

Journal homepage:http://www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

Business intelligence from customer review management using rough set model.

ZIME SONGBIAN.

University of Electronic Science and Technology of China.

Manuscript Info	Abstract
Manuscript History:	Many people examine online opinions written by different users to analyze
Received: 18 January 2016 Final Accepted: 19 February 2016 Published Online: March 2016	more about a product or venue. However, the overwhelming quantity of user- generated opinions and variance in length, detail and quality across the reviews make it difficult to collect useful information. In this paper, we present a structure which can be used for the purpose of discovering business
Key words:	intelligence with the help of decision rules induced from the client reviews of
Business intelligence, rough set, rule induction, learning from	a product or a service posted online. We have clearly the analysis of the review for I W Hotel Ankara. Our proposed framework has been designed by
Module algorithm.	collecting the reviews from the JW hotel official website, preprocessing data and inducing rules by using rough set based Learning from Module (LEM2)
*Corresponding Author	algorithm. The prompted rules could be useful for business analyst in
ZIME SONGBIAN.	understanding the product dimensions, attributes and inherent affiliation among them.
	Copy Right, IJAR, 2016,. All rights reserved.

Introduction:-

As customers search online for product statistics and to assess product alternatives, they often have get right of entry dozens or hundreds of product evaluations from different clients. These clients' reviews are provided in addition to product descriptions, reviews from experts, and customized advice generated with the aid of automatic advice structures. Every of those options has the capacity to feature the price for the potential client. Past studies has substantially examined the position of professional opinions (Chen and Xie 2005), and the role of online recommendation systems (Bakos 1997; Chen et al. 2004; Gretzel and Fesenmaier 2006), and the effective impact feedback mechanisms could have on buyer agree with (Ba and Pavlou 2002; Pavlou and Gefen 2004). Further recently, studies has examined the position of online purchaser product reviews, mainly searching on the characteristics of the reviewers (Forman et al. 2008, Smith et al. 2005) and self-choice bias (Hu et al. 2008; Li and Hitt 2008). Latest research has also shown that client opinions may have a fantastic effect on sales (see Chen et al. 2008; Chevalier and Mayzlin 2006; Clemons et al. 2006; Ghose and Ipeirotis 2006). Specifically, Clemons et al. (2008) located that the exceptional of the review as measured by way of helpfulness votes additionally definitely influences sales.

According to two surveys, 81% of net users have done online research on a product at least once; between 73% and 87% report that reviews had a substantial impact on their buying (Pang and Lillian Lee, 2008).

Online client overview are two types, a textual description about the product which consists of revel in, emotions, sentiments, opinion and the other type is a collective summarization of usual satisfaction approximately the product's capabilities and performance inside the form of ratings. Consumer evaluations can be described as peer-generated product evaluations posted on organization or third party web sites. Retail web sites provide consumers the possibility to submit product reviews with content material in the form of numerical star ratings (generally rating from 1 to 5 stars) and open-ended patron-authored remarks about the product.

A product having opinions with high star ratings has constantly been the preferred because it builds an assurance for buying that product. However the bad critiques, negative sentiments make the client to desist from shopping for or keep on his choices of purchasing.

Nerveless, from the previous research, the findings do no longer lead to a definitive expectation of whether or not opinions control reaction are greater helpful. This ambiguity can be partly improve by using the statement that preceding research on online review control.

The purpose of this paper is to apply rough set principle for coming across business intelligence with the help of decision policies took on from the client opinions of a product or a service posted on-line.

To gain meaningful decision regulations, we experienced the subsequent levels. First of all, the statistics data is preprocessed. After, the rule induction algorithm ELEM is applied to the preprocessed data. This international cover, also called relative reduce, based totally on rule induction algorithm generates selection guidelines, that could display profound information and offer new insights of the effect of the review in constructing the marketplace section, growth in sales, client retention and also adding new customer service.

Related work:-

Even though consumers overview summarization has been investigated heavily in computational linguistics (Hu and Liu, 2004), consumers interfaces it haven't now been studied as considerably. One technique for client evaluate summarization is to gather reviews on unique functions in an entity (e.g., food or provider in a restaurant evaluate). (Liu et al., 2005) advanced a machine to visualize how many nice or poor evaluations were posted approximately capabilities of an entity using bar graphs, but the method not evaluated from the user interface perspective. (Carenini et al., 2006) (Carenini et al., 2009) used a tree map visualization to reveal the records extracted from user-generated evaluations organized in a tree shape based totally on the capabilities. But, their user have a look at showed that the members were regularly pressured by way of the tree map visualization, and desired text-based summarization. They also developed a device just like Liu et al.'s paintings (2005), however they tailored it to allow the person to evaluate a couple of entities (Carenini et al., 2009). Their consumer interface indicates the distribution of advantageous and negative evaluations along exclusive features. But, they now not formally studied the efficacy of their user interface. A computational linguistics approach often used for summarizing consumer-generated critiques the sentiment evaluation, which determines the semantic orientation of a given textual content. (Turney and Pang et al. 2002) carried out a sentiment analysis approach to investigate review textual content. Both systems used machine learning techniques to pick out the semantic orientation of the terms extracted using n-gram methods.

Until now, past research on reviews provides findings with conflicting implications for review analytic and helpfulness. For opinions of movies with moderate star ratings, (Schlosser, 2005) observed that two -sided arguments have been extra credible and led to greater effective attitudes about the movie, but within the case of movies with intense ratings, two-sided arguments have been less credible. Different research on online review presents insights on the connection among assessment diagnosticity and overview extremity. (Pavlou and Dimoka,2006) discovered that the extreme scores of eBay dealers had been more influential than moderate ratings, and (Forman et al.,2008) observed that for books, moderate reviews had been less helpful than excessive opinions. One possible explanation for the issue is the customer's initial attitude. For example, (Crowley and Hoyer, 1994) discovered that two-sided arguments are more persuasive than one-sided high-quality arguments when the initial attitude of the customer is neutral or negative, however no longer in different conditions.

Preliminaries:-

Business Intelligence:-

Business Intelligence is set connecting humans using a right information infrastructure and performance driven culture, enabling them running smarter and greater and more closely together towards company and personal goals.

Business Intelligence, bringing the right data at the proper time to the proper people in the proper format. It's a fivestep manner to run your enterprise smarter, starting with registering the right data successfully, collecting the statistics from several sources, transforming, combining and storing it in a data warehouse. This data need to be described, analyzed and distributed to the right users on the proper time with the right format. The figure below shows these steps.



Fig1: business intelligence process

Rough set theory:-

Rough Set Theory (RST) is a mathematical technique developed by Pawlak in 1982 (Pawlak, 1982). This method has been developed to control uncertainties from data that presents a few inexactitude, incompleteness and noises. When the available statistics data is insufficient to determine the precise value of a given set, lower and upper approximations may be used by rough set for the illustration of the giving set. The approximation synthesis of principles from the acquired records is the primary objective of the rough set analysis.

In decision making, it has showed that rough set methods have an effective essence in dealing with uncertainties. Rough set concept proposed by the author in (Pawlak. Z, 1991) gives nevertheless every other attempt to this trouble. The concept has attracted attention of many researchers and engineer all around the world, who contributed essentially to its development and packages.

The rough set philosophy is based on the belief that with every item of the universe of discourse we associate some information i.e., understanding is associated, through which class can be accomplished.

Objects characterized by the same data are similar in view of the available data information. The indiscernibility relation generated in this way is the mathematical foundation of rough set principle. Any set of all indiscernible (comparable) items is called a basic set (community), and forms a simple granule (atom) of know-how about the universe. Any union of simple units is known as crisp (unique) set - otherwise the set is rough (obscure, indistinct). A number of the rough set associated phrases are provided underneath (B. Walczak, D.L. Massart, 1999; Silvia Rissino, Germano Lambert-Torres,2009).

The rough sets approach to data evaluation hinges on two fundamental concepts, specifically the upper and the lower.

Upper Approximation ($\overline{A}(x)$)

Upper Approximation is a description of the objects that possibly belong to the subset of interest. The upper approximation of the set X, denoted as BX, is the union of these elementary, which have a non-empty intersection with X: $BX = \{xi \in U | [xi] Ind(B) \subset X\}$.

Lower Approximation (A(x))

It consists of those objects that can be with certainty classified as belonging to X. It is also known as positive POS(X).

Let X denote the subset of elements of the universe $U(X \subset U)$. The lower approximation of X in B (B \subseteq A), denoted as BX, is defined as the union of all these elementary sets which are contained in X. More formally: $BX = \{xi \in U | [xi] Ind(B) \cap X \neq 0\}$.

Boundary Region

A set is said to be rough if its boundary region is nonempty, otherwise the set is crisp. It is also known as BR(X) Whereas U - $\overline{A}(x)$ is known as negative NEG(X). If the boundary region is a set $X = \emptyset$ (empty), then the set is considered "Crisp", otherwise, if the boundary region is a set $X \neq \emptyset$ the set X "rough" is considered.



Rule Induction:-

Rule induction (Burgess, Heidi, guy Burgess and Michelle Maiese, 2004 and Grzymala-Busse J W, 1992) is one of the most essential techniques of machine learning. The rule induction is one of the fundamental tools of data mining. Rules are generally in the following form

If (attribute 1, value 1) and (attribute2, value2) and

(attribute n, value n) then (decision, value)

Information from which rules are induced are usually presented in a form similar to a table in which cases are labels for rows and variables are labeled as attributes and a decision. Attributes are non-dependent variables and the choice is a dependent variable. The set of all cases labeled by the same decision value is called a concept.

Proposed approach:-

We built our framework by collecting online reviews posted from http://www.marriott.com/hotels/hotel-reviews.

Modules:-

The project consist of following five modules:-

a. Attribute Selection:-

A huge volume of textual data is posted on the JW-Marriott Hotel website and it has a description about large number of attributes. Attribute selection means picking relevant and important attributes out of a large number of attributes such that further analysis would be proved efficient. We have selected six attributes as cleaning, dining, location, service, amenities and value of the money.

	July 27, 2015	
CharlesG Age: 60+	"Exceptional Service at JW Marriott - Ankara"	
Rewards Level: Gold	From the first month they open until now, every time I travel to Ankara Turkey I	Cleanliness
Traveler Type: Solo	stay at the JW Marriott. No where in the world that I have been, have I stayed in a	•••• 5 .0
	hotel with the level of customer service that you experience at this JW Marriott. ABSOLUTELY the BEST!!!!	Dining
	Yes, I recommend this hotel.	Location
	I recommend this hotel for:	Service
	✓ Business Travel, Leisure, Families, Couples, Weekend, Groups, Shopping	5.0
		Amenities

b. Attribute Value Extraction:-

A review is associated with a 'star' rating that is one of those values: 1,2,3,4 and 5. The 5-star means excellent, 1star means poor. Each of the attributes cleaning, dining, location, service, amenities and value of the money are rated out of 5 by their owners and the date of posting of the review is also recorded from the website.

After preprocessing, the data set is formatted as a decision table required for the rule induction algorithm.TABLE 1 gives the decision table. Rows of the decision table represent various objects and columns represent the set of conditional attributes {Excellent, very good, average, poor, terrible} and the decision attribute {Rating}.

Table. I shows information about a couple of reviews and then overall rating.									
User	Date of posting	Customer rating							
		Cleanliness	Dining	Location	Service	amenities	Value of money	Start	
nomuniquement	August 27, 2015	5	5	5	5	5	5	5	
YKfromIreland	July 29, 2015	4	4	4	5	4	3	4	
CharlesG	July 27, 2015	5	5	4	5	5	5	5	
Alp	July 27, 2015	5	5	4	5	5	4	4	
Om	July 2, 2015	4	4	3	4	4	5	4	
unhappy	September 27, 2014	2	2	1	4	1	1	1	

Table.1 shows information about a couple of reviews and their overall rating

Table 1: Example of some attribute and they rating.

c. Decision Table Construction

Value for the Money

The decision table is a collection of data of a particular problem. It records values of attributes and decisions in a large number of samples. In the following Table attributes A1, A2, A3, A4, A5, A6 represents feature, performance, design and value respectively. The decision attribute is modeled as A5 and the star rating is numerical. Table. 2 is a decision table which would be further processed for computing decision rules.

Review	A1	A2	A3	A4	A5	A6	Decision D
1	5	5	5	5	5	5	5
2	4	4	4	5	4	3	4
3	5	5	4	5	5	5	5
4	5	5	4	5	5	4	4
5	4	4	3	4	4	5	4
6	2	2	1	4	1	1	1

Table 2: Part of decision table

Then calculated the condition attribute reduction relative to decision-making attributes, based on reduction of rough set theory and method. We used a program written in MATLAB redu.m, to achieve reduction of rough functions. Specific procedures are as follows:

```
function y=redu(c,d,x)%
y=core(c,d,x); q=ind(d,x); p=ind(c,x);
pos_cd=pospq(p,q); re=y; red=ind(y,x); pos_red=pospq(red,q);
while pos_cd~=pos_red
  cc=setdiff(c,re); [c1,c2]=size(cc);
  for i=1:c2, yy(i)=sgf(cc(i),re,d,x); end
  cd=setdiff(c,y); [d1,d2]=size(cd);
  for i=d2:-1:c2+1, yy(i)=[]; end
  [zz,ii]=sort(yy);
  for v1=c2:-1:1
     v2=ii(v1); re=cat(2,re,cc(v2)); red=ind(re,x); pos red=pospq(red,q);
  end
end
[re1,re2]=size(re);
for qi=re2:-1:1
  if ismember(re(qi),core(c,d,x)), y=re; break; end
  re=setdiff(re,re(qi)); red=ind(re,x); pos_red=pospq(red,q);
  if pos_cd==pos_red, y=re; end
end
[y1,y2]=size(y); j=1;
for i=1:y_2, [y,j]=redu_2(j,y,c,d,x); end
l{displaedu2(i,re,c,d,x)%i index,re reduce
yre=re; [re1,re2]=size(re); q=ind(d,x);
p=ind(c,x); pos\_cd=pospq(p,q);
for qi=i:re2
  re=setdiff(re,re(qi)); red=ind(re,x); pos_red=pospq(red,q);
  if pos_cd==pos_red
     y=re; j=i; break
  else
     y=yre; j=i+1; break
  end
end
```

Here only four of the attribute reduction results: location, service, amenities and value of money (A3, A4, A5, A6) to illustrate the classification rule extraction process, eliminate redundant attributes and merge the same line, the example of part of the decision table simplified as shown in Table 3.

Lubie et Entimple of the part of the simplify devision table .								
Review	A3	A4	A5	A6	Decision D			
1	5	5	5	5	5			
2	4	5	4	3	4			
3	4	5	5	5	5			
4	4	5	5	4	4			
5	3	4	4	5	4			
6	1	4	1	1	1			

Table 3: Example of the part of the simplify decision table.

d. Decision rule induction:-

A decision rule is a logical implication of the form 'IF..THEN' which belongs to entirely different domain. Typically a decision rule is of the form (a1 = v1) III. . .III(an= vn) where a1, a2,...,an are the attributes and v1,v2,...vn are their values respectively. For inducing decision rule from decision table, we have implemented Rough set theory algorithm LEM2 by using Matlab 2012a.

e. Analysis and Discovering Business Intelligence:-

Currently, most rule reduction methods focus on consistent decision-making system, the rule reduction methods applied inconsistent decision-making system is less. However, there is often inconsistent information in the current data storage systems. Therefore, we studied the direct access rules in an inconsistent system. Complete rules from the results of this attribute reduction. The rules generated out of the data set are being analyzed further for discovering business intelligence.



f. Intelligence business system architecture

Fig 3: Rule induction process.

Results and discussion:-

Result

In this experimentation, we selected 45 reviews posted from J.W Marriott hotel Ankara official website and thus a decision table is being constructed. The table has attributes X1, X2, X3, X4 and a decision attribute D. All these attributes have range of values between 1 and 5. Classes = $\{1, 2, 3, 4, 5\}$.

	Amenities X1	Money X2	Location X3	ServiceX4	DecisionD
Rule 1	-	5	5	-	5
Rule 2	-	5	4	-	5
Rule 3		3	-	5	5
Rule 4	-	4	4	5	4
Rule 5	-	5	-	-	4
Rule 6	-	4	5	-	5
Rule 7	-	-	3	-	5
Rule 8	-	3	3	-	4
Rule 9	3	-	4	-	4
Rule 10	5	-	-	4	4
Rule 11	4	4	-	4	4
Rule 12	1	-	1	-	1
Rule 13	-	-	2	-	1
Rule 14	4	3	4	4	5
Rule 15	-	-	-	1	4

Table 4: reules table

The above result shows that there are 15 rules.

In this phase we have grouped all the rules based on the decision attributes outcome. For example, from first ten rules it is clear that the rating is 1 (D=1). So the rules producing decision as 1(very low) are grouped. Rule 11 and 12 are the minimal decision rule.

Rule 1. (A2=5) & (A3=5) \Rightarrow (D = 5); 8	[17,7%]
Rule 2. (A2=5) & (A3=4) => (D = 5); 8	[17,7%]
Rule 3. $(A2=4) \& (A3=5) \Longrightarrow (D=5); 7$	[15,5%]
Rule 4. (A2=4) & (A3=4) & (A4=5) =>	(D = 4); 6 [13.33%]
Rule 5. (A $1=5$) & (A $4=4$) => (D = 4); 2	2 [4,44%]
Rule 6. (A 1= 4)& (A2=4) & (A4=4) =>	(D = 4); 2[4,44%]
Rule 7. $(A2=5) \Rightarrow (D = 4); 2$	[4,44%]
Rule 8. $(A3=3) \Rightarrow (D = 5); 2$	[4,44%]
Rule 9. (A2=3) & (A3=3) \Rightarrow (D = 4); 2	[4,44%]
Rule 10. (A $1=3$)& (A $3=4$) => (D = 4);	1 [2,22%]
Rule 11. (A $1=1$)& (A $3=1$) => (D = 1);	1 [2,22%]
Rule 12 (A3=2) => (D = 1); 1	[2.22%]
Rule 13. $(A2=3) \& (A4=5) \Rightarrow (D=5);$	1 [2.22%]
Rule 14. (A 1= 4)& (A2=3) & (A3=4) &	$x (A4=4) \Longrightarrow (D = 5); 1 [2,22\%]$
Rule 15. $(A4=1) \Rightarrow (D = 4); 1$	[2,22%]

Out of the above 15 rules, we have filtered top 5 rules based on the number of matches in the data set. Table 3. Shows top 4 rules in decreasing order of frequency. From these rules many interesting association between the attributes is discovered.

	<u>.</u>		
No	Rules	Attributes	Match
1	(Value of money=5) & (Location=5) \Rightarrow (Ranking = 5)	2	8
2	(Value of money=5) & (Location=4) $=>$ (Ranking = 5)	2	8
3	(Value of money=4) & (Location=5) \Rightarrow (Ranking = 5)	2	7
4	(Value of money=4) & (Location=4)&(service=5) => (Ranking = 4)	3	6
5	(Amenities =5) & (service=4) \Rightarrow (Ranking = 4)	2	2

Table 6: The top 5 rules

Discussion:-

The result showed clearly behind the classic business how far good business intelligence can be used for purpose of discovering business intelligence with the help of decision rules induced from the customer reviews of a product or a service posted online. However, numerical rating (1, 2, 3, 4, and 5) only does not explain the customer

dissatisfaction about the product. For example if a customer has given poor rating '1' for service, the managers, marketers cannot conclude that which of the feature made the customer dissatisfied. J.W Marriott hotel include lot feature for client satisfaction such as amenities, location, service, price, the cleanliness and the dinning. So, for an in depth analysis the textual reviews have to be studied and related keywords can be picked. e.g. a review with 4 star ranking is associated with the following comments "Only comment is the place of the hotel, sometimes it needs to be city centered for better nearby locations...!" helps the Business Intelligence system to address those issues.

Conclusion and future work:-

We have formalized the business intelligence for customer review using the theory of Rough set. Based on the proposed architecture of our business intelligence analysis system, we conducted an experiment for a specific cooperation. We've got the optimum decision making combination guidelines by using rough set model based on rule induction LEM2 algorithm. Furthermore, we've diagnosed top-five rules on the basis of number of matches. The managers could get an insight out of these rules and can further analyze it in different dimensions. They are able to find out the knowledge that makes the customer to give a high star ranking feedback about the product. This technique could be helpful in analyzing the user online review in automatic fashion rather than going through high volume of reviews manually. Business intelligence can generate more possibility in less time from a huge volume of text data recorded.

Most of the online postings are in the form of text. The future work will be extend to posted text analysis combine with the online review.

References:-

- 1. Chen, Y., and Xie, J. 2005. "Third-Party Product Review and Firm Marketing Strategy," Marketing Science (24:2), pp. 218-240.]
- 2. Bakos, J. 1997. "Reducing Buyer Search Costs: Implications for Electronic Marketplaces," Management Science (43:12), pp. 1676-1692.
- 3. Bo Pang and Lillian Lee, "Opinion mining and sentiment analysis", Foundations and Trends in Information Retrieval, Vol. 2(1-2), pp.1–135, 2008.
- 4. Hu and Liu, 2004 Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In Proceedings of the International Conference on Knowledge Discovery and Data Mining, pages 168–177, 2004
- 5. Giuseppe Carenini and Lucas Rizoli. A multimedia interface for facilitating comparisons of opinions. In Proceedings of the International Conference on Intelligent User Interfaces, pages 325–334, 2009.
- Schlosser, A. 2005. "Source Perceptions and the Persuasiveness of Internet Word-of-Mouth Communication," in Advances in Consumer Research (32), G. Menon and A. Rao (eds.), Duluth, MN: Association for Consumer Research, pp. 202-203.]
- 7. Pavlou, P., Liang, H., and Xue, Y. 2007. "Uncertainty and Mitigating Uncertainty in Online Exchange Relationships: A Principal– Agent Perspective," MIS Quarterly (31:1), pp. 105-131.]
- 8. Burgess, Heidi, Guy Burgess and Michelle Maiese. "Incrementalism." Beyond Intractability, Conflict Research Consortium, University of Colorado, Boulder (2004) source link. http://www.beyondintractability.org/essay/incrementalism
- Guo, S Wang, Z Y Wu, Z C & Yan, "A novel dynamic incremental rules extraction algorithm based on Rough set theory". In proceedings of the fourth International Conference on machine learning and cybernetics pp 18 – 21
- 10. Grzymala-Busse J W, "Rule Induction", Chapter 1, pp 01-19 in Intelligent Decision support Handbook of Application and Advances of the Rough set Theory (Ed : Slowinski. R), Volume 11,(1992).
- 11. Düntsch, I., Gediga, G. Rough Set Data Analysis A Road to Non-invasive Knowledge Discovery. Angor: Methodos, 2000, 107 p,14
- [12]- Komorowski, J., Pawlak, Z., Polkowski, L., Skowron, A. Rough sets: A tutorial. In: S.K. Pal and A. Skowron (Eds.), Rough-Fuzzy Hybridization: A New Trend in Decision-Making. Singapur :Springer Verlag. 1998, pp.3-98.