OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

Extending Market Basket Analysis Using Multidimensional Customer Based Profile In Graph Mining

[1]M.Santhosh, [2]K.Anisha,[3]B.Priyanka,[4]K.Sirisha
[1]Asst. proff(Dept of Computer science GVP college of Engg for Women)
[2](Dept of Computer science GVP college of Engg for Women)
[3](Dept of Computeerscience GVP college of Engg for women)
[4](Dept of Computeerscience GVP college of Engg for women)

ABSTRACT:-A common problem for many companies, like retail stores, it is to find sets of products that are sold together. The only source of information available is the history of sales transactional data. Common techniques of market basket analysis fail when processing huge amounts of scattered data, finding meaningless relationships. We propose approach for market basket analysis based on graph mining techniques, multidimensional data with customer profiling which is able to process millions of scattered transactions. We proposed the effectiveness of our approach in a wholesale supermarket chain and a retail supermarket chain, processing large number of transactions respectively compared to classical approach. The proposed work also is develop a cloud based parallel algorithm for efficient sub graph mining and association rule generation

Keywords:-Market basket analysis, Graph mining, Transactional data, Customer Profile, Multidimensional data

Existing System

Market basket analysis is a mathematical modelling techniques based upon the theory that if you buy a certain group of items, you are likely to buy another group of items. It is used to analyze the customer purchasing behaviour and helps in increasing the sales and maintain inventory by focusing on point of sale transactional data.

The purpose of market basket analysis is to get a customer to spend more money based on two different principles: the first one is Up-Selling, which consists in buying a large quantity of the same product, or adding new features or warranties. The second way is Cross-Selling, which consists in adding more products from different categories. The main purpose to discover frequent item sets. Also known as the discovery of if-then rules called Association rules [1] The form of an association rule is I !j where I is a set of items (products) and j is a particular item. The process consist of finding sets of products (items) presents in a large number of transactions (basket).

The main disadvantages are poor quality information was generated disabling decisions such as finding customers profiles, discount offers generation, supermarket products layout etc.,Common techniques of market basket analysis fail when processing huge amounts of scattered data, finding meaningless relationships.

I. INTRODUCTION

Business Analytics

In addition to customer lifetime value analysis, organizations in the retail, financial services, insurance, telecommunications, health care and information communication and technology industries are using Market Basket Analysis modelling to deploy and improve:Product recommendations and promotions (couponing and discounts)

Cross-sell and Upsell strategiesProduct placement (store layout, shelf, flyer, website)Product and service bundlingNext Best Offer strategiesAffinity marketing programs (loyalty and retention)Inventory selection and optimization (supply chain)

Market Basket Analysis is a critical tool to help improve the efficiency of marketing and sales strategies by analyzing customer (CRM) data collected during sales transactions. It is equally effective in streamlining and improving business operations in the areas of inventory control and distribution or channel optimization.

Market Basket Analysis can also uniquely be used to identify items that are **unlikely** to occur together. This can be useful when designing marketing strategies that avoid promoting high margin items with unassociated or low selling products thus leading to poor sales of the high margin items. Likewise, it is useful when detecting anomalies and unexpected purchase patterns that point to potential sales opportunities and new marketing campaigns. Anomaly detection also helps to identify fraudulent behaviours.

II. MARKET BASKET ANALYSIS

Today, leading companies are looking to improve business performance via faster, better decision making by applying advanced predictive modelling to their vast and growing volumes of data. Business analytics, whether for marketing, CRM, loyalty or operations, provides organizations with valuable insights from their data - allowing them to uncover and action new opportunities to increase revenue and profitability.

Market Basket Analysis is a data modelling technique used to find associations between items or events by determining the likelihood for them to occur together. Taking its name from the concept of identifying products that customers are putting into their shopping cart, Market Basket Analysis is also commonly referred to as Product Affinity Analysis or Association Rule Learning.

A typical use case of Market Basket Analysis is to leverage large amounts of customer transaction data to determine products that are most commonly purchased together. Understanding these purchasing patterns empowers marketing and sales organization to make more informed decisions about how and where to deploy their efforts and resources.

An obvious application of Market Basket Analysis is in the retail sector where retailers have large amounts of transactional data and often thousands of products. One of the recognizable examples is the **Amazon.com** recommendation system: “Customers who bought this item also bought items A, B and C”. Market Basket Analysis is applicable in many other industries and use cases, many of which this paper will illustrate and explore.

III. GRAPH MINING

A network is defined as a set of elements interconnected between each other. A common way to represent a network is using a graph. A graph is a manner to specify relationships between sets of items. Formally, a graph consists of a set of objects, called nodes with some pairs of them connected by links called edges.

A product network is defined as a network where nodes representproducts and edges represent relationships between a pair of them. We have to define what kind of relationship is represented by an edge. In this case, an edge between two products represents that both products are present in the same ticket from the same buyer opportunity.

A classical approach to getting information from data in retailand department stores is through market basket analysis (MBA),frequent item set discovery and clustering techniques such as

K-means[2]and FP-growth Algorithm. Themain idea behind this is to discover purchasing patterns fromtransactional data. However, when we used these techniques toprocess real supermarket chain data, the results obtained were ofmain idea behind this is to discover purchasing patterns fromtransactional data. However, when we used these techniques toprocess real uppermarket chain data, the results obtained were of very poor quality. For example, with K-means techniques onlyone cluster grouped 93% of transactions and the 7% remaining isnot meaningful. Therefore, poor quality information was generated disabling decisions such as finding customers profiles, discountoffers generation, supermarket products layout, etc. Thus, wedeveloped a novel approach to perform MBA based on graph miningtechniques; specifically using overlap communities,[3][4] that allowsto generate highly related products to each other within theimmunity. We benchmarked our method using severaltraditional approaches applied over millions of transactional data.The results of our evaluation show that our approach out–performsthe traditional methods.

Finally the given task is divided into sub tasks as well as the given transactional graph is also partitioned and assigned to each task. The sub tasks are executed in parallel at cloud nodes and evaluated local support at each node. Then all the local results are joined or merged at a joiner or server in cloud process to generate global support. Based on the global support final association rules will be generated at server. The process executed in single system is replicated in more than one system and final result will be merged.

IV. PROPOSED SYSTEM

Multidimensional Data Storage in Analytic Workspaces

In the logical multidimensional model, a cube represents all measures with the same shape, that is, the exact same dimensions. In a cube shape, each edge represents a dimension. The dimension members are aligned on the edges and divide the cube shape into cells in which data values are stored.In an analytic workspace, the cube shape also represents the physical storage of multidimensional measures,[5][6] in contrast with two-dimensional relational tables. An advantage of the cube shape is that it can be rotated: there is no one right way to manipulate or view the data. This is an important part of multidimensional data storage[7], calculation, and display, because different analysts need to view the data in different ways. For example, if you are the Sales Manager for the Pacific Rim, then you need to look at the data differently from a product manager or a financial analyst.Assume that a company collects data on sales. The company maintains records that quantify how many of each product was sold in a particular sales region during a specific time period. You can visualize the sales measure as the cube shown in Figure 2-3.

Figure 2-3 Comparison of Product Sales By City

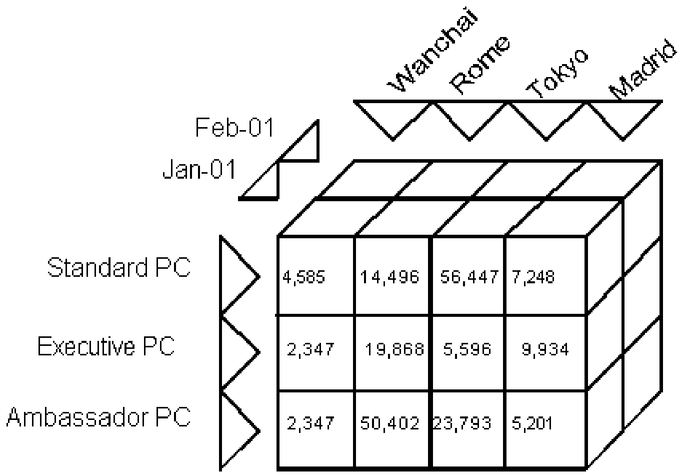
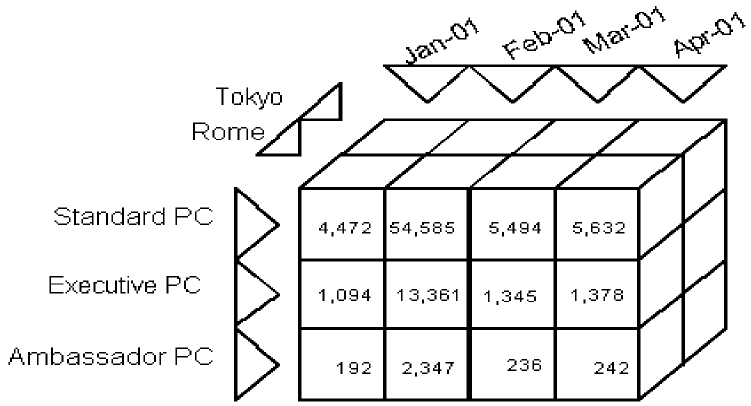


Figure 2-3 compares the sales of various products in different cities for January 2001 (shown) and February 2001 (not shown). This view of the data might be used to identify products that are performing poorly in certain markets. Figure 2-4 shows sales of various products during a four-month period in Rome (shown) and Tokyo (not shown). This view of the data is the basis for trend analysis.A cube shape is three dimensional. Of course, measures can have many more than three dimensions, but three dimensions are the maximum number that can be represented pictorially. Additional dimensions are pictured with additional cube shapes.

Figure 2-4 Comparison of Product Sales By Month



Understanding Customer Purchase Behaviour

Market Basket Analysis results in a better understanding of your customers and their purchasing behavior by allowing you to explore associations and predict the likelihood of a customer's future purchase behavior based on associations. It is one of many advanced business analytics tools that can help organizations optimize marketing and sales operations for improved performance.Marketing and sales organizations across all industries are looking to analyze, understand and predict customer purchase patterns towards achieving strategic goals to reduce churn rates and maximize customer lifetime value (CLV). Selling additional products and services to existing customers over their lifetime is key to optimizing revenues and profitability. Market Basket Analysis association rules identify the products and services that customers typically purchase together, empowering organizations to offer and promote the right products to the right customers.Moreover, with predictive analytics organizations are able to promote their most profitable products and services to the most likely buyers. They can also encourage additional purchases by introducing new targeted products, products with high margin, or high performing products that may not have otherwise been an obvious next purchase.

Market Basket Analysis case study

Retail Cross-sell and Upsell

Product recommendations and promotions Bussiness challenge A big box superstore wants to encourage customers to branch out and buy products in departments they do not regularly shop in. The retailer has a customer loyalty card program and is able to leverage this to track and monitor purchases by customer. Analysis of customer purchases indicates many customers only shop in one or two departments. The retailer wants to increase sales by encouraging customers to purchase items outside of their normal habits by identifying the department(s) most likely to be shopped in outside of the ones they frequent by providing customers with a promotion incentive to purchase products in new department(s).

Solution

Design a promotional campaign across departments that encourage customers to shop in new departments and increases market share by capturing purchases from competitive retailers.

The solution comprises the development of a cross-sell model and deployment of a promotional strategy:

1. Market Basket Analysis is used to develop a cross-sell model that recommends new departments that are most likely to be of interest to customers.
2. Strategy Trees in StrategyBUILDER is used to develop and deploy a promotional campaign to encourage sales in the department(s) that customers are most likely to shop in outside of their normal habits.

The workflow to develop and deploy the solution comprises the following:

1. Use Market Basket Analysis to find out what departments customers typically shop in together.
2. Deploy the association rules discovered by Market Basket Analysis to produce a list of recommended departments personalized by customer.
3. Create a promotional campaign that leverages Market Basket Analysis recommendations using the StrategyBUILDER module.

The data prepared for analysis contains the customer ID and a record for each department where the customer shops. In the example below, the customer with ID 308459 purchased items in two departments: Accessories and Home.

Customer ID	Department
308459	Accessories
308459	Home
308468	Home
308522	Sports
308580	Accessories
308580	Home
308580	Luggage

When building the Market Basket Analysis model, it is important to make sure that the number of resulting association rules is manageable and there are no rules based on too few transactions (there are enough customers exhibiting each particular shopping pattern). This is achieved by adjusting model parameters, such as **support**, which filters out cases that are too rare.

The Market Basket Analysis model in this case may discover association rules such as **{*Women, Accessories, Home*} {*Kids*},** which means that customers who shop in the Women, Accessories, and Home departments also purchase items in the Kids department with a certain likelihood.

The business analyst can automatically select rules by using any of the available filters at the top of the Rules tab and can also compliment automatic selections by manually selecting or unselecting rules. Once the desired rules have been selected for deployment, apply them to new data using the Tools | Score command to produce recommendations.

Outcomes:

The exhibited Market Basket Analysis will recommend departments likely to be shopped in by the given customer. Customers may have zero, one or multiple recommendations. The default scoring option "Recommend New Items Only" will restrict results to the traditional cross-sell option and provide recommendations for departments not previously shopped in. This option can be unselected if the user wishes to extend recommendations.

The other scoring default will exclude records for customers with no recommendations; a simple click on the check box will change the default and provide records for all customers. There are various reasons a customer may not receive a recommendation. In this example, a customer who has only purchased luggage is unlikely to have any recommendations since we have already discovered earlier that luggage is an exceptional purchase not typically associated with other departments.

A sample of deployment results are illustrated below. Recommendations produced by the most predictive rules have the highest rank (*Rank_Lift* =1)

CustId	recommendations	lift	Rank	confi	rank_confi
308459	health &beauty	2.766	1	0.149	5
308459	women	2.219	3	0.237	
308459	baby	1.791	4	0.324	1
308468	kids	1.677	1	0.124	5
308468	accessories	1.673	2	0.360	1

V. CONCLUSION

Business analytics, whether for marketing, CRM, loyalty or operations, provides organizations with valuable insights from their data - allowing them to uncover and action new opportunities to increase revenue and profitability. Market Basket Analysis allows organizations to optimize sales and marketing efforts. By discovering and applying a deeper level of insight into customer purchase patterns, organizations can make more informed, better decisions about how and where to deploy their efforts and resources. Market Basket Analysis results in a better understanding of your customers and their buying behaviour, allowing you to explore associations, predict the likelihood of a future customer response based on associations, and ultimately tailor your marketing and sales strategies and operations for improved performance. Market Basket Analysis is applicable in many industries and across use cases, many of which this paper has illustrated and defined.

REFERENCES

[1]. Agrawal, R., &Srikant, R. (1994).Fast algorithms for mining association rules.InProceedings of 20th international conference on very large data bases, VLDB (pp. 1–32).

[2]. Agrawal, R., Imielinski, T., & Swami, A. (1993). Database mining: A performanceperspective. IEEE Transactions on Knowledge and Data Engineering, 5(6),914–925.

[3]. Arenas, A., Danon, L., Díaz-Guilera, A., Gleiser, P. M., &Guimerà, R. (2004).

[4]. Community analysis in social networks. The European Physical Journal B –Condensed Matter and Complex Systems, 38(2), 373–380.

[5]. Clauset, A., Newman, M., & Moore, C. (2004). Finding community structure in verylarge networks. Physical Review E, 70(6), 066111.

[6]. Faloutsos, M., Faloutsos, P., &Faloutsos, C. (1999).On power-law relationships ofthe internet topology. ACM SIGCOMM Computer Communication, 29(4), 251–262.

[7]. Fortunato, S. (2010).Community detection in graphs. Physics Reports, 486(3),75–174.

[8]. Gregory, S. (2010). Finding overlapping communities in networks by labelpropagation. New Journal of Physics, 12(10), 103018.

[9]. Hartigan, J., & Wong, M. (1979). Algorithm AS 136: A K-means clustering algorithm.

[10]. Journal of the Royal Statistical Society. Series C (Applied Statistics), 28(1), 100–108.

[11]. Kim, H. K., Kim, J. K., & Chen, Q. Y. (2012).A product network analysis for extendingthe market basket analysis.Expert Systems with Applications, 39(8), 7403–7410.

[12]. Kohonen, T. (1990).The self-organizing map. Proceedings of the IEEE, 78(9),1464–1480.

[13]. Lancichinetti, A., &Fortunato, S. (2009). Benchmarks for testing community detection algorithms on directed and weighted graphs with overlappingcommunities. Physical Review E, 1–9.

[14]. Lu, W., Janssen, J., Milios, E., Japkowicz, N., & Zhang, Y. (2006). Node similarity in the Moore, C., & Newman, M. (2000).Epidemics and percolation in small-worldnetworks. Physical Review E, 61(5), 5678.

[15]. Nepusz, T., Petróczi, A., Négyessy, L., &Bazsó, F. (2008). Fuzzy communities and theconcept of bridgeness in complex networks. Physical Review E, 1, 1–13.

[16]. Newman, M., & Girvan, M. (2004). Finding and evaluating community structure innetworks. Physical Review E, 69(2), 026113.Raeder, T., &Chawla, N. V. (2009). Modeling a store's product space as a socialnetwork. In 2009 International conference on advances in social network analysis and mining (pp. 164–169). IEEE.

[17]. Raghavan, U., Albert, R., & Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. Physical Review E, 1–12.
[18]. Rajaraman, A., & Ullman, J. D. (2012). Mining of massive datasets. CambridgeUniversity Press.
[19]. Multi-dimensional data
[20]. Buzydlowski, J. W., Song, II-Y., Hassell, L.: A Framework for Object-Oriented On-Line Analytic Processing. DOLAP 1998.
[21]. Cabibbo, L., Torlone, R.: A Logical Approach to Multidimensional Databases. EDBT 1998
[22]. Chaudhuri, S., Dayal, U.: An Overview of Data Warehousing and OLAP Technology. SIGMOD Record Volume 26, Number 1, September 1997.
[23]. Hartigan& Wong, 1979), SOM (Kohonen, 1990).
[24]. Gregory, S. (2010). Finding overlapping communities in networks by label Fortunato, S. (2010).Community detection in graphs. Physics Reports, 486(3),75–174