MICANS INFOTECH

MINING COMPETITORS
FROM LARGE UNSTRUCTURED DATASETS

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 tha 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 custome reviews, an abundant source of information that is available in a wide range of domains.

CONT..

- We propose a unique perspective to achieve this goal, which is quantifying user vitality by analyzing the dynamic interactions among users on social networks. Examples of social network include but are not limited to social networks in micro blog sites and academicals collaboration networks.

- If a user has many interactions with his friends within a time period and most of his friends do not have many interactions with their friends simultaneously,

- Based on this idea, we develop quantitative measurements for user vitality and propose our first algorithm for ranking users based vitality. Also we further consider the mutual influence between users while computing the vitality measurements and propose the second ranking algorithm, which computes user vitality in an iterative way.

EXISTING SYSTEM

- 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.

DISADAVANTAGE

- The efficiency of our methodology was verified via an experimental evaluation on real datasets from different 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.

- We are presentedwith a market with a set of n items I and a set of featuresF. Then, given a single item i 2 I, we want to identify the k items from I that maximize CF(i,).

- In addition, a naive MapReduce implementation would face the bottleneck of passing everything through the reducer to account for the self-join included in the computation.

PROPOSE SYSTEM

- A formal definition of the competitiveness between two items, based on their appeal to the various customer segments in their market.

- Our approach overcomes the reliance of previous work on scarce comparative evidence mined from text.

- A formal methodology for the identification of the different types of customers in a given market, as well as for the estimation of the percentage of customers that belong to each type.

- A highly scalable framework for finding the top-k competitors of a given item in very large datasets

ADVANTAGES

- Such as the position of the items in the multi-dimensional feature space and the preferences and opinions of the users

- For a user who likes an item i, an item j that is far superior than i with respect to the user's requirements (and thus quite different) is a better recommendation candidate than an item j′ that is highly similar.

SOFTWARE REQUIREMENT

- **HARDWARE REQUIREMENT:**
- System   :   Pentium IV 2.4 GHz.
- Hard Disk   :   40 GB.
- Floppy Drive   :   1.44 Mb.
- Monitor   :   15 VGA Colour.
- Mouse   :   Sony.
- Ram   :   512 Mb.

Cont..

- **SOFTWARE REQUIREMENT:**
- Operating system   :   Windows XP.
- Coding Language   :   ASP. Net with C#
- Data Base   :   SQL Server 2005.

MICANS INFOTECH

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 beenlargely 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.