Advances in Intelligent Systems Research, volume 172
Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019)

Online Retail Marketing Recommendation System Based on Generalized Sequential Pattern Algorithm and FP-Growth Algorithm

Destrilia[1], Rifkie PRIMARTHA[1*], Sukemi[1], and Adi WIJAYA[2]

[1]*Faculty of Computer Science, Sriwijaya University, Indonesia*
[2]*Informatics Engineering Departement, Universitas MH Thamrin, Jakarta, Indonesia*
**Corresponding Author : rifkie77@gmail.com*

ABSTRACT
Data mining association is a technique to find the relationship between items where the function can help sellers in determining their sales strategy. The algorithm used in this data mining techniques are Generalized Sequential Pattern Algorithm and FP-Growth Algorithm. Generalized Sequential Pattern Algorithm is an algorithm based on sequential patterns in the formation of rules, while FP-Growth Algorithm is a tree-based algorithm in the formation of rules. This research produces a comparison of the computation time of each algorithms in carrying data mining process associated with the data that has been determined. The result of computational time comparisons show that FP-Growth Algorithm is 11.97% faster than Generalized Sequential Pattern Algorithm based on 30 tests. Generalized Sequential Pattern Algorithm produces 2 rules and FP-Growth Algorithm produces 8 rules by testing 500 transaction data and minimum support value is 3. Where the rules obtained is evaluated using the lift ratio techniques to calculate the value of the rule accuracy generated from each algorithms.
Keywords: *Generalized Sequential Pattern Algorithm, FP-Growth Algorithm, lift ratio, rule*

Introduction

In the competition of online retail business world, sales strategy is the main goal to achieve success. To get a good sales strategy the concept of data mining can be applied by using previous transaction data. In this case, we use the association method to find associative rules between a combination of items in a transaction data. The association also checks all possible *if-then* relationship between items and selects only those that are most likely to be indicators of the relationship between items.

Apriori Algorithm is a basic algorithm proposed by Agrawal and Srikan in 1994 to find frequent item sets in Boolean association rules. Process of Apriori Algorithm divided into two, there are *join step* and *prune step*. In this study using a comparison of Generalized Sequential Pattern Algorithm and FP-Growth Algorithm.

Generalized Sequential Pattern Algorithm processes and discovers all existing sequential and non sequential patterns to form the association rule and sequential pattern rules of the frequent sequence patterns that have been found. FP-Growth Algorithm can be used to determine the set of data that most often appears in a set of data. The data structure used in this algorithm is in the form of a tree commonly referred to as *fp-tree*, where with the existence of this fp-tree, FP-Growth Algorithm can directly extract frequent item set from fp-tree.

Related Works

Author use another references in this research, there are research by Supardi, Dian Eka Ratnawati and Wayan Firdaus Mahmudy (2017) regarding the introduction of book circulation transaction patterns in library databases using Generalized Sequential Pattern Algorithm. This study calculates the value of support so that a candidate 1-itemset (C1) is found, also found a large 1-itemset (L1) with repeated iteration until it reaches a large 3-itemset (L3), every iteration there is a join and prune process. The results of this study created a pattern of borrowing books based on book categories, there are Criminal Law and Civil Law books which are often borrowed in the same time.

Research by Rama Novta Miraldi, Antonius Rachma and Budi Susanto (2017) on the implementation of FP-Growth Algorithm for the book recommendation system in UKDW Library. The result of this study found the highest support value is a rule that has a number of occurrences in the data 20% by 5, while the rule with the lowest occurrence value is 1. The test produces rules that meet 20% of transaction data by 31 rules, so that the accuracy of the FP-Growth Algorithm calculation is obtained by 60.78%.

There for, its important for us to research on the comparison of Generalized Sequential Pattern Algorithm and FP-Growth Algorithm to finding the right sales

strategy in the case of online retailing with computational time comparison parameters of each algorithm, so it can be concluded which algorithm is more efficient to calculate and each algorithm will produce rules from testing predetermined transaction data.

MATERIAL AND PROPOSED METHOD

The dataset used in this paper was obtained from donation in UCI public repository. That data are transaction data of online retail that selling unique gift for all event. Author using 500 data transaction in November. The online retail dataset has 5 attributes there are InvoiceNo, StockCode, Description, Quantity, CustomerID. In this paper, the author using minimum support value is 3 and the algorithms produce the rules than to make the rules are accuracy is using lift ratio method.

EXPIRIMENTAL RESULT

This paper was conducted using computer calculation with the details of Intel ® Core ™ i3, CPU @ 2.30GHz, RAM 4 GB, Hard disk 500 GB, Windows 8.1 Pro 32-bit and Java with IDE Neatbeans 8.2 for make the program execution the algorithms and testing the algorithms.

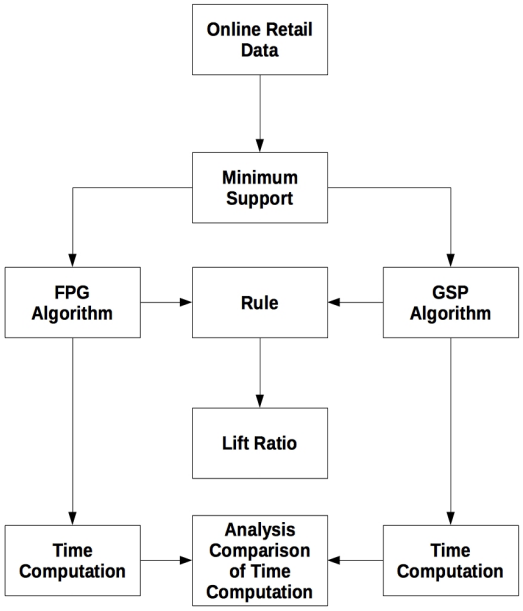

Figure 1. Testing Squence Diagram

The result from the computation of Generalized Sequential Pattern Algorithm with 500 transaction data and minimum support value is 3 produce 2 rules. And the result from the computational of FP-Growth Algorithm with same data and minimum support value is produce 8 rules. That rules show in this table below.

Table 1. The Rules from Generalized Sequential Patern Algorithm Table

Data	Minimum Support	Antecedent	Consequent	Support	Confidence	Expected Confidence	Lift Ratio
Novem ber	3	BLUE HARMONIC A IN BOX	RED HARMONICA IN BOX	0.0724	0.8334	0.0869	9.5834
		HAND WARMER RED LOVE HEART	ZINC T-LIGHT STARS SMALL	0.0869	0.8571	0.0869	9.8571

Table 2. The Rules from FP-Growth Algorithm Table

Data	Minimum Support	Antecedent	Consequent	Support	Confidence	Expected Confidence	Lift Ratio
November	3	PLASTERS IN TIN SKULLS	PLASTERS IN TIN CIRCUS PARADE	0.0434	1	0.0579	17.25
		PAPER CHAIN KIT 50'S CHRISTMAS	RABBIT NIGHT LIGHT	0.0434	0.6	0.0724	8.28
		JUMBO BAG 50'S CHRISTMAS	JUMBO BAG VINTAGE DOILY	0.0434	0.75	0.0869	8.625
		WOODEN STAR CHRISTMAS SCANDINAVIAN	WOODEN TREE CHRISTMAS SCANDINAVIAN	0.0434	1	0.0434	23
		MOODY BOY DOOR HANGER	MOODY GIRL DOOR HANGER	0.0724	1	0.0724	13.799
		VINTAGE DOILY JUMBO BAG RED	JUMBO BAG PAISLEY PARK	0.0434	1	0.0579	17.25
		6 GIFT TAGS VINTAGE CHRISTMAS	SCOTTIE DOG HOT WATER BOTTLE	0.0434	0.75	0.0579	12.9375
		PAPER CHAIN KIT VINTAGE CHRISTMAS	PAPER CHAIN KIT 50'S CHRISTMAS	0.0434	0.75	0.0724	10.35

For make the stabilized comparison, so the algorithms are testing in 30 testing which is time computing from each algorithm is different. The time computing result is show in that table below.

Table 3. Comparison of Time Computation Table

Data	Minimum Support	Computation Time		Efficient Algorithm
		GSP (ms)	FPG (ms)	
November	3	156	141	FPG
November	3	94	94	-
November	3	94	94	-
November	3	110	93	FPG
November	3	63	47	FPG
November	3	109	93	FPG
November	3	110	79	FPG
November	3	94	78	FPG
November	3	94	94	-
November	3	110	78	FPG
November	3	79	93	GSP
November	3	94	78	FPG
November	3	109	93	FPG
November	3	94	93	FPG
November	3	94	93	FPG
November	3	79	78	FPG
November	3	94	78	FPG
November	3	94	93	FPG
November	3	94	78	FPG
November	3	93	78	FPG
November	3	93	78	FPG
November	3	110	78	FPG
November	3	109	78	FPG
November	3	78	94	GSP
November	3	93	93	-
November	3	109	79	FPG
November	3	93	79	FPG
November	3	109	78	FPG
November	3	78	93	GSP
November	3	94	78	FPG

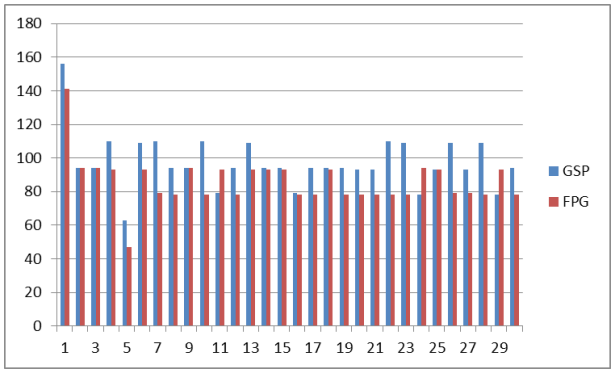

Figure 2. Comparison time of GSP Algorithm and FPG Algorithm

Based on Figure 2, for 30 times average testing time computation of Generalized Sequential Pattern Algorithm is 97.467ms while average testing time computation of FP-Growth Algorithm is 86.8ms. based on the result of average comparison time, we can calculate like this:

$$\frac{\bar{X}ET1 - \bar{X}ET2}{\bar{X}ET1} \times 100\% \quad (1)$$

$$\frac{97.467 - 86.8}{97.467} \times 100 = 11.97\%$$

With *ET1* means GSP Algorithm and *ET2* means FPG Algorithm. The result is show that FP-Growth Algorithm

computation time is 11.97% more faster than Generalized Sequential Pattern Algoritm computation time.
Based on the result of the Algorithms computation with data transaction are 500 data. These are the rules of Generalized Sequential Pattern Algorithm:
IF Blue Harmonica in Box THEN Red Harmonica in Box
IF Hand Warmer Red Love Heart THEN Zinc T-Light Stars Small

While these are the rules of FP-Growth Algorithms:
IF Plasters in Tin Skulls THEN Plasters in Tin Circus Parade
IF Paper Chain Kit 50'S Christmas THEN Rabbit Night Light
IF Jumbo Bag 50'S Christmas THEN Jumbo Bag Vintage Doily
IF Wooden Star Christmas Scandinavian THEN Wooden Tree Christmas Scandinavian
IF Moody Boy Door Hanger THEN Moody Girl Door Hanger
IF Vintage Doily Jumbo Bag Red THEN Jumbo Bag Paisley Park
IF 6 Gift Tags Vintage Christmas THEN Scottie Dog Hot Water Bottle
IF Paper Chain Kit Vintage Christmas THEN Paper Chain Kit 50'S Christmas

CONCLUSION AND FUTURE WORKS

The computation with 500 data online retail transaction and with minimum support value is 3 and for 30 times average testing time computation of Generalized Sequential Pattern Algorithm is 97.467ms and produce 2 rules while average testing time computation of FP-Growth Algorithm is 86.8ms and produce 8 rules. Based on the result of average comparison time, the result is show that FP-Growth Algorithm computation time is 11.97% more faster than Generalized Sequential Pattern Algoritm computation time.
Future work can be focused on that the data is not only use an online retail data, focused to compare another algorithms in data mining association and make an optimization on the algorithms that author used or another algorithms.

REFERENCES

[1] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques.* San Francisco,

CA, Itd: Morgan Kaufmann, 745. https://doi.org/10.1016/B978-0-12-381479-1.00001-0

[2] Fadlina. 2014. Data Mining untuk Analisa Tingkat Kejahatan Jalanan dengan Algoritma

Association Rule Metode Apriori (Studi Kasus di Polsekta Medan Sunggal). *Informasi dan*

Teknologi Ilmiah (INTI), Vol: III No: 1, 144-154.

[3] Kusumo, D., Bijaksana, M., & Darmantoro, D. (2016). Data Mining dengan Algoritma Apriori pada RDBMS Oracle. Jurnal Penelitian Dan, 1-5. Retrieved from http://www.tektrika.org/index.php/tektrika/article/download/10/2

[4] Kusrini, Luthfi, E. T. (2009). *Algoritma Data Mining*. Andi Yogyakarta.

[5] Krutchen, P. (2000). The Rational Unified Process, An Introduction, 2nd edition.

[6] Tan, Pang-Ning., Steinbach, Michael., & Kumar, Vipin. (2004). *Introduction of Data Mining*.

[7] Zaki, Mohammed J. (1997). Fast Mining of Sequential Patterns in Very Large Databases. *The University of Rochester Computer Science Department Rochester* , New York 14627.

[8] North Matthew (2012). *Data Mining For The Masses.pdf*

[9] Han, Jiawei "*Data Mining Concept and Technique*", Presentation.

[10] Larose, Daniel T, 2005, *Discovering Knowledge in Data: An Introduction to Data Mining*, ohn Willey & Sons. Inc