International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 11 (2015) pp. 30029-30043

http://www.ripublication.com

An Intelligent Business Expert System For Predicting Sales Patterns Using Temporal Association Rule Mining

Mr. C. Perkinian
Research scholar,
School of Management Studies, Coimbatore Regional Centre,
Anna University, Chennai-25.
E-Mail: perksphd@gmail.com, cperk81@yahoo.com
Dr. P. Vikkraman
Associate Professor, School of Management Studies,
Coimbatore Regional Centre, Anna University,
Chennai-25. E-Mail: dr.p.vikkraman@gmail.com

Abstract

In this paper, we propose a new intelligent business expert system for helping the seller by providing feedback on sales with respect to time for new products or items. For this purpose, this system uses Temporal Association Rule Mining Technique and business rules with time constraints to predict the sales patterns. This helps to provide feedback to the manufacturer and the sales persons on the quantity of items to be produced at each season of a year. In addition, we use a market transaction dataset having a large number of features for effective analysis. In order to make the inference faster, a feature selection algorithm called Intelligent Conditional Random Field based Feature Selection Algorithm (ICRFFSA) is used for selecting the optimal number of features. The experimental results obtained from this work show that the performance of the proposed expert system is more accurate and correlates with the survey obtained using a questionnaire from a large sample of customers who performed online shopping for ordering the products.

Keywords: Temporal Association Rule Mining, Conditional Random Field, Business Expert System

Introduction

Business expert systems are widely used by various business groups in different areas and time to improve the business strategy on the items sold. The Business Expert Systems are used by the clients and these systems attract the customers by producing sufficient details about the items needed by them. The various features that affect the

consumer interest are categorized to form business rules. Since different individuals have different types of interests, it is necessary to make prediction on purchase of items. Expert systems, best known for their use in medical applications can be applied in business to provide suggestions. Modern business expert systems are facing challenges such as collection of data and analysis with rules to produce high-quality results. Though, already several statistical methods are available. Association rule mining is one of the most used data mining and analysis technique for predicting the sales patterns.

Association rule mining finds interesting associations and/or correlation relationships among large set of data items. Association rules show attributes value conditions that occur frequently together in a given dataset. Association rules provide information of this type in the form of "if-then" statements. These rules are computed from the data and, unlike the if-then rules of logic, association rules are probabilistic in nature. In addition to the antecedent and the consequent, an association rule has two numbers that express the degree of uncertainty about the rule. In association analysis the antecedent and consequent are sets of items that are disjoint. The mining association rule is a data mining task that aims to discover relationships among items in a transactional database [5]. This task has been studied widely in the literature for its benefit in many application domains, such as Web usage mining, recommender systems, and intrusion detection. The support is simply the number of transactions that include all items in the antecedent and consequent parts of the rule. Confidence is the ratio of the number of transactions that include all items in the consequent as well as the antecedent to the number of transactions that include all items in the antecedent. Lift is nothing but the ratio of confidence to expected confidence. Lift is a value that gives us information about the increase in probability of the "then" (consequent) given the "if" (antecedent) part. Time is an important constraint over the sales (dynamic) data in this business environment. To improve the prediction accuracy on sales patterns, the combination of association rule mining and time is very important today. For this purpose this paper propose a new intelligent business expert system for predicting the sales patterns on the expected sales of new product or items using feature selection and association rule mining with time constraints. This expert system uses the existing algorithms with temporal constraints to find the sales prediction over the transactions and products. Rest of this paper is organized as follows: Chapter 2 discusses about various past works done in this direction. Chapter 3 explains the overall system architecture. Chapter 4 described the proposed method. Chapter 5 contains the results and discussion. Chapter 6 gives conclusion and future works.

Literature Survey

Many works have been done by the various researchers during the past decade. Among them, Ali Gohary et al [1] investigated the impact of consumer behaviors on online shopping and motivations. They conducted t-test to find the differences between male and female personalities. Their investigation is useful for managers to set a goal and choose suitable customer on products. Saim and Vijay [2] explored

whether family firms reveal unique purchase behavior and whether their unique behavior in turn helps them outperform non-family firms during periods of economic contraction in their study. Asem et al [3] introduced a method to design an improved and well-structured website design for an E-shop in the design phase with the help of Association Rule Mining. Moreover, they have a physical grocery store which has no website, but it has a dataset that records the transactions of its customers. Association Rule Mining techniques are applied on this dataset. The extracted interesting Association Rules from the transactions dataset of the grocery store are took into account in the process of designing a website for the grocery store. The extracted Association Rules are invested to support the website design from the beginning (i.e. in the design phase). Many improvements and modifications are done to the website's design, such as adding/modifying links, and/or creating index pages. They introduced a technique to evaluate our method by comparing the navigation efficiency among different website designs.

Surendren and Bhuvaneswari [4] introduced a framework for Recommender System incorporating cognitive dissonance a psychological factor. The recommender system is designed as a hybrid system which combines both content and collaborative concepts using association rule mining concept a data mining technique. Hu Le-wei and Li Yu-Sheng [5] proposed a method which divides the customer acts data into browsing, evaluation, selection and purchasing four act data, and combines these data to the goods data, thus it can provides to commerce web site managers the results that more close to actual acts of the relevant rules. Md. Rezaul Karim et al [6] proposed a complete solution of predicting e-shopper's purchase rules by using maximal frequent patterns and sequence close level. Their experimental results indicates their proposed approach can give more accurate measurement in discovering e-shoppers purchase rules and useful for e-marketers to make marketing decisions.

R.M.Somasundaram and Lakshmanan [7] proposed a new intrusion detection system using the temporal association rules for effective classification. Ganapathy et al [8] proposed a new pattern classification system by combining Temporal features with Fuzzy Min–Max (TFMM) neural network based classifier for effective decision support in medical diagnosis. Intelligent fuzzy rules are extracted from the temporal features with Fuzzy Min–Max neural network based classifier, and then PSO rule extractor is also used to minimize the number of features in the extracted rules. Yoon Kyoung Choi et al [9] implemented a recommendation system, which are suitable for products showing repetitive purchase pattern and this system recommends three sets of products, by applying user-based Collaborative Filtering approach, item-based Collaborative Filtering approach, and by finding Associate Products that are frequently bought at the same time. Finds Associate Products for each user and analyze to consider baskets for each user, or consider baskets for similar user groups respectively. Xianzhen Deng et al (2009) [10] presented a recommendation method which is based on rank correlated association rules in which this method's recommendation quality is measured by recall rate, and is better than that of the traditional strategy.

Item-based CF recommendation algorithm proposed by G. Karypis et al [12] determine the similarities between the various items, and use them to identify the set

of items to be recommended. One of the methods to computing the item-to-item similarity models the items as vectors in the user space and uses the cosine function to measure the similarity. Another method is to use a measure that is based on the conditional probability of purchasing one of the items given that the other items has already been purchased. Y. K. Choi et al [11] implemented Recommendation algorithms for periodically repetitive purchasing shopping malls. Once a new purchase occurs, the recommendation system in [11] updated "purchase frequency" of the corresponding product by off-line, followed by re-calculating a "Favorite-Items Set" per user, and re-calculate "user similarity" based on the Favorite-Items Set. The Favorite-Items Set was defined as frequently purchased products per user. Whenever a user logged in, products in Favorite-Items Sets of other users who have similar purchase patterns were recommended using a user-based Collaborative Filtering method.

System Architecture

The overall architecture of the proposed system is shown in figure 3.1. The proposed system architecture consists of eight major components such as Super market dataset and dresses sales data, User Interface Module, Business expert system and recommended transactions (products). The business expert system consists of three modules namely Feature selection, temporal association rule mining and Recommendation manager. Super market & sales dataset contains the detailed dress sales data and super market sales data which is collected from all over the world.

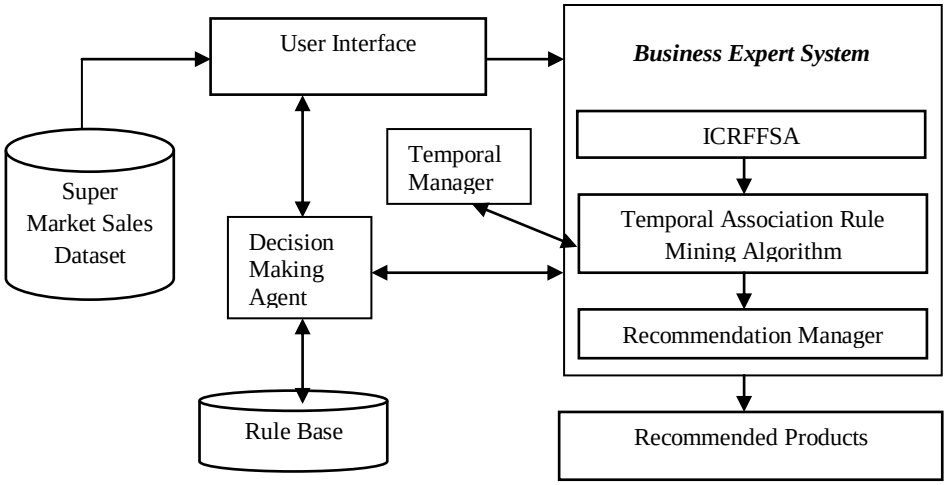

Figure 3.1: System Architecture

The user interface module collects the necessary data from the standard bench mark dataset and sent it to the business expert system for predicting the products. The business expert system contains three sub modules such as Temporal Association Rule Mining, Feature Selection and Recommendation Manager. Among them, the Temporal Association rule mining module selects the necessary combination of items

from the given dataset based on the support, confidence and lift with the help of temporal manager. Then the Temporal Apriori classifier classifies the suitable transactions from the selected features of the given bench mark dataset for further process. Finally, the recommendation manager analyzes the recommended transactions and generates the recommended lists of products as a result.

Proposed Work

This work proposes a new business expert system for predicting new items using association rules with temporal constraints and effective feature selection technique. This section discusses about the uses of Temporal Apriori algorithm [7] and the effective feature selection algorithm called Intelligent CRF based Feature Selection Algorithm [13] in this business expert system.

A. Feature Selection

For effective feature selection, we have used an effective feature selection algorithm called Intelligent CRF based Feature Selection Algorithm [13] according to [14] for selecting appropriate features for the various season and improve the prediction accuracy using the classification method. We have selected optimal number of features from the standard benchmark dataset for different season. We have introduced a new parameter is time to predict appropriate sales pattern for the particular season. This algorithm will be chosen the various optimal numbers of features for three major seasons such as summer, winter and monsoon.

B. Association Rules

An association rule [3] is an implication expression of the form A⇒B, where A and B are disjoint item-sets. The support denoted ($A \cup B$) of an association rule A⇒B is the percentage of instances that contain all the items included in the association rule:

$$S(A \cup B) = \frac{k}{n} \tag{4.1}$$

where k is the number of instances containing all the items of A and B; n is the total number of instances of the dataset.

The confidence [3] of an association rule is a fraction that shows how frequently B occurs among all the instances containing A:

$$confidence = \frac{S(A \cup B)}{S(A)} \tag{4.2}$$

where S(A∪B) is support of the rule; S(A) is support of A. The confidence value indicates how reliable the rule is. Confidence provides an estimate of ($B|A$) the conditional probability of B given A assuming B statistically depends on A, and is used to measure the reliability or interestingness of the rule.

The lift value [3] of an association rule is the ratio of the confidence of the rule to the support of B:

$$lift = \frac{confidence}{S(B)} \tag{4.3}$$

where S(A) is the support of A; S(B) is the support of B. Lift is a measure of the deviation of the rule from the statistical independency of A and B. The lift is a value between 0 and ∞: a value greater than 1 indicates that A and B appear together more often than expected; a value less than 1 indicates that A and B appear together less often than expected; a value close to 1 indicates that A and B appear together almost as often as expected.

C. Temporal Association Rule Mining

In this work, we have used an effective association rule mining algorithm with time constraints called Temporal Apriori algorithm [7] for decision making over the large amount of transactions in various season in all over the world. It is useful choose most wanted transactions over the items on market in each different time periods. We have used two bench mark datasets such as dresses sales data and dresses image database for decision making over the sales pattern.

Result and Discussion

This section discusses about the dataset which is used for evaluating the proposed system and the experimental analysis.

Data Set

For carrying out the analysis, data were obtained from five different shops in each city and five cities namely Chennai, Bangalore, Mumbai, New Delhi and Kolkata were considered in this work. Table 5.1 shows the item-sets sold in different dates of 2014. However, this analysis was carried out for five years starting from 2010 in the same cities and same shops. In this paper, a small subset of the sample data are listed and shown in tables for the purpose of illustration.

Table 5.1: Sample Transactions From Dresses Sales Dataset With Highest Number of Itemsets

T Id	Items	Date
T1	{Pant, T-Shirt, Shirt}	6/5/2014
T2	{Pant, T-Shirt}	8/5/2014
T3	{Saree, Blouse}	17/5/2014
T4	{Pant, Long dress}	3/08/2014
T5	{Kurthi, Jean, T-Shirt}	13/8/2014
T6	{Shirt, Pant}	18/8/2014
T7	{Saree, T-Shirt, Jean}	27/08/2014
T8	{Saree, Pant, Shirt}	5/11/2014
T9	{Trouser, T-Shirt, Shirt}	14/11/2014
T10	{Shirt, Pant, Trouser}	27/11/2014

Questionnaire

In this work, survey was conducted from 10000 people from each city and was conducted in 5 different cities of India namely Chennai, Mumbai, Kolkata, New Delhi and Bangalore. Table 5.4 shows that the sample questions and the responses provided by the participants for the questions provided in the questionnaire.

Table 5.4: Top Ten Questions From The Questionnaire

S. No.	Questions	Item Name	May 2014	August 2014	November 2014
1.	What garment items will you purchase during this month? (Ladies)	Saree	27%	24%	16%
		Chudidhar	42%	33%	30%
		Jeans Pant	31%	43%	54%
2.	What garment items will you purchase during this month? (Gents)	Dhoti-Shirt	28%	25%	16%
		Pant-Shirt	40%	34%	33%
		Jeans-T-Shirt	32%	41%	51%
3.	What type of prices preferred?	<500 Rupees (Rs.)	35%	32%	34%
		500 – 1500 Rs.	52%	54%	56%
		>1500 Rs.	13%	14%	10%
4.	What type of cloth quality the customers preferred?	Low	33%	35%	34%
		Medium	55%	52%	54%
		High	12%	13%	12%
5.	How did you know the items which you purchased recently?	Friends & Own experience	33%	27%	31%
		TV Advt.	45%	52%	59%
		Posters & News papers	22%	21%	10%
6.	What percentage of people liked to see more number of varieties in garments while purchasing?		82%	75%	98%
7.	Do you like to buy garments with new design?	Yes	76%	81%	85%
		No	24%	19%	15%
8.	What type of cloth you prefer?	Cotton	51%	39%	25%
		Silk	8%	12%	20%
		Others	41%	49%	55%
9.	Do you give preference to color?	Yes	92%	94%	97%
		No	8%	6%	3%
10.	For whom do you buy garments as gift?	Children	40%	38%	42%
		Elders	18%	17%	40%
		All	42%	45%	18%

Experimental Results

A new intelligent business expert system has been proposed and implemented in MATLAB. We have considered market transactions dataset for experiments and analysis. We have compared the prediction result of the proposed system with actual sold. Finally, we have also made comparative analysis between questionnaires and actual sold.

Table 5.5, Table 5.6 and Table 5.7 show the sales prediction analysis for the month of May, August and November 2014 which fall in summer, winter and monsoon seasons. Here, five experiments have been conducted by considering 50000, 60000, 70000, 80000 and 90000 transactions respectively. Moreover, we have compared the number of Sarees purchase predicted and actual number of Saree purchased.

Table 5.5: Sales Prediction analysis for the month of May 2014 (summer)

Ex. No.	Number of transactions considered	Number of Saree purchases predicted	Actual Number of Saree Purchases
1	50000	49925	49931
2	60000	59915	59920
3	70000	69910	69916
4	80000	79907	79913
5	90000	89905	89912

Table 5.6: Sales Prediction Analysis For The Month of August 2014 (Monsoon)

Ex. No.	Number of transactions considered	Number of Saree purchases predicted	Actual Number of Saree Purchases
1	50000	49955	49823
2	60000	59923	59700
3	70000	69942	69698
4	80000	79937	79599
5	90000	89910	89623

Table 5.7: Sales Prediction analysis for the month of November 2014 (winter)

Ex. No.	Number of transactions considered	Number of Saree purchases predicted	Actual Number of Saree Purchases
1	50000	49910	49915
2	60000	59943	59960
3	70000	69912	69918
4	80000	79972	79963
5	90000	89956	89952

From tables 5.5, 5.6 and 5.7, it is observed that the performance of the proposed intelligent business expert system is capable of predicting the sales in all seasons of

the year with more accuracy. Moreover, experiments were conducted to include all types of garment items sold in 25 shops which are located in five different cities. In each city, five shops were considered for computing the actual number of purchases. From these experiments, it is proved that the correlation between the predicted number of sales and the actual number of sales is more than 95%. Hence, the proposed system is capable of predicting the sales more accurately.

Figures 5.1, 5.2 and 5.3 show the performance of comparative analysis between the Apriori Algorithm [5] and the Temporal Apriori Algorithm [7]. We have used the same number of transactions for all experiments. Therefore, five experiments have been conducted with 500 transactions for summer, winter and monsoon seasons.

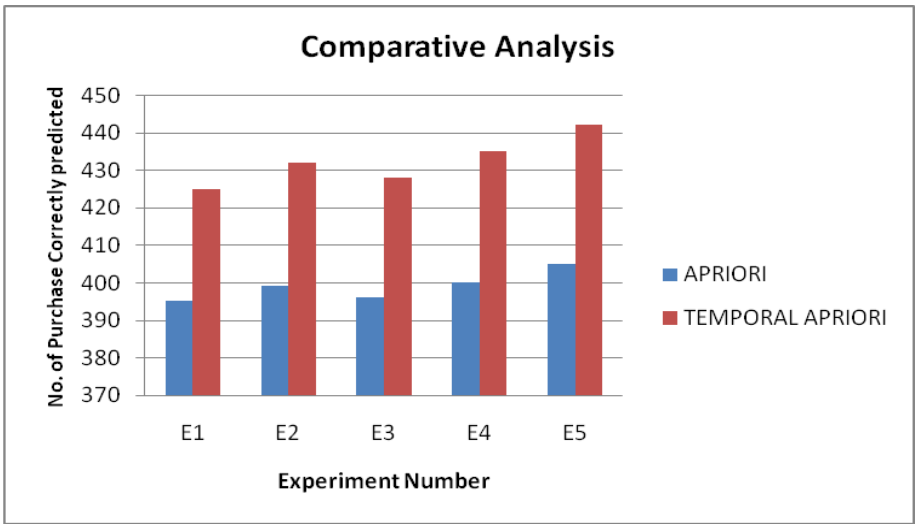

Figure 5.1: Classification Accuracy analyses for summer data

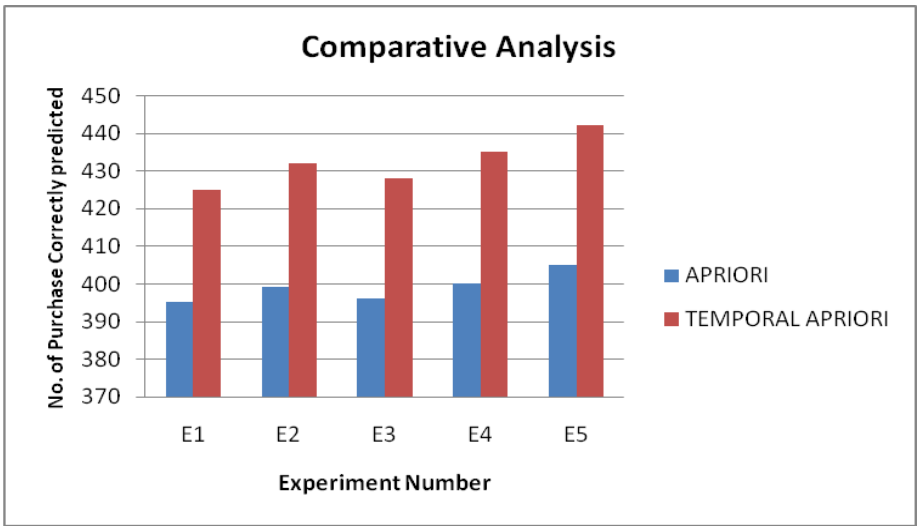

Figure 5.2: Classification Accuracy analysis for Monsoon data

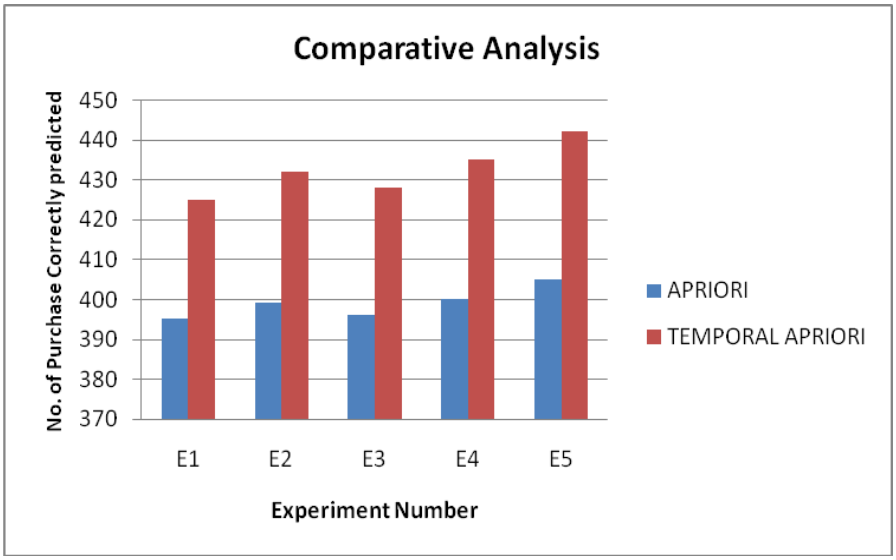

Figure 5.3: Classification Accuracy analysis for Winter data

From figures 5.1, 5.2 and 5.3, it is observed that the performance of the Temporal Apriori algorithm with time constraints is better than the standard Apriori algorithm due to the use of time constraints.

Tables 5.8, 5.9 and 5.10 show the performance analysis between the Temporal Apriori Algorithm [7] along with feature selection [13]. Five experiments have been conducted with 500, 1000, 2000, 5000 and 10000 transactions respectively. We have used the same number of transactions have been used for every experiment.

Table 5.8: Comparative analysis of classification algorithm with full and selected features for May 2014 (Summer)

Ex. No.	**Number of transactions considered**	**Temporal Apriori with Full features**	**Temporal Apriori + ICRFSA**
1	500	462	494
2	1000	935	964
3	2000	1876	1953
4	5000	4673	4912
5	10000	9482	9878

Table 5.9: Comparative analysis of classification algorithm with full and selected features for August 2014 (Winter)

Ex. No.	Number of transactions considered	Temporal Apriori with Full features	Temporal Apriori + ICRFFSA
1	500	452	496
2	1000	945	974
3	2000	1886	1961
4	5000	4682	4922
5	10000	9471	9798

Table 5.10: Comparative analysis of classification algorithm with full and selected features for November 2014 (Monsoon)

Ex. No.	Number of images considered	Temporal Apriori with Full features	Temporal Apriori + ICRFFSA
1	500	468	491
2	1000	962	991
3	2000	1855	1983
4	5000	4654	4956
5	10000	9451	9893

From tables 5.8, 5.9 and 5.10, it is observed that the performance of the Temporal Apriori with feature selection is better than Temporal Apriori classification algorithm due to the effective feature selection.

Figure 5.4, shows the sales analysis for sarees sold in various seasons namely Monsoon, winter and summer at Chennai. We have considered the number of sarees sold at Chennai for 9 years from 2006 to 2014. Finally, the sarees sold at Chennai in the three seasons considered in this work are provided in the comparative analysis.

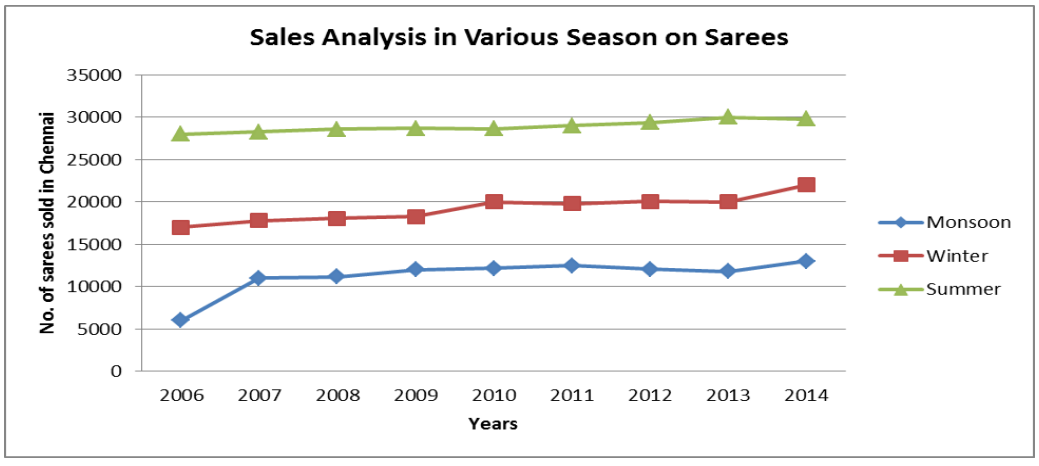

Figure 5.4: Saree sold at Chennai in various seasons

From figure 5.4, it can be observed that the number of sarees sold at Chennai in summer season is high, winter season is medium and monsoon season is low.

Figure 5.5 shows the sales analysis for chudidhar sold in various seasons namely Monsoon, winter and summer at Bangalore. We have considered the number of chudidhar sold at Bangalore for 9 years from 2006 to 2014. Finally, the chididhar sold at Bangalore in the three seasons considered in this work are provided in the comparative analysis.

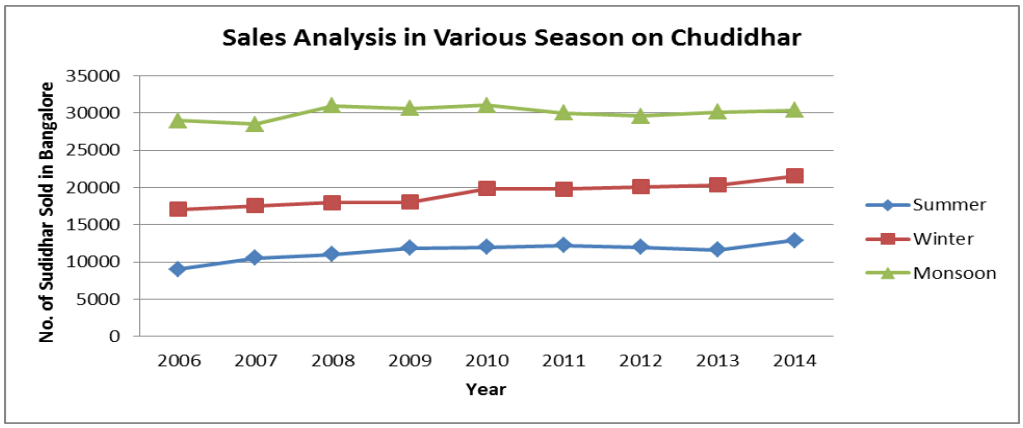

Figure 5.5: Chudidhar sold at Bangalore in various seasons

From figure 5.5, it can be observed that the number of chudidhars sold at Bangalore in summer season is high, winter season is medium and monsoon season is low.

Figure 5.6 shows the sales analysis for Jeans Pant sold in various seasons namely Monsoon, winter and summer at Mumbai. We have considered the number of Jeans Pant sold at Mumbai for 9 years from 2006 to 2014. Finally, the Jeans Pant sold at Mumbai in the three seasons considered in this work are provided in the comparative analysis.

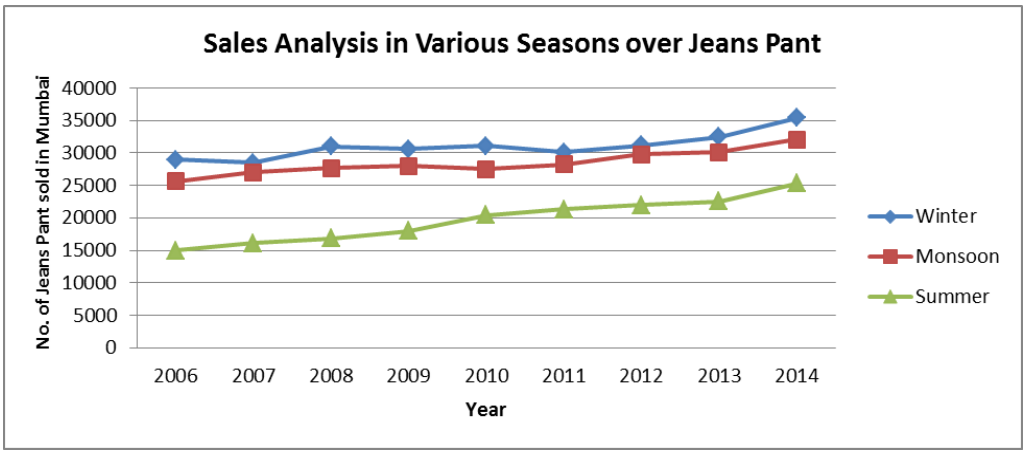

Figure 5.6: Jeans Pant Sold At Mumbai In Various Seasons

From figure 5.6, it can be observed that the number of jeans pant sold at Mumbai in summer season is high, winter season is medium and monsoon season is low.

Figure 5.7 shows the comparative analysis of items sold in a month between Actual and predicted based on questionnaire by one vendor in all cities. The various items of the vendor have been considered for comparative analysis.

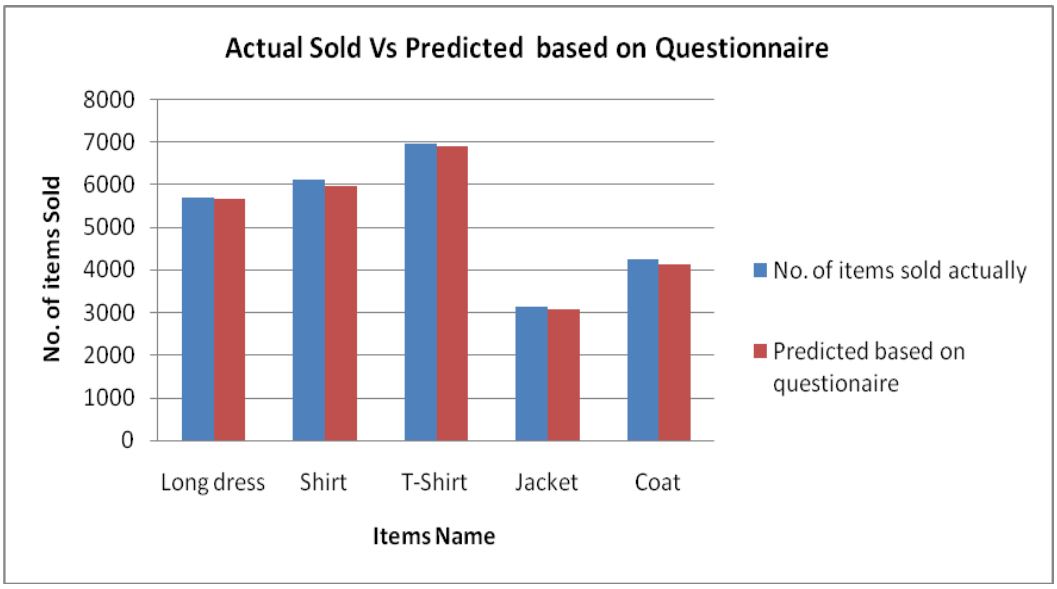

Figure 5.7: Comparative Analyses Between Actual Sales And Predicted Sales Based On Questionnaire

From figure 5.7, it can be observed that the performance of the sales prediction based on questionnaire is almost near to the actual value. The reason for this sales achievement is based on the effectiveness of the questionnaire and the prediction performed by the expert system.

The proposed intelligent business expert system takes necessary decision over shopping of garment items during different seasons. The analysis uses the feature selection with time constraint, classification algorithms and the effective questionnaire.

Conclusion

In this paper, a new business expert system is proposed and implemented for predicting the sales patterns of products using association rule mining and feature selection methods with temporal constraints. This work introduced time constraints for selecting appropriate features over the sales of dresses in various seasons. The system used support and confidence values with temporal rules for predicting the sales pattern. It is also used conditional probability with time constraint for selecting suitable features over the benchmark dataset for various seasons. The experimental results show that the performance of the proposed expert system is more accurate.

Future works in this direction could be the introduction of effective fuzzy rules for better sales pattern prediction over the dataset.

References

[1] Ali Gohary, Kambiz Heidarzadeh Hanzaee, "Personality Traits as Predictors of Shopping Motivations and Behaviors: A Canonical Correlation Analysis", Arab Economics and Business Journal, Vol.9, pp.166–174, 2014.

[2] Saim Kashmiri, Vijay Mahajan, "Beating the recession blues: Exploring the link between family ownership, strategic marketing behavior and firm performance during recessions", Intern. J. of Research in Marketing, Vol.31, pp. 78–93, 2014.

[3] Asem Omari, Stefan Conrad and Sadet Alcic, "Designing a Well-Structured E-Shop Using Association Rule Mining", 4th International Conference on Innovations in Information Technology, pp. 6-10, 2007.

[4] D. Surendren, V.Bhuvaneswari, "A Framework for Analysis of Purchase Dissonance in Recommender System Using Association Rule Mining", 2014 International Conference on Intelligent Computing Applications, 2014 International Conference on Intelligent Computing Applications (ICICA'14), pp. 153-157, 2014.

[5] HU Le-wei, Li Yu-sheng, "Exploration in commercial websites on the association rules base on the consumers' behavior", 2011 6th IEEE Joint International Conference on Information Technology and Artificial Intelligence Conference (ITAIC), pp. 42-44, 2011.

[6] Md. Rezaul Karim, Jae-Hyun Jo, Byeong-Soo Jeong, and Ho-Jin Choi, "Mining E-shopper's Purchase Rules by Using Maximal Frequent Patterns: An E-commerce Perspective", 2012 International Conference on Information Science and Applications (ICISA), pp. 1-6, 2012.

[7] R.M. Somasundaram, K.Lakshmanan, "An Intrusion Detection System for MANET using CRF Based Feature Selection and Temporal Association Rules", International Journal of Soft Computing, Vol.8, No.6, pp. 400-405, 2013.

[8] S Ganapathy, R Sethukkarasi, P Yogesh, P Vijayakumar, A Kannan, "An Intelligent Temporal Pattern Classification System Using Fuzzy Temporal Rules And Particle Swarm Optimization", Sadhana, Vol. 39, Part 2, Pp. 283–302, 2014.

[9] Yoon Kyoung Choi and Sung Kwon Kim, "An Auxiliary Recommendation System for Repetitively Purchasing Items in E-Commerce" 2014 International Conference on Big Data and Smart Computing (BIGCOMP), pp. 96-98, 2014.

[10] Xianzhen Deng and Xinwei Wang, "Mining Rank-Correlated Associations for Recommendation Systems", International Conference on Web Information Systems and Mining, pp. 625-629, 2009.

[11] Y. K. Choi and S. K. Kim, "Recommendation algorithms for online shopping malls with periodically purchasing users" Journal of KIISE: Software and Applications, vol.40, no.8, 2013. (in Korean).

[12] G. Karypis, "Evaluation of item-based top-N recommendation algorithms," Proc. of CIKM '01 Conference, pp.247-254, 2001.

[13] S.Ganapathy, P.Vijayakumar, P.Yogesh, A.Kannan, "An Intelligent CRF based Feature Selection for Effective Intrusion Detection", International Arab Journal of Information Technology – Article in online.

[14] Kapil Kumar Gupta, Baikunth Nath, Ramamohanarao Kotagiri, Layered Approach Using Conditional Random Fields for Intrusion Detection, IEEE Transactions On Dependable And Secure Computing, Vol. 7, No. 1, pp. 35-49, 2010.