Clickstream Data Mining and Market Segmentation

Jay Gandhi
Huntington Park, California, United States.
Corresponding author email: jaygandhi@gmail.com

Kushal Dalal
Huntington Park, California, United States

Hitesh Ravani
Huntington Park, California, United States

Abstract---*Clickstream data is one of the most important sources of information in websites usage and customers' behavior in Banks e-services. A number of web usage mining scenarios are possible depending on the available information. While simple traffic analysis based on clickstream data may easily be performed to improve the e-banks services. The banks need data mining techniques to substantially improve Banks e-services activities. The relationships between data mining techniques and the Web usage mining are studied. Web structure mining has three types of these types are web usage structure, mining data streams, and web content. The integration between the Web usage mining and data mining techniques are presented for processes at different stages, including the pattern discovery phases, and introduces banks cases, that have analytical mining technique. A general framework for fully integrating domain Web usage mining and data mining techniques are represented for processes at different stages. Data Mining techniques can be very helpful to the banks for better performance, acquiring new customers, fraud detection in real time, providing segment based products, and analysis of the customers purchase patterns over time. And in Market Segmentation the importance of data mining techniques for market segmentation is becoming indispensable in the field of marketing research. This is the first identified academic literature review of the available data mining techniques related to market segmentation. This research paper provides surveys of the available literature on data mining techniques in market segmentation. Eight online journal databases were used for searching, and finally, 103 articles were selected and categorized into 13 groups based on data mining techniques. The utility of data mining techniques and suggestions are also discussed. The findings of this study show that neural networks is the most used method, and kernel-based method is the most promising data mining techniques. Our research work provides a comprehensive understanding of past, present as well as future research trend on data mining techniques in market segmentation.*
Keywords---association rules, clickstream, clustering, customer profiles, K-means.

Introduction

Clickstream mining is a record of a user's activity on the internet, including every website and every page of every website that the users visits, how long the user was on a page or site, in what order the pages were visited, any newsgroups that the user participates in and even the email-addresses of mail that the users send and receive. Both ISPs and individual websites are capable of tracking a user's clickstream. Clickstream data is becoming increasingly valuable to internet marketers and advertisers. Be aware of the big amount of data a clickstream generates. These 'footprints' visitors leave at a site grown wildly - large businesses may gather a terabyte of it every day. But the ability to analyse such data hasn't kept pace with the ability to capture it. The next frontier of web data analysis is better integration of clickstream data with other customer information such as purchase history and even demographic profiles, to form what's often called a "360-degree view" of a site visitor. Clickstream analysis can be

seen as a four-stage process of collection, storage, analysis, and reporting. The first two concentrate on gathering and formatting information, and the latter two on making sense of it. Market segmentation is a marketing concept which divides the complete market set up into smaller subsets comprising of consumers with a similar taste, demand and preference. A market segment is a small unit within a large market comprising of like-minded individuals. One market segment is totally distinct from the other segment. A market segment comprises of individuals who think on the same lines and have similar interests. The individuals from the same segment respond in a similar way to the fluctuations in the market.

Research Methods

Classification in Business Intelligence

Business intelligence can be applied to the following business purposes, in order to drive business value.
Following are some classification techniques:
   a) Measurement – program that creates a hierarchy of performance metrics (see also Metrics Reference Model) and benchmarking that informs business leaders about progress towards business goals (business process management).
   b) Analytics – program that builds quantitative processes for a business to arrive at optimal decisions and to perform business knowledge discovery. Frequently involves data mining, process mining, statistical analysis, predictive analytics, predictive modeling, business process modeling, data lineage, complex event processing and prescriptive analytics.
   c) Reporting/enterprise reporting – program that builds infrastructure for strategic reporting to serve the strategic management of a business, not operational reporting. Frequently involves data visualization, executive information system, and OLAP.
   d) Collaboration/collaboration platform – a program that gets different areas (both inside and outside the business) to work together through data sharing and electronic data interchange.
   e) Knowledge management – program to make the company data-driven through strategies and practices to identify, create, represent, distribute, and enable adoption of insights and experiences that are true business knowledge. Knowledge management leads to learning management and regulatory compliance.

Business intelligence can provide a pro-active approach, such as alert functionality that immediately notifies the end-user if certain conditions are met. For example, if some business metric exceeds a pre-defined threshold, the metric will be highlighted in standard reports, and the business analyst may be alerted via e-mail or another monitoring service. This end-to-end process requires data governance, which should be handled by the expert.

Results and Analysis

Clickstream Datamining and Market Segmentation

All Clickstream Datamining and Market Segmentation were first introduced to solve the pattern classification and regression problems by Vapnik and his colleagues.

a) Overview of SVM

Clickstream or click path data had to be gleaned from server log files. Because human and machine traffic were not differentiated, the study of human clicks took a substantial effort. Subsequently, Javascript technologies were developed which use a tracking cookie to generate a series of signals from browsers. In other words, information was only collected from "real humans" clicking on sites through browsers. It was not possible to identify the click path. Market segmentation is the process of dividing a broad consumer or business market, normally consisting of existing and potential customers, into sub-groups of consumers (known as segments) based on some type of shared

characteristics. In dividing or segmenting markets, researchers typically look for common characteristics such as shared needs, common interests, similar lifestyles or even similar demographic profiles.

Figure 1. Marketing program is designed with the needs of the target market in mind

b) Classification in

We can classify Clickstream Mining and Market Segmentation using Business Intelligence.

1) Types of Clickstream Mining
    (a) Web traffic analysis
        (1) Web traffic analysis operates at the web server level and concentrates on how visitors navigate through the site.
        (2) It measures the number of pages delivered to the customer as opposed to pages sent by the server. It determines how often visitors hit the browser stop button, how much of the page was delivered until they hit the button and how long they waited before they hit it.
    (b) E-business feedback
        (1) The e-business analysis cycle is more sophisticated. This process combines website activity with data from other sources, such as visitor profile information, sales databases, and campaigns that include links to the website.
        (2) It provides higher-level information, more focused answers and information that can be used to enhance e-commerce activities across the business as well as improving the website.
2) Types of Market Segmentation
    (a) Psychographic segmentation
        The basis of such segmentation is the lifestyle of the individuals. The individual's attitude, interest, value help the marketers to classify them into small groups.
    (b) Behaviouralistic Segmentation
        The loyalties of the customers towards a particular brand help the marketers to classify them into smaller groups, each group comprising of individuals loyal towards a particular brand.
    (c) Geographic Segmentation
        Geographic segmentation refers to the classification of the market into various geographical areas. A marketer can't have similar strategies for individuals living in different places.
        Nestle promotes Nescafe all through the year in cold states of the country as compared to places which have well-defined summer and winter season.
        McDonald's in India does not sell beef products as it is strictly against the religious beliefs of the countrymen, whereas McDonald's in the US freely sells and promotes beef products.

44

Identifying the market to be segmented

The market for a given product or service known as the *market potential* or the *total addressable market (TAM)*. Given that this is the market to be segmented, the market analyst should begin by identifying the size of the potential market. For existing products and services, estimating the size and value of the market potential is relatively straightforward. However, estimating the market potential can be very challenging when a product or service is totally new to the market and no historical data on which to base forecasts exists.

A basic approach is to first assess the size of the broad population, then estimate the percentage likely to use the product or service and finally to estimate the revenue potential.
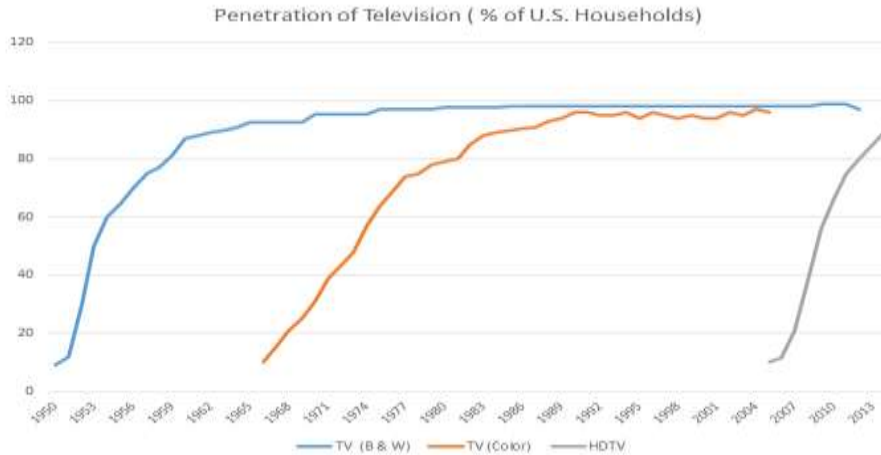

Figure 2. To estimate market size, a marketer might evaluate adoption and growth rates of comparable technologies

A more robust technique for estimating the market potential is known as the Bass diffusion model, the equation for which follows:[27]

$N(t) - N(t-1) = [p + qN(t-1)/m] \times [m - N(t-1)]$

Where:

a) $N(t)$ = the number of adopters in the current time period, (*t*)
b) $N(t-1)$ = the number of adopters in the previous time period, (*t-1*)
c) p = the coefficient of innovation
d) q = the coefficient of imitation (the social contagion influence)
e) m = an estimate of the number of eventual adopters

Conclusion

Clickstreams can also be used to allow the user to see where they have been and allow them to easily return to a page they have already visited, a function that is already incorporated in most browsers. Unauthorized clickstream data collection is considered to be spyware. However, authorized clickstream data collection comes from organizations that use opt-in panels to generate market research using panelists who agree to share their clickstream data with other companies by downloading and installing specialized clickstream collection agents.

Most segmentation approaches have retained this tactical perspective in that they address immediate short-term decisions; such as describing the current "market served" and are concerned with informing marketing mix decisions. However, with the advent of digital communications and mass data storage, it has been possible for marketers to conceive of segmenting at the level of the individual consumer. Extensive data is now available to support segmentation at very narrow groups or even for the single customer, allowing marketers to devise a customized offer with an individual price which can be disseminated via real-time communications.

References

Beane, T. P., & Ennis, D. M. (1987). Market segmentation: a review. *European journal of marketing*, *21*(5), 20-42.

Evelson, B., & Norman, N. (2008). Topic overview: Business intelligence. *Forrester Research*, *61*.

Moe, W. W., & Fader, P. S. (2004). Capturing evolving visit behavior in clickstream data. *Journal of Interactive Marketing*, *18*(1), 5-19.

Omer, A. M. (2017). Identifying, Developing, and Moving Sustainable Communities through Application of Bioenergy for Energy or Materials: Future Perspective through Energy Efficiency. *International Journal of Life Sciences (IJLS)*, *1*(1), 9-39.

Omer, A. M. (2017). Sustainable Development and Environmentally Friendly Energy Systems. *International Journal of Physical Sciences and Engineering (IJPSE)*, *1*(1), 1-39.

Rowley, J. (2016). *Information marketing*. Routledge.

Rud, O. P. (2009). *Business intelligence success factors: tools for aligning your business in the global economy* (Vol. 18). John Wiley & Sons.

Wedel, M., & Kamakura, W. A. (2012). *Market Segmentation: Conceptual and methodological foundations* (Vol. 8). Springer Science & Business Media.