

Abduganieva M.

"Digital economy and information technologies" department

senior teacher

ASSOCIATIVE RULES AND MARKET BASKETS

Abstract: In this article, modern service and commercial organizations collect accurate information about each order made through plastic cards and control computer systems, using data recording and storage technology. collecting a large amount of information about purchases, orders and services made by consumers, identifying patterns in the behavior of buyers by experts in the field of management and marketing, their consumer knowledge, behavior in managing the organization's marketing and product policy and increasing the organization's income and competitiveness , in the field of modern information technologies, the issues of the analysis of statistical data analysis tools collected using the intellectual analysis of data are highlighted.

Key words: Associative rules, Apriora, Market baskets, Genetic algorithms, Cell processors, Intel processors.

INTRODUCTION

The development of computer technology leads to an increase in the amount of data that needs to be stored. This makes it difficult for a person to work with data. The importance of analytics in working with data is undoubtedly huge, because it leads to the construction of knowledge among the "unprocessed data". This knowledge can be used in decision making. Therefore, recently the direction of "knowledge discovery in databases" is developing rapidly. Nowadays, the size of the data warehouse is the main reason for the emergence of new scalable algorithms.

Data Mining is the process of finding useful, previously unknown, practically useful and interpretable knowledge from "unprocessed" data. This knowledge plays an important role in the decision-making process in various areas of human life. [1]

The information revealed by the application of data mining methods should be non-trivial and unfamiliar, for example, average sales cannot be an example of this. Identification of new relationships between the properties of knowledge, one predetermines the properties of the other, and so on. The established knowledge must be applied to new information with some degree of reliability.

Every knowledge should be useful in some direction. This knowledge should be understandable and simple to view for the non-mathematical user. For example, logical constructions easily accepted by humans "if ... then ...". In addition, these rules can be applied via SQL queries to different MBBTs. If the data obtained at the beginning is not comprehensible, there must be methods of processing that bring the data to a comprehensible level for the user. Algorithms used in Data Mining require a large number of calculations. Previously, this fact was seen as a difficult problem for Data Mining, but nowadays the development of modern processors has reduced the importance of this problem. [2]

Data Mining in Problem Solving different algorithms and methods are used. Among them, the most widely used are: neural networks, tree-based solutions, clustering algorithms, including scalable algorithms that identify associative relationships between events and causality.

Effectively scalable algorithms for finding associative rules are required as the data warehouse expands day by day. These rules make it possible to solve problems quickly and easily.

used knowledge discovery methods is the associative rule discovery algorithm. First associative rule discovery method called AIS 1993, IBM Almaden Research Center developed by employees. Since then, great attention has been paid to this dish. The mid -90s is the peak period of openings in this direction.

Today, mainly Apriori algorithm is used to determine associative rules it is used . Its author is Rakesh Agrawal (Rakesh Agrawal). [3]

are used to identify regularities between events . As an example of this, it is possible to cite the following statement : a consumer who wants to buy bread has a 75% chance of buying milk as well .[4]

Market basket analysis (market basket analysis) is the search for the most typical, patterned purchases in supermarkets (search for imaginary rules). Market basket analysis is done through database analysis to identify related product combinations. In other words, it defines "paired goods". One of these pairs of goods is key, and the goods purchased with it are companion goods. This analysis helps to determine the rate of purchase of paired products and the probability that the companion product will be purchased with key products.

Literature analysis on the research topic

Foreign scientists on the analysis of associative rules and market baskets, including R. Agrawal, T. Imielinski, A. Swami, R. Srikant. A. Savasere, E. Omiecinski, and S. Navathe, JS Park, M.-S. Chen, and SY Philip, J. Hipp, U. Guntzer, and G. Nakaeizadeh . many results are given in their works.

As one of the effective algorithms A priori and G genetic algorithms are shown.

Research methodology

To solve the given problem, associative rules and algorithms in data analysis and a parallel algorithm for analysis of market baskets adapted for a computing system based on Cell processors were used.

The purpose and objectives of the research. Optimization of analysis of market baskets based on selected methods and algorithms. To achieve this goal, the following issues were resolved:

- Study associative rules and gather theoretical information from existing literature;

- Apriora - analysis of the algorithm of the associative rule and theoretical information about it;
- Analysis of the Genetic Algorithm of solving problems of analysis of market baskets;
- Analysis of a parallel algorithm for solving problems of analysis of market baskets;
- Analysis of calculations based on learned algorithms;

Development of reasonable proposals and recommendations for improvement of algorithms and methods .

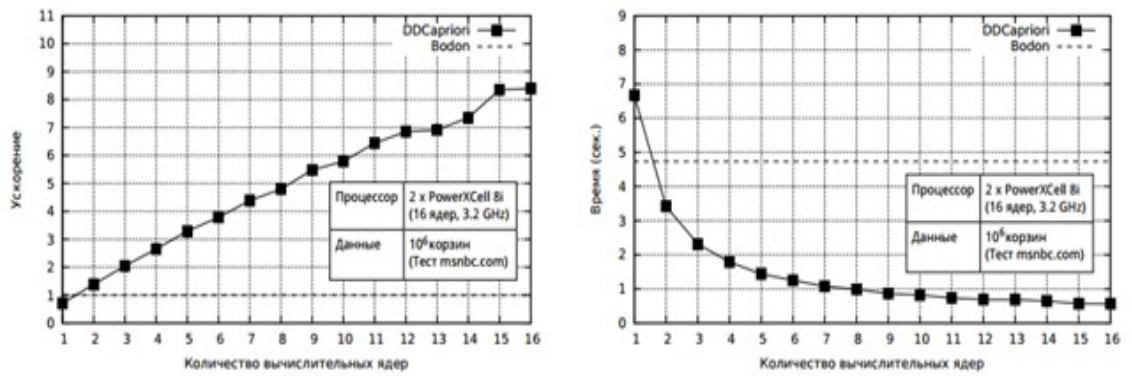
Research results

The theoretical significance of the research is the search for associative rules and the development of algorithms for solving the problems of market basket analysis.

of the dissertation , creation of a parallel algorithm developed on the basis of associative rules and the implementation of data analysis of this algorithm and calculations based on Cell processors.

To evaluate the effectiveness of the developed algorithm, we conducted three series of computational experiments. As external data of the experiments, a standard test set of data on visits to the pages of the website msnbc.com, which was used to evaluate the effectiveness of the Data Mining algorithm, was obtained. Set B test task displays records of visits to site pages. Each record has a label indicating which content category it belongs to. In the experiments, search sets of frequently visited pages were performed.

In the first series of experiments, we determined the speedup of the algorithm depending on the computing core. It is considered one of the best algorithms for solving tasks of market basket analysis today when calculating the acceleration per unit of performance of consistent algorithms. Experiments showed that the DDCapriori algorithm showed a near-linear acceleration.



a) acceleration b) analysis over time

Figure 1. Performance of DDCapriori algorithm

In addition, we compared the dimensions of the developed algorithm and the Count Distribution algorithm for Cell Count Distribution. The comparison showed that the developed algorithm has several good dimensions

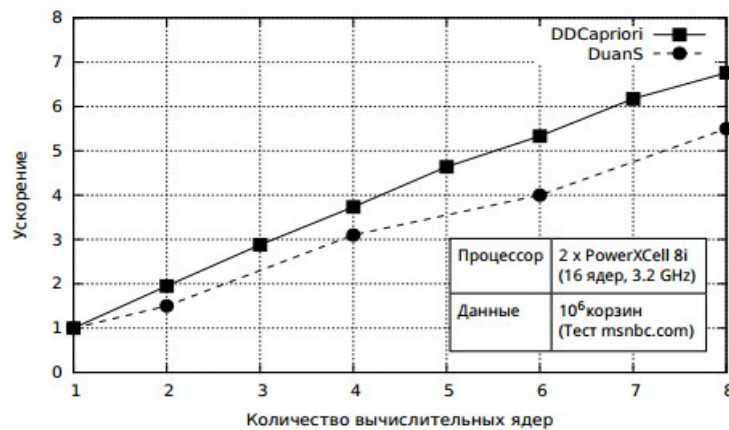
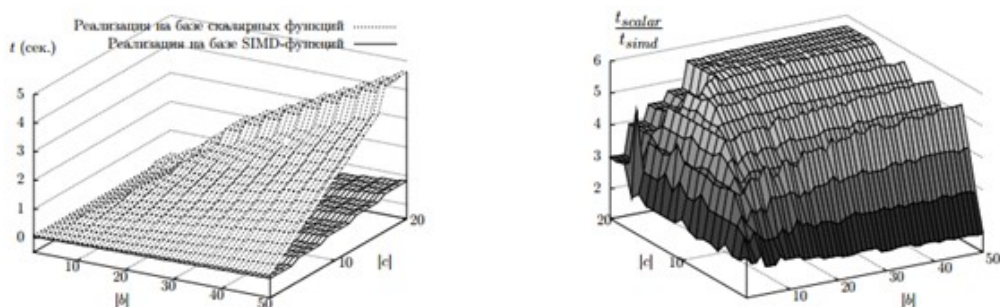


Figure 2. Comparison of DataDistribution and CountDistribution implementation.

In the second series of experiments, we compared the gain from using vector operations with scalar in checking candidate access to the basket as a function of candidate length and basket. The results of the experiments showed that the gain from vector operations is directly proportional to the length of the candidate and the basket.



a) Time check b) vectorial back check _

Figure 3. Checking the candidate's access to the basket with a vector function .

The third series of experiments is aimed at comparing the performance of the developed algorithm on the Cell and Intel platforms. To conduct these experiments, we developed the DDCapriori algorithm for the Intel processor. Together with the SPE-system, POSIX-system is used to do this and vector functions are not used. The results of the experiments are presented in Figure 10.

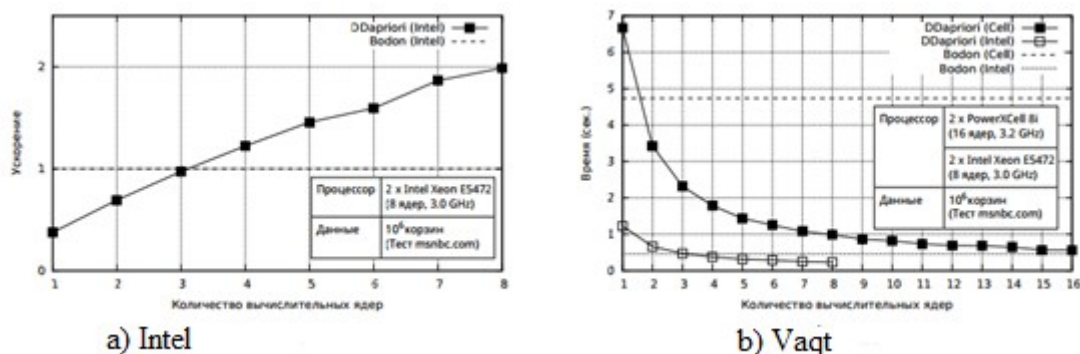


Figure 3. Compare the performance of Cell and Intel processors

Experiments showed that the algorithm on Cell processors showed several better accelerations than the algorithm on Intel processors. But Intel processors provide much faster performance than Cell processors.

Summary

This research leads to the construction of knowledge among the raw data. This knowledge can be used in decision making. Therefore, we can see as a clear

example that in recent years, the direction of "knowledge discovery in databases" has been developing rapidly.

Data Mining is the process of finding useful, previously unknown, practically useful and interpretable knowledge from "unprocessed" data. This knowledge plays an important role in the decision-making process in various areas of human life.

In particular, the main concepts of the market basket are explained, Data, which allows processing large amounts of data and finding the necessary information effectively. As one of the levers of mining, associative rules are considered. As mentioned above, the problem of searching for associative rules was originally presented for the analysis of the market basket.

used in purchases, analysis of customer preferences, planning of placement of goods in supermarkets, cross-marketing, segmentation of customer behavior in address shipping . However, the field of application of these algorithms is not limited to trading.

A general method of parallelization was created using the data from the research results.

The basic concepts for the description of the above-mentioned algorithms are expressed.

Practical examples of the directions of market baskets are presented.

A parallel algorithm for solving market basket analysis tasks adapted for a computing system based on Cell and Intel processors is presented. Parallel algorithms are achieved by dividing the set into groups and distributing these groups across computing cores. In this case, the basket set is transferred to each computing core.

The results of computational experiments showing the effectiveness of the proposed algorithm are presented.

Many experiments were conducted with the help of the given algorithm and model, and final conclusions were drawn based on the experimental results.

List of used literature

1. Wang, N. _lay, and V. Liu. interestingness-Based Interval Merger for Numeric Association Rules. In *Proc. of the 4 th Intl. Conf. on Knowledge Discovery and Data Mining* , pages 121-128, New York, NY, August 2014.
2. I. Webb. Preliminary investigations into statistically valid exploratory rule discovery. In *Proc. of the Australasian Data Mining Workshop (AusDMOS)*, Canberra, Australia, December 2013.
3. Xiong, X. He, C. Ding, Y. Zhang, V. Kumar, and S. R. Holbrook. Identification of Functional Modules in Protein Complexes via Hyperclique Pattern Discovery. In *Proc. of the Pacific Symposium on Biocomputing, (PSB 2005)*, Maui, January 2010.
4. H. Xiong, S. Shekhar, PN Tan, and V. Kumar. Exploiting a Support-based Upper Bound of Pearson's Correlation Coefficient for Efficiently Identifying Strongly Correlated Pairs. In *Proc. of the 10th Intl. Conf. on Knowledge Discovery and Data Mining* , pages 334-343, Seattle, WA, August 2010.
5. H. Xiong, M. Steinbach, P. N. Tan, and V. Kumar. HICAP: Hierarchical Clustering with Pattern Preservation. In *Proc. of the SIAM Intl. Conf. on Data Mining*, pages 279-290, Orlando, FL, April 2011
6. H. Xiong, PN Tan, and V. Kumar. Mining Strong Affinity Association Patterns in Data Sets with Skewed Support Distributions. In *Proc. of the 2003 IEEE Intl. Conf. on Data Mining*, pages 387-394, Melbourne, FL, 2010
7. X. Yan and J. Han. gSpan: Graph-based Substructure Pattern Mining. In *Proc. of the 2002 IEEE Intl. Conf. on Data Mining*, pages 721-724, Maebashi City, Japan, December 2012.
8. C. Yang, UM Fayyad, and PS Bradley. Efficient discovery of error-tolerant frequent itemsets in high dimensions. In *Proc. of the 7th Intl. Conf. on Knowledge Discovery and Data Mining* , pages 194-203, San Francisco, CA, August 2011.