Clustering Web Users' Access Pattern Using Agglomerative Naïve Bayes' With Laplacian Smoothing

Olukumoro S. O.
Department of Computer Technology
Yaba College of Technology
E-mailL gbengaolukumoro@gmail.com
Phone: +234703-388-6271

Arotimi Busayo
National Open University of Nigeria (NOUN)
Lagos, Nigeria
E-mail: arotimi.busayo@gmail.com
Phone: +234810-494-5994

ABSTRACT

Due to exponential rate growth in the complexity of website and in the volume of traffic of the World Wide Web (WWW), it has become very important to analyze the usage of the web sites and the web traffic. Because of this growth, different client side and server side tools that mine knowledge from information resources has been developed. Organizations discover the lifetime value of their clients and also supply them with a more sophisticated structure of the web site and services by analyzing this data. Web usage mining, which involves using the interaction generated by the users in the form of access logs, proxy-server logs, browser logs, user session data, cookies to extract knowledge. The goal of this work is to use the mean shift algorithm to cluster in order to study their use of web resources and the users' navigation patterns. The primary focus of this project is the Web Usage Mining, also through this project more light will be shed on the different stages involved in the mining process. This project is concluded by results and analysing of the experiment carried on the NYC (New York City) Social Media Usage web access logs.

Keywords: Website, traffic, mean shift algorithm, Mining

1. INTRODUCTION

Due to exponential growth of information sources present on the World Wide Web (WWW), there is an urgent need for users to employ the use of automated tools in order to get the required information resources, and at same time track and analyse usage pattern of users. These factors brought about the necessity to create server side and client side intelligent systems for mining useful knowledge in order to analyse the lifetime value of customers and increase their cross marketing strategies. The need to revisit and review techniques, and the traditional strategies used for market analysis because there is an increase in organizations relying on the World Wide Web (WWW) to conduct business. Nowadays, a large collection of unstructured data are being analysed by the revisited traditional techniques and strategies. The data stored in the transaction logs are been formed by collecting the data on the web in the form of server access logs which are clients interaction with the web site. Generally, information are been stored automatically by the Web server, either as a server or access logs.

For different reasons, this data is being analysed by different genres of the organization. A better structure for the effective use of the web site and organization's benefit can be achieved by getting more information from analysing the server logs. The infrastructure of an organization and communication workgroup can be well managed as a result obtained from analysing the server logs by organizations that work with the intranet technologies. By analysing the user access patterns by the advertising organizations, this will help in focusing on or targeting a specific customer group. By Web Usage Mining, we aim to know the rate at which a page is being accessed by different clients and thereby getting the popular users path traversal. From the analysis, It can therefore be deduced in an intuitive manner that if there is long and convoluted user access paths alongside with a low use of web page shows that the web is not laid or well structured. Re-structuring of the web site alongside its navigation result based on the analysis is being carried out. Clustering, rule generation, sequential pattern generation are the algorithms that have been employed in the mining process.

1.1 Statement of Problem

The need for mining users' access pattern is required for organisations so as to improve their customers' lifetime value and so as to increase cross market strategies and also to make them provide the necessary structure for effective use of web site for their customers.

1.2 Problem Statement

To develop a web mining application that can cluster users' access pattern using the Agglomerative Naïve Bayes with Laplacian smoothing.

1.3 Objectives

The aim of this work is to provide a framework of Agglomerative Naïve Bayes with Laplacian smoothing that generate web users' access pattern from a web data in terms of platforms used in accessing the site.

1.4 Methodology

There were two methodologies employed for this project; one of this is the object modelling technique, used for the software design and also Research methodology was used for the research being done. It was developed (Rumbaugh 1991) as a method to develop object-oriented systems, and to support object-oriented programming. The object-modelling technique is an object modelling language for software modelling and designing. Three main types of models was proposed

Object model: In (Totland 1997), the object model depicts the static and most stable phenomena in the modelled domain. Main concepts are classes and associations, with attributes and operations. Aggregation and generalization (with multiple inheritances) are predefined relationship (Totland 1997). In this project, the object model is depicted with the class diagram.

Dynamic model: This depicts a state or view of transition of state on the model. The transitions made between states, events to trigger transitions and also the state are the main concept. Each action taken between states can be modelled. Generalization and aggregation (concurrency) are predefined relationships (Totland 1997). The sequence diagram was used to show the implementation of dynamic model as related to this project.

Functional model: The process perspective of the model is being handled by the functional model, which corresponds to the data flow diagrams. In functional model, main concepts are process, data store, data flow, actors (Totland 1997). The functional model is a predecessor of the Unified Modelling Language (UML). In this project, the activity diagram was used to show the functional model.

2. LITERATURE REVIEW

2.1 Bayesian Theory

The Bayesian classification is a supervised learning method based on Bayes Theorem. It is also a method of statistical classification. With the Bayesian, uncertainty can be captured by evaluating probabilities of the outcomes in a disciplined way following fundamental probabilistic model.

Bayes Theorem was proposed by a British mathematician, Thomas Bayes (1702-1761), also the Bayesian classification was given after him. Bayesian Classification combines observed data, prior knowledge and independence assumption.. A useful viewpoint of evaluation of many algorithms and understanding of many learning algorithms are being provided by Bayesian classification. It is potent for noisy input data and also evaluates probabilities explicitly. Laplacian smoothing was applied in this project in the case where the outcome of probability results to zero.

2.2 History of Naïve Bayesian

Independence Bayes and simple Bayes are among the names which the naive Bayes is known according to different literature. Apparently, all of these names are referenced to the use of the Bayes' theorem which is a decision rule classifier, though, naive Bayes is not necessarily a Bayesian method (Hand; Yu). In (Russell & amp; Peter, 2003, Russell and Norvig emphasised that "[naive Bayes] is sometimes called a Bayesian classifier, a somewhat careless usage that has prompted true Bayesian to call it the idiot Bayes model.

2.3 Naïve assumption for Bayes model

For any class variable, a naive Bayes classifier assumes that the value of a particular feature is unrelated to the absence of any other feature. In a supervised learning, naive Bayes classifiers can be trained to work effectively for some probability models types. Maximum likelihood method is being used by naive Bayes model for parameter estimation in numerous practical applications.

$$p(y \vee x_1, \cdots, x_n) = \frac{p(y)p(x_1, \cdots, x_n \vee y)}{p(x_1, \cdots, x_n)}$$

using the naive independence assumption which is simplified to

$$p(y \mid x_1, \cdots, x_n) = \frac{p(y)\prod_{i=1}^{n} p(x_1 \vee y)}{p(x_1, \cdots, x_n)}$$

where

$p(y \vee x_1, \cdots, x_n)$ is the posterior probability of class given prediction (attribute)

p(c) /p(x) p(y) is the prior probability of class

$p(x_1, \ldots., x_n \vee y)$ is the likelihood which is the probability of predictor given class

$p(x_1, \ldots., x_n)$ is the prior probability of predictor

A naive Bayesian model is easy to build with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. Despite its simplicity, the Naive Bayesian classifiers often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods.

How it works

Bayes theorem provides a way of calculating the posterior probability, $P(c \vee x)$, from $P(c)$, $P(x)$, and $P(x \vee c)$.

Naive Bayes Classifier assumes that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors.

Probability estimation

By applying Bayes' theorem, P(N|M) , a posterior probability is gotten. To get the value of P(N|M), the values of each of P(M|N), P(N) and P(M) is first derived or known.

Suppose, given a dataset T with a set of feature $M = \{m_1, \ldots, m_t\}$, the estimation of $(m_i|N)$ is given as

$$P(m_i|N) = \frac{freq(m_i, N)}{t}$$

From the equation given above, the number of m_i in N is represented by $freq(m_i, N)$, where the total number of instances of N is represented by M. This illustrated estimator is known as maximum likelihood estimate (MLE).

A zero frequency problem results in some instances where a given feature m_i in N may be absent, when using the maximum likelihood estimate to compute $P(m_i|N)$. In such cases, the probability of the unseen item is taken to be zero; this cannot be used as an input into the Bayesian classifier this is because

everything will be reduced to zero. Besides this problem, If the training data is minute, the maximum likelihood estimate (MLE) also tend to give poor estimate (John Kelleher & Brian Mac Namee, 2012).

By using smoothing, the zero frequency problems can be taken care of so as to get non-zero probabilities. Adjusting the probability of an unseen event from seen is the aim of smoothing.

2.4 Applications of Naïve Bayes classification

1. Naive Bayes text classification: One of the most successful algorithm used for learning text documents classification is Naive Bayes. Naive Bayes text classification is a probabilistic learning method that is derived from bayesian classification.
2. Spam filtering : Naive Bayes classifier has been used for identifying spam e-mail. Naive Bayesian text classification is mostly popular for spam filtering. A genuine email is differentiated from an ingenuine spam email using the popular Bayesian spam filtering (Chai & H. T. Hn, August 2012). Also, Server-side email filters like Spam Bayes, DSPAM, Bogo filter, Spam Assassin can also be installed separately, which are also email filtering programs, all employ the use of Bayesian spam techniques for filtering.
3. Hybrid Recommender System using Naive Bayes Classifier and Collaborative Filtering. To filter information not seen, the Recommender systems uses data mining techniques and machine learning, also It can predict if a suggested resource will be liked by a user.
4. Online applications: Supervised machine learning and non cognitive computing has being used for online applications. This is done by differentiating if a training set is nice, nasty or neutral sentiments. A single words and word pairs is used as a feature by using the Naive Bayes classifier. +1, -1, and 0 are used respectively as a label and are mapped to utterances either nice, nasty and neutral classes being made by user.

2.5 Web Mining

In (Hsinchun Chen, 2001),Web mining was defined as the use of data mining, text mining, and information retrieval techniques to extract useful patterns and knowledge from the Web [(Etzioni, 1996), (Chau, 2003)], it has been frequently used in real world applications, such as business intelligence (Chau 2003), Website design (Chau 2003), and customer opinion analysis (Fang 2006).With Web mining, our aim is to discover knowledge and useful information from the page content, Web hyperlink structure and Web usage. Web mining is an aspect of data mining; Web mining employs many data mining techniques. Because of the heterogeneity and semi-structures or unstructured nature of Web data (Shaheen Parveen1 2012), Web mining is therefore not a pure application of traditional data mining. Web mining can be classified into three types, which are; Web structure mining, Web usage mining and Web content mining.

2.5.1 Web Structure mining:

According to (Shaheen Parveen1, 2012), in web structure mining, we uncover useful knowledge from the hyperlinks, which stand for the structure of the Web. For instance, important web pages can be discovered from the links which is an important technology employed in search engines. With this also, Communities of users that shares common interests are discovered. Since, there is no link structure in a relational table, traditional data mining cannot perform such task.

2.5.2 Web content mining:

Mining or Extraction of useful knowledge or information from Web page contents is done by Web content mining (Shaheen Parveen1, 2012).. For instance, Web pages can be automatically clustered or classified based on their topics. Useful data like postings of forums, descriptions of products can be extracted from Web pages for many purposes. In addition, Consumer sentiments can also be discovered by mining their forum postings and customer reviews. This discourse is not a traditional data mining task.

2.5.3 Web Usage mining:

From web usage logs which collects and stores record every user made by each user, with Web usage mining we can discover user access patterns (Shaheen Parveen1, 2012).. Many data mining algorithms are applied to Web usage mining. In order to get the right data for mining, pre-processing of click- stream data in usage logs turn out to be a bottleneck and a key issues in Web usage mining.

Data collection is the difference between Web mining process and data mining process, though Web mining process is similar to the data mining process. For a traditional data mining, the data warehouse is where the data is often already stored and collected. But in Web mining, data collection can be substantial task, which involves the crawling a large number of target Web pages. We step through the same three-step process as soon as the data is being collected. Which include data pre-processing, Pattern discovery and Pattern analysis.

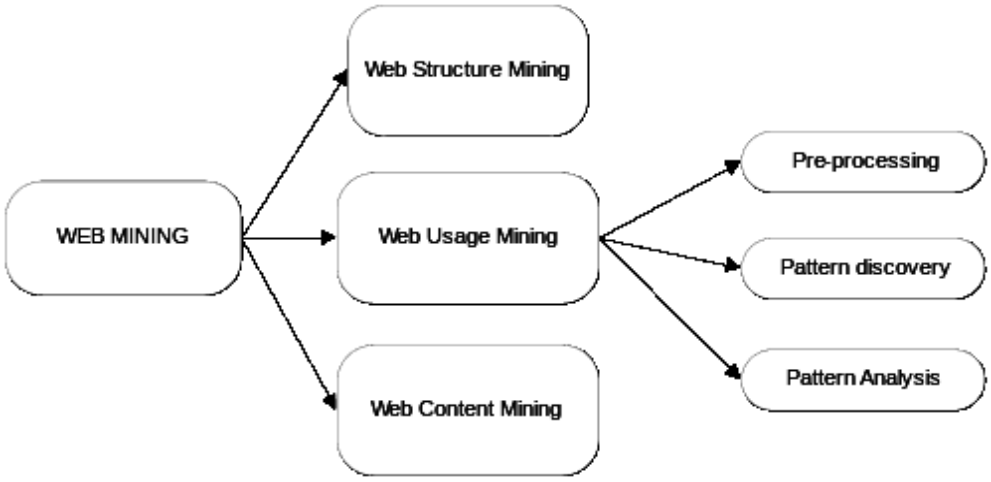

Fig 2.1: Web mining Architecture

Source: Introduction to Data Mining (Chris Clifton]

By discovering patterns and trends in large amounts of complex data, data mining can help promote decision-making. According to (Hsinchun Chen, 2001), by performing analysis on Web usage log data, Web mining systems can discover knowledge about a system's usage characteristics and the users' interests. Such knowledge has various applications, such as personalization and collaboration in Web-based systems, marketing, Website design, Website evaluation, and decision support (Armstrong, 1995; Cooper, 2001; Fang, 2006; Marchionni, 2002).

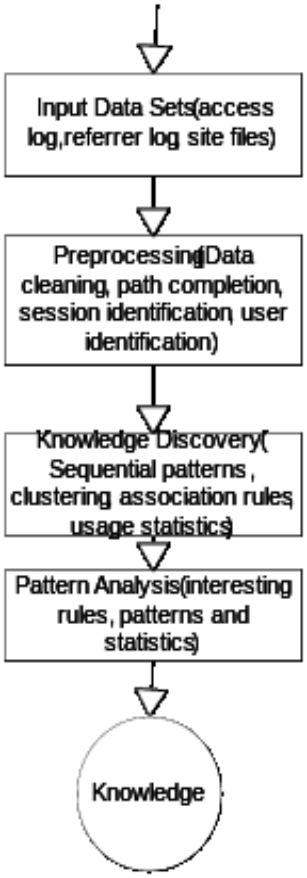

Fig. 2.2: Detailed Flowchart of Web Usage Mining.
Source: Introduction to Data Mining [Chris Cliftono]

2.6 Existing systems

2.6.1 WebLogMiner:

WebLogMiner (Zaiane et al., 1998) , a knowledge discovery tool which interactively extract implicit knowledge from a very large Web log files which benefits from data mining techniques and multidimensional data cubes, and OLAP (Online analytical processing) . The increasingly enlarging of the enormous sizes log files, the techniques used to analyze this data, and the types of data collected in the Web Logs do not restrict or limit the WebLogMiner

2.6.2 LogMiner

LogMiner is a powerful log analysis package for Apache and IIS (or other web servers employing the combined or Extended W3C log formats). The LogMiner can extract and present different reports, about visits, hit, traffic, requests, navigation paths, browsers and OS's used by users and so on. With LogMiner is easier to index previous months thereby ensuring that no unused files. One of the key feature of LogMiner is the ability to presents reports in a more extensible way, usually in a format supported by a PL/pgSQL. LogMiner expresses its output in an Extended W3C log format.

2.6.3 Webalizer

Webalizer, web log analysis software is a General Public License (GPL) that spawn analysis for web pages. With Webalizer, reports including the hits, visits, referrers , amount of data downloaded and the visitors' countries can be generated graphically in various time frames like as by month, day and hour.

2.7 Architecture of a Typical Data Mining System

It may have the following components:

- Clustering: This is an unsupervised classification technique in which we seek to describe dataset in terms of natural clusters of cases.
- Decision tree: The Decision tree is a tree-shaped structure which depicts sets of decisions. Rules are being generated for classifying a data set.
- Nearest neighbour method: The nearest neighbour method employs techniques that classify every record in a database based on a combination of the classes of the k records most similar to it in a historical dataset. It is sometimes called the k-nearest neighbour technique.
- Rule induction: The extraction of useful if-then rules from data based on statistical significance.
- Data visualization: This involves the visual interpretation of complex relationships in multi-dimensional data. Graphic tools are used to illustrate data relationships.

Data mining is also known as knowledge discovery in databases (KDD).It is best defined as the process of discovering useful knowledge or patterns from data sources, for instance from text, databases, image, the Web and so on. The patterns discovered must be valid, easily understand and potentially useful. Data mining consist of vast discipline field which involve statistics, machine learning, artificial intelligence, visualization and information retrieval.

There are numerous number of data mining tasks. Which include common ones such as the unsupervised learning, association rule mining, and sequential pattern mining and supervised learning. Since to develop a solution to any problem, the problem domain must be well understood, which applies to data mining application. Data mining application begins usually understanding of the domain of the application by data analysts (data miners); those who go on to identify best data sources and the target data. When the data is being collected, data mining can be carried out on the data collected; this is usually carried out orderly in three main steps:

1. **Pre-processing**: By pre-processing we mean the collected raw data needs to be cleaned so as to remove abnormalities or noises. This is done because for some various reasons most of the raw data collected is always not suitable for mining. In the data, there may be some irrelevant attributes or may be the data is too gigantic (large), which results in attribute selection and also sampling.
2. **Data mining**: Then, the data processed is then input into a data mining algorithm in which will discover patterns or knowledge.

3. **Post-processing**: Albeit, not all discovered patterns in many applications are useful. Here, the useful patterns are identified for the applications. Also, decisions are made by using various evaluation and visualization.

The entire process, i.e. the data mining process is almost always iterative. The desired or satisfactory result that is being incorporated into real -world operational normally requires many rounds to achieve. Web mining and text mining are becoming popular and important owing to the growth of the Web and text documents. Though, traditional data mining uses structured data stored in relational spread sheets, relational tables, or flat files in the tabular form.

The diagram below shows the architecture of a typical data mining system

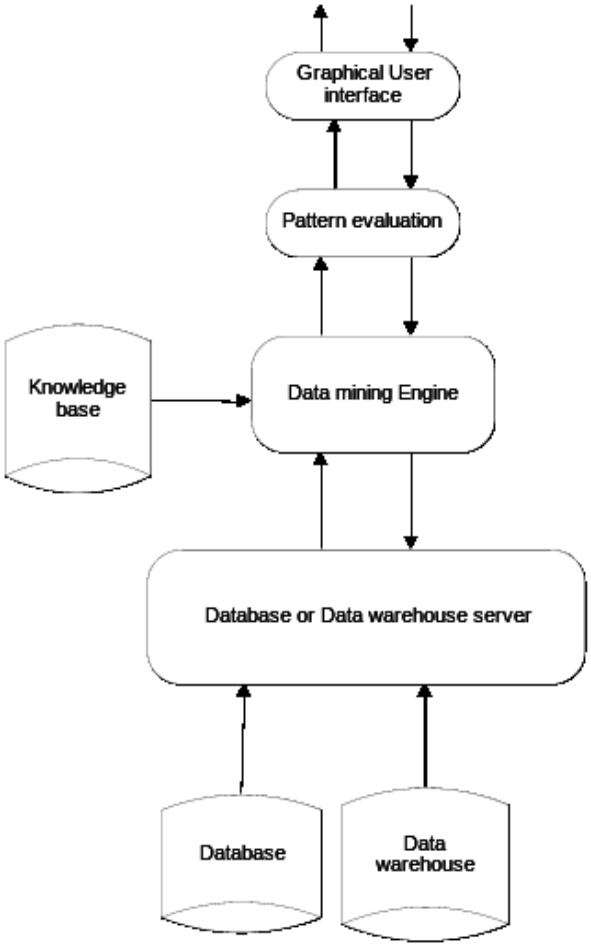

Fig 2.3: Architecture of a typical data mining system

Source: Introduction to Data Mining [Chris Clifton]

3. METHODOLOGY (NAÏVE BAYES)

Naive Bayes is a stochastic technique that evaluates the similarities among objects by computing individual or partial probability and fed it into the conditional probabilities of items used for basis of similarities. Naïve categorizes N number of feature vectors into n number of identified definite and unique clusters. The small n number of unique clusters may be part of the features or separated. They may be considered as seeded feature vector. The task of the algorithm is to compute the partial probabilities of individual vector and then compute their conditional probabilities in relation with the seeded vectors. Finally, the cluster with the maximum conditional probability is selected as the most appropriate cluster for the vector.

3.1 Naïve Bayes' Policy

- The seeded vectors or cluster is determined.
- The partial probabilities of input vectors are evaluated.
- The conditional probabilities of vectors in relation to clusters are evaluated.
- Then, the cluster with the highest conditional probability is selected in relation to the vector under investigation.
- Repeat process until you investigate all vectors.

3.2 Naïve Bayes Formula

$$P(C_i \vee V_k) = \frac{PP(C_i \vee V_k)}{\sum_{i=1}^{n} PP\left(\frac{Ci}{Vk}\right)} \forall \quad z = 1, \ldots, n \qquad \text{(i)}$$

The equation above can be well simplified to:

$$P_p(C_i \vee V_k) = \Pi_{j=1}^{m} P(T_j \vee C_i) \qquad \text{(ii)}$$

$$P_p(C_i \vee V_k) = \Pi_{j=1}^{m} P(T_j \vee C_{(c_i+n)}) T_j \geq 1 \qquad \text{(iii)}$$

Note: Equation (iii) is a variant of equation (ii); the latter is applied to avoid zeroing of terms or probability values. We hereby apply Laplacian smoothing initializing all T_j = 1 and adding n to Ct for every partial probability computed.

3.3 Laplacian Smoothing

So far, the parameters computed are unsmoothed. Owing to the fact that the estimates of $P(C_i \vee V_k)$ are barely adequate in real systems. This is because we've been using the empirical estimates of the parameters $P(C_i \vee V_k)$ which can result into zero. So, a need to ensure that there is no zero estimate result. With good smoothing, you can reduce overfitting thereby increasing the accuracy.

if k=0, the probabilities are unsmoothed if k=0. Note that the smoothing of probabilities occur as value of k grows larger. The good value of k can be gotten from validation set.

Laplacian smoothing is employed in this project, which includes k to counts every possible observation value.

$$P(C_t=c_t \vee V_k=v_k) = \frac{\varepsilon\{0,1\}(c_i^t,y)+k}{\sum_{c_i^t} c(c_i,y)+k}$$

Where:

V_k=A particular input vector or feature

C_i=A particular n seeded vector or cluster

C_z=A particular input vector

n=number of clusters or seeded vector

m= number of terms in an input vector or feature

C_t=number of terms in V_k

P=Conditional probability

P_p=Partial probability

Dataset Extraction

The dataset used in this project is for NYC social media usage and was gotten online. We evaluate our analyzing procedure from the displayed charts in the application after running the algorithm on the dataset provided or supplied.

3.4.1 Activity Diagram

Activity diagrams are dynamic and behavioural models of the activities performed by. It depicts the flow of action per system.

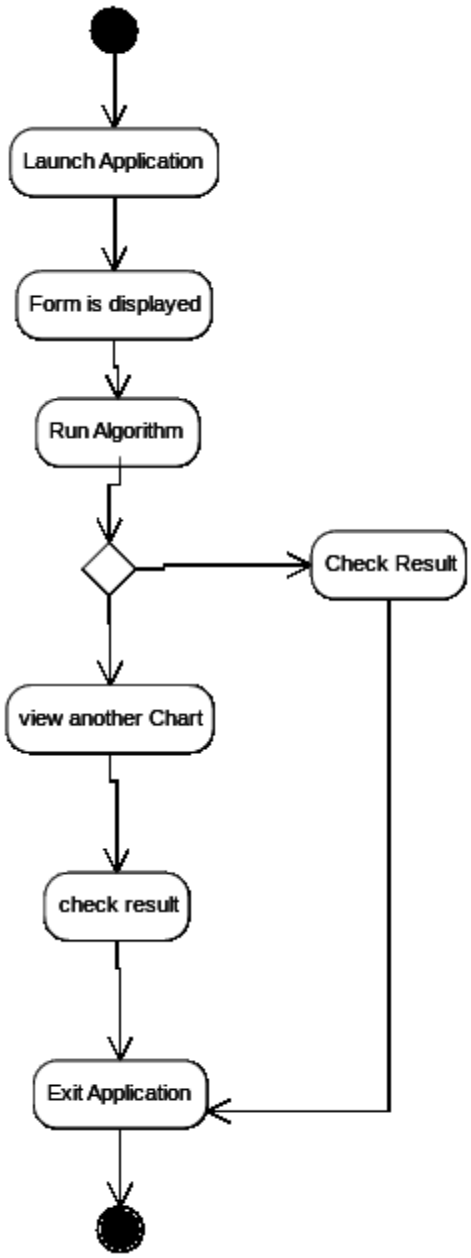

Fig 3.0 Activity Diagram

3.4.2 Class Diagram

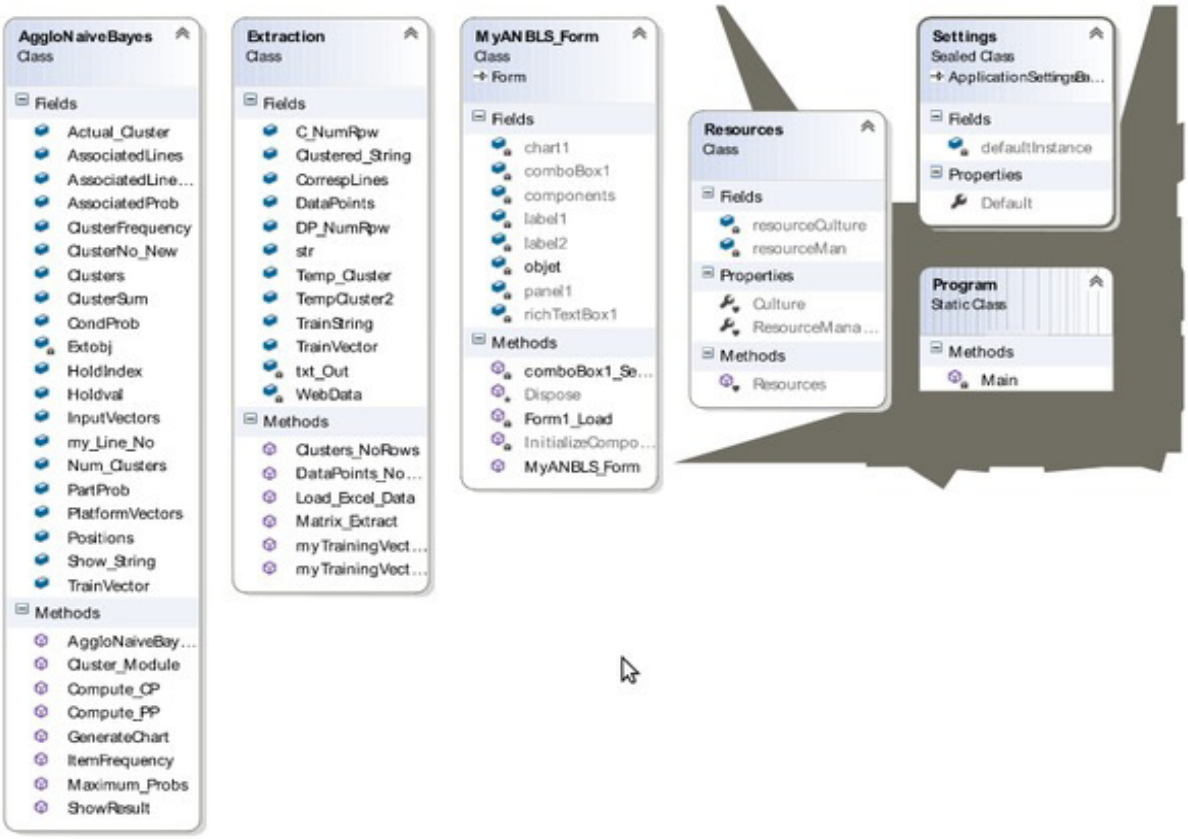

Fig 3.2 Class Diagram

4. IMPLEMENTATION

4.1 Procedural Design

AggloNaïveBayes, Extraction and MyANBLS_Form are the three main classes implemented in this application. The Extraction is the input layer for AggloNaïveBayes class; it reads from a text file which contains the NYC Social Media Usage (2011-2012) dataset with five thousand nine hundred records (5900). Each record also refers to as Vector. The records contain both seeded and investigated vectors.

4.2 Feature Extraction

The Extraction class has five class attributes (WebData, TrainVector, str, DP_NumRpw and C_NumRpw) and there are six methods (matrix_Extract, myTrainingVector, myTrainingVectorSTring, DataPoint_NoRows, Clusters_NoRows, Load_Excel_Data). The path to the source file is being held by the string variable WebData. TrainingVector is a jagged array (an array of arrays) that holds the split items in a vector, str is a string variable that returns an attack vector, DP_NumRpw is class attribute that stores the number of investigated vectors while C_NumRpw holds the number of seeded vectors.

The dataset used for this application is being loaded by a primary method; the two methods C_NumRpw and DP_NumRpw are assigned values in the methods are being assigned values .In matrix_Extract . the vectors were separated into jagged array.TrainVector and TrainingVectorString return the split investigated and the seeded vectors and the investigated split are being passed to AggloNaiveBayes class respectively. The number of seeded and investigated vectors respectively are being returned by both DataPoints_NoRows and Clusters_NoRows.

The extraction class directly communicates with the AggloNaiveBayes class for rudimentary information regarding the input vectors. The communication is seamless such that the structure of the program does not need to change even when new investigated vectors are to be supplied. This ability represents the scalability of our design, such that new records can be added without extra cost in terms of restructuring the old implementation, just the source needs to be populated.

4.3 Algorithm Design

The algorithm design systematically explains the logical steps to achieving this analyzing system.

1 Extract features
2 Compute partial probability of investigated vectors
3 Compute the conditional probability of investigated vectors in relation to seed vectors
4 Select the seeded vector with the maximum conditional probability
5 Assign investigated vector to a seeded vector i.e. cluster
6 Repeat 1 to 5 until investigated vectors are exhausted.
7 End

Interface design

The verified and functional software system is a user friendly which is made up of efficient interfaces that are well linked together. The interface was designed in a simple, functional and intuitive way.

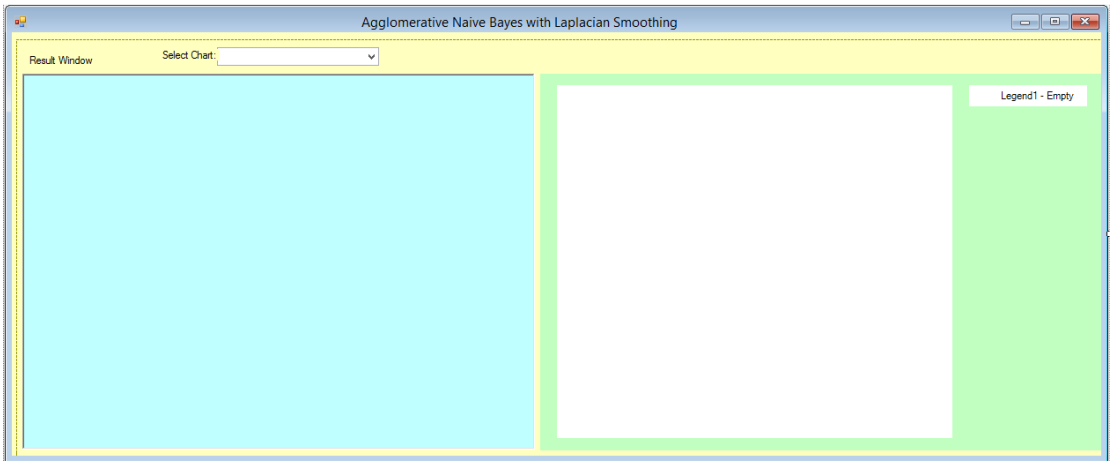

Fig 4.0 Screen shot for Defaultform

The Fig 4.0 pictured above which is the default form; showing a left window that displays the result of the clustering exercise, a combo box to select the type of chart and the lower pane to display the result in the selected chart. There are three available charts: bubble, pie and point charts.

4.4 Results

The diagrams below (from to Fig 4.1) depicts the results when the application is run, it displayed the clustered access pattern in the left layer of window and the chart that will show the analyzes of the result when a selection is made from the combo box above the three layers.

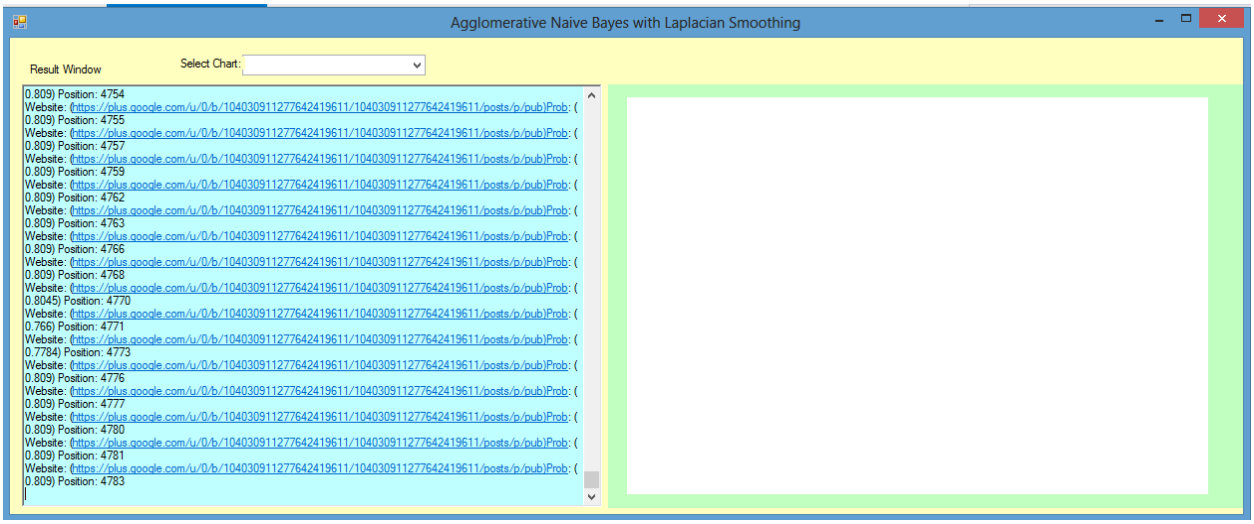

Fig 4.1: Screenshot of output when the application is run.

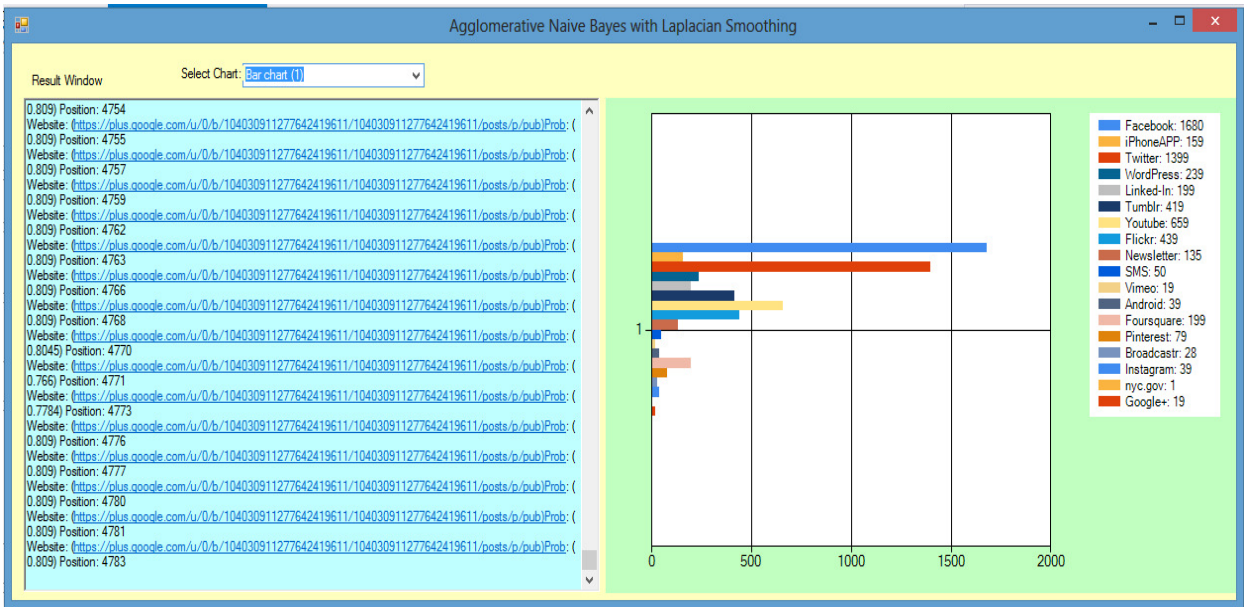

Fig 4.2: Output when the Bar chart is selected from the combo box.

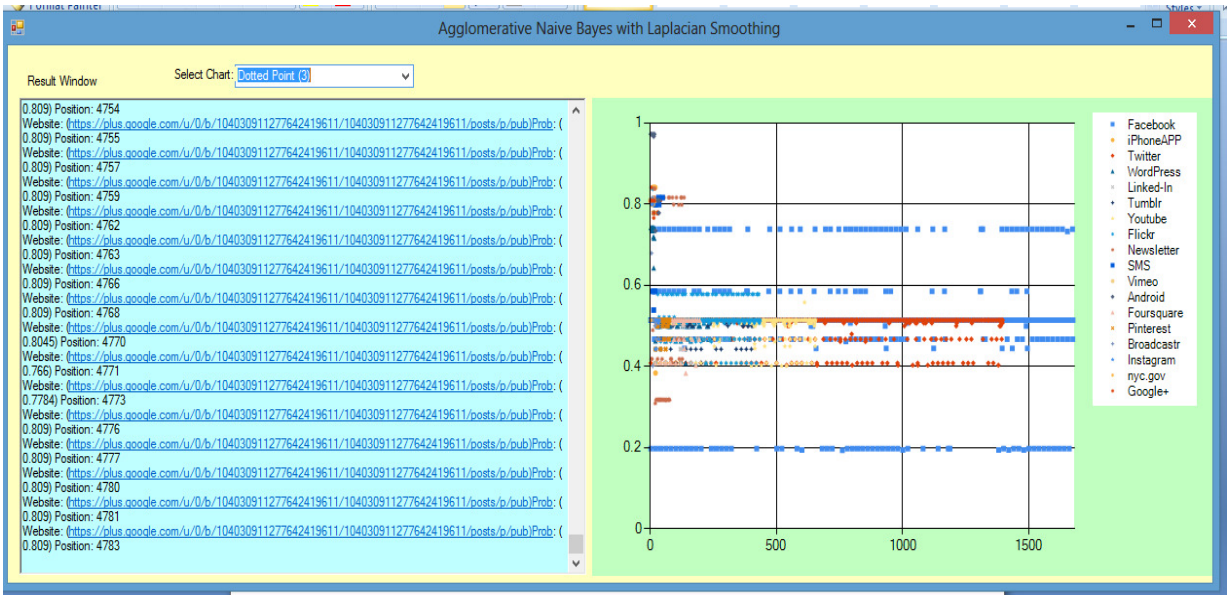

Fig 4.3: Output when the Dotted Point chart is selected.

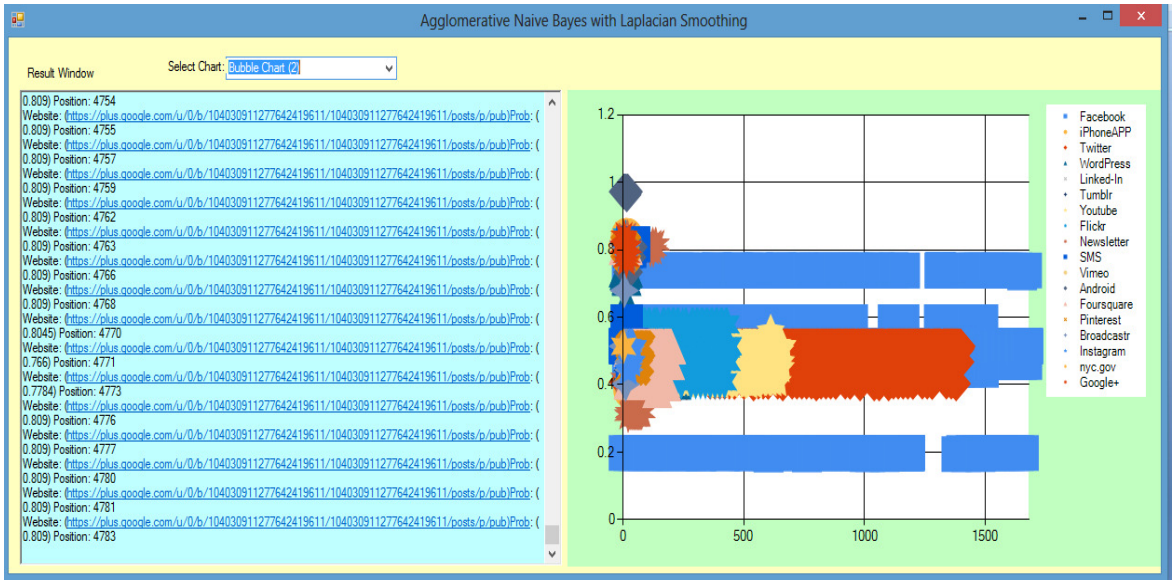

Fig 4.4: Output when the Bubble Chart is selected.

5. CONCLUSION

Having implemented and tested this web mining solution, which is an aspect of data mining, it can be concluded that using this techniques on web access log files is a very useful tool that help in analyzing users' access pattern using different platforms.

We have seen that with Naive Bayes with Laplacian smoothing, it is conceivable to actualize the human intelligence in web usage mining. One important reason of using Naive Bayes is that the system will mirror the capacity being showed with an adequate level of preparing. In addition, with Naive Bayes, peradventure irrelevant inputs that are few are being supplied to the system, during the pre-processing, the data will be cleaned so that the system will figure out how to overlook that don't help the yield. Apparently, if there are empty cells or field(s) in the input file, then the system will neglect to join on an answer.

REFERENCES

1. Armstrong, R., Freitag, D., Joachims, T. and Mitchell, T. (1995). WebWatcher: A Learning Apprentice for the World Wide Web. In Proceedings of the AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments, Mar 1995.
2. Chai, K.; H. T. Hn, H. L. Chieu; (2002) "Bayesian Online Classifiers for Text Classification and Filtering", Proceedings of the 25th annual International ACM SIGIR Conference on Research and Development in Information Retrieval, August 2002, pp 97-104
3. Chau, M and Chen, H. (2003). Comparison of Three Vertical Search Spiders. IEEE Computer 36(5), 56-62.
4. Chau, M., Shiu, B., Chan, I., and Chen, H. (forthcoming) Redips: Backlink Search and Analysis on the Web for Business Intelligence. Journal of the American Society for Information Science and Technology, accepted for publication.
5. Chen, H. M. and Cooper, M. D. (2001). "Using Clustering Techniques to Detect Usage Patterns in a Web-Based Information System," Journal of the American Society for Information Science and Technology 52(11),888-904.
6 en.wikipedia.org/wiki/Naive_Bayes_classifier
7. Etzioni, O.(1996). The World Wide Web: Quagmire or Gold Mine. Communications of the ACM39(11), 65-68
8. Fang, X., Chau, M., Hu, P. J., Yang, Z., Sheng,). R. L. (2006). Web Mining-Based Objective Metrics for Measuring Website Navigability," in Proceedings of the International Conference on Information Systems, Milwaukee, Wisconsin, USA, December 2006.
9. Hand, D. J.; Yu, K. (2001). "Idiot's Bayes — not so stupid after all?" International Statistical Review 69 (3): 385–399
10. Hsinchun Chen, Xin Li, Michael Chau, Yi-Jen Ho , Chunju Tseng (2001).Using Open Web APIs Teaching Web Mining. The University of Arizona, The University of Hong Kong. 1073-0516/01/0300-0034
11. http://en.wikipedia.org/wiki/Webalizer#Log_file_types
12. http://logminer.sourceforge.net
13. Rumbaugh, J (1991). Object Oriented Modeling and design. Englewood Cliffs, N.J.: Prentice Hall.
14. Russell Stuart and Norvig Peter (2003). Artificial Intelligence: A Modern Approach (2nd ed.). Prentice Hall. ISBN 978-0137903955.
15. Totland, T. (1997). Toronto Virtual Enterprise, Thesis: Norwegian University of science and Technology, Trondheim.
16. Web Mining: Information and Pattern Discovery on the World Wide Web * R. Cooley, B. Mobasher, and J. Srivastava Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455, USA.
17. Web Usage Mining: Data Preprocessing, Pattern Discovery and Pattern Analysis on the RIT Web Data. Anand Sharma, Rochester Institute of Technology, Rochester, NY,aps2177@rit.edu www.cs.rit.edu/~aps2177