Journal of Engineering Sciences
Vol 10, Issue 12, DEC/2019
ISSN NO:0377-9254

MINING CONTEST FROM LARGE TRADITIONAL INFORMATION

Shabeer Mohammed Abdul[1], Dr.Thayyaba Khatoon Mohammed [2]

[1]M.Tech Student, Dept of CSE, Malla reddy college of Engineering and Technology, Hyderabad, T.S, India

[2] Associate Professor, Dept of CSE, Malla reddy college of Engineering and Technology, Hyderabad, T.S, India

ABSTRACT:

In any competitive business, success depends on the ability to make the article more attractive to customers than the competition. Several questions arise in the context of this mission: How can we formalize and measure competitiveness between two elements? Who are the main competitors of a particular article? What are the advantages of the article that most influence your competitiveness? Despite the impact and importance of this problem in many areas, only a limited amount of work has been dedicated to an effective solution. In this document, we present a formal definition of competitiveness between two elements, based on the market segments that both can be covered. Our competitiveness assessment uses customer opinions and is an abundant source of information available in a wide range of areas.

Keywords: Data mining, Web mining, Information Search and Retrieval, Electronic commerce

1. INTRODUCTION:

There is a lot of personal information in online text reviews, which plays a very important role in decision-making processes. For example, the customer will decide what to buy if he sees valuable comments posted by others, especially the trusted friend. We believe that reviews and reviewers will help predict ratings based on the idea that high star ratings can be attributed largely to good reviews. Therefore, how to conduct review reviews and the relationship between auditors on social networks has become a major problem in web mining, machine learning and natural language processing. We focus on the task of evaluating the prediction. However, the star rating level information is not always available to the user in many review sites. On the other hand, the reviews contain sufficient detailed product information and user opinion information, which have a great reference value for the user's decision. Above all, a specific user on the website cannot rate each item. Consequently, there are many unclassified elements in the user element classification matrix. It is inevitable in many classification prediction methods, for example [1], [4]. Review / Comment, as we all know, is always available. In this case, it is appropriate and necessary to take advantage of user opinions to help predict unclassified elements. Boarding sites and other review sites provide a broad idea to extract user preferences and forecast user ratings. In general, the interests of the users are stable in the short term, so the user topics of the reviews can be representative. For example, in the category of cups and cups, different people have different tastes. Some people care about quality, some focus on price and others can evaluate comprehensively. Whatever it is, everyone has their own themes. Most theme templates provide interest to users as theme distributions according to the content of the reviews. It is widely applied in the analysis of the feeling of travel recommendation and the analysis of social networks. Sentiment analysis is the most important and essential work to extract user preferences. In general, emoticons are used to describe a user's attitude towards the elements. We note that in many practical cases it is more important to provide numerical degrees rather than binary decisions. In general, the reviews are divided into two groups, positive and negative. However, it is difficult for customers to choose when all candidate products

reflect a positive or negative feeling. To make a purchase decision, customers not only need to know if the product is good, but they also need to know how good the product is. It is also agreed that different people may have different preferences for emotional expression. For example, some users prefer to use the word "good" to describe an "excellent" product, while others prefer to use "good" to describe a "very good" product. In our daily lives, customers are more likely to buy these products with very favorable reviews. That is, customers are more interested in the reputation of the item, which reflects the overall assessment of consumers based on the intrinsic value of a particular product. To get the reputation of a product, you need to feel the criticism. Usually, if item reviews reflect positive emotions, then the article can be a very good reputation. On the contrary, if the reviews of the articles were full of negative feelings, then the article would have a bad reputation. For a particular product, if we know the user's feeling, we can deduct the reputation and even the general qualifications. When we look for online purchases, both positive and negative reviews are valuable as a reference. For positive reviews, we can see the advantages of the product. For negative reviews, we can get shortcomings if they cheat you. Therefore, it is worth exploring these reviewers who have a clear and objective position on the elements. We note that the morale of the reviewers will affect others: if the reviewer has a similar feeling and hates the feelings, other users will be very interested in it. However, user morale is difficult to predict, and the inability to predict the emotional impact among people makes it very difficult to explore social users. In addition to extracting user preferences, there are many

2. TERMINOLOGY AND PROBLEM STATEMENT

Despite the impact of this problem and its relevance in many areas, a limited amount of work has been devoted to an effective solution. In this document, we present a formal definition of competitiveness between two elements, based on the market segments that both can be covered. Our competitiveness assessment uses customer opinions and is an abundant source of information available in a wide range of areas. We provide effective methods to assess competitiveness in large audited datasets and address the natural problem of finding large competitors in a particular component. Finally, we evaluate the quality of our results and the possibility of developing our approach using multiple data sets from different areas. The effectiveness of our methodology has been validated by the experimental evaluation of real data sets from different fields. We offer effective methods to assess competitiveness in large audit data sets and address the natural problem of finding the best competitors in a particular component. We are presented to the market a set of elements n I and a set of characteristics F. Then, looking at a component i 2 I, we want to define the elements k of I that increase the CF (i,). In addition, a simplified MapReduce application will face a bottleneck to pass everything through the reducer to the self-registration account included in the account.

A. COLLABORATIVE FILTERING

CF's task is to predict user preferences for unclassified items, after which a list of most users' favorite items can be recommended. To improve the performance of the recommendation, several CF algorithms have been proposed. One of the best known CF algorithms is the CF algorithm based on the proposed user. The basic idea is that people who have expressed similar preferences in the past would prefer to buy similar products in the future, suggest a general method that allows you to incorporate tags into standard CF algorithms and merge 3D connections between users, articles and tags. In addition, the element-based CF algorithm produces a user-to-article classification based on average ratings of similar or related articles by the same user. Improves the calculation of similarities between elements, a collaborative recommendation algorithm for audit experts based on the assumption that those projects / experts that have similar themes have vectors of similar characteristics suggest a CF-based service

3. LITERATURE SURVEY

A. Competitive Strategy: Techniques for Analyzing Industries and Competitors

Five competitive forces act on an industry:
(1) threat of new entrants,
(2) intensity of rivalry among existing firms, (3) threat of substitute products or services, (4) bargaining power of buyers, and

(5) bargaining power of suppliers. Looking at industry structure provides a way to consider how value is created and divided among existing and potential industry participants. One competitive force always captures essential issues in the division of value.

B.Competitive analysis

A competitive analysis is a critical part of your company marketing plan. With this evaluation, you can establish what makes your product or service unique--and therefore what attributes you play up in order to attract your target market.

Evaluate your competitors by placing them in strategic groups according to how directly they compete for a share of the customer's dollar. For each competitor or strategic group, list their product or service, its profitability, growth pattern, marketing objectives and assumptions, current and past strategies, organizational and cost structure, strengths and weaknesses, and size (in sales) of the competitor's business.

C. Managerial Identification of Competitors

Despite extensive academic research on how to identify competitors objectively, marketers know relatively little about how managers identify competitors in practice. The authors bring together diffuse literature in this area and propose a cognitive framework for managerial identification of competitors. They report the results of two studies that examine the attributes managers use in deciding who their competitors are.

4. IMPLEMENTING DYNAMIC FACETED SEARCH

A formal definition of competitiveness between two elements, based on its appeal to various customer sectors in the market. Our approach overcomes the dependence of previous work on rare comparative evidence extracted from the text. A formal methodology to identify different types of customers in a given market, as well as to estimate the percentage of customers that belong to each type. Significantly extensible framework to find the most sought-after competitors for a particular element in very large data sets, such as the position of the elements in the space of multidimensional features and user preferences and opinions for a user who loves element i, element j which is much higher than i in relation to the user's requirements (and therefore varies absolutely) is a better recommendation filter than the very similar j element.

A.EXTRACTING PRODUCT FEATURES

The characteristics of the product mainly focus on the problems discussed in the product. In this document, we are extracting the product features of text revisions using LDA. We primarily want to obtain product features, including some named entities and some product / item / service attributes. LDA is a Bayesian model, used to model the relationship of revisions, themes and words in which the shaded variables refer to observed variables and the non-shaded variables refer to underlying variables. The arrow indicates a conditional dependence between the variables and the panels represented by the square. The definition of terms in the LDA is described below:

1.**V**: vocabulary, has Nd unique words. Each word is presented by the corresponding classification {1, 2, ⋯, N}. ∈ ∈ {1, 2, ⋯, N}: the word, each word of a revision is set to $\boldsymbol{V}$ whose size is matching the letters. $\boldsymbol{d}m$: The document / user revision corresponds to a set of words of the revision. User with only one document. All documents refer to D = {d1 ,, 2, ⋯, dM}.

2.**Γ**: number of subjects (numerical const). **m m**: polynomial distribution of specific topics of the document m. A relation for each document, Θ = {θ m} m = 1 M (matrix M x Γ) $\vec{\boldsymbol{\varphi}}$: the component for each subject, Φ = {$\vec{\varphi}$ k} k = 1 Γ (matrix Γ x k) $\boldsymbol{zm}$, $\boldsymbol{n}$: the subject associated with the symbol n-th in document m. $\boldsymbol{a}$, $\boldsymbol{b}$: Dirrichlet providers of the polynomial distribution θ m and $\vec{\varphi}$.

B. BOOSTING THE CMINER ALGORITHM

Next, we describe several improvements we have made to CMiner to achieve mathematical savings while maintaining the exact nature of the algorithm

C. QUERY ORDERING

Our complexity analysis is based on the premise that CMiner evaluates all queries Q for each candidate

item j. However, this assumption naively ignores the algorithm's pruning ability, which is based on using lower and upper bounds on competitiveness scores to eliminate candidates early. Next, we show how to greatly improve the algorithm's pruning effectiveness by strategically selecting the processing order of queries (line 23 of CMiner). CMiner uses the following update rules for the lower and upper bounds for a candidate j:

$$low(j) \leftarrow low(j) + p(q) \times V\ q\ i,j$$

$$up(j) \leftarrow up(j) - p(q) \times V\ q\ i,i + p(q) \times V\ q\ i,j$$

By expanding the sequences and using the initial values low(j) = 0 and up(j) = CF (i, i), we can re-write the bounds:

$$lowm(j) = \sum m\ 1\ p(qm) \times V\ qm$$

$$upm(j) = CF\ (i, i) - \sum m\ 1\ p(qm) \times V\ qm\ i,i + \sum m\ 1\ p(qm) \times V\ qm\ i,j\ ,$$

where lowm(j) and upm(j) are the values of the bounds after considering the mth query qm. We can then define a recursive function T(j) = up(j) − low(j) as follows:

$$T(j) \leftarrow T(j) - p(q) \times V\ q\ i,i$$

T(j) captures the margin of error for the competitiveness between the item of interest i and a candidate j. As more queries are evaluated and the two bounds are updated, the margin decreases. Finally, it becomes equal to zero when we have the final CF (i, j) score. We hypothesize that the ability to minimize this margin faster can increase the number of pruned candidates due to the existence of stricter bounds in early iterations the value of T(j) after considering m queries can be re-written as follows:

$$T\ m(j) = CF\ (i, i) - \sum m\ \ell{=}1\ p(q\ell) \times V\ q\ell\ i,i\ ,$$

where qℓ is the ℓth query processed by the algorithm. Given above it is clear that we can optimally minimize the margin between the lower and upper bounds on the competitiveness of a candidate by processing queries in decreasing order of their p(q) × V q i,i values. We refer to this ordering scheme as COV. We evaluate the computational savings achieved by COV in our experiments, where we also compare it with alternative approaches.

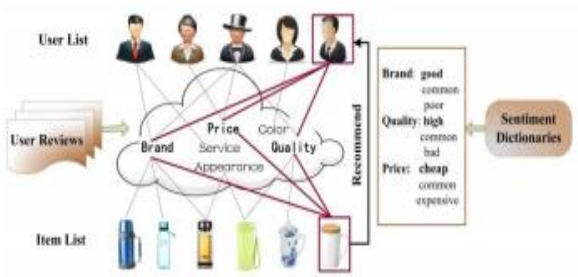

Figure 1: System Architecture

5. CONCLUSION:

We provide a formal definition of competitiveness between two elements, which we verify in quantitative and qualitative terms. Our formalization applies to areas, overcoming deficiencies in previous approaches. We consider a series of factors that have been overlooked in the past, such as the position of the elements in the space of multidimensional characteristics and the preferences and opinions of the users. Our work provides a comprehensive methodology to extract this information from large data sets for customer reviews. According to our definition of competitiveness, we address the problem of the computational challenge of finding the best competitors in a particular element. The proposed framework is effective and applicable to areas with a large number of elements. The effectiveness of our methodology was verified by experimental evaluation of real data sets from various fields. Our experiences also revealed that only a small number of reviews is sufficient to estimate differently the different types of users in a given market, as well as the number of users belonging to each type.

REFERENCES:

[1] M. E. Porter, Competitive Strategy: Techniques for Analyzing Industries and Competitors. Free Press, 1980.

[2] R. Deshpand and H. Gatingon, "Competitive analysis," Marketing Letters, 1994.

Journal of Engineering Sciences

[3] B. H. Clark and D. B. Montgomery, "Managerial Identification of Competitors," Journal of Marketing, 1999.
[4] W. T. Few, "Managerial competitor identification: Integrating the categorization, economic and organizational identity perspectives," Doctoral Dissertaion, 2007.
[5] M. Bergen and M. A. Peteraf, "Competitor identification and competitor analysis: a broad-based managerial approach," Managerial and Decision Economics, 2002.
[6] J. F. Porac and H. Thomas, "Taxonomic mental models in competitor definition," The Academy of Management Review, 2008.
[7] M.-J. Chen, "Competitor analysis and interfirm rivalry: Toward a theoretical integration," Academy of Management Review, 1996.
[8] R. Li, S. Bao, J. Wang, Y. Yu, and Y. Cao, "Cominer: An effective algorithm for mining competitors from the web," in ICDM, 2006.
[9] Z. Ma, G. Pant, and O. R. L. Sheng, "Mining competitor relationships from online news: A network-based approach," Electronic Commerce Research and Applications, 2011.
[10] R. Li, S. Bao, J. Wang, Y. Liu, and Y. Yu, "Web scale competitor discovery using mutual information," in ADMA, 2006

Journal of Engineering Sciences