. The collected data was mined by sequence clustering.

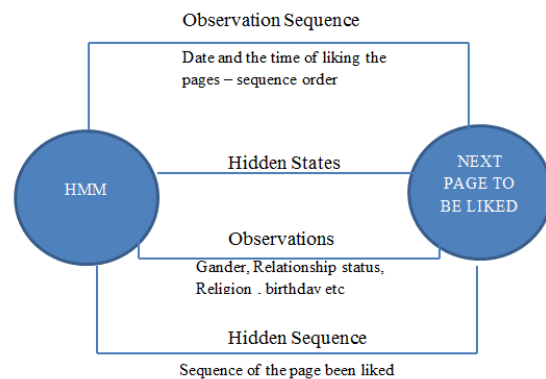


Fig.6 Facebook page like - states

It would be possible only if the users reveal their details public. Company can view the reports by using the interface given in figure 7. They can select the category they want to predict. Then the stored information is mined and the results will be displayed.

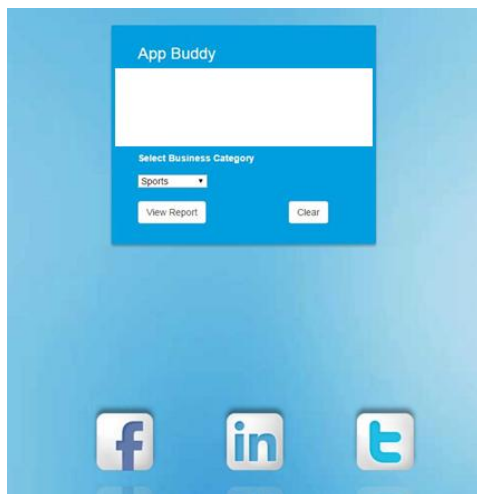


Fig. 7. Report Interface

The report would be displays as follows. The sample report given below highlights the possible Facebook profile users who would be interested in photography.

Photography

facebook id	name	gender	relationship status	religion	hometown
640980000	Sanadha Jayasekara	male	married	Catholic	Colombo Sri Lanka
640980100	Tharindri Rupasinghe	female	single	Buddhist	Kandy
640980737	Sajev Lucksman	male	married	Catholic	Colombo Sri Lanka
640980738	Prasadini Ariyaratne	female	single	Buddhist	Colombo Sri Lanka
640982020	Anna Amache	female	married	Catholic	Colombo Sri Lanka
645980000	Narmatha Rajenthiran	female	single	Hindu	Kotahena
670980000	Ayodya Piyaruwan	female	single	Buddhist	Colombo Sri Lanka
100000712428651	Hashan Jye	male	single	Buddhist	Gampaha
100000712428659	Dimal Bandara	male	single	Catholic	Colombo Sri Lanka
1000022327016	Prasad D	male	single	Catholic	Nenamho

Fig.8 . Sample Report

Customer profile generated by the mining model includes the data of the profile holders who might like company page in future. Those information can be used for personalization, when creating and placing advertisements. If the customer has given his/her email address, then the company can even email the personalized advertisement. So the predicted results can be effective used in personalized marking via social media.

5. LIMITATIONS

It is a constraint that the company should obtain application key from the social website in order to scrape information. Few limitations were identified from system perspective. The major limitation of the system would be the scalability. The social network data should be accessed via the API. Accessing the social network data without the APIs provided would not be ethical. The user of the system should have a basic knowledge on the social network platforms in order to configure the system. The next limitation is the company should have a page in the social network. The page should have at least 500 followers or likes. The model is tested using the profile data of those users. If there are less users or followers there would be less records of

the data warehouse. Then the accuracy of the results would decrease. The result accuracy depends on the truthfulness of the data published by the profile holders. If they have provided incorrect data, the predictions also would be inaccurate. If many profile holders are reluctant to provide information publically, the system might not be that useful.

The mining model uses the HMM. So the limitations of the HMM model would apply for the proposed system as well. Limited number of attributes can be found that having a sequence order from the attributes of profile holders. For an example Facebook data has sequence of liking pages, sequence of going to events. When the number of states increases in the model would increase when number of variable increase as well. This will make algorithm performance to decrease. The other limitation would be emission function and the transition function are independent and both depend only on the hidden state. System has not been tested on real world scenario due to the technical issues faced by the researcher. It has been tested on localhost in a virtual machine. The exact performance level of the commercial product cannot be determined. This is an experiment study only.

6. CONCLUSION

The research proposes a framework intends to mine social network data and use that knowledge in social network marketing. As discussed earlier there are few limitations. It is recommended to run the proposed system with a real business. In order to achieve the objective high-end server machine has to be used. A marketing research can be carried out by introducing the system to several businesses. The performance can be tracked. The researcher can check whether there is an impact of marketing the products to the target customers via social networks to the business profit. As an additional feature the web data can be merged with company database. By using a primary key like the email address, the data can be linked with each other. Interesting pattern can be mined by analyzing those data. To handle the dynamic behavior of the web data, it would be an advantage if the possible states are determined before inserting to the database. Since there are many HMM library files available a small application can be written to read the data from the crawler then predict the possible state and finally insert to the database. That component would be an artificial intelligent unit. The researcher believes that it would increase the accuracy level of the predictions.

The proposed system can be modified to add more social network sites and provides a general platform to collect data from the three social networks. A new research trend would be to implement a common platform for all social networks, so data can be gathered from any social network once the application ID is provided. It would make the system more scalable. This research test the applicability of using the hybrid approach of web content mining and web structure mining in the field of marketing. The same logic can be applied to any field where you need to identify the patterns of social network data. One example would be in social science researches, it can be used to user profiling. It can used to analyses life patterns and behaviors of users.

7. ACKNOWLEDGEMENTS

We sincerely thank Ms Lekhani Pathirana for the help and the guidance given. Further we would like to extend the gratitude to Professor Dr. Ali Khatibi all the encouragement given.

8. REFERENCES

- [1] D. Boyd and N. Ellison, "Social network sites: definition, history, and scholarship", *IEEE Engineering Management Review*, vol. 38, no. 3, pp. 16-31, 2010.
- [2] The Facebook effect: the inside story of the company that is connecting the world", *Choice Reviews Online*, vol. 48, no. 04, pp. 48-2179-48-2179, 2010.
- [3] A. Montgomery and M. Smith, "Prospects for Personalization on the Internet", *SSRN Electronic Journal*.
- [4] Rennie and G. Zorpette, "The social era of the web starts now", *IEEE Spectrum*, vol. 48, no. 6, pp. 30-33, 2011.
- [5] M. Hao, H. Yang, M. R. Lyu, and I. King. "Mining social networks using heat diffusion processes for marketing candidates selection." *Proceedings of the 17th ACM conference on Information and knowledge management*, pp. 233-242. ACM, 2008.
- [6] Tucker, E Catherine. "Social networks, personalized advertising, and privacy controls." *Journal of Marketing Research* 51, no. 5 (2014): 546-562.
- [7] D. Ediger, K.. Jiang ,J. Riedy D. A. Bader, C. Corley, R. Farber, W. N. Reynolds, " Massive Social Network Analysis: Mining Twitter for Social Good", *IEEE 39th International Conference on Parallel Processing*, pp 583-593, 2010.
- [8] S. Abrol, L. Khan, "TweetHood: Agglomerative Clustering on Fuzzy k- Closest Friends with Variable Depth for Location Mining", *IEEE International Conference on Social Computing / IEEE International Conference on Privacy, Security, Risk and Trust*, pp 153-160, 2010.
- [9] J. Chen, J. Fagnan, R. Goebel, R. Rabbany, F. Sangi, M.Takaffoli, E. Verbeek, O. Zaiane, "Meerkat: Community Mining with Dynamic Social Networks" , *IEEE International Conference on Data Mining Workshops*, pp 1377-1380, 2010.
- [10] N. Koochakzadeh, A. Sarraf , K. Kianmehr , J. Rokne, R. Alhaji, " NetDriller: A Powerful Social Network Analysis Tool", *11th IEEE International Conference on Data Mining Workshops*, pp 1235-1238, 2011.
- [11] H. Asadi, C. Martenson, P. Svenson, M. Skold,, "The HiTS/ISAC Social Network Analysis Tool", *IEEE European Intelligence and Security Informatics Conference*, pp 291-296, 2012.
- [12] M. Bilgic , L. Getoor . B. Shneiderman, "D-Dupe: An Interactive Tool for Entity Resolution in Social Networks", *IEEE Symposium on Visual Analytics Science and Technology*, pp 43-50 2006.
- [13] Xue, Wei, JuWei Shi, and Bo Yang. "X-rime: cloud-based large scale social network analysis." *Services Computing (SCC), 2010 IEEE International Conference on*. IEEE, 2010.
- [14] Wikipedia, 'Hidden Markov model', 2015. [Online]. Available: http://en.wikipedia.org/wiki/Hidden_Markov_model. [Accessed: 14- Apr- 2015].
- [15] Centrifugesystems.com, 2018. [Online]. Available: <http://centrifugesystems.com/>. [Accessed: 14- Feb- 2018].
- [16] "Cuttlefish - Network visualization and workbench", *Cuttlefish.sourceforge.net*, 2018. [Online]. Available: <http://cuttlefish.sourceforge.net/>. [Accessed: 14- Feb- 2018].
- [17] Da Silva, G. Aires, and D. R. Ferreira. "Applying hidden Markov models to process mining." *Sistemas e Tecnologias de Informação: Actas da 4ª Conferência Ibérica de Sistemas e Tecnologias de Informação, AISTI/FEUP/UPF*. 2009.
- [18] M. Zaki, C. Carothers and B. Szymanski, 'VOGUE', *ACM Transactions on Knowledge Discovery from Data*, vol. 4, no. 1, pp. 1-31, 2010.
- [19] Kemeny, John G., and James Laurie Snell. "Finite Markov Chains, Undergraduate Texts in Mathematics." (1976).
- [20] Eirinaki, Magdalini, and Michalis Vazirgiannis. "Web mining for web personalization." *ACM Transactions on Internet Technology (TOIT)* 3.1 (2003): pp 1-27.
- [21] Crnovrsanin, T., Muelder, C., Correa, C., & Ma, K. L. (2009, October). Proximity-based visualization of movement trace data. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on* (pp. 11-18). IEEE.
- [22] Chang, Peng, Mei Han, and Yihong Gong. "Extract highlights from baseball game video with hidden Markov models." *Image Processing. 2002. Proceedings. 2002 International Conference on*. Vol. 1. IEEE, 2002.
- [23] G. Fernando, M. Gapar and M. MdJohar, "Framework for Social Network Data Mining", *International Journal of Computer Applications*, vol. 116, no. 18, pp. 7-10, 2015.
- [24] B. Stone, "Advertising on Facebook Strikes Some as Off-Key", *Nytimes.com*, 2017. [Online]. Available: http://www.nytimes.com/2010/03/04/technology/04facebook.html?_r=0. [Accessed: 01- Jun- 2017].