ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 9, March 2018

MINING COMPETITIVE STRATEGY AND ANALYSIS USING CLUSTERING ALGORITHM

Mrs.Vidhya[1],
Teaching fellow,
Computer Science and Engineering,
University college of Engineering,
Villupuram.

S.Karthish[2],M.Dinesh[3],C.Suriya[4]
R, Mohammedthaga[5],
UG students/ CSE
UCE,Villupuram.

Abstract- In any competitive business, success is based on the ability to make an item more appealing to customers than the competition. A number of questions arise in the context of this task: how do we formalize and quantify the competitiveness between two items? Who are the main competitors of a given item? What are the features of an item that most affect its competitiveness? Despite the impact and relevance of this problem to many domains, only a limited amount of work has been devoted toward an effective solution. In this paper, we present a formal definition of the competitiveness between two items, based on the market segments that they can both cover. Our evaluation of competitiveness utilizes customer reviews, an abundant source of information that is available in a wide range of domains. We present efficient methods for evaluating competitiveness in large review datasets and address the natural problem of finding the top-k competitors of a given item. Finally, we evaluate the quality of our results and the scalability of our approach using multiple datasets from different domains.

Index Terms—Data mining, Web mining, Information Search and Retrieval, Electronic commerce.

1]Introduction

Along line of research has demonstrated the strategic importance of identifying and monitoring a firm's competitors.Motivated by this problem, the marketing and management community have focused on empirical methods for competitor identification,as well as on methods for analyzing known competitors.Extant research on the former has focused on mining comparative expressions (e.g. "Item A is better than Item B") from the Web or other textual sources Even though such expressions can indeed be indicators of competitiveness, they are absent in many domains. For instance, consider the domain of vacation packages (e.g flight-hotel-car combinations). In this case, items have no assigned name by which they can be queried or compared with each other. Further, the frequency of textual comparative evidence can vary greatly across domains. For example,when comparing brand names at the firm level (e.g. "Google vs Yahoo" or "Sony vs Panasonic"), it is indeed likely that comparative patterns can be found by simply querying the web. However, it is easy to identify mainstream domains where such evidence is extremely scarce, such as shoes,jewelery, hotels, restaurants, and furniture. Motivated by these shortcomings, we propose a new formalization of the competitiveness between two items, based on the market segments that they can both cover. Our competitiveness paradigm is based on the following observation: the competitiveness between two items is based on whether they compete for the attention and business of the same groups of customers (i.e. the same market segments). For example, two restaurants that exist in different countries are obviously not competitive, since there is no overlap between their target groups.

2]Literature Review

Estimating aggregate consumer preferences from online product reviews. An econometric framework is presented that can be applied to the mentioned type of data after having prepared it using natural language processing techniques. The suggested methodology enables the estimation of parameters, which allow inferences on the relative effect of product attributes and brand names on the overall evaluation of the products. Specifically, we discuss options for taking opinion heterogeneity into account.Both the practicability and the benefits of the suggested approach are demonstrated using product review data from the mobile phone market.This paper demonstrates that the review-based results compare very favorably with consumer preferences obtained through conjoint analysis techniques.

A probabilistic rating inference framework for mining user preferences from reviews

We propose a novel Probabilistic Rating inference Framework, known as Pref, for mining user preferences from reviews and then mapping such preferences onto numerical rating scales.Pref applies existing linguistic processing techniques to extract opinion words and product features from reviews. It then estimates the sentimental orientations (SO) and strength of the opinion words using our proposed relative-frequency-based method.This method allows semantically similar words to have different SO, thereby addresses a major limitation of existing methods.Pref takes the intuitive relationships between class labels, which are scalar ratings, into consideration when assigning ratings to reviews.

Identifying Customer Preferences about Tourism Products Using an Aspect-based Opinion Mining Approach

An approach for considering a new alternative to discover consumer preferences about tourism products, particularly hotels and restaurants, using opinions available on the Web as reviews.An experiment is also conducted, using hotel and restaurant reviews obtained from Trip Advisor, to evaluate our proposals.Results showed that tourism product reviews available on web sites contain valuable information about customer preferences that can be extracted using an aspect-based opinion mining approach.The proposed approach proved to be very effective in determining the sentiment orientation of opinions, achieving a precision and recall of 90%. However, on average, the algorithms were only capable of extracting 35% of the explicit aspect expression

3]Proposed work

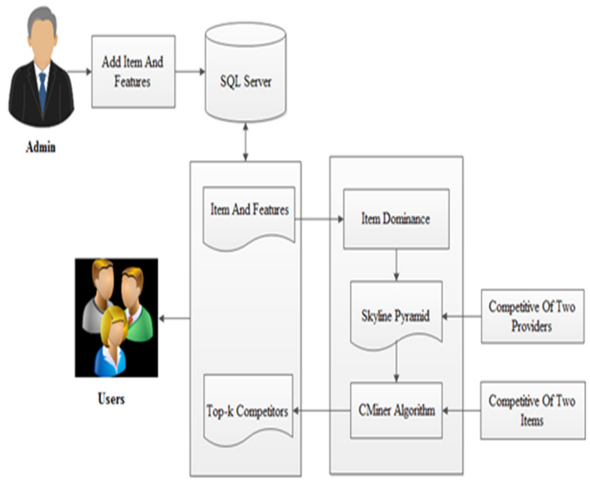
Competitor Mining Algorithms, identify key competitive measures (e.g., market share, share of wallet) and showed how a firm can infer the values of these measures for its competitors by mining,

a. Its own detailed customer transaction data and
b. Aggregate data for each competitor.

Finding Competitive Products has explored competitiveness in the context of product design. The first step in these approaches is the definition of a dominance function that represents the value of a product.

Skyline Computation, Our work leverages concepts and techniques from the literature on skyline computation and has ties to the recent publications in reverse skyline queries.

The **CMiner Algorithm**, for finding the top-k competitors of a given item. Our algorithm makes use of the skyline pyramid in order to reduce the number of items that need to be considered.



4]Experimental Results

In this section we describe the experiments that we conducted to evaluate our methodology. All experiments were completed on an desktop with a Quad-Core 3.5GHz Processor and 2GB RAM. Our experiments include four datasets, which were collected for the purposes of this project. The datasets were intentionally selected from different domains to portray the cross-domain applicability of our approach. In addition to the full information on each item in our datasets, we also collected the full set of reviews that were available on the source website. These reviews were used to (1) estimate queries probabilities, as described in Section 2.2 and (2) extract the opinions of reviewers on specific features.The highly-cited method by Ding et al. is used to convert each review to a vector of opinions, where each opinion is defined as a feature-polarity combination (e.g. service+, food-). The percentage of reviews on an item that express a positive opinion on a specific feature is used as the feature's numeric value for that item. We refer to these as *opinion features*. Table 4 includes descriptive statistics for each dataset, while a detailed description is provided below.

CAMERAS: This dataset includes 579 digital cameras from Amazon.com. We collected the full set of reviews for each camera, for a total of 147192 reviews. The set of features includes the *resolution* (in MP), *shutter speed* (in seconds),*zoom* (e.g. 4x), and *price*. It also includes opinion features on *manual, photos, video, design, flash, focus, menu options, lcd*m *screen, size, features, lens, warranty, colors, stabilization, battery life, resolution*, and *cost*.

HOTELS: This dataset includes 80799 reviews on 1283 hotels from Booking.com. The set of features includes the *facilities*, *activities*, and *services* offered by the hotel. All three of these multi-categorical features are available on the website. The dataset also includes opinion features on *location, services, cleanliness, staff*, and *comfort*.

RESTAURANTS: This dataset includes 30821 reviews on 4622 New York City restaurants from TripAdvisor.com. The set of features for this dataset includes the *cuisine types* and *meal types* (e.g. lunch, dinner) offered by the restaurant, as well as the *activity types* (e.g. drinks, parties) that it is good for. All three of these multi-categorical features are available on the website. The dataset also includes opinion features on *food,service, value-for-money, atmosphere*, and *price*.

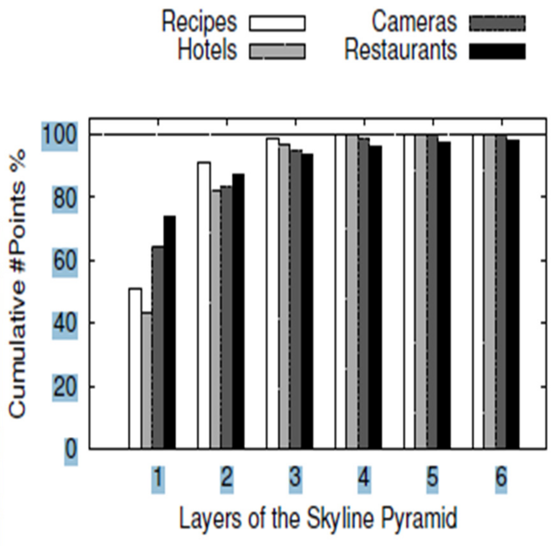
RECIPES: This dataset includes 100000 recipes fromSparkrecipes.com. It also includes the full set of reviews on each recipe, for a total of 21685 reviews. The set of featuresfor each recipe includes the *number of calories*, as well as the following nutritional information, measured in grams:*fat, cholesterol, sodium, potassium, carb, fiber, protein, vitamin A, vitamin B12, vitamin C, vitamin E, calcium, copper, folate,magnesium, niasin, phosphorus, riboflavin, selenium, thiamin,zinc*. All information is openly available on the website.

Skyline

Dataset	**#Items**	**#Feats.**	**#Subsets**	**Layers**
CAMERAS	579	21	14779	5
HOTELS	1283	8	127	5
RESTAURANTS	4622	8	64	12
RECIPES	100000	22	133	22

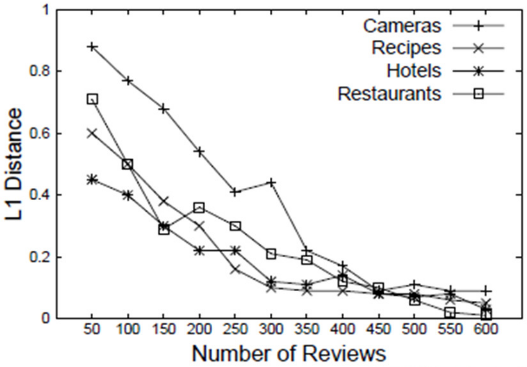
For each dataset, the 2nd, 3rd, 4th and 5th columns

All datasets, nearly 99% of the items can be found within the first 4 layers, with the majority of those falling within the first 2 layers. This is due to the large dimensionality of the feature space, which makes it difficult for items to dominate one another. As we show in our experiments, the skyline pyramid enables CMiner to clearly outperform the baselines with respect to computational cost. This is despite the high concentration of items within the first layers, since CMiner can effectively traverse the pyramid and consider only a small fraction of these items



Convergence of Query Probabilities

In this we described the process of estimating the probability of each query by mining large datasets of customer reviews. The validity of this approach is based on the assumption that the number of available reviews is sufficient to allow for confident estimates. Next, we evaluate this assumption as follows. First, we merge all the reviews in each dataset into a single set, sort them by their submission date, and split the sorted sequence into fixed-size segments. We then iteratively append segments to the review corpus considered and re-compute the probability of each query in the extended corpus. The vector of probabilities from the *ith* iteration is then compared with that from the (*i*□1)*th* iteration via the L1 distance: the sum of the absolute differences of corresponding entries (i.e. the two estimates for the same query in both vectors).We apply the process for segments of 25 reviews.The x-axis of each plot includes the number of reviews, while the y-axis is the respective L1 distance.



The convergence of the probabilities is an especially encouraging outcome that (i) reveals a stable categorical distribution for the preferences of the users over the various queries, and (ii) demonstrates that only a small seed of reviews, that is orders of magnitude smaller than the thousands of reviews available in each dataset, is sufficient to achieve an accurate estimation of the probabilities.

5] Conclusion

We presented a formal definition of competitiveness between two items, which we validated both quantitatively and qualitatively. Our formalization is applicable across domains, overcoming the shortcomings of previous approaches. We consider a number of factors that have been largely overlooked in the past, such as the position of the items in the multi dimensional feature space and the preferences and opinions of the users. Our work introduces an end-to-end methodology for mining such information from large datasets of customer reviews. Based on our competitiveness definition, we addressed the computationally challenging problem of finding the top-k competitors of a given item. The proposed framework is efficient and applicable to domains with very large populations of items. The efficiency of our methodology was verified via an experimental evaluation on real datasets from different domains.

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.

[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.