

CONTENT-BASED IMAGE RETRIEVAL USING QUADRANTS OF IMAGE: A SURVEY

N.Danapaquame¹, M.Aravindhan², D.P.Karthikayan³, R.Kathiravan⁴, R. Rajadurai⁵
Department Of CSE^{1, 2, 3, 4, 5}, Sri Manakula Vinayagar Engineering College
n.danapaquame@gmail.com¹, aravindhanmurali@gmail.com², dpkarthik2012@gmail.com³,
r.kathiravan95@gmail.com⁴, rajadurai@smvec.ac.in⁵

Abstract— Data mining is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Cloud computing is popularizing the computing paradigm in which data is outsourced to a third-party service provider (server) for data mining. An efficient probabilistic and deterministic verification approaches has been proposed to check whether the server has returned correct and complete frequent itemsets. The probabilistic approach can catch incorrect results with high probability, while the deterministic approach measures the result correctness with 100% certainty. In our project, we are going to use clustering approach to find the frequent item sets which increases the probability rate.

Keywords— Data mining, probabilistic, deterministic, clustering.

1. Introduction

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information. Data mining is sorting through data to identify patterns and establish relationships. Data mining is the extraction of hidden information from large databases. Data mining is a process used by companies to turn raw data into useful information.

1.1 Architecture

Fundamentally, data mining is about processing data and identifying patterns and trends in that information so that you can decide or judge. Data mining principles have been around for many years, but, with the advent of big data, it is even more prevalent. Big data caused an explosion in the use of more extensive data mining techniques,

partially because the size of the information is much larger and because the information tends to be more varied and extensive in its very nature and content. With large data sets, it is no longer enough to get relatively simple and straightforward statistics out of the system. With 30 or 40 million records of detailed customer information, knowing that two million of them live in one location is not enough. You want to know whether those two million are a particular age group and their average earnings so that you can target your customer needs better. These business-driven needs changed simple data retrieval and statistics into more complex data mining. The business problem drives an examination of the data that helps to build a model to describe the information that ultimately leads to the creation of the resulting report. Figure outlines the process.

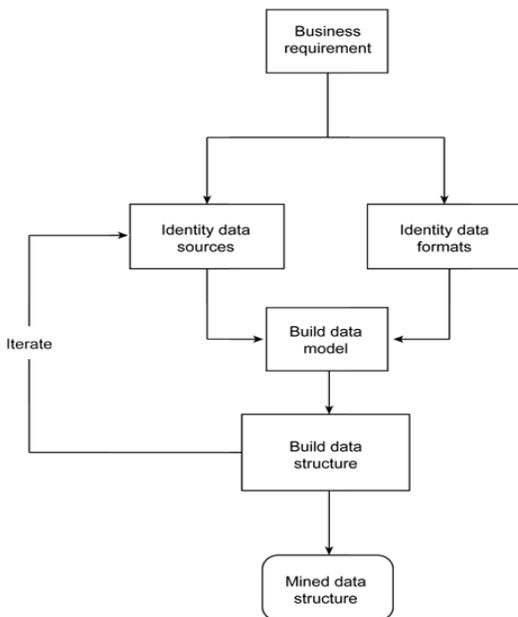


Figure 1.1 Outline of the data mining process

The process of data analysis, discovery, and model-building is often iterative as you target and

identify the different information that you can extract. You must also understand how to relate,

map, associate, and cluster it with other data to produce the result. Identifying the source data and formats, and then mapping that information to our given result can change after you discover different elements and aspects of the data.

1.2 Organisation

This paper is organized as follows:

Section [1] describes the data mining with frequent item sets. Section [2] deals with the related works. Section [3] deals with the research directions and the proposed work which will help in overcoming the problem of the existing system. Section [4] deals with the research directions and the Section [5] contains the conclusion.

2. Related works

Doug Burdick, Manuel Calimlim, Johannes Gehrke presented a algorithm for mining

maximal frequent itemsets from a transactional database. Their algorithm is especially efficient when the itemsets in the database are very long. The search strategy of their algorithm integrates a depth-first traversal of the itemset lattice with effective pruning mechanisms. Their implementation of the search strategy combines a vertical bitmap representation of the database with an efficient relative bitmap compression schema. In a thorough experimental analysis of their algorithm on real data, we isolate the effect of the individual components of the algorithm. Their performance numbers show that their algorithm outperforms previous work by a factor of three to five.

3. Research Directions

If we use clustering to find the frequent data sets, the time frequency will be very less and the memory will be consumed low more accurate results can be gained. By using this we can able to do the mining of data from the data sets easier. The correctness and completeness are measured against the proofs with 100% certainty. The efficiency and effectiveness of our approaches. An interesting direction to explore is to extend the model to allow the client to specify her verification needs in terms of budget (possibly in monetary format) besides precision and recall threshold.

4. Discussions

We ran experiments to compare the performance of our probabilistic and deterministic approaches. Along with it we clustered the itemsets. The deterministic approach brings higher overhead at the server side than the probabilistic approach. We used rapidminer tool for mining.

5. Conclusion

In this paper, we present two integrity verification approaches for outsourced frequent itemset mining. The probabilistic

verification approach constructs evidence (in)frequent itemsets. In particular, we remove a small set of items from the original dataset and insert a small set of artificial transactions into the dataset to construct evidence (in)frequent itemsets. The deterministic approaches requires the server to construct cryptographic proofs of the mining result. The correctness and completeness are measured against the proofs with 100% certainty. By clustering we increased the efficiency compared to the existing system.

References

- [1] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules in large databases. In Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), pages 487–499, 1994.
- [2] Laszlo Babai, Lance Fortnow, Leonid A. Levin, and Mario Szegedy. Checking computations in polylogarithmic time. In STOC, pages 21–32, 1991.
- [3] Ran Canetti, Ben Riva, and Guy N. Rothblum. Verifiable computation with two or more clouds. In Workshop on Cryptography and Security in Clouds, 2011.
- [4] Kun-TaChuang, Jiun-LongHuang, andMing-SyanChen. Power-law relationship and self-similarity in the itemset support distribution: analysisandapplications. TheVLDBJournal,17:1121–1141, August 2008.
- [5] Rosario Gennaro, Craig Gentry, and Bryan Parno. Non-interactive verifiable computing: outsourcing computation to untrusted workers. In CRYPTO, pages 465–482, 2010.