

Data Mining Based Social Network Analysis from Online Behaviour

Hamza Ahmed

Abstract : Data mining has evolved into a complex knowledge-seeking venture that provides variable perceptions of viewing data. Previously data mining was intended for extracting useful and extensive information to be used for potentially any application. Several databases allowed discovery of past information through complicated software and hardware reading to make any application understandable without excessive learning about how the system really operated. Ironically data mining has been created by the human mind yet the concepts transitioning into practice has been complicated through its evolution through the usage of data. But sometimes when data systems are subject to improvement and evolution, the concept becomes lost in translation and created a number of challenges instead. Large organizations and associations have been faced with challenges of amassing data of information into servers required to be accessible by all sorts of company personnel. The majority of challenges encountered while attempting to resolve complex data mining emerged in the form of designing high speed data streams, mining sequence, network settings, and security integrity concerns. This report is outlined to the issues of data mining through notable uses and their processes in various applications mostly for computer, business and visual data. A few details of obtaining company data is illustrated and how multiple networking systems are compromised, although not necessarily intruded by external users. The data is described through open-source and proprietor sources where commercial data is frequently shared in networking systems.

Introduction

Data mining has become the central and pivotal process accommodating researchers to analyze behavior on social media. Since social media is flooded with unlimited and various activities from sharing opinions to posting articles to name a few activities, studies apparently indicate the direction of social media is wide open to online predictability. Virtually every video, comment or photograph posted is always subject to comment and sharing because it is the quickest way to transmit through social media. But many glitches have been discovered to be evidently mistaken and researchers need to revisit the claims of online human behavior through social media channels (Chipello, 2014). Hence social media information may not be the most reliable source for information since such information is subject to sharing. At the most the information is opinionated rather than revealing anything factual. Any social media-based material is utilized for research that render online users to make decisions, some of them life-altering,

and become dependent without doing their own due diligence.

Contrary to popular belief, several data mining applications are invalidated by online customer behavior with the fallacy that behavior is self-starting. But as social media is a massive conglomerate of identical thoughts and interrelated presentations that data mining becomes convoluted by conflicting effects among online users. Researchers and media developers though have the ability to access the misleading data, but the difficulties can be enormous (Jensen, 2002). The decisions have to be made whether to analyze data as a whole or spend time extravagantly to break down data into components for analysis to centralize any relevant information. Data mining places companies in positions uncertainty when attempting to categorize their customer base among a multitude of subscribers and users that may not be relevant to their community (Boorman, 2011). Researchers representing companies face the problem of

converting data into useful information for company use because social media data utilizes different technology that deviates from storage format especially from master data formats. The remainder of the report identifies the problems of data mining and how it relates to the conversion of social media behavior. Some related areas include online transactions and security issues between various social media channels and the online behavioral relationships between companies and their customers.

Framework for Social Network Analysis

Whether online or offline, people of like or unlike groups still share things in common and maintain contact with one another regarding what they have in common. Usually various social groups are drawn together following an interesting article or video stream based on a particular topic that draws these groups virtually or physically. Data mining manages to identify active members from the less active members. Physical gatherings often occur consisting people of common interests and knowledge, although the environment can vary bringing together groups with differing opinions yet they remain interested in the same topics. For example, public debates on specific topics may attract people of differing opinions even if they do not frequently agree on the topic at hand (Takaffoli, Sangi, Fagnan & Zaiane, 2010). Yet they often communicate their opinions and express their ideas on various levels.

Online behavior with the comfort of social media expresses similar behaviors where they are characterized by their comments and the topics rather than by individual personalities. Hence this can be defined as a data mined community because individuals are still brought together. Sometimes communal groups are divided into smaller groups. The results are the main categories characterized by two opposing sides of a discussion somewhat buffered by a neutral group. All it takes is a method of slight persuasion based on a single comment or written piece of information. If a group or individual suddenly expresses a strong opinion against another group the focus then shifts forming a new train of thought regardless of who belongs to the group. The framework follows with the diverse attitudes

of social behaviors when monitored under diverse environments. This is one indication as to why data mining is challenging for researchers because of the constantly changing elements (Stucchio, 2011). If illustrated on a daily basis, the results of data mining would be rarely seen as uniform because the framework is always changing by way of each individual's attitude.

A tedious data mining method would be to enlarge the subject social media channels and measure regularly the common interests of each group or person. Even measuring the behavior of an individual person would produce variable daily results on social media. Surprisingly the data could be large because the aggregate computation of this person's presence can be very frequent. The data mining process would need to distinguish this person from others to gain accurate readings. A database could be produced to perform an analysis marking the annotations created by each group or person. But sometimes researchers prefer to select their own database to perform a data mining analysis and then choosing how to resolve the data at their own discretion. The solution they select need not be absolute because all data is collected for comparison to avoid confusion with unusual sets of data records eliminating excessive investigation. The framework is more presentable when pairing or grouping related variables on commenting habits on the most popular social topics. Similar comments or the people frequenting on topics could be clustered for classification. Some data on the other hand may be unknowingly filtered depending on the software or application platform used. The issue with filtering is the data could relevant and important and misses the opportunity to be recorded for data mining. In fact much data on social media is filtered and, this defeats the purpose of searching through computerized data needed to characterize any trends.

Classical Social Network Analysis

Studies in social networking analysis indicated a geometric change in the data mining process compared to the pioneer periods, where in the past personal communication first dominated social structure. Interestingly social media had existed in ancient times regarding social

networking and is quite similar to the modern computerized framework. One major difference is how communication brought common interests together. A few examples of ancient data mining are artifacts that are associated with particular social aggregates and tell of its origins about the culture it represents (Cline, 2014). Archaeologically social networking had its own form of media channels between various groups of ancient history. One example is the interaction among members of a royal family, peasants, between the royal family and the peasants. Historically it shares almost a direct link to modern social media with an identical approach to studying relationships between people and their masters, similar to relationships between social media users and their administrators.

As for customer relationships, data mining techniques are developed when information was solid enough to maximize customer interaction with management techniques. Unlike ancient interaction, much of data mining has actually been a hidden process used in past decades under different channels. Researchers and developers understand how users gathering sufficient information may hinder the performance of data mining software vendors. Most business scenarios when encountering problems utilize proprietary techniques drawn from specific software vendors rather than open-source techniques. The accelerated growth of data to researchers can be both rewarding and challenging because it illustrates how online business social media is required to assist with competitive advantages. Hence data mining is important for deriving information for virtually most large corporations to maintain leads in competition. Financial institutions search for data to discern loopholes to qualify mortgage and interest rates for their home-buying customers (Berson, Smith & Thearling, 2014). The data records searched could also be a fallback technique to predict when and where their clients will purchase property or when they withhold their investments. Vast amounts of data require different server sizes and storage capacity and are arranged to keep them categorized for specific investment qualifications.

In spite of data mining having been developed to handle efficiently several tasks that would otherwise take months for conventional

computing, predictive formulation is the chosen technique analogous with the laws of physics. This is to measure numerous possibilities where a number of inputs determining the outputs to base to build a model sustaining a variety of tasks. Naturally the outputs are reflected by the inputs entered. The model quality depends on the most expedient desired outputs in order to eliminate the inputs not to be explored (Galloway & Simoff, 2007). A data mining technique can also be observed as a trial and error run by exploring every input relevant to the informational scenario in order to predict social behavior in social media. Sometimes a single variable can alter the direction of outputs. When a feature in a social media feature is altered, what customers may be accustomed to in a transaction may alter the direction of navigation through the data stream. This change also alters customer behavior. What happens afterward is this change becomes recorded throughout social media and is detected either as problematic or beneficial depending on who receives the results. What does this have to do with classic social networking analysis? Historically people were always used to customary and routine procedures. Any change would bring about resistance or satisfaction.

Data Mining for Social Network Analysis

Researchers of data mining have described their experiences in climbing the obstacles of informational challenges. When illustrated graphically or pictorially, data mining regardless of its complexity can be understood in laymen's terms in business and how data mining delivers the desired results. The problem discovered is the rate at which the problem persists over the time it takes to find a solution to informational challenges (Rexar, 2011). Management of large corporations usually only expresses the terms of data mining and briefly explaining it by definition. But the time and scheduling are deemed too complicated to take any stakeholders through detailed walkthroughs and case studies. Even following a data mining presentation, employees surrounded and affected

by data mining do not possess the background. However they find the results more attractive. This is obscurely another technique in social networking behavior. Often stakeholders only discuss among themselves the complexities of data mining rather than forwarding their concerns on the lack of knowledge to the presenter or developer of data mining. For example, stakeholders may understand how to operate specific work-related computer functions imbedded in a vast software program. But the majority of the functions they either do not understand or they are irrelevant to their performance.

A few more challenges reported by users of data mining have emerged regarding social networking analysis. Difficulty accessing data can pose a problem and relations between businesses and customers. When data is inaccessible it could indicate the needs of social media users go unchecked. Companies have to rely on what little data they can find, but may not be relevant. Additionally the company should have prepared data profiling to render quicker and easier access to the data needed. What causes grievances on a larger corporate scale is some social media has only limited data storage or navigational contact. In other words, when searching for information on an internal or proprietor source, the information found is not satisfactory.

Normally a search reveals the highest to the lowest match for finding useful information. But the percentage correspondence is not very high and searching further would only be futile. This may imply the multitude of social networks are not properly interlinked and cannot recognize each other's databases. A search in each unique database may bring about different results. If there is no data mining analysis in place the information search becomes futile. The networking analysis may not even be frequently analyzed. The approach has to be revisited to identify if the data embedded in the social network is based on generalized findings or if they are outdated. When there is a lack of linkage, the implications can be that no networking is present and people cannot share the desired information that may prove

useful in future applications. The same occurs with social media where customers searching for data on a product may not be abundant. They expect to discern what they are looking for. When data is difficult to collect or is very limited during a search, customers and sharers of information will either reject sources on the thought of being reliable and useful or navigate to other sources.

Data mining is essential for information sharing and evidently reveals the copious amounts of data marking the socio-cognitive assessment of commentary and email networks. Social network analysis has catapulted its presence in popular online applications including marketing, testimonials and search engine optimization. Previously Google was one of the largest search engines with concise webbing for data mining. Now almost every informational search involves social networking via Google and launches a viral marketing campaign. Data mining if updated is a reliable technique for high capacity datasets that are otherwise too cluttered for traditional data searching methods. The reason why some data searched amount to unrelated information or blank searches is because the datasets are not updated to reflect the data mining upgrade protocols. As for social media networking, the same result of the inability to pinpoint the desired information brings unwanted and irrelevant searches. For example, when online users are searching for a specific popular music video, the video cannot be found on the primary search. It may instead be found on the secondary or tertiary search because the pathway linkage is not updated.

Application of Data Mining Based Social Network Analysis Technique

Data mining for social networking analysis may utilize a quantitative analysis, also known as quantification, of connecting specific groups of people. The groups can connect randomly and arbitrarily exchange information or join specifically titled groups for a particular purpose. Some virtual social groups are focused on structural patterns with a defined purpose. Other

people use social media for personal networking. In certain virtual scenarios some interactions occur between an individual and a group where one person leads the network and others follow (Srivastav, Ahmed, Patjak & Hsu, 2008). This concept was derived and adapted through the studies of psychology and anthropology by examining various social patterns among certain groups of people what the drivers that triggered their choices to associate with certain groups. The data mining techniques can actually be applied similarly to examine a social pattern to measure their trends.

The most popular social networking tool is the computer network. Computers and mobile devices are the basic building blocks of social networks that people use to connect and link through threads of online conversations. Data sources use public commentaries and instant messaging for quick information transportation and information sharing. Data mining allows each message and comment sent to be tracked and saved until users decide to delete them. The data embedded into this social network is measured by databases and statistically prepares related messages to categorize the thoughts and people's behavior. Examples of search messaging platforms are Blogger, Yahoo, and Skype where online conversations provide exchange of information. In the last few recent years social sites have been data mined for popular gaming communities and the latest gaming updates. This gives customers opportunities to play alone or play in multimedia environments with other players. Similarly to information sharing on product marketing and product reviews, players often forward comments about their thoughts on games and leave helpful comments for other players to refer to for assistance in proceeding throughout the games.

Data mining application not only provided more free movement throughout social networking, but this gave users the power to feel like they in control of their social environment. Social sites allow posting of real-time events that practically travels viral in many times faster than any traditional news source. YouTube for example is one viral tool that transmits daily current events and also provides news events from many years

past that was feared lost. Like forums and gaming sites, users are able to express their views and whether their views are agreeable, the growing sociality remains drawing users together (Adedoyin-Olowe, Gaber & Stahl, 2009). Regardless of popularity though, online social developers and data mining communities have become wary of specific comments posted. This includes comments and customer promotions of their products and services offered over the media. Some social networks have strict protocols against specific products and services without authorization. The reason for such protocols is because many users have abused social systems. Simultaneously some activities are permitted for target audience reaction. But with large volumes of data created and shared in social networks, data mining techniques also filter and classify specific content with computational processes.

Emerging Applications

Currently many searches for useful data still need to eliminate most of the glitches associated with the returns of unwanted and unrelated information. For security purposes, which has been another growing trend that continues to persist online are being sought for counteraction with anti-money laundering, credit risk, identity theft and closed-network information sharing. In some instances when data mining results in revealing a mostly unmatched data search, the reason is to limit the access in case any security breach occurs. Data can be accidentally accessed under certain search engines if the safeguards are not updated regularly even if the data is meant to be private and only exchanged between intended recipients. A closed network may not always recognize data mining with limited access and the software may not include security protocols (SAS Talks, 2012). To prevent security leaks, an application incorporating local information only in network learners shall be executed in a closed network systems. This is similar to allowing data accessible for a community data mining system where for example, a members' site has accessible data to members only. At the most other users may only

access data through a quasi-social network only to prevent any malware from entering the system. Data mining a social network would define the differences between local and network variables to not only improve dataset searching, but to ensure the security protocols are in place.

Social media trends are leaning further toward identifying datasets collaborated in any informational source that prioritizes data accumulation when specific topics are searched. In a business sense, customers prefer bookmarking and indexing information that is popular and viral when searching for vital information. If the information is captured by any online business, it could prove useful and help an advantage over its competition in the industry. This may be circumstantially an informational security breach, but if the information is sent viral freely, it can also be argued that this information is sent by consent. Evidently this incident provides essential action to model this data and the tools needed to analyze have to be carefully selected. The organizations whether they are the hosting application platform or online companies themselves need to turn to data mining to search where the trends are headed. Seeking the trends alone may not always suffice because of substantial amounts of data released. If data cannot be properly exchanged in a timely or efficient manner, the organizations will be unable to improve their methods of interactive trends. The only result they may derive is the trends that actually grow interactions among customers and companied further apart.

Ironically the pioneer days of online marketing and social media lacked sufficient information thus also lacking efficient storage for data analysis. However, current data mining techniques have a similar share of issues. Companies have experienced informational and financial losses because of lack of information systems. Currently information systems are in place but the issue with receiving unmatched information for analysis is persisting. Irrelevant information cannot evolve into patterns of data searching and monitoring customer behavior would be inaccurate. The reason why data mining is essential as a marketing and measuring tool for behavior is because it expresses media

applications. Although they seem and operate independently, they remain under human interaction and control, involving data mining techniques at rudimentary levels (Aldana, 2000). Advanced databases capture activities of automated search patterns interpretations of frequented data sought but are not essentially considered data mining techniques because it is mainly dealing with discovery of knowledge processing rather than extracting data like databases do. This type of database is known as the Knowledge Discovery Database (KDD). Instead of dataset searching, a KDD only deals with data readily available from previous records searches implemented in marketing and planning realms. If researchers intend to use data mining through KDDs, the data obtained would only be inconclusive and scattered, and as mentioned many times before, irrelevant. But data mining can bypass KDDs and instead strategize the data processing for business data.

Data mining companies are now actually into competition to providing optimization processing of all business data. The plan is to strengthen interactions between customers and data mining companies integrate their sub-companies and ensure the best services reach out to their customers. Specific fields of industry are focalized in order for their customers to understand how the industries and their products and services by means of information sharing through integrated markets. Based on customer interaction on data validly extracted by the industry that was unavailable or unknown for public use, the data analysis gathers all the information needed for companies to make effective decisions to improve their customer base. Whenever data mining allows new information built upon the complex datasets, it becomes a library of useful information positively affecting customer interaction. Social media becomes optimized through the highest ranking search engines and businesses would gain recognition with every piece of vital information stored for others to use.

In the last few decades high volume data amounts of company information used to be stored only remotely although

electronically. With the data mining process, information will continue to grow but capacity will be more flexible with the growth of networking data because information always accumulates as long as the servers are able to contain data. But only a few companies have been able to make best on their investments on data. The reason has to do with examining the complexity of the data to put it to better use to determine the trends they can monitor. Otherwise data is useless if companies cannot follow the trends. Data mining is observed as self-operating in many instances, but cannot always compute the percentage abundance for each trend. For example, large organizations such as supply chain stores have various methods of receiving customer information especially from those who frequently buy every visit. From the frequency of buying customers, the inventory is important to keep track of if the client purchases several items. What happens if the inventory is not tracked? Personnel will not know how much inventory will be left. Analogously data mining would be a futile application if there were lack of data. As with supply chain stores multitudes of customers arrive everyday. If supply chains are familiar with frequent customers, personnel will be able to forecast what customers will buy on their subsequent visits. Data mining would be a useful tool for making customer predictions (Rajan, Hegadi & Kumara, 2012). Data mining also somewhat guesses what customers will buy before they arrive. Similarly in other words, when customers search online for information on search engines they frequent, the information automatically displays as if reading customers' minds unless the cookies are disabled. Hence data mining techniques are similar to supply chain sales techniques because they already in many instances understand customer-buying behavior and know the kind of inventory to be readily supplied and available. With databases, all this information can geometrically bring the servers with this data mass over capacity. Although, this may not cause the system to crash, but would be difficult to identify all this information. This is why data mining techniques are expedient to help organize patterns of customer behavior and what they frequently buy.

As opposed to the last decade, online sources are user-friendly and ease of navigation where previously users had to depend on experienced users or web developers. But user-friendliness has its unintentional disadvantages as it causes security breaches into remote systems and corrupted programs with malware and viruses. With sophisticated software traffic into websites hackers attacked the most popular websites because they study areas of high traffic volume. Since customers rely heavily on information on products they seek to purchase, hackers learn when to strike with their virus plans. Since information tracking and collecting are the main ingredients of data mining, its techniques are equipped to track the origins of these attacks in order to protect online accounts, medical records, and business transactions. But the combined forces of human interaction and automated data mining are what make possible high security measures against attacks. The analytical components within the server are equipped with data mining safeguards applied to detect any suspicious activity or hacking attempts. The search channels for data mining to operate at full efficiency can be very large and time consuming because the attacks can appear to be very miniscule. In some software and monitoring detection systems data mining repeats its scans over a few extra counts to hone in on the attack. Over time the data mining technique learns its own protective routine periodically.

Conclusion

Data relationships can range from simple and primitive societal interactions to complex online and internet-based information sharing. This can also be classified as cluster based techniques to derive relationships from any relationships between customers or online users who do not have much in common which is a good practice of data mining as well. The future of data mining for social networking includes magnifying its presence in informational extraction in various ways. The most prominent methods still holding their discoveries are technical and empirical. To understand how structures data mining functions in the real world, or inasmuch, the virtual world, the theory needs to be understood (Memon, Xu, Hick & Chen, 2010). Social networking has

advanced geometrically that whatever new techniques are innovated, they become just as obsolete almost immediately after they are launched into the virtual world. Yet more comprehensive frameworks are required to research as data mining progresses and transitions form one application to another. This will become a new task for researchers and companies that employ them to draw findings that belong to the framework. With uncertain fates of software and viral attacks, data mining may need to devolve because of information overload. The information may need to be dissolved or diffused to begin research anew.

Given that existing data mining techniques have demonstrated to be less than perfectly efficient, in the least it served its purpose as far as connecting people and customers. Information itself can be readily available to build the latest models and test new phenomena in social networking. Possibly social networking is been the ongoing communication trend that user ma one day digress into personal and social communication as their ancestors have done and understand where and how interactions between people began. Network evolution can technically experience further development by adapting techniques for data mining and following social patterns in the business and personal world that would be shared by online and offline characteristics. The significant impact of the future of data mining will depend on decision-making processes based on communicative and customer patterns. Subsequently each decision made will also depend on the decision made by its predecessor. Testing the theory of transient outcomes, the decisions may be altered to determine new possibilities of information extraction.

Whether it evolved or devolves, data mining will hold the answers for decision-making for collaboration among various field disciplines. As much interaction takes place over social media streams, a new trend can flourish interactions between experts and professionals of academic fields. Professionals share as much information through data mining as frequently as social gatherings. But professionals in fact other frequenters in other fields do not know do data mine as frequently except it. This once again

demonstrates that similar minded groups only follow behavioral patterns like their own. As the social nature of data mining persists in shaping behavioral patterns, its will continue to offer a variety of directions applied to social media.

References

- Chipello, C. (2014). Social media data pose pitfalls for studying behavior. McGill.
- Jensen, D. (2002). Data Mining in Social Networks. University of Massachusetts – Computer Science Department Faculty Publication Series.
- Boorman, C. (2011). Why Data Mining is the Next Frontier for Social Media Marketing. Mashable.
- Takaffoli, M., Sangi, F., Fagnan, J. & Zaiane, O.R. (2010). A Framework for Analyzing Social Networks. Department of Computing Science – University of Alberta.
- Stucchio, C. (2011). Which the Best Data Mining Framework? Quora.
- Cline, D. (2014). Social Network Analysis and Ancient History. Society for Classical Studies.
- Berson, A., Smith, S. & Thearling, K. (2014). An Overview of Data Mining Techniques. Building Data Mining Applications for CRM.
- Galloway, J. & Simoff, S.J. (2007). Network Data Mining: Methods and Techniques for Discovering Deep Linkage between Attributes. Complex Systems Research Centre, University of Technology Sydney.
- Rexar, K. (2011). Data Miner Survey Examines Trends and Reveals New Insights. Analytics Magazine.

Srivastava, J., Ahmad, M., Pathak, N. & Hsu, D.K. (2008). Data Mining Based Social Network Analysis from Online Behaviour. Society for Industrial and Applied Mathematics.

Adedoyin-Olowe, M., Gaber, M.M. & Stahl, F. (2009). A Survey of Data Mining Techniques for Social Network Analysis. School of Computing Science and Digital Media, Robert Gordon University.

SAS Talks. (2012). Social Networks in Data Mining: Challenges and Applications. SAS – The Power to Know.

Aldana, W.A. (2000). Data Mining Industry: Emerging Trends and New Opportunitites. B.S. Electrical Engineering and Computer Science.

Ranjan, A., Hegadi, R.S. & Kumara, P. (2012). Emerging Trends in Data Mining for Intrusion Detection. International Journal of Advanced Research in Computer Science.

(Memon, N., Xu, J.J., Hick, D.L. & Chen, H. (2010). Data Mining for Social Network Data. Springer Science & Media.