

# Business Strategies for Fashion Industry: Mines and preserves high utility items

Nandhini SS<sup>1</sup>, Priya R<sup>2</sup>, Saranya DV<sup>3</sup>, Shanmuga Priya V<sup>4</sup>, Sountharrajana S<sup>5</sup>

<sup>1,2,3,4</sup>Bannari Amman Institute of Technology, Sathyamangalam, Tamilnadu, India  
<sup>2</sup>VIT Bhopal University, Madhya Pradesh, India

Email: <sup>1</sup>nandhiniss@bitsathy.ac.in

**Abstract:** *The Global Apparel Sector projects a higher growth value of the overall economy. Fashion apparel business is always an ongoing process as long as mankind is existent on earth since clothing is one of our essentials. But the only thing that keeps this industry on spree is the changing consumer behavior patterns. This change in buying pattern is immensely influenced by the demographics and psychographics. Brands are forced to formulate strategies that best serve the needs and expectations of consumers. Satisfaction with emotional bonding is the key to success of brands these days. Also the buying behavior of the customer is widely affected by many factors. With all these measures, it really makes the process tough to bring in a strategy for fashion business based on the customer's buying patterns. Mining and preserving high utility itemsets from fashion data with all the dimensions, allows a business to plan a strategy to improve the profit and also it is important to preserve such valuable information from being discovered. The paper discusses the algorithms that work perfectly on fashion data being affected by multiple factors with the combination of privacy preservation.*

**Keywords:** *Fashion industry, fashion data, mining, privacy preservation, preservation, data mining, high utility itemsets, multi-dimensional.*

## 1. INTRODUCTION

### 1.1 Data Analysis in Fashion Industry

Since the last decade, there is an enormous growth in the data in all the industries. Fashion industry, being popular from human existence, has started working on their sales data and it is really not new to the fashion industries working on their data to come up with business models. But the data to be processed now becomes multi-dimensions and that makes the fashion data to stand out of the crowd, whereas the sales data from other industries involves quantity and combination of buying behavior as the major factor to be considered. In fashion data, major factors to be considered are the buying behaviour, factors affecting the buying behavior, the demographic scenarios affecting the behavior and a lot more. In recent years, the change in fashion is always suddenly happening and finding the reason behind will be done only after the change has made its complete move. So it is always important to forecast the trend and buying behaviors of the customer based on the sales data available.

### 1.2 Consumer behavior

Consumers are well aware of the recent trends and the phenomenal approaches of the industry as the impact of technological advancements are enormous. Customer engagement is

most important now-a-days as people look for the emotional connection with the brand they prefer. That is one of the reasons why loyalty programs are a mandate aspect in fashion retail. There could be numerous reasons for a consumer visiting a store to purchase something, wherein few end up purchasing and few don't. Such instances happen when customers are not met with their style preferences, colour choices, improper sales assistance and many more which is unpredictable.

Only when a customer is met with all his requirements irrespective of the aspects of sales to service and feels satisfied, becomes a consumer. Retaining existing consumers and attracting new customers are equally important to run any business. This is a huge process which involves a huge data tracking and not a one-time task but rather a repetitive process. But for all this, data is essential and gathering it is apparently quick and easier as all brands have a variety of Point-of-sale (POS) applications. The data are automatically stored in the database and is managed on a day-to-day or a weekly basis depending on the Standard Operating Procedures (SOP) of the brand. With the increase in organized Fashion retail sector over the two decades, analysis of the historic data in terms of sales, demographics and merchandise allocations are mandatory to study the trend movement along the strata

### **1.3 Challenges**

The season always affects the sales in the fashion industry and also the buying patterns will have a complete shift when there is a new season and that is obvious in all geographical locations and the mass production can be made based on the seasonal needs. Fashion is merely an expression which will be admired and accepted by a wide range of people over the years.

Marketing factors like trend prediction, purchase impulse, life cycle of a trend, demand in the market, satisfying the market demand all depends on the time taken by the industry to reach the customers. Nowadays, the fashion industry follows "Fast Fashion", the term that represents the flow of trends from launching in fashion shows to being available in a retail store and again that affects the buying behavior of the customers. So, the time to market plays an important role. The apparel retailers need to make their products available on time in the market which is not in the stores already and before their competitors. Further, based on the response once the product is released should be again taken for consideration for the design of the next product.

## **2. LITERATURE SURVEY**

Utility mining is the process of analyzing the data and discovering itemsets that satisfies the utility given by the user. The utility is classified as internal utility and external utility, where the internal utility mostly represents the quantity of the item and the external utility represents the profit obtained from the itemsets [1]. The high utility itemset mining (HUIM) using genetic algorithm [2] handles the large search space and it proves that the genetic algorithm is robust among others. The major challenge in HUIM is always the exponentially larger search space and that is addressed by HUI-GA.

The algorithm iFUM is the improved version of the UMining algorithm and FUM algorithm and that proved to be more efficient than these algorithms on IBM synthetic dataset [3]. The iFUM shows improvement in time and even it scales well with more number of unique items when gets added up in the dataset.

The challenge with HUIM is finding the itemsets from a database which has a very high range of utility when compared to the other itemsets and the challenge will be higher when the dataset has much more distinct items in it and it also affects the execution time to discover the

high utility itemsets (HUI). Grey wolf optimization [4] provides an optimized way to mine HUI using boolean operations. To mine HUI from parallel and scalable datasets, HDFS - based mining [5] was proposed to address the challenges in parallel datasets. With exponential increase in the data, and the analysis being made on the data for knowledge extraction compromises the security and privacy of the sensitive data present in the database. It is necessary to preserve the sensitive data from the competitors and hence privacy preservation is an important task in the knowledge extraction process. Genetic algorithm [6] used to preserve the HUIs in the dataset which enables HUIs being not discovered by the third party.

### **HUIM - GA**

The novel approach to discover high utility itemset mining using genetic algorithm [2], identifies the HUIs using ranked mutation method by following all the steps in the genetic algorithm and this approach is suggested for the datasets with large search space which is the case in datasets from fashion industry and hence requires more memory usage. Also the minimum utility value as threshold values needs to be given to discover the HUIs and another approach does not require the threshold value instead it optimizes the HUIs.

### **Privacy Preserving**

Privacy preservation in utility mining using genetic algorithm [6], utilizes various operators of genetic algorithms to preserve the HUIs. This approach reduces the scanning time required over the large dataset and also decreases the number of scans for the algorithm to hide the HUIs from the transactional database [7][8]. Also, it is possible for any algorithm to fail to cover the HUIs during privacy preservation [9][10] and these are the challenges when hiding the HUIs in a transactional database and producing a new sanitized database [11][12].

### **HUIM-PP GA**

The proposed approach is to combine the mining process of discovering HUIs using the genetic algorithm and the privacy preservation process using the genetic algorithm[13]. Once a sanitized database is created and again the database is scanned for rare HUIs and the same will be hidden using the genetic algorithm approach. As specified in [6], the fitness function may be defined as:

$$\text{Fitness function } f = ( \alpha * 15 )$$

where, 15 is the constant taken for calculating fitness function and  $\alpha$  represents the maximum number of bits present in the lengthiest combination of items from items. Similarly, the utility of an item is calculated by using the formula as in [6]:

$$u(i_v) = \sum ex(i_v) * c(i_v, T_q)$$

where,

$u(i_v)$  - utility of an itemset  $i_v$

$ex(i_v)$  - external utility of an itemset  $i_v$

$c(i_v, T_q)$  - summation of profit of each item in itemset \* quantity of that item in the itemset under consideration

Using this fitness function in the utilities of the itemset in the fashion database, the HUIs are discovered and the same will be hidden by altering the utility values[14]. For all HUIs discovered, the external utility of all the items are reduced to '0' to hide the HUIs and the remaining items' external utility will not be changed even though they are present in the itemset.

### 3. RESULTS AND DISCUSSION

This hybrid approach helps in finding the HUIs that need to be examined for predicting the fashion trend and customer behavior. Also hides the HUIs to be able to be predicted by the third party members who have access to the data set since the items' external utility values are reduced to '0' in the sanitized database that is kept in the place of the original database. The CPU usage with the number of records is plotted below.

**Table 1. CPU time and No. of records**

No. of Records (*10 <sup>4</sup> )	CPU Time (ms)
0.25	20
0.5	22
0.75	81
1	108
1.25	124

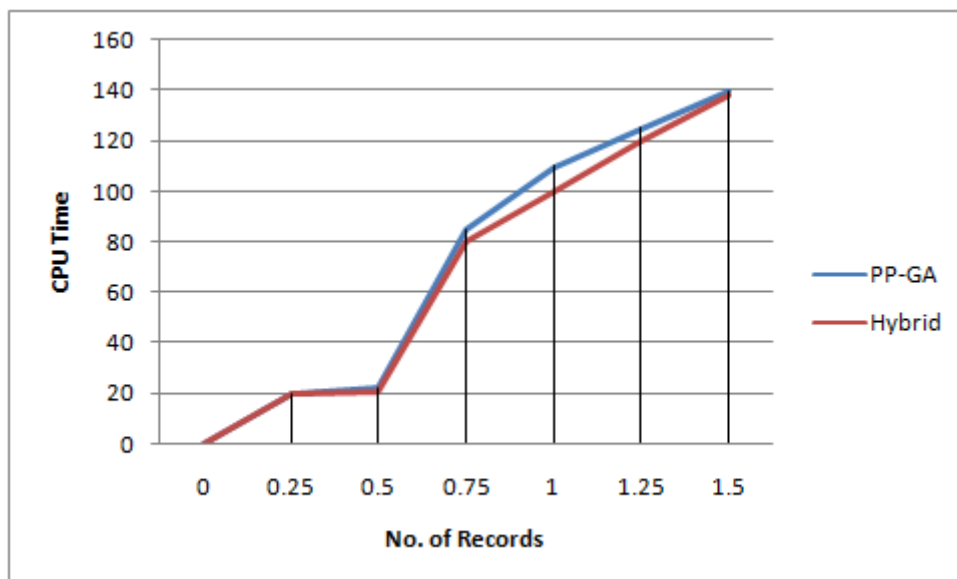


Figure 1. Comparison CPU time and No. of Records with PP-GA and Hybrid approach

### 4. CONCLUSION

The hybrid approach proves to be efficient when the dataset is mediumly large and multi-dimensional. Also to proceed with this work, the fashion data can be cleansed before applying this approach. The cleansing may be like stemming down all the categories of items into basic items. This may miss out the major expectation in predicting the trend but provides

better results for privacy preservation. Based on the context, the approaches may be varied and preprocessing on the fashion data provides a way to make the work efficient.

## 5. REFERENCES

- [1] Philippe Fournier-Viger, Jerry Chun-Wei Lin, Quang-Huy Duong, Thu-Lan Dam, “PHM: Mining Periodic High-Utility Itemsets”, *Advances in Data Mining. Applications and Theoretical Aspects. ICDM 2016*.
- [2] S. Kannimuthu & K. Premalatha (2014), “Discovery of High Utility Itemsets Using Genetic Algorithm with Ranked Mutation”, *Applied Artificial Intelligence: An International Journal*, 28:4, 337-359, DOI: 10.1080/08839514.2014.891839.
- [3] S. Kannimuthu, K. Premalatha & S. Shankar, “iFUM - Improved Fast Utility Mining”, *International Journal of Computer Applications (0975 - 8887)*, Volume 27 - No.11, August 2011.
- [4] Pazhaniraja, N., Sountharajan, S. & Sathis Kumar, B. High utility itemset mining: a Boolean operators-based modified grey wolf optimization algorithm. *Soft Comput* 24, 16691–16704 (2020).
- [5] S. Sountharajan, E. Suganya, N. Aravindhraj, S. Sankarananth, C. Rajan, “HDFS-based parallel and scalable pattern mining using clouds for incremental data”, *International Journal of Computer Aided Engineering and Technology*, Vol. 13, Nos.1/2, 2020.
- [6] Sugandha Rathi, Rishi Soni, Virendra Singh Kushwah, “A New Algorithm for Privacy Preservation in Utility Mining Using Genetic Algorithm”, *Publication:ICTCS '16: Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies*, March 2016.
- [7] Fayyad, U.M., Shapiro, G. P.and Smyth, P. *Data Mining to Knowledge Discovery in Databases*. ISSN 0738-4602. *AI Magazine*, 1996, 37–53.
- [8] Han J and Kamber M. *Data Mining: Concepts and Techniques*. Second edition Morgan Kaufmann Publishers.
- [9] Lindell Y., Pinkas B.: *Privacy-Preserving Data Mining*. CRYPTO, 2000.
- [10] Bertino, E., Fovino, I. N., Provenza, L. P. A framework for evaluating privacy preserving data mining algorithms, *Data Mining and Knowledge Discovery*, 2005, Vol. 11, No. 2, 121-154.
- [11] Mukherjee, S., Chen Z., Gangopadhyay, A. A privacy preserving technique for Euclidean distance-based mining algorithms using Fourier-related transforms, *The VLDB Journal*, Vol: 15, No. 4, 2006, 293–315.
- [12] Agrawal, R. and Srikant, R. Privacy-preserving data mining, In *Proc. of ACM SIGMOD'00*, 2000, 439–450, Dallas, Texas, USA.
- [13] Sujatha krishnamoorthy, Muthukumar, Balamuraugan “Effective data Access provision with advanced Security Features in Cloud Computing (SUSCOM 2019/Feb (26-28 India)
- [14] Viji, C., Rajkumar, N., Suganthi, S.T. et al. An improved approach for automatic spine canal segmentation using probabilistic boosting tree (PBT) with fuzzy support vector machine. *J Ambient Intell Human Comput* (2020).