Consumer Behavior Analysis In Modern Environment

Dieudonne Ouedraogo

Abstract

Social networks activities are parts of our lives, and we interact with others through those means more often.

As the number of users grows and technology improve, a new reality is taking place: even though most people identify themselves as unique, they tend to show preferences that could be clustered into groups.

As technology continues to develop and we are more willing to share our information on the social network, one may argue that we are becoming predictable than ever before.

We can defined groups or communities by labeling reference group as an instance where an individual is using perspectives or values as the basis for his or her current behavior. A reference group could guide the action in some situations.

Primary groups: family and close friends where there is a lot of interaction lead to considerable influence. Social media networks were considered secondary groups as they used to involve relatively weak ties and less frequent communication, but this is has begun to change as members are interacting more frequently.

Within groups either directly or indirectly more experienced members serve as experts and leaders and newer members seek advice and information.We can also argue that the collective behavior or properties influence members.

It is now indisputable that great insights could be extracted from social networks to drive business decisions and gain competitive advantage.

Recent studies in network science reveal the presence of well-defined structures in social networks; an example is the presence of homophily which shows individuals with similarity tending to connect more than with the dissimilar individuals. This paper examines the structural qualities of Social Networks towards the identification of trends in consumers behavior The goal of this work is to lend insight to the characterization of consumer behavior, particularly in the area of predictive analytic.

Keywords:

Consumer Behavior, Network Science,Data Mining,Trend Discover,Predictive Analytics,Personalization

Literature Review

Twitter has been extensively used on predicting trends in many research because of the relatively small size of its attributes. In a broader sense of social network being used to forecast events Achrekar et. al.[1] used Twitter to predict the trend of the flu virus. They successfully used auto-regression models on tweets to accurately predict the numbers published by the Center for Disease Control (CDC).While the CDC wait to collect actual cases to generate figures,their model could quickly predict outbreak and could be used to save lives.

(Iyengar et. al.) were able to predict the start and the end of a set of sports, weather and social activities using SVM classifier and hidden Markov Model on Twitter data.[2]

(Peng et al.) investigated re-tweets patterns using conditional random fields.They defined features as content influence,network influence and temporal decay factor .The results showed that re-tweet predictions could be substantially improved under social relationships compare to a baseline environment[3].

(Gloor, Nann, Schoder) used structural qualities to find betweenness centrality of actors by weighing the context of their positions in the network, they successfully predict long-term trends on the popularity of movies and politicians[4].

Understanding the underline structure of social networks and their relationship to each other is important on predicting the behavior of nodes, in that sense (Mislove et al.) studied the structure of different online social networks. Their results confirm the presence of power-law, small-world, and scale- free properties of online social networks they observe that the in-degree of user nodes tends to match the out-degree; that the networks contain a densely connected core of high-degree nodes; and that this core links small groups of strongly clustered, low-degree nodes at the fringes of the network[5].

(Cantonese et al.) analyzed the properties of social networking graphs.They examined scaling laws distribution of friendship and centrality measurements [6].

A useful tool in consumer behavior prediction is "Link Mining". Algorithms using this technique are designed to support performance among some activities including question answering, information retrieval and web-based data warehousing [7].

(Erbs et al.) proved that training data and data volume improve performance in link discovery with text-based approaches[8].

(Qian et al.) used link mining techniques on the Enron mail corpus data and were able to show communities within linked nodes, they were able also to identify 'common friends' using cluster analysis.[9].

Other methods explored link-predictions with applications for exploring data, distributed environments, and spam analysis. [10][11][12]. Research in Social Networks is also visible on current search technology including Page Rank and HITS [13][14][15]. Using these techniques, (Bharat, Henzinger, and Chakraborti) presented variations that utilize web page context to weight pages and links based on relevance.[16][17].(Sugiyama et al.) used the topological structure of a graph to successfully combine few methods including network, quantitative, semantic, data processing, conversion and visualization-based components [18]. Research in Semantic Web technologies also yielded development in Social Networks. In that sense (Zhou, Chen, and Yu) combined an ontology-based Social Network along with a statistical learning method towards Semantic Web data. This involved using an extended FOAF (friend-of-a-friend) ontology applied as a mediation schema to integrate Social Networks and a hybrid entity reconciliation method to resolve entities of different data sources [19]. (Thushar and Thilagam) used Semantic Web technology for the identification of associations between multiple domains within a Social Network [20].

Several Relational Learning methods have supported Social Network analysis predicated on the concept of homophily-based associations to support learning. In that context, we have the application of probabilistic modeling [21] collaborative relationship [22] and inference-based approaches [23].

Visualization techniques are being used, and it substantially helps in studying Social Networks dynamics. (Batajelj and Mrvar) created tools for the visualization of large-scale networks where it is possible to identify vertices and relations between clusters [24].

(Noel et al.) calculated inter-item distances among combinations of elements from which hierarchical clustering dendrograms are visualized to enhance measurement consistency between clusters and frequent item-sets [25].

They introduced an application of association mining to the visualization of link structures. Important (frequently occurring) higher-order item-sets are often obscured by the mere pairwise treatment of traditional analysis. The approach they take here involves the discovery of frequently occurring item-sets of arbitrary cardinalities, and the assigning of importance to them according to their support frequencies[25].

(Levng et al.) created Social Viz which provided users with a means to view frequency relationships among multiple entities in a network [26]. They used frequent pattern mining and visualization techniques. a visualizer called SocialViz is developed for providing users with frequency information on social relationship among multiple entities in the networks. SocialViz could be used a standalone visualization tool, or as an additional tool to existing visualizers, for social networks exploration[26].

(David Alfred et al.)used a collection of twitter message to extract metrics that determine the effect of key players and find a correlation between their graph structure and the market share of three major mobile Operating System[27].

(Sharad Goel, Daniel Goldstein) used retail data and applied logistic regression and five-fold cross validation to compute the likelihood of an individual making a purchase based on his contacts previous activities.The results show that individuals with contacts who made previous purchase are more likely to make purchase than individuals with contact who did not have any previous purchase[28]. (Yoon et al.) used S&P companies data from 2010-2015 and math them to 24 millions user comments directed at those companies's Facebook posts.They tested hypotheses using fixed effect(FE) and random effects(RE) and dynamic (generalized method of moments) and reached to the conclusion that digital engagement volume has significant positive effect on revenue[29]

(John et al.)cautioned the translation of "liking" on social media into an indicator of an intend to make a purchase.Through their study they discover based on more than 14000 cases,that more features in addition to the button "Like" are needed to make the accurate prediction on the purchase[30].

Chong et al.) conducted an experimental study on the consumer engagement behavior(CEB) and were able to show using ordinary least square (OLS) regression models that Facebook and YouTube activities positively correlate with box-office revenue,however their results are not conclusive for twitter.They proposed and tested a set of metrics[31].

(Ding et al.)used Pre-released movies "likes" data from Facebook and discovered that more campaign on those Pre-released data increase revenue for movies[32]. (Yung et al.) propose an experimental model where businesses can target new customers.When the customers visit the store recommendations are guided using the user social media data. The experimental results show that companies can generate substantial revenue using this process[33].

(Hyunmi et al.) investigated the ewom(electronic word-of mouth) of different social networks using Roger's innovation diffusion model on collected daily data from movies and the results pertain that Twitter influence on box office revenue is greater in the initial opening stage[34].

Limits of previous approach.

Most of the work elaborated above even though spread around different techniques approach Social Network analysis in a static fashion where nodes or actors dynamics are not really take into account,the arrival or departure time of agents have been account,but we believe this feature can lead to better insights,the geo-spatial distribution of the contacts are somehow also neglected in most study.We are living in a world now where our contacts and network is spread around the globe and depending of the geographical position,reality and culture can create barrier that are hard to overcome so we believe a segmented approach can be helpful. most of the work above also are somehow single network a approach,individual may have preference using different social networks and we believe an approach that combine multiple networks analysis may be more conclusive. An handicap with most of those research is a the limited scope into Social Networks analysis only.Most of the data available in Social Network are unstructured, a hybrid approach where also structured transaction data are used also can lead to better predictions.If a new node join a network we can do recommendation based on the his contacts preference,but an old node which has transaction data available in retailer database could have better recommendation based on both analysis.

Our Approach

We defined a framework where nodes are individuals or or user. We consider the network to be temporal and defined the time a node enter or exit the network. We try to determine the dynamics of each node (how important or Active a node is in the network) in a dynamical setting. We identify links that connect nodes as shared characteristics or values,We combine the usage of multiple networks data which are available online through their API (Facebook,Instagram,Twitter). We define a second set of nodes as products and we specify characteristics that differentiate products. We study the emergent proprieties of nodes or group of nodes as well as their edges(relationships). we build hybrid models for purchase predictions based on the presence or absence of transaction data. We measure the accuracy of the models proposed on the actual data.

[1] Achrekar, H.; Gandhe, A.; Lazarus, R.; Ssu-Hsin Yu; Benyuan Liu; Predicting Flu Trends using Twitter data Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on Publication Year: 2011 , Page(s): 702 - 707

[2] Peng, Huan-Kai; Zhu, Jiang; Piao, Dongzhen; Yan, Rong; Zhang, Ying; Retweet Modeling Using Conditional Random Fields Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on Publication Year: 2011, pp 336-343

[3] Iyengar, Akshaya; Finin, Tim; Joshi, Anupam; Content-Based Prediction of Temporal Boundaries for Events in Twitter Privacy, Security, Risk and Trust (PASSAT), 2011 IEEE Third International Conference on and 2011 IEEE Third International Conference on Social Computing (SocialCom) Publication Year: 2011, pp 186-191

[4] Wasserman, Faust, "Social network analysis: methods and applications" (structural analysis in the social sciences), Cambridge University Press, Cambridge.

[5] Measurement and analysis of online social networks by Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, Bobby Bhattacharjee In Proceedings of the 7th ACM SIGCOMM conference on Internet measurement (2007), pp. 29-42,

[6] Cantonese, Salvadore, De Meo, Pasquale, Ferrara, Emilio, Fiumara, Giacomo, Provetti, Alessandro, Crawling Facebook for Social Network Analysis, WIMS'11 May 25-27, 2011 Sogndal Norway

[7] Knowledge Discovery and Retrieval on World Wide Web Using Web Structure Mining

Boddu, Sekhar Babu; Anne, V.P Krishna; Kurra, Rajesekhara Rao; Mishra, Durgesh Kumar; Mathematical/Analytical Modelling and Computer Simulation (AMS), 2010 Fourth Asia International Conference on, 2010, pp: 532-537

[8] Erbs, Nicolai, Zesch, Torsten, Gurevych, Iryna, Link Discovery: A Comprehensive Analysis, 2001 Fifth IEEE International Conference on Semantic Computing

[9] Acar, E.; Dunlavy, D.M.; Kolda, T.G.; Link Prediction on Evolving Data Using Matrix and Tensor Factorizations Data Mining Workshops, 2009. ICDMW '09. IEEE International Conference on 2009, pp 262-269

[10]Cai-Rong Yan; Jun-Yi Shen; Qin-Ke Peng; Ding Pan; Parallel Web mining for link prediction in cluster server Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on Volume: 4: 2005 , Page(s): 2291 - 2295 Vol. 4

[11] Caverlee, J.; Webb, S.; Ling Liu; Rouse, W.B.; A Parameterized Approach to Spam-Resilient Link Analysis of the Web Parallel and Distributed Systems, IEEE Transactions on Volume: 20, Issue: 10 2009, pp 1422-1438

[12] Rong Qian; Wei Zhang; Bingni Yang; Detect community structure from the Enron Email Corpus Based on Link Mining, Intelligent Systems Design, and Applications, 2006. ISDA '06. Sixth International Conference on Volume: 2Publication Year: 2006 , Page(s): 850 -855

[13] Web structure mining: an introduction da Costa, M.G., Jr.; Zhiguo Gong; Information Acquisition, 2005 IEEE International Conference on 2005

[14]L. Page, S. Brin, R. Motwani, T. Winograd, The PageRank citation ranking: Bringing order to the Web., Technical Report, Stanford Univesity, 1998

[15]NMF: Network Mining Framework Using Topological Structure of Complex Networks Sugiyama, K.; Ohsaki, H.; Imase, M.; Yagi, T.; Murayama, J.; Congress on Services Part II, 2008. SERVICES-2. IEEE Publication Year: 2008, pp 210-211

[16] J. Kleinburg, Authoritative sources in a hyperlinked environment. Journal of the ACM 46(5): 604-632 1999

[17]K. Bharat , M.R. Henzinger, Improved algorithms for topic distillation in a hyperlinked environment. In ACM SIGIR International Conference on Research and Development in Information Retrieval, pages 104-111, 1998

[18]S. Chakrabarti, B.Dom, and P.Indyk Enhanced hypertext categorization using hyperlinks. In SIGMOD International Conference on Management of Data pp 307- 318, 1998

[19]Semantic Message Link Based Service Set Mining for Service Composition Anping Zhao; Xiaoyong Wang; Ke Ren; Yuhui Qiu;

Semantics, Knowledge and Grid, 2009. SKG 2009. Fifth International Conference on 2009 , Page(s): 338 - 341

[20]Thushar, A.K.; Thilagam, P.S.; An RDF Approach for Discovering the Relevant Semantic Associations in a Social Network Advanced Computing and Communications, 2008. ADCOM 2008. 16th International Conference on Publication Year: 2008, pp 214- 220

[21] Achim Rettinger Matthias Nickles, Volker Tresp Statistical Relational Learning with Formal Ontologies, ECML PKDD '09 Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases: Part II

[22] Kirsten, Mathias, Wrobel, Stefan, Inductive Logic Programming, Lecture Notes in Computer Science, 1998, Volume 1446/1998, 261-270, DOI: 10.1007/BFb0027330

[23] Chunying Zhou; Huajun Chen; Tong Yu; Learning a Probabilistic Semantic Model from Heterogeneous Social Networks for Relationship Identification Tools with Artificial Intelligence, 2008. ICTAI '08. 20th IEEE International Conference on Volume: 1

[24] Batagelj, Vladimir, Mrvar, Andrej, Pajek: Analysis and visualization of large networks, Graph Drawing Software Book. Junger, P. Mutzel, editors 2003

[25]Noel, S.; Raghavan, V.; Chu, C.-H H.H.; Visualizing association mining results through hierarchical clusters, Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on Publication Year: 2001, pp 425 - 432

[26]Leung, Carson Kai-Sang, Carmichael, Christopher L., Exploring Social Networks: A Frequent-Pattern Visualization Approach, IEEE International Conference on Social Computing, 2010

[27] David Alfred Ostrowski. "Social Network Analysis for Consumer Behavior Prediction". Accessible at: http://worldcomp-proceedings.com/proc/p2012/ICA3445.pdf

[28]Sharad Goel, Daniel C. Goldstein. "Predicting Individual Behavior with Social Networks"

[29] Attracting Comments: Digital Engagement Metrics on Facebook and Financial Performance Gunwoo Yoon, Cong Li, Yi (Grace) Ji, Michael North, Cheng Hong & Jiangmeng Liu Pages 1-14 | Received 28 Apr 2017, Accepted 07 Nov 2017, Published online: 24 Jan 2018 https://doi.org/10.1080/00913367.2017.1405753

[30]Leslie K. John, Oliver Emrich, Sunil Gupta, and Michael I. Norton (2017) Does "Liking" Lead to Loving? The Impact of Joining a Brand's Social Network on Marketing Outcomes. Journal of Marketing Research: February 2017, Vol. 54, No. 1, pp. 144-155. https://doi.org/10.1509/jmr.14.0237

[31]Chong Oh, Yaman Roumani, Joseph K. Nwankpa, Han-Fen Hu Beyond likes and tweets: Consumer engagement behavior and movie box office in social media Information & Management Volume 54, Issue 1, January 2017, Pages 25-37 10.1016/j.im.2016.03.004

[32]Chao Ding,Hsing Kenneth Cheng,Yang Duan,YongJin The power of the "like" button: The impact of social media on box office Decision Support Systems Volume 94, February 2017, Pages 77-84 https://doi.org/10.1016/j.dss.2016.11.002

[33]Yung-Ming Li, Lien-Fa Lin, Chun-Chih Ho A social route recommender mechanism for store shopping support Author links open overlay panel Decision Support Systems Volume 94, February 2017, Pages 97-108

https://doi.org/10.1016/j.dss.2016.11.004

[34]Hyunmi Baek,Sehwan Oh,Hee-DongYang,JoongHo Ahn Electronic word-of-mouth, box office revenue and social media Electronic Commerce Research and Applications Volume 22, March-April 2017, Pages 13-23 https://doi.org/10.1016/j.elerap.2017.02.001