

RELATIVE EXPLORATION OF WEBSITE RANKING USING CLUSTERING ALGORITHMS

A. Angel, Research Scholar

R.Uma. MCA., M.Phil., M.Ed.,

Department of computer application,

Department Of Computer Science and Information Technology,

Nadar Saraswathi College of Arts and Science, Theni

ABSTRACT

As an internet is spreading out of its bound the demand of online transaction is also getting considerably increased. Now the demand is fast and direct to home service very flexibly. So the ultimate research is online shopping. Online shopping is the method of effective transaction between money and goods .Which is done by user without spending large time of span. Products of Flip kart, Amazon, Snap deal, big basket are analyzed with opinion mining. Customer spends more time when purchasing the product on online by reading all the reviews before buying these products. This paper proposed novel approaches to rank the website efficiently by the mining of genuine reviews of the above specified websites. The deciding factors of these websites include user rating and reviews. In this paper intentional to the website reviews datasets persist analyzed by using simple k-means and x-means clustering algorithms are compared to find the time and cluster instance

Keyword: - Data Mining, K- means and x-means cluster algorithm, unsupervised learning, time, and instance, opinion mining.

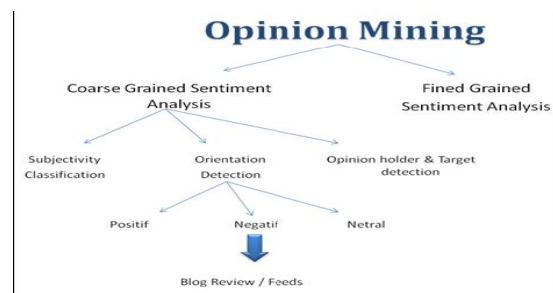
1. INTRODUCTION

Data mining is defined as extracting or mining knowledge from large amount of data. ^[2]Opinion mining is the method of determining the approach of the speaker with respect to the product .In general opinion of the user is most important for organization or individual to increase the performance of the service .So opinion mining is the technique to extract the data about particular things based upon the review. ^[3]The opinion mining is becoming very interesting area of research due to the

emerging web technology. The machine learning is used to cluster the opinion text.

In old days shopping was a concept in which a customer used to buy a product from a mall or from shop. And customer was paying money to supplier at the time of shopping. For traditional shopping. Customer needs to be physically present at shop or at mall .Also there was no review system available to describe the quality criteria of product. Customer used to buy product on the basis of retailers opinion .Sometimes retailers give fake feedback to sell their product. But in early days internet has done massive amount of evolution. Every activity is getting associated with internet. .In online shopping customer can buy or sell products by sitting at home and using digital devices like smart phones, laptops, computers etc. Here user is doing payment by means of credit card or net banking systems .There is no need of customer to be physically present over shop or at mall for purchasing product and paying money. For selecting good quality products among all products, online shopping provides review of each product given by various customers. Normally customer refers this review before buying and product.

But it consumes more time for customer to read each and every review of the product given by the other customers and then take decision for purchasing product. As some reviews are good and bad, so customer has to examine each and every review before selecting that product ^[4]A proposed methodology to guide customer for choosing a best in

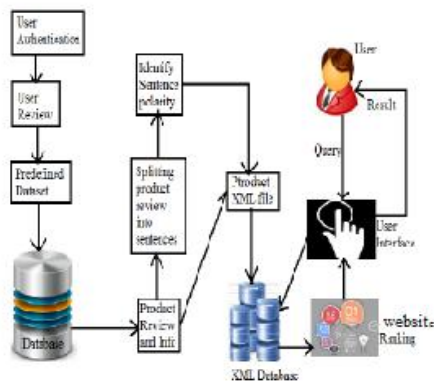


websites and sort out the useful reviews of website by using opinion mining concept.

2. METHODOLOGY

2.1. DATA COLLECTION

A pre-defined dataset of particular E-shopping website [5], on the website, user can specify their product brand and information of the product and all data has been saved into the database. Data



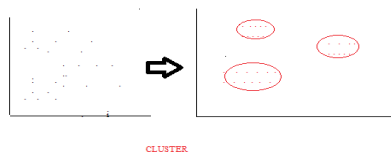
sets are collected to the customers reviews in websites. In the since of amazon, flip kart, snap deal, big basket.

2.2 PREPROCESSING

Preprocessing or cleaning work begin with collected data set [7], data preprocessing means filling in missing values smoothing noisy data identifying or removing outliers and resolved inconsistencies.

2.2.1. CLUSTER

Cluster is therefore [8] a collection of object which is similar between them and is dissimilar to the objects belonging to other cluster



Clustering [4] techniques contains many algorithms clope, cob can, dBs can, em, farthest first,

filter cluster, hierarchical cluster, make density based cluster, optics, sib, x means, simple k-means. Such as paper based on website reviews datasets are analyzed by using simple k-means and x-means clustering algorithms are compared to find the time and instance.

3. PROPOSED WORK

3.1. K-MEANS

K-means clustering is a method of classifying/grouping items into k groups (where k is the number of pre-chosen groups). The grouping is done by minimizing the sum of squared distances (Euclidean distances) between items and the corresponding centroid.

A centroid is "the center of mass of a geometric object of uniform density", though here, we'll consider mean vectors as centroids.

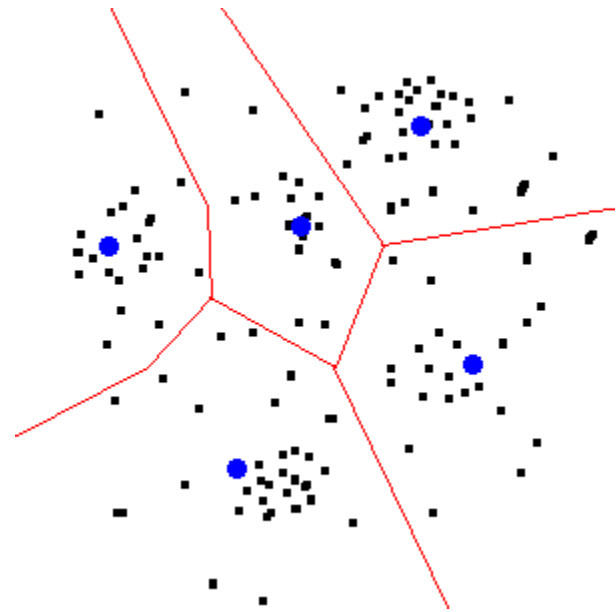


Figure 1. A clustered scatter plot. The black dots are data points. The red lines illustrate the partitions created by the k-means algorithm. The blue dots represent the centroids which define the partitions.

The initial partitioning can be done in a variety of ways.

- **Dynamically Chosen:** This method is good when the amount of data is expected to grow. The initial cluster means can simply be the first few items of data from the set. For instance, if the data will be grouped into 3 clusters, then the initial cluster means will be the first 3 items of data.

- **Randomly Chosen:** Almost self-explanatory, the initial cluster means are randomly chosen values within the same range as the highest and lowest of the data values.
- **Choosing from Upper and Lower Bounds:** Depending on the types of data in the set, the highest and lowest (or at least the extremities) of the data range are chosen as the initial cluster means. The example below uses this method.

3.2. X-MEANS

Cluster data using the X-means algorithm

X-Means is K-Means extended by an Improve-Structure part in this part of the algorithm the centers are attempted to be split in its region. The decision between the children of each center and itself is done comparing the BIC-values of the two structures

X-Means clustering algorithm, an extended K-Means which tries to automatically determine the number of clusters based on BIC scores. Starting with only one cluster, the X-Means algorithm goes into action after each run of K-Means, making local decisions about which subset of the current centroids should split themselves in order to better fit the data. The splitting decision is done by computing the Bayesian Information Criterion (BIC).

X-Means is K-Means extended by an Improve-Structure part In this part of the algorithm the centers are attempted to be split in its region. The decision between the children of each center and itself is done comparing the BIC-values of the two structures.



The result of running K-means with three centroids.



Each original centroid splits into two children.

Fig2 denotes clustering data using the x-means algorithm.k-means result is three centroids.and the x-means centroid in 4 childrens .

PARAMETER

Time

This is referred to as the time required to complete training or modeling of a data set .it is represented in seconds .

Instance

Class weka ,core .instance class for handling an instance .all values (numeric,and nominal)are intrnally stored as floating -point number.if an attribute is nominal (or a string),the stored values is the index of the corress ponding nominal or string value in the attribute.

4.EXPERIMENTAL RESULT

In this paper compared to the two algorithms and two parameters and to take the data set since the repository and applying the tool for weka.thus the two algorithm are calculated for time and instance values for normal clustering outputs .then finaly compared k-means or x-means to it .

VISUALIZE THE DATA SETS

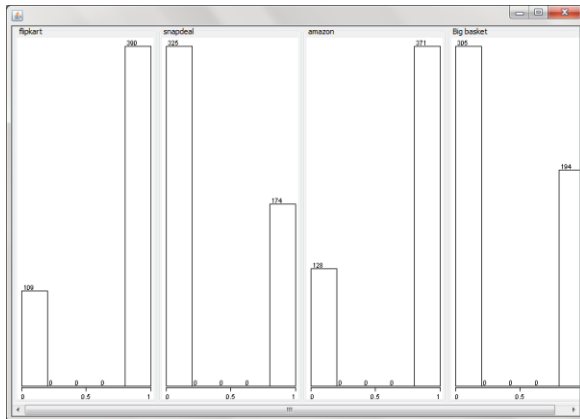


Fig 3 shows over all visualization in data set website to be ranked.

The k-means clustering assignment

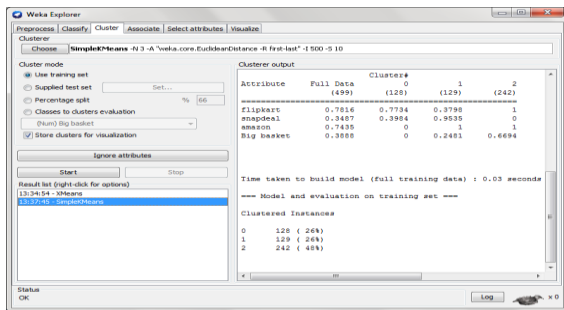
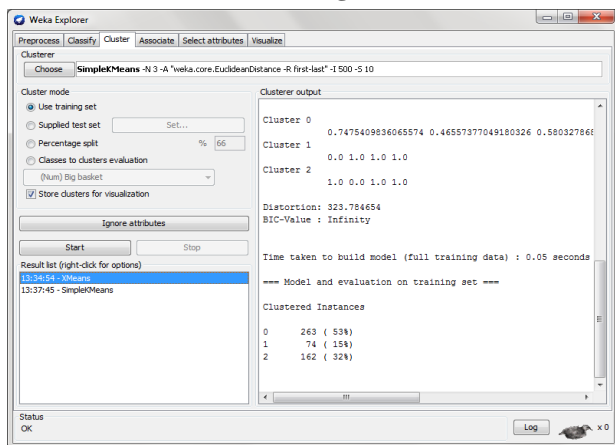


Fig4contains overall execution time and number of clusters. Over all time taken build model contains overall execution 0.03 sec k-means time.

The x means cluster assignment



03fig 5 execution time and number of clusters. Over all time taken build model (full raining data):0.05seconds in x-means

COMPARISON OF TWO CLUSTERING RESULT

Sno	Algorithm	Cluster instance	Time	Cluster instance			Percentage		
				0	1	2	0	1	2
1	Simple k-means	3	0.03 (sec)	128	129	242	26	26	48
2	Simple x-means	3	0.05 (sec)	263	74	162	53	15	32

Tab 6 show overall result in comparision of two clustering result

CLUSTERING INSTANCE PERCENTAGE

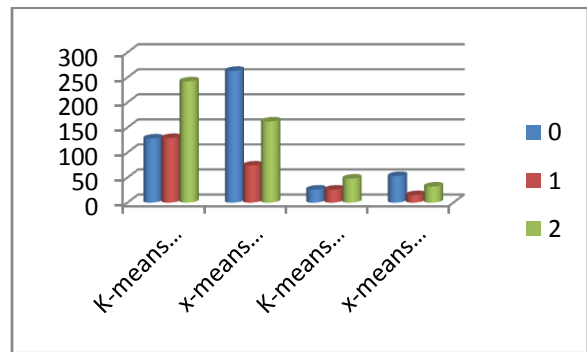


Fig7shows that clustering instance percentage

TIME PERCENTAGE

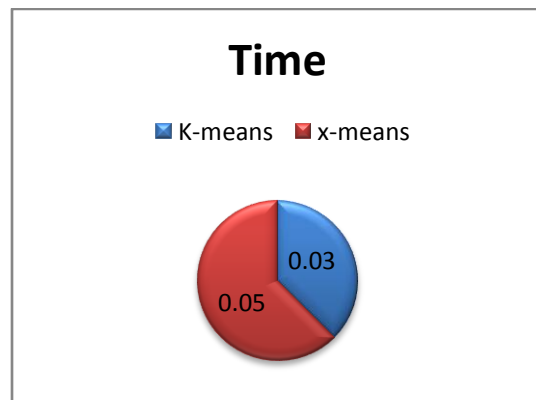


Fig8shows that timing percentage of k-means and clustering algorithms

WEBSITE RESULT IN K-MEANS

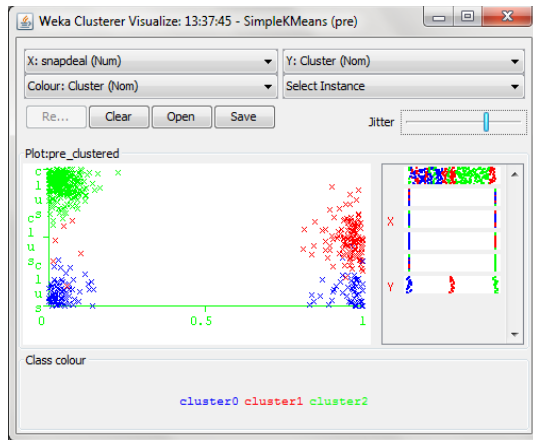


Fig 9 shows that data are grouped into two clusters 0 and 1 using simple k-means algorithm.

WEBSITE RESULT IN X-MEANS

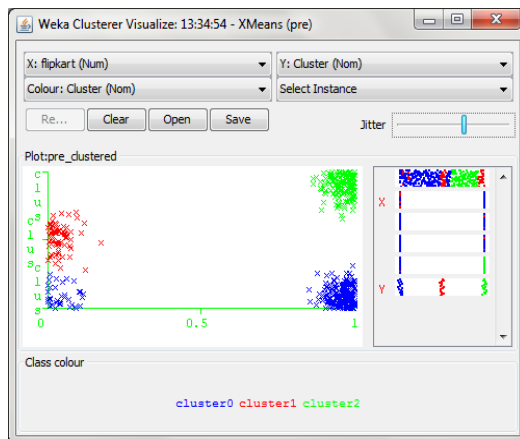


Fig10 shows that data are grouped into two clusters 0 and 1 using simple k-means algorithm.

6. CONCLUSION

In this paper proposed a website ranking system which can take the data set from customer reviews. The experimental will helps to find best website in online with less time. Clustering algorithms used to predict the instance and time. Currently this paper applied two algorithms x-means, and k-means to analyze the time and instance level. Finally k-means clustering algorithm gives the result in less time and fewer instances. In future, focus will be on improving the construing method.

7. FUTURE WORK

Online shopping makes use of digital technology for managing the flow of information, products, and payment between consumer, site owners and suppliers. Online shopping can be either B2B or B2C Shopping cart is one of the important facility provided in online shopping this lets customer to browse different goods and services and once they select an item to purchase they can place the item in shopping cart, and continue browsing till the final selection. Customers can even remove the items from shopping cart that were selected earlier before they place the final order. It reminds us of shopping basket that we carry in departmental store

8. REFERENCE

- [1]. G.kgupta introduction to data mining with case studies paretic-hall of indiaptv.limited 2006.
- [2]. H.w.ian”datamining–practical machinelearningtoolsand techniques”, 2nd edition-elsevier-2008.
- [3]. Jiaweihan,michlinekambar and jianpie,”data mining :concepts and techniques “third editionelesver,2012
- [4]. A.ahamadyfardandh.modares”combinipso and k-means to enhance data clustering”proceeding of international symposium to telecommunication, pp 688691, 2008.
- [5]. Jiaweihan,michlinekambarand pie,”datamining:conceptandtechniques”third edition ,morgankavalmano publisher,2011.
- [6]. n.d.d.douldmisp.kokkinos and e.varvarigos, resource selection fortail with time requirement using spectral clustering “ieeetransaction on computer, vol63pp.461-474, 2014.
- [7]. Data mining introductory and advanced topicsperseduation.margath.dunham,.
- [8]. Effective product ranking methodbasedonopinionminingmadhavkulka nistudent,professormayurilingayat
- [9]. Implementation of data mining in online shopping system using tanagra tool dataset .nov2014
- [10]. Importance of online product reviews from a consumer perspective