



# Digital Receipt

This receipt acknowledges that **Turnitin** received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author: Ridowati Gunawan  
Assignment title: Periksa similarity  
Submission title: Online Retail Pattern Quality Improvement From Frequent S...  
File name: equent\_Sequential\_Pattern\_to\_High-Utility\_Sequential\_Patter...  
File size: 167.51K  
Page count: 5  
Word count: 4,674  
Character count: 25,543  
Submission date: 07-Mar-2022 11:28AM (UTC+0700)  
Submission ID: 1778182441

2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)

## Online Retail Pattern Quality Improvement: From Frequent Sequential Pattern to High-Utility Sequential Pattern

Ridowati Gunawan  
Informatics Study Program  
Sesana Dharma University  
Yogyakarta, Indonesia  
rdo@sud.ac.id

*Abstract*—There has been a change in people's shopping behavior, especially during the Covid-19 pandemic, from what traditionally requires direct face-to-face meetings between sellers and buyers, to virtual face-to-face through various shopping media. All activities carried out by customers, click streams performed, items purchased, the number of items including the price will be recorded in a log. Activity records in the log are very useful to be able to find out the pattern of activity sequences from customers, especially the order of items purchased by customers. However, the management certainly needs more knowledge, not just the order of goods that are often purchased by customers. Does the order of items purchased also provide maximum profit? There have been many methods to get frequent sequential patterns from customer activities, but getting a pattern that chooses more quality by adding utility value needs to be considered. In this research, the method used to obtain frequent sequential patterns is using Profibusan (Prefix-projected Sequential Pattern) and the method used to obtain a high-utility sequential pattern is the USpan (Utility Sequential Pattern) method. USpan is applied to the BMS (Blue Martini Database) dataset, which is the dataset used in KDD (Knowledge Discovery in Databases) CUP 2000 which consists of clickstream data from an e-commerce. The experimental results show that the frequent sequential pattern will always appear in the high-utility sequential pattern but not vice versa. It is certain that a high-utility pattern must be sequential, but a sequential pattern is not necessarily a high-utility sequential. From the results of the high-utility sequential pattern, it can be used as input to provide recommendations to customers to carry out the shopping process on items that can provide greater profits. The conclusion of the research conducted is that the high-utility sequential pattern mining can produce a higher quality pattern than just getting a frequent sequential pattern.

*Index Terms*—online retail, frequent sequential pattern, high-utility, high-utility sequential pattern, e-commerce

### I. INTRODUCTION

One of the impacts of increasing internet access and adoption is the increasing number of digital buyers every year. In 2020, more than two billion people bought goods or services online. The transaction value of e-retail sales exceeds US\$ 4.2 trillion. The COVID-19 pandemic has had a significant impact on e-commerce and online consumer behaviour worldwide. Digital channels are becoming the most crowded stores amid the people's need to stay at home during the pandemic. In June

2020, e-commerce activity reached a record 22 billion visits per month. During the pandemic, e-commerce sales increased by more than 25 percent. In the 3rd quarter of 2020 Indonesia was reported as the country that ranks the highest in terms of online purchases using mobile devices [1].

E-commerce activities and online shopping are certainly interesting to observe. All online shopping activities are recorded in a system, whether carried out by online shop owners or using third party services such as marketplace sites. Activities in online shopping will certainly be interesting to note. Customer shopping behaviour, the order of the items purchased by the customer, what items contribute the greatest profit, all these will be a concern for online business owners to further develop their business.

Several methods have been developed to observe the shopping activities of customers on the online system. [2] provides the concept of sequential pattern mining for the market basket analysis by using the General Sequential Pattern (GSP). This research is used as a reference for the sequential pattern mining. The process begins by determining the minimum support value. The sequential pattern obtained is a pattern that has a support value higher than the specified minimum support value. This study has not considered the benefits obtained for each sequence found, only focusing on searching for frequent sequences.

Customer behaviour in the entire shopping process can be analysed using sequential mining and clustering. The method used to perform the analysis is to adopt maximal repeat patterns (MRPs) and lag sequential analysis (LSA) to analyse sequential searches and identify significant repeating patterns [3]. The results show that customers who have definite goals will use sequential search using a shorter path than customers who only do exploration. Customers who only do exploration have a more complicated sequential pattern. However, although sequential patterns have been obtained, it is not yet known how much profit can be obtained from each sequential pattern. Would a shorter sequential pattern get a bigger advantage?

[4] proposed a system that recommends items based on matching rules obtained from frequent sequential purchase

# Online Retail Pattern Quality Improvement From Frequent Sequential Pattern to High-Utility Sequential Pattern

*by* Gunawan Ridowati

---

**Submission date:** 07-Mar-2022 11:28AM (UTC+0700)

**Submission ID:** 1778182441

**File name:** equent\_Sequential\_Pattern\_to\_High-Utility\_Sequential\_Pattern.pdf (167.51K)

**Word count:** 4674

**Character count:** 25543

# Online Retail Pattern Quality Improvement: From Frequent Sequential Pattern to High-Utility Sequential Pattern

Ridowati Gunawan

Informatics Study Program  
Sanata Dharma University  
Yogyakarta, Indonesia  
rido@usd.ac.id

**Abstract**—There has been a change in people's shopping behavior, especially during the Covid-19 pandemic, from what traditionally requires direct face-to-face meetings between sellers and buyers, to virtual face-to-face through various shopping media. All activities carried out by customers, click streams performed, items purchased, the number of items including the price will be recorded in a log. Activity records in the log are very useful to be able to find out the pattern of activity sequences from customers, especially the order of items purchased by customers. However, the management certainly needs more knowledge, not just the order of goods that are often purchased by customers. Does the order of items purchased also provide maximum profit? There have been many methods to get frequent sequential patterns from customer activities, but getting a pattern that chooses more quality by adding utility value needs to be considered. In this research, the method used to obtain frequent sequential patterns is using PrefixSpan (Prefix-Projected Sequential PatterN) and the method used to obtain a high-utility sequential pattern is the USpan (Utility Sequential PatterN) method. USpan is applied to the BMS (Blue Martini DataSet) dataset, which is the dataset used in KD [36] Knowledge Discovery in Databases) CUP 2000 which consists of clickstream data from an e-commerce. The experimental results show that the frequent sequential pattern will always appear in the high-utility sequential pattern but not vice versa. It is certain that a high-utility pattern must be sequential, but a sequential pattern is not necessarily a high-utility sequential. From the results of the high-utility sequential pattern, it can be used as input to provide recommendations to customers to carry out the shopping process on items that can provide greater profits. The conclusion of the research conducted is that the high-utility sequential pattern mining can produce a higher quality pattern than just getting a frequent sequential pattern.

**Index Terms**—online retail, frequent sequential pattern, high-utility, high-utility sequential pattern, e-commerce

## I. INTRODUCTION

One of the impacts of increasing internet access and adoption is the increasing number of digital buyers every year. In 2020, more than two billion people bought goods or services online. The transaction value of e-retail sales exceeds US\$ 4.2 trillion. The COVID-19 pandemic has had a significant impact on e-commerce and online consumer behaviour worldwide. Digital channels are becoming the most crowded stores amid the people's need to stay at home during the pandemic. In June

2020, e-commerce activity reached a record 22 billion visits per month. During the pandemic, e-commerce sales increased by more than 25 percent. In the 3rd quarter of 2020 Indonesia was reported as the country that ranks the highest in terms of online purchases using mobile devices [1].

E-commerce activities and online shopping are certainly interesting to observe. All online shopping activities are recorded in a system, whether carried out by online shop owners or using third party services such as marketplace sites. Activities in online shopping will certainly be interesting to note. Customer shopping behaviour, the order of the items purchased by the customer, what items contribute the greatest profit, all these will be a concern for online business owners to further develop their business.

Several methods have been developed to observe the shopping activities of customers on the online system. [2] provides the concept of sequential pattern mining for the market basket analysis by using the General Sequential Pattern (GSP). This research is used as a reference for the sequential pattern mining. The process begins by determining the minimum support value. The sequential pattern obtained is a pattern that has a support value higher than the specified minimum support value. This study has not considered the benefits obtained for each sequence found, only focusing on searching for frequent sequences.

Customer behaviour in the entire shopping process can be analysed using sequential mining and clustering. The method used to perform the analysis is to adopt maximal repeat patterns (MRPs) and lag sequential analysis (LSA) to analyse sequential searches and identify significant repeating patterns [3]. The results show that customers who have definite goals will use sequential search using a shorter path than customers who only do exploration. Customers who only do exploration have a more complicated sequential pattern. However, although sequential patterns have been obtained, it is not yet known how much profit can be obtained from each sequential pattern. Would a shorter sequential pattern get a bigger advantage?

[4] proposed a system that recommends items based on matching rules obtained from frequent sequential purchase

patterns. The combination of semantic context and sequential historical purchase has given good results and can reflect user preferences through similar product recommendations and semantic sequences. However, the observed sequential purchases have not considered whether the sequential provides an advantage because it only follows the sequence. The recommendation pattern will be more meaningful if the observed sequence also considers the advantages of the sequence.

Comparison of the Generalized Sequential Pattern and FP-Growth sequential pattern algorithms from online retailers to get products recommendation to customer has been successfully carried out [5], but only limited to getting frequent sequential patterns. The use of sequential patterns for various fields has been carried out, such as for online retail [6]–[8] and for the field of e-learning [9], [10]. The use of sequential patterns will have more quality when the pattern sought is not limited to frequent sequential patterns but adds utility value in the pattern search process.

To further improve the quality of the patterns obtained, one way to do this is to develop the search for frequent sequential patterns into high utility sequential patterns so that they have more impact and benefit value for decision makers, especially in the online retail sector. The purpose of this study is to obtain patterns that are more qualified and useful for decision makers using a high-utility sequential pattern approach. The analysis carried out is to compare the pattern results obtained using the frequent sequential pattern and high-utility sequential pattern. For frequent sequential pattern mining we use PrefixSpan (Prefix-projected Sequential pattern mining) and using USpan (Utility Sequential Pattern Mining) for high-utility sequential pattern mining.

The rest of this paper is organized as follows: Related work is briefly reviewed in Section II. Methodology is presented in Section III. Experimental design, result, and analysis are provided in Section IV. The paper is concluded in Section V and also for future works.

## II. RELATED WORK

In this section, work related to methods for sequential pattern mining, high-utility itemsets mining and high-utility sequence itemsets mining will be briefly reviewed.

Sequential pattern mining is the process of searching for frequent sequences in a sequence databases. Sequential pattern mining has been developed for a wide variety of applications such as web click-stream analysis, medical data, biological data, e-learning and e-commerce. Initially, the algorithm for sequential pattern search used the Apriori approach, based on the Apriori property for frequent itemset search [11]. Another alternative, besides using Apriori based, is using the pattern-growth approach FP-growth algorithm to search for frequent itemset [12]. These two approaches are the basis for sequential pattern mining. For Apriori-based searches, sequential pattern mining generally uses the [13], [2], [14] algorithm. Based on the input format, the Apriori-based sequential pattern is divided into two categories, namely the Apriori-based with horizontal format as in the General Sequential Pattern (GSP)

algorithm [2] and the Apriori-based with vertical format as in the SPADE (Sequential Pattern Discovery using Equivalence classes) [15] and SPAM (Sequential Pattern Mining) [16] algorithms. The first algorithm to use the item-growth concept for sequential pattern mining is the FreeSpan (Frequent Pattern-Projected Sequential Pattern Mining) algorithm [17]. [18] proposes PrefixSpan (Prefix-projected sequential Pattern), which uses prefix-projection to mine a complete set of sequential patterns. In mining large databases, PrefixSpan outperforms both the GSP algorithm and FreeSpan.

Research on frequent itemset mining and frequent sequential mining both assume that each item appears once in every transaction and has the same level of importance [19]. This limitation has given rise to research on high-utility itemset mining. Consideration of the usefulness of each item becomes the main consideration. Like sequential pattern mining, high-utility itemset mining also uses an Apriori-based approach [20]–[22], pattern-growth [23], [24]. The search for high-utility itemset has also been carried out using a tree approach [25]–[27], based on indexing [28], using computational intelligence based on genetic algorithms [29], particle swarm optimization [30]–[33], ant colony system [34], bee colony [35], and bio-inspired [36]. The merging of the concepts of sequential itemset and high-utility itemset led to research on high-utility sequential mining (HUSPM), which is to get frequent sequential patterns that have high-utility at the same time. [37] revealed the concept of high-utility sequential pattern mining (HUSPM), which is a search for sequential patterns that also has high utility. The input from HUSPM is the minimum utility limit and the sequence itemset that has a utility value for each item. The output of HUSPM is a set of high-utility sequences that have a utility value higher than the specified minimum utility. HUSPM was developed using the basic sequential pattern algorithm based on the general sequential pattern or using the pattern growth approach. [38] improved pattern search based on a vertical database format named USpan (Utility Sequential Pattern). Generate itemset is composed of a lexicographic q-sequence tree, two concatenation mechanisms, and two pruning strategies. Meanwhile, the strategy to find high utility pattern is using depth-first search strategy. This approach is used in this study. There are several other algorithms to get high-utility sequential pattern including [39]–[41].

## III. METHODOLOGY

Our proposed method consists of four processes, namely: pre-processing, sequential pattern using PrefixSpan algorithm, high-utility sequential pattern mining using USpan algorithm, and last process is evaluation pattern.

### A. Pre-Processing

The pre-processing phase is carried out to prepare the collected data so that it is ready to be processed into the algorithm. Because there are two algorithms that will be applied, at this stage the data that has been collected will be replicated into two new datasets. The dataset for the frequent sequential



pattern is only the sequence database, while the dataset for obtaining the high-utility sequential itemset is a utility-based sequence dataset. Each item in the sequence in the database has utility information in the form of  $item[utility]$ .

#### B. PrefixSpan (Prefix-projected Sequential pattern)

The sequential pattern mining algorithm used is the PrefixSpan algorithm because it has better results than GSP and FreeSpan [18]. The pseudo-code of PrefixSpan is shown in Algorithm 1.

#### Algorithm 1 PrefixSpan( $S, minsup$ )

```

1: Input : A sequence databases  $S$ , minimum support thresh-
   35  $minsup$ 
2: Output : The complete set of sequential patterns
3: Parameter:  $\alpha$  : a sequential pattern;  $l$  : the length of  $\alpha$ ;
    $S|\alpha$ : the  $\alpha$  - projected database, if  $\alpha \neq \langle \rangle$ ; otherwise,
   the sequence database  $S$ 
4: Method:
5: Scan  $S|\alpha$  once, find the set of frequent items  $b$  such that
6: (a)  $b$  can be assembled to the last element of  $\alpha$  from a
   sequential pattern; or
7: (b)  $\langle b \rangle$  can be appended to  $\alpha$  from a sequential pattern.
8: for each frequent item  $b$  do
9:   append it to  $\alpha$  to form a sequential pattern  $\alpha'$  and
   output  $\alpha'$ ;
10: end for
11: for each  $\alpha'$  do
12:   construct  $S|\alpha'$  projected databases  $S|\alpha'$ ,
13:   and call PrefixSpan( $\alpha', l+1, S|\alpha'$ ).
14: end for

```

#### C. USpan (Utility Sequential Pattern)

The pseudo-code of USpan is shown in Algorithm 2 [38].

#### D. Pattern Evaluation

Pattern evaluation is done by comparing the results of the itemset formed between those generated with PrefixSpan and USpan. The measurement for PrefixSpan is the support value limit, while for USpan it is the utility value limit. The assumption used is that the pattern obtained from PrefixSpan has the same utility value, namely one.

### IV. EXPERIMENTS, RESULT AND ANALYSIS

#### A. Dataset

The dataset used is the BMS (Blue Martini Dataset). This dataset was used in KDD CUP 2000. It contains clickstream data from an e-commerce. Characteristics of the datasets used in the experiments as seen in Table I.

#### B. Experimental setup

All algorithms are implemented in Java, executed on a PC with AMD Ryzen 7 2700 Eight-Core Processor, 3200 Hz, 8 cores, 16 logical processors, and 16.0 GB of RAM. Implementation of the algorithms is based on SPMF, the open-source data mining library [42].

#### Algorithm 2 USpan( $t, v(t)$ )

```

1: put : A sequence  $t$ ,  $t$ 's utility  $v(t)$ , a utility-based se-
   quence database  $S$ , the minimum utility threshold  $minutil$ 
2: output : All high-utility sequential pattern
3: if  $p$  is a leaf node then return
4: end if
5: Scan the projected database  $S(v(t))$  once to:
6: a) put I-Concatenation items into  $ilist$ , or
7: b) put S-Concatenation items into  $slist$ 
8: move unpromising items in  $ilist$  and  $slist$ 
9: for each item  $i$  in  $ilist$  do
10:    $(t', v(t')) \leftarrow I - Concatenate(p, i)$ 
11:   if  $v\_max(t') \geq minutil$  then
12:     output  $t'$ 
13:   end if
14:   USpan( $t', v(t')$ )
15: end for
16: for each item  $i$  in  $slist$  do
17:    $(t', v(t')) \leftarrow S - Concatenate(p, i)$ 
18:   if  $v\_max(t') \geq minutil$  then
19:     output  $t'$ 
20:   end if
21:   USpan( $t', v(t')$ )
22: end for
23: return

```

TABLE I  
CHARACTERISTICS OF THE DATASETS USED IN THE EXPERIMENTS

Characteristics	Value
Number of sequence	59,601
Number of distinct item	497
Average number of itemsets per sequence	2.5
Average distinct item per sequence	2.5
Average number of items	1
Total utility	2,477,599
Average utility per sequence	41.57

To get a sequential pattern from PrefixSpan, experiments were carried out using various minimum support values, namely 0.6%, 0.7%, 0.8%, 0.9%, and 1.0%. The observed results are frequent sequence count, total time, and memory used.

The minimum utility values used to obtain high-utility sequential patterns from USpan are 3000, 4000, 5000, 6000, and 7000. The observed results are high-utility sequence count, total time, and memory used.

#### C. Result and analysis

The experimental results using the PrefixSpan algorithm as in Table II. The higher the minimum support value, the fewer the number of patterns formed. The obtained subsequences are not too long, because the average number of itemsets per sequence is only 2.5, resulting in a maximum of 3 items being obtained (see the initial data in Table I).

TABLE II  
EXPERIMENT RESULT FOR PREFIXSPAN

No	Minsup (%)	Minsup Absolute	Number of Patterns	Total Time (ms)	Memory (mb)
#1	0.6	358	162	1437	125
#2	0.7	418	133	1060	244
#3	0.8	477	105	906	131
#4	0.9	537	90	744	58
#5	1	597	77	909	134

The experimental results using the USpan algorithm as in Table III. The higher the minimum utility value, the fewer the number of patterns formed. When the minimum utility value is 3000, it means that the minimum profit obtained is 3000, the number of itemset obtained is very large and most of them are 1 item. Memory usage is quite stable, which is 538 megabytes, while the time required to obtain a pattern is proportional to the number of itemset obtained and inversely proportional to the specified minimum utility value.

TABLE III  
EXPERIMENT RESULT FOR USpan

No	Minutil (%)	Minutil Absolute	Number of Patterns	Total Time (ms)	Memory (mb)
#1	0.12	3000	106127	385988	538
#2	0.16	4000	5175	319671	538
#3	0.20	5000	1151	288730	538
#4	0.24	6000	559	263402	538
#5	0.28	7000	342	252002	538

Comparison of patterns obtained from PrefixSpan and USpan as in Table IV. The patterns shown in Table IV are part of the patterns obtained with the minimum limit of the number of items in the pattern formed is 2 items and the minimum utility is 6000. The experimental results show that the pattern formed in PrefixSpan always exists in the pattern from the USpan algorithm, but not vice versa. This shows that although it is not frequent, the pattern has a high-utility value. In the USpan algorithm, the number of items in each sequence is more than in the PrefixSpan algorithm. For example, it can be seen in the pattern that begins with the item 10295. There are four patterns formed in PrefixSpan, each pattern consists of two items. Meanwhile, for USpan, six patterns can be formed, there are two patterns that have three items and four pattern that have two items. There is an increase in the number of patterns obtained using USpan.

## V. CONCLUSIONS AND FUTURE WORKS

Searching for sequence patterns by entering utility values can improve the quality of the obtained sequence patterns. Patterns that have high utility can be frequent sequences but not vice versa. The addition of utility values is needed to improve the quality of the pattern so that it can better assist decision makers.

For future work, high-utility sequential pattern mining can be applied to various cases such as in e-learning. Not just

TABLE IV  
COMPARISON OF PATTERNS

PrefixSpan		Uspan	
Pattern	Support	Pattern	Utility
10295 , 10307	916	10295 , 10307	35418
10295 , 10311	738	10295 , 10311	24203
10295 , 10315	722	10295 , 10315	23962
10307 , 10311	621	10295 , 10307 , 10311	18322
10307 , 10315	496	10295 , 12895	15736
10311 , 10315	771	10295 , 10299 , 10307	14739
10311 , 12487	615	10311 , 32213	18214
10311 , 12703	576	12487 , 32213	22826
12483 , 12487	877	12487 , 12703 , 32213	16166
12487 , 12703	631	12679 , 12683	15342
12487 , 32213	506	12679 , 12895	15909
12695 , 12703	615	12703 , 32213	22258
12703 , 32213	571	12815 , 12895	18088
12815 , 12895	552	12827 , 12895	23029
12827 , 12895	590	33433 , 33469	27398
33433 , 33469	509	33433 , 33449	22780
33449 , 33469	1204	32205 , 32213	20848
		33433 , 33449 , 33469	19934
		33433 , 33453	16360
		33449 , 33469	39699

looking for material but can find out the search for material that can be used to get good learning values (utility).

Sequential pattern mining does not only start from the process after the consumer has made the purchase process, but can also start from "digital journey" (what is interesting thing for the customer) until finally making the purchase process. Combining digital journey and high-utility sequential pattern mining is a research that will try to be developed.

## REFERENCES

- [1] D. Coppola, "E-commerce worldwide - statistics & facts," 2021. [Online]. Available: <https://www.statista.com/topics/871/online-shopping/#dossier-chapter1>
- [2] R. Srikant and R. Agrawal, "Mining sequential patterns: Generalizations and performance improvements," in *Apers P., Bouzeghoub M., Gardarin G. (eds) Advances in Database Technology — EDBT '96*. Springer, Berlin, Heidelberg, 1996, pp. 1–17. [Online]. Available: <http://link.springer.com/10.1007/BFb0014140>
- [3] I.-C. Wu and H.-K. Yu, "Sequential analysis and clustering to investigate users' online shopping behaviors based on need-states," *Information Processing & Management*, vol. 57, no. 6, p. 102323, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306457320308189>
- [4] M. Nasir, C. I. Ezeife, and A. Gidado, "Improving e-commerce product recommendation using semantic context and sequential historical purchases," *Social Network Analysis and Mining*, vol. 11, no. 1, p. 82, 2021. [Online]. Available: <https://doi.org/10.1007/s13278-021-00784-6>
- [5] Destilia, R. Primartha, Sukemi, and A. Wijaya, "Online Retail Marketing Recommendation System Based on Generalized Sequential Pattern Algorithm and FP-Growth Algorithm," in *Proceedings of the Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019)*, vol. 172, no. Siconian 2019, 2020, pp. 353–357.
- [6] V. Kakollu, M. Srivenkatesh, K. N. Soujanya, and B. Sridevi, "Sequential Pattern Mining With Variables Used Log , Retail and Medical Analysis," vol. 27, no. 6, pp. 481–486, 2021.
- [7] G. Aloysius and D. Binu, "An approach to products placement in supermarkets using PrefixSpan algorithm," *Journal of King Saud University - Computer and Information Sciences*, vol. 25, no. 1, pp. 77–87, 2013. [Online]. Available: <http://dx.doi.org/10.1016/j.jksuci.2012.07.001>

- [8] M. Kanaan, H. Kheddouci, M. Patterns, W. Durations, and E.-c. Dataset, "Mining Patterns With Durations from E-commerce Dataset," *Complex Network*, vol. Dec, no. hal-01960321, 2018.
- [9] E. Doko, L. A. Bexheti, M. Hamiti, and B. P. Etemi, "Sequential pattern mining model to identify the most important or difficult learning topics via mobile technologies," *International Journal of Interactive Mobile Technologies*, vol. 12, no. 4, pp. 109–122, 2018.
- [10] Q. Li, R. Baker, and M. Warschauer, "Using clickstream data to measure, understand, and support self-regulated learning in online courses," *Internet and Higher Education*, vol. 45, no. July 2019, p. 100727, 2020. [Online]. Available: <https://doi.org/10.1016/j.iheduc.2020.100727>
- [11] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proceedings of the 20th International Conference on Very Large Data Bases*, vol. 1215, 1994, pp. 487–499.
- [12] J. Han, P. Jian, and Y. Yin, "Mining frequent patterns without candidate generation," in *SIGMOD '00 Proceedings of the 2000 ACM SIGMOD International Conference on Management of data*, vol. 1. Dallas, Texas, USA: ACM Press, New York, NY, USA, 2000, pp. 1–12.
- [13] R. Agrawal and R. Srikant, "Mining sequential patterns," in *Proceedings of the Eleventh International Conference on Data Engineering*. Taipei, Taiwan, Taiwan: IEEE Comput. Soc. Press, 1995, pp. 3–14. [Online]. Available: <http://ieeexplore.ieee.org/document/380415/>
- [14] P. Fournier-Viger, J. C. W. Lin, R. U. Kiran, Y. S. Koh, and R. Thomas, "A Survey of Mining sequential patterns," *Data Science and Pattern Recognition Ubiquitous International*, vol. 1, no. Februari, 2017.
- [15] M. J. Zaki, "SPADE: An efficient algorithm for mining frequent sequences," *Machine Learning*, vol. 42, no. 1-2, pp. 31–60, 2001.
- [16] J. Ayres, J. Flannick, J. Gehrke, and T. Yiu, "Sequential Pattern mining using a bitmap representation," in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '02*. New York, New York, USA: ACM Press, 2002, p. 429. [Online]. Available: <http://portal.acm.org/citation.cfm?doid=775047.775109>
- [17] J. Han, J. Pei, B. Mortazavi-Asl, Q. Chen, U. Dayal, and M. Hsu, "Freespan: Frequent pattern-projected sequential pattern mining," in *Proceeding of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. Proceeding of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, R. Ramakrishnan, S. Stolfo, R. Bayardo, I. Parsa, R. Ramakrishnan, S. Stolfo, R. Bayardo, and I. Parsa, Eds., Dec. 2000, pp. 355–359, proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2001) ; Conference date: 20-08-2000 Through 23-08-2000.
- [18] J. Pei, J. Han, B. Mortazavi-Asl, J. Wang, H. Pinto, Q. Chen, U. Dayal, and M. C. Hsu, "Mining sequential patterns by pattern-growth: The prefixspan approach," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 11, pp. 1424–1440, 2004.
- [19] H. Yao, H. Hamilton, and C. Butz, "A Foundational Approach to Mining Itemset Utilities from Databases," in *Proceedings of the 2004 Society for Industrial and Applied Mathematics (SIAM) International Conference on Data Mining*. Lake Buena Vista: SIAM, 2004, pp. 482–486.
- [20] Y. Liu, W.-k. Liao, and A. Choudhary, "A two-phase algorithm for fast discovery of high utility itemsets," in *Proceeding PAKDD'05 Proceedings of the 9th Pacific-Asia conference on Advances in Knowledge Discovery and Data Mining*. Springer-Verlag Berlin, Heidelberg, 2005, pp. 689–695.
- [21] —, "A fast high utility itemsets mining algorithm," in *Proceedings of the 1st International Workshop on Utility-Based Data Mining - UBDM '05*. Chicago, Illinois, USA: ACM, 2005, pp. 90–99.
- [22] Y. Liu, W.-k. Liao, A. Choudhary, H. Yao, and H. J. Hamilton, "Mining itemset utilities from transaction databases," in *Data and Knowledge Engineering*, vol. 59. Chicago, Illinois, USA: ACM, 2005, pp. 603–626.
- [23] V. Tseng, C. Wu, B. Shie, and P. Yu, "UP-Growth: an efficient algorithm for high utility itemset mining," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 253–262, 2010.
- [24] A. Erwin, R. P. Gopalan, and N. Achuthan, "CTU-Mine: An Efficient High Utility Itemset Mining Algorithm Using the Pattern Growth Approach," *7th IEEE International Conference on Computer and Information Technology (CIT 2007)*, pp. 71–76, 2007.
- [25] C.-W. Lin, T.-P. Hong, and W.-H. Lu, "An effective tree structure for mining high utility itemsets," *Expert Systems with Applications*, vol. 38, no. 6, pp. 7419–7424, 2011.
- [26] M. Liu and J. Qu, "Mining High Utility Itemsets without Candidate Generation," in *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM)*, 2012, pp. 55–64.
- [27] P. Fournier-Viger, C. W. Wu, S. Zida, and V. S. Tseng, "FHM: Faster high-utility itemset mining using estimated utility co-occurrence pruning," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8502 LNAI, no. June. Springer International Publishing, 2014, pp. 83–92.
- [28] G.-C. Lan, T.-P. Hong, and V. S. Tseng, "An efficient projection-based indexing approach for mining high utility itemsets," *Knowledge and Information Systems*, vol. 38, no. 1, pp. 85–107, 2014.
- [29] S. Kannimuthu and K. Premalatha, "Discovery of high utility itemsets using genetic algorithm," *International Journal of Engineering and Technology*, vol. 5, no. 6, pp. 4866–4880, 2013.
- [30] J. C.-W. Lin, L. Yang, P. Fournier-Viger, J. M. T. Wu, T. P. Hong, L. S. L. Wang, and J. Zhan, "Mining high-utility itemsets based on particle swarm optimization," *Engineering Applications of Artificial Intelligence*, vol. 55, pp. 320–330, 2016.
- [31] J. C.-W. Lin, L. Yang, P. Fournier-Viger, J. Frnda, L. Sevcik, and M. Voznak, "An Evolutionary Algorithm to Mine High-Utility Itemsets," *Advances in Electrical and Electronic Engineering*, vol. 13, no. 4, pp. 392–398, 2015.
- [32] J. C. W. Lin, L. Yang, P. Fournier-Viger, T. P. Hong, and M. Voznak, "A binary PSO approach to mine high-utility itemsets," *Soft Computing*, vol. 21, no. 17, pp. 5103–5121, 2017.
- [33] R. Gunawan, E. Winarko, and R. Pulungan, "A BPSO-based method for high-utility itemset mining without minimum utility threshold," *Knowledge-Based Systems*, vol. 190, 2020. [Online]. Available: <https://doi.org/10.1016/j.knsys.2019.105164>
- [34] J. M. T. Wu, J. Zhan, and J. C. W. Lin, "An ACO-based approach to mine high-utility itemsets," *Knowledge-Based Systems*, vol. 116, pp. 102–113, 2017. [Online]. Available: <http://dx.doi.org/10.1016/j.knsys.2016.10.027>
- [35] W. Song and C. Huang, "Discovering High Utility Itemsets Based on the Artificial Bee Colony Algorithm," in *Advances in Knowledge Discovery and Data Mining. PAKDD 2018. Lecture Notes in Computer Science*, D. Phung, V. Tseng, G. Webb, H. B. G. M., and L. Rashidi, Eds., vol. 10939, no. June. Cham: Springer, Cham, 2018. [Online]. Available: <http://link.springer.com/10.1007/97861>
- [36] —, "Mining High Utility Itemsets Using Bio-Inspired Algorithms: A Diverse Optimal Value Framework," *IEEE Access*, vol. 6, no. April, pp. 19 568–19 582, 2018.
- [37] B. Zhang, J. C. W. Lin, P. Fournier-Viger, and T. Li, "Mining of high utility-probability sequential patterns from uncertain databases," *PLoS ONE*, vol. 12, no. 7, pp. 1–21, 2017.
- [38] J. Yin, Z. Zheng, and L. Cao, "USpan: An efficient algorithm for mining high utility sequential patterns," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '12*. New York, New York, USA: ACM Press, 2012, p. 660. [Online]. Available: <http://dl.acm.org/citation.cfm?doid=2339530.2339636>
- [39] J.-Z. Wang, J.-L. Huang, and Y.-C. Chen, "On efficiently mining high utility sequential patterns," *Knowledge and Information Systems*, vol. 49, no. 2, pp. 597–627, nov 2016. [Online]. Available: <http://link.springer.com/10.1007/s10115-015-0914-8>
- [40] O. K. Alkan and P. Karagoz, "CRoM and HuspExt: Improving Efficiency of High Utility Sequential Pattern Extraction," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 10, pp. 2645–2657, oct 2015. [Online]. Available: <http://ieeexplore.ieee.org/document/7080884/>
- [41] J. C. W. Lin, Y. Li, P. Fournier-Viger, Y. Djenouri, and J. Zhang, "Efficient Chain Structure for High-Utility Sequential Pattern Mining," *IEEE Access*, vol. 8, pp. 40 714–40 722, 2020.
- [42] P. Fournier-Viger, C. Lin, A. Gomariz, T. Gueniche, A. Soltani, Z. Deng, and H. T. Lam, "The SPMF Open-Source Data Mining Library Version 2," in *Proc. 19th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD 2016) Part III*. Springer LNCS 9853, 2016, pp. 36–44. [Online]. Available: <http://www.philippe-fournier-viger.com/spmf/index.php>

# Online Retail Pattern Quality Improvement From Frequent Sequential Pattern to High-Utility Sequential Pattern

## ORIGINALITY REPORT

18%  
SIMILARITY INDEX

14%  
INTERNET SOURCES

12%  
PUBLICATIONS

4%  
STUDENT PAPERS

## PRIMARY SOURCES

1 Submitted to Universitas Sanata Dharma 2%  
Student Paper

2 [www-staff.it.uts.edu.au](http://www-staff.it.uts.edu.au) 2%  
Internet Source

3 [download.atlantis-press.com](http://download.atlantis-press.com) 1%  
Internet Source

4 Yin, Junfu, Zhigang Zheng, and Longbing Cao. "USpan : an efficient algorithm for mining high utility sequential patterns", Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 12 KDD 12, 2012. 1%  
Publication

5 Jun-Zhe Wang, Jiun-Long Huang, Yi-Cheng Chen. "On efficiently mining high utility sequential patterns", Knowledge and Information Systems, 2016 1%  
Publication

6 [deepai.org](http://deepai.org) 1%  
Internet Source



7	Kybernetes, Volume 41, Issue 9 (2012-10-27) Publication	1 %
8	Tai Dinh, Van-Nam Huynh, Bac Le. "Chapter 51 Mining Periodic High Utility Sequential Patterns", Springer Nature, 2017 Publication	1 %
9	<a href="http://www.philippe-fournier-viger.com">www.philippe-fournier-viger.com</a> Internet Source	1 %
10	Submitted to University of Sheffield Student Paper	1 %
11	<a href="http://link.springer.com">link.springer.com</a> Internet Source	1 %
12	Bahareh Rahmati, Mohammad Karim Sohrabi. "A Systematic Survey on High Utility Itemset Mining", International Journal of Information Technology & Decision Making, 2019 Publication	<1 %
13	Submitted to Texas A & M University, Kingville Student Paper	<1 %
14	<a href="http://coek.info">coek.info</a> Internet Source	<1 %
15	<a href="http://vn.booksc.eu">vn.booksc.eu</a> Internet Source	<1 %
16	Submitted to University of Hull Student Paper	<1 %

17	<a href="http://journalofbigdata.springeropen.com">journalofbigdata.springeropen.com</a> Internet Source	<1 %
18	<a href="http://bmcsystbiol.biomedcentral.com">bmcsystbiol.biomedcentral.com</a> Internet Source	<1 %
19	<a href="http://chbrown.github.io">chbrown.github.io</a> Internet Source	<1 %
20	<a href="http://docplayer.net">docplayer.net</a> Internet Source	<1 %
21	<a href="http://etd.lib.nsysu.edu.tw">etd.lib.nsysu.edu.tw</a> Internet Source	<1 %
22	<a href="http://export.arxiv.org">export.arxiv.org</a> Internet Source	<1 %
23	Chia-Ming Wang, Shyh-Huei Chen, Yin-Fu Huang. "A fuzzy approach for mining high utility quantitative itemsets", 2009 IEEE International Conference on Fuzzy Systems, 2009 Publication	<1 %
24	Guo-Cheng Lan, Tzung-Pei Hong, Hong-Yu Lee. "An efficient approach for finding weighted sequential patterns from sequence databases", Applied Intelligence, 2014 Publication	<1 %
25	Y.-L. Chen, T.C.-K. Huang. "Discovering Fuzzy Time-Interval Sequential Patterns in Sequence Databases", IEEE Transactions on Systems,	<1 %

# Man and Cybernetics, Part B (Cybernetics), 2005

Publication

26

[cs.nju.edu.cn](http://cs.nju.edu.cn)

Internet Source

<1 %

27

[dokumen.pub](http://dokumen.pub)

Internet Source

<1 %

28

[dspace.vutbr.cz](http://dspace.vutbr.cz)

Internet Source

<1 %

29

[propertibazar.com](http://propertibazar.com)

Internet Source

<1 %

30

[shahrooz-abghari.github.io](http://shahrooz-abghari.github.io)

Internet Source

<1 %

31

[tel.archives-ouvertes.fr](http://tel.archives-ouvertes.fr)

Internet Source

<1 %

32

[www.ijert.org](http://www.ijert.org)

Internet Source

<1 %

33

[www.ijicic.org](http://www.ijicic.org)

Internet Source

<1 %

34

Anita Bai, Parag S. Deshpande, Meera Dhabu.  
"Selective Database Projections Based  
Approach for Mining High-Utility Itemsets",  
IEEE Access, 2018

Publication

<1 %

35

Jian Pei, Jiawei Han, B. Mortazavi-Asl, Jianyong Wang, H. Pinto, Qiming Chen, U. Dayal, Mei-Chun Hsu. "Mining sequential patterns by pattern-growth: the PrefixSpan approach", IEEE Transactions on Knowledge and Data Engineering, 2004

Publication

&lt;1 %

36

Wang, Jun-Zhe, Zong-Hua Yang, and Jiun-Long Huang. "An Efficient Algorithm for High Utility Sequential Pattern Mining", Lecture Notes in Electrical Engineering, 2014.

Publication

&lt;1 %

37

Lecture Notes in Electrical Engineering, 2014.

Publication

&lt;1 %

Exclude quotes On

Exclude matches &lt; 5 words

Exclude bibliography On