Reprinted From

ISSN: 0973-4562, E-ISSN: 0973-9769

International Journal of Applied Engineering Research

Volume 11, Number 8 (2016) pp 5369-5375
www.ripublication.com

EASY UNDERSTANDING OF ATTRIBUTE ORIENTED INDUCTION (AOI) CHARACTERISTIC RULE ALGORITHM

[1]Harco Leslie Hendric Spits Warnars, [2]Muhamad Iskandar Wijaya, [3]Hok Bun Tjung, [4]Dendy Fransiskus Xaverius, [5]Dedy Van Hauten and [6]Sasmoko

[1] Doctor of Computer Science, Bina Nusantara University, Jakarta, Indonesia, 11480
[2]PT. Central Santosa Finance, Jakarta, Indonesia, 11480
[3]Master of Information Technology, Bina Nusantara University, Jakarta, Indonesia 11480
[4]Yayasan Pendidikan St. Leo, Jakarta, Indonesia, 11170.
[1,5]Database, Data-warehouse & Data Mining Research Center, Surya University, Tangerang, Indonesia, 15810.
[6]Research Interest Group in Education Technology, Bina Nusantara University, Jakarta, Indonesia, 11480.

URL: *http://www.ripublication.com/Volume/ijaerv11n8.htm*

Document details

International Journal of Applied Engineering Research
Volume 11, Issue 8, 2016, Pages 5369-5375

Easy understanding of Attribute Oriented Induction (AOI) characteristic rule algorithm (Article)

Warnars, H.L.H.S.[a,d], Wijaya, M.I.[b], Tjung, H.B.[a], Xaverius, D.F.[c], Van Hauten, D.[d], Sasmoko[e]

[a]Bina Nusantara University, Jakarta, Indonesia
[b]PT. Central Santosa Finance, Jakarta, Indonesia
[c]Yayasan Pendidikan St. Leo, Jakarta, Indonesia

View additional affiliations

Abstract

View references (10)

This paper is continuation from previous research, where selective generalize attributes is executed in order to find final characteristic rule with execution of Attribute Oriented Induction (AOI) characteristic rule algorithm between line number 9 and 12. This paper gives easy understanding for bachelor and master degree students to understand about AOI characteristic rule algorithm which only implement AOI characteristic rule algorithm between line number 1 and 8. Meanwhile, using threshold in AOI has been known as tedious and consuming work, however in our future research this limitation will be resolved where maximum number of threshold will depend on maximum number of distinct values in dataset. © Research India Publications.

Author keywords

AOI | Attribute Oriented Induction | Characteristic rule | Concept Hierarchy | Data Mining

ISSN: 09734562
Source Type: Journal
Original language: English
Document Type: Article
Publisher: Research India Publications

References (10)

View in search results format

All Save to Mendeley Print E-mail Save to PDF Create bibliography

1 Han, J., Cai, Y., Cercone, N.
Knowledge discovery in databases: An attribute-oriented approach
(1992) *Proceeding of 18Th International Conference Very Large Data Bases*, pp. 547-559. Cited 204 times.
Vancouver, British Columbia

2 Warnars, S.
Mining Patterns with Attribute Oriented Induction
(2015) *Proceeding of International Conference on Database, Data Warehouse, Data Mining and Big Data (DDDMBD)*, pp. 11-21. Cited 3 times.
Tangerang, Indonesia

3 Warnars, S.
Measuring Interesting rules in characteristic rule
(2010) *Proceeding of the 2Nd International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT)*, pp. 152-156. Cited 3 times.
Bali, Indonesia

4 Wu, Y., Chen, Y., Chang, R.
Generalized knowledge discovery from relational databases
(2009) *International Journal of Computer Science and Network*, 9 (6), pp. 148-153.

5 Han, J., Cai, Y., Cercone, N.
Data-Driven Discovery of Quantitative Rules in Relational Databases
(1993) *IEEE Transactions on Knowledge and Data Engineering*, 5 (1), pp. 29-40. Cited 219 times.
doi: 10.1109/69.204089

View at Publisher

6 Han, J., Fu, Y.
Exploration of the power of attributeoriented induction in data mining
(1995) *Advances in Knowledge Discovery and Data Mining*, pp. 399-421. Cited 99 times.
U. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, eds

7 Cai, Y.
(1989) *Attribute-Oriented Induction in Relational Databases*. Cited 2 times.
Master thesis, Simon Fraser University

8 Cheung, D.W., Hwang, H.Y., Fu, A.W., Han, J.
Efficient rule-based attribute-oriented induction for data mining
(2000) *Journal of Intelligent Information Systems*, 15 (2), art. no. 267357, pp. 175-200. Cited 17 times.
doi: 10.1023/A:1008778107391

View at Publisher

9 Warnars, S.
Mining frequent pattern with Attribute Oriented Induction High Level Emerging Pattern (AOI-HEP)
(2014) *2014 2nd International Conference on Information and Communication Technology, ICoICT 2014*, art. no. 6914056, pp. 149-154. Cited 7 times.
ISBN: 978-147993581-9
doi: 10.1109/ICoICT.2014.6914056

View at Publisher

10 Warnars, S.
Mining Frequent and Similar Patterns with Attribute Oriented Induction High Level Emerging Pattern (AOI-HEP) Data Mining Technique
(2014) *International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)*, 3 (11), pp. 266-276. Cited 9 times.

Related documents

Mining similar pattern with attribute oriented induction high level emerging pattern (AOI-HEP) data mining technique
Warnars, H.L.H.S. , Anwar, N. , Randriatoamanana, R.
(2017) Jurnal Teknologi

Mining job logs using incremental attribute-oriented approach
Adewale, I.O. , Alhajj, R.
(2005) Lecture Notes in Computer Science

Attribute Oriented Induction High Level Emerging Pattern (AOI-HEP) future research
Warnars, S.
(2014) Proceedings of 2014 International Conference on Information, Communication Technology and System, ICTS 2014

View all related documents based on references

Find more related documents in Scopus based on:

Authors > Keywords >

Research India Publications

International Journal of Applied Engineering Research (IJAER)

Editorial Board Members

[UP]

Home Journals Contact

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 8 (2016) pp 5369-5375

Easy Understanding of Attribute Oriented Induction (AOI) Characteristic Rule Algorithm

Harco Leslie Hendric Spits Warnars[1], Muhamad Iskandar Wijaya[2], Hok Bun Tjung[3], Dendy Fransiskus Xaverius[4], Dedy Van Hauten[5], and Sasmoko[6]

[1]Doctor of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia.
[2]PT. Central Santosa Finance, Jakarta 11480, Indonesia.
[3]Master of Information Technology, Bina Nusantara University, Jakarta 11480, Indonesia.
[4]Yayasan Pendidikan St. Leo, Jakarta 11170, Indonesia.
[15]Database, Data-warehouse & Data Mining Research Center, Surya University, Tangerang 15810, Indonesia.
[6]Research Interest Group in Education Technology, Bina Nusantara University, Jakarta 11480, Indonesia.

Abstract
This paper is continuation from previous research, where selective generalize attributes is executed in order to find final characteristic rule with execution of Attribute Oriented Induction (AOI) characteristic rule algorithm between line number 9 and 12. This paper gives easy understanding for bachelor and master degree students to understand about AOI characteristic rule algorithm which only implement AOI characteristic rule algorithm between line number 1 and 8. Meanwhile, using threshold in AOI has been known as tedious and consuming work, however in our future research this limitation will be resolved where maximum number of threshold will depend on maximum number of distinct values in dataset.

Keywords: Data Mining, Attribute Oriented Induction, AOI, Characteristic rule, Concept Hierarchy.

Introduction

This paper will give easy understanding for those who want to learn and understand data mining, particularly to understand Attribute Oriented Induction (AOI) data mining technique with intention to AOI characteristic rule algorithm. Data mining has been interested research topics in any kind of science disciplines such as economy, education, biology, social, medicine, banking and so on. Data mining has been recognizing as powerful weapon for dealing with huge data and complicated attributes. Data mining has been known for finding interesting patterns which can be used for description or prediction.

For example, frequent pattern has been using as strong discriminator which discriminates between frequent and non frequent patterns[9,10]. Market basket analysis in the form of association rules has been known as frequent pattern which synonym with large pattern. Meanwhile, similar pattern is interesting to be discovery as well since similar pattern show the similarity which can be used for prediction purposes[10].
This paper gives research idea for bachelor and master degree students particularly in Computer Science or Information Technology degrees in order to have data mining topic for study case idea in their bachelor or master thesis. Using AOI data mining technique particularly AOI characteristic rule algorithm with many different data resources or study case will help students to understand basic knowledge to find patterns in data mining.
At the end, people can easily to understand and implement data mining which can make people realize that data mining can help human activities in order to find patterns in huge data, in easy, fast and intelligent way. People can realize that data mining is easy to implemented, just only pieces of algorithm which implemented with any kind of language software on text dataset or database, help human with artificial intelligence activity such as think like a human.

AOI Characteristic Rule Understanding

AOI was developed to learning different kinds of knowledge rules such as characteristic rule, discrimination rules, classification rules, data evolution regularities[5], association rules and cluster description rules[6]. Characteristic rule in AOI is an assertion which characterizes the concepts which satisfied by all of the data stored in database. This rule provides generalized concepts about a property which can help people recognize the common features of the data in a class, for example the symptom of the specific disease [7].
Characteristic rule has been used by AOI in order to recognize, learning and mining as a specific character for each of attribute as their specific mining characterization. Characteristic rule process the generalization with help of concept hierarchy as the standard saving background knowledge to find target class as a positive learning[2]. AOI characteristic rule algorithm has 7 strategy steps in process of generalization[1] and they are:

1) Generalization on the smallest decomposable components where generalization should be performed on the smallest decomposable components on data relation.
2) Attribute removal, will be executed if there is distinct values for an attribute but there is no higher level concept for the attribute, then the attribute should be removed during generalization.

3) Concept tree Ascension, will be executed if there is higher level concept in the concept hierarchy for an attribute value of a tuple, then the substitution of the value by its higher level concept would generalize the tuples.
4) Vote propagation, will be executed based on the value of the vote is the value of counted tuples where the vote will be counted when merging identical tuples in the generalization.
5) Threshold control on each attribute, will be executed if the number of distinct values in a resulting relation is larger than the specified threshold value, then further generalization on this attribute should be performed.
6) Threshold control on generalized relations, will be executed if the number of tuples are larger than the specified threshold value, then further generalization will be done based on the selected attribute and the merging of the identical tuples should be performed.
7) Rule transformation, will be executed by changing final generalization to quantitative rule and qualitative rule from a tuple (conjunctive) or multiple tuples (disjunctive).

AOI characteristic rule algorithm [3] as shown below has 2 sub processes such as control number of distinct attributes and number of tuples[2]. Firstly, this AOI characteristic algorithm will be run by input number of threshold with range integer number >0 and this threshold as limitation for rules result which is shown in line number 2 and 9. Using threshold in AOI is recognized as disadvantage of AOI algorithm where small number of threshold will lead to simple rule with more ANY value which is recognized as uninteresting value. On other hand, large threshold will lead to complex rule with many tuples result. The number which is assigned to threshold will influence the number of distinct value in each attribute as shown in line number 2 and the number of tuples as shown in line number 9. Threshold for line number 2 and 9 can be made different such as attribute threshold and rule threshold.

```
1.  For each of attribute Ai (1≤i≤n, where
    n= # of attributes) in the generalized
    relation GR
2. { While
      #_of_distinct_values_in_attribute_Ai >
      threshold
3.    {If no higher level concept in concept
         hierarchy for attribute_Ai
4.     Then remove attribute Ai
5.     Else substitute the value of Ai by
            its corresponding minimal
            generalized concept
6.     Merge identical tuples
7.    }
8.  }
9.  While #_of_tuples in GR > threshold
10. {Selective generalize attributes
11.  Merge identical tuples
12. }
```

The first line of AOI characteristic rule algorithm above shows looping as many as number of attribute/column in dataset. The looping process between line number 2 and 8 show the process to check number of distinct values for each attribute in dataset. If the number of distinct value in attribute <= threshold which is inputted as attribute threshold, then the process will not execute process between line number 3 and 7 and will loop to the 1st line and assigned to the next attribute in dataset.

Meanwhile, the process in line number 3 shows if each attribute has higher level concept hierarchy file and if there is no higher level in concept hierarchy file then line number 4 will be executed where will remove the attribute. On other hand, if the attribute has higher level in concept hierarchy file then line number 5 will be executed where will substitute all value of the attribute with corresponding high level concept in concept hierarcy file. Afterward, the line number 6 will be executed in order to merge identical tuples and if there are some identical tuples then the number of tuples will be reduced.

Meanwhile, line number between 9 and 12 had been discussed in previous research[3] where line number 9 will check if number of tuples > rules threshold and line numbers 10 and 11 will be executed if number of tuples > rules threshold. Line number 10 shows the specific algorithm which can be used to reduce the number tuples until <= rules threshold, by selecting generalize attribute. In previous research, line number 10 where the selective generalize attributes was proposed based on the simplicity of concept hierarchy, where the simpler concept hierarchy more interesting result are likely to be found[3]. Finding interesting rule is influenced with wider of length and depth or level of concept tree as shown in figure 1[3]. Figure 1 shows that small number of depth/level and length/amount of concepts in concept tree will have simple generalization process but large number will have more generalization process[3].

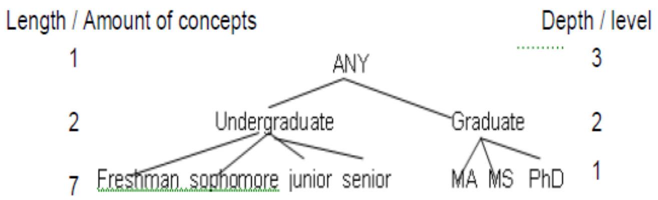

Figure 1: Depth and length of Concept tree

Running AOI Characteristic Rule Algorithm

In this paper, AOI Characteristic rule algorithm will be run with 12 tuples of student dataset as shown in table 1 refer to previous research[1] and will saved as text file. Using 12 tuples in this paper will show the easy understanding of this AOI characteristic rule algorithm rather than using huge tuples. In this paper, the AOI Characteristic rule algorithm which is shown above will only applied between line number 1 and 8 while line number between 9 and 12 had been discussed on previous research [3]. The AOI characteristic rule algorithm will be run with threshold=2 and the first line of AOI characteristic rule algorithm will be executed with n=5 since the number of attribute in table 1 is 5. The next explanation will mention for each of attribute in table 1, start from 1st attribute to 5th attribute since there are 5 attributes in table 1.

Process on first attribute or attribute name:
The second line of AOI characteristic rule algorithm is executed with i=1, where attribute name is the first attribute in table 1 and the second line has meaning to check number of distinct value in attribute. In attribute name there are 12 distinct values and they are: {Anderson, Bach, Carey, Fraser, Gupta, Hart, Jackson, Liu, Meyer, Monk, Wang, Wise}. Because distinct value in attribute name=12 and greater than 2 as threshold. Next, the third line of AOI characteristic rule algorithm will be run in order to check if attribute name name has higher concept in concept hierarchy.

Table 1: Example of 12 tuples of student dataset

Name	Category	Major	Birthplace	GPA
Anderson	M.A.	History	Vancouver	3.5
Bach	Junior	Math	Calgary	3.7
Carey	Junior	Liberal arts	Edmonton	2.6
Fraser	M.S.	Physics	Ottawa	3.9
Gupta	Ph.D.	Math	Bombay	3.3
Hart	Sophomore	Chemistry	Richmond	2.7
Jackson	Senior	Computing	Victoria	3.5
Liu	Ph.D.	Biology	Shanghai	3.4
Meyer	Sophomore	Music	Burnaby	3.0
Monk	Ph.D.	Computing	Victoria	3.8
Wang	M.S.	Statistics	Nanjing	3.2
Wise	Freshman	Literature	Toronto	3.9

Table 2: Removing attribute name

Name	Category	Major	Birthplace	GPA
	M.A.	History	Vancouver	3.5
	Junior	Math	Calgary	3.7
	Junior	Liberal arts	Edmonton	2.6
	M.S.	Physics	Ottawa	3.9
	Ph.D.	Math	Bombay	3.3
	Sophomore	Chemistry	Richmond	2.7
	Senior	Computing	Victoria	3.5
	Ph.D.	Biology	Shanghai	3.4
	Sophomore	Music	Burnaby	3.0
	Ph.D.	Computing	Victoria	3.8
	M.S.	Statistics	Nanjing	3.2
	Freshman	literature	Toronto	3.9

Because the first attribute or attribute name does not have higher concept then this attribute will be removed as implementation line number 4 of AOI characteristic rule algorithm. The current student dataset should be as shown in table 2 and afterward, line number 6 of AOI characteristic rule will be executed in order to merge identical tuples. However, there is no identical tuples in current student dataset in table 2. Thereafter, the process will loop and executed line number 2 of AOI characteristic rule algorithm with still i=1, where attribute name has 0 distinct value because attribute name values had been removed. Because line number 2 has return false value for 0>2 , where 0 for distinct value in attribute name and 2 for threshold, then the process will loop to line number 1 of AOI characteristic rule algorithm with next i=2 and attribute category is the second attribute in student dataset as shown in table 2.

Process on second attribute or attribute category:
Later, the second line of AOI characteristic rule algorithm is executed with i=2 to check distinct value in second attribute or attribute category and table 2 shows there are 7 distinct values such as {M.A., Junior, M.S., Ph.D, Sophomore, Senior, Freshman}. Because distinct value in attribute category=7 and greater than 2 as threshold, then the third line of AOI characteristic rule algorithm will be run in order to check if attribute category has higher concept in concept hierarchy. Because the second attribute or attribute category has higher concept, then line number 5 of AOI characteristic rule algorithm is run in order to substitute the value of attribute category with corresponding minimal generalized concept. The minimal corresponding generalized concept for attribute category can be seen on concept tree for category in figure 2 on previous research[3] and shown in figure 2.

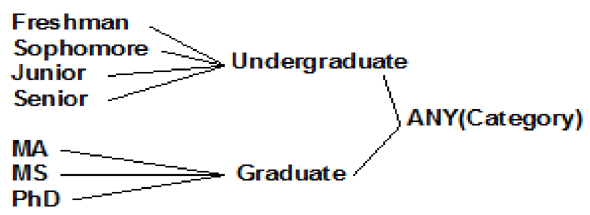

Figure 2: Concept tree for category

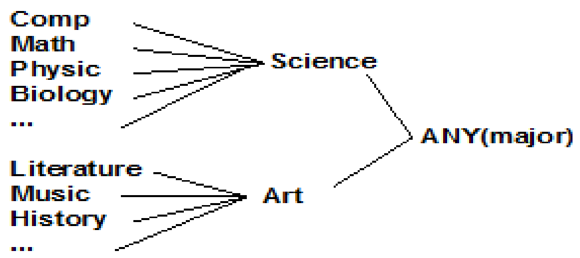

Figure 3: Concept tree for major

Table 3 shows student dataset after attribute category first generalized concept and afterward, line number 6 of AOI characteristic rule will be executed in order to merge identical tuples. However, there is no identical tuples in current student dataset in table 3.

Table 3: First generalization on attribute category

Name	Category	Major	Birthplace	GPA
	Graduate	History	Vancouver	3.5
	Undergraduate	Math	Calgary	3.7
	Undergraduate	Liberal arts	Edmonton	2.6
	Graduate	Physics	Ottawa	3.9
	Graduate	Math	Bombay	3.3
	Undergraduate	Chemistry	Richmond	2.7
	Undergraduate	Computing	Victoria	3.5
	Graduate	Biology	Shanghai	3.4
	Undergraduate	Music	Burnaby	3.0
	Graduate	Computing	Victoria	3.8
	Graduate	Statistics	Nanjing	3.2
	Undergraduate	literature	Toronto	3.9

Thereafter, the process will loop and executed line number 2 of AOI characteristic rule algorithm with still i=2 which means to check number of distinct value in 2nd attribute in table 3, where the 2nd attribute or attribute category in table 3 has 2 distinct values such as {Graduate, Undergraduate}. Because line number 2 has return false value for 2>2 , where 2 for distinct value in attribute category and 2 for threshold, then the process will loop to line number 1 of AOI characteristic rule algorithm with next i=3 and attribute major is the third attribute in student dataset as shown in table 3.

Process on third attribute or attribute major :
Next, the second line of AOI characteristic rule algorithm is executed with i=3 to check distinct value in third attribute or attribute major and table 3 shows there are 10 distinct values such as {History, Math, Liberal arts, Physics, Chemistry, Computing, Biology, Music, Statistics, literature}. Because distinct value in attribute major=10 and greater than threshold=2, then the third line of AOI characteristic rule algorithm will be run in order to check if attribute major has higher concept in concept hierarchy. Because the third attribute or attribute major has higher concept, then line number 5 of AOI characteristic rule algorithm is run in order to substitute the value of attribute major with corresponding minimal generalized concept. The minimal corresponding generalized concept for attribute major can be seen on concept tree for major in figure 1 on previous research[3] and shown in figure 3.
Table 4 shows student dataset after attribute major first generalized concept and afterward, line number 6 of AOI characteristic rule will be executed in order to merge identical tuples. However, there is no identical tuples in current student dataset in table 4. Thereafter, the process will loop and executed line number 2 of AOI characteristic rule algorithm with still i=3, which means to check number of distinct value in 3rd attribute in table 4, where attribute major in table 4 has 2 distinct values such as {Science, Art}. Because line number 2 has return false value for 2>2 , where 2 for distinct value in attribute major and 2 for threshold, then the process will loop to line number 1 of AOI characteristic rule algorithm with next i=4 and attribute birthplace is the fourth attribute in student dataset as shown in table 4.

Table 4: First generalization on attribute major

Name	Category	Major	Birthplace	GPA
	Graduate	Art	Vancouver	3.5
	Undergraduate	Science	Calgary	3.7
	Undergraduate	Art	Edmonton	2.6
	Graduate	Science	Ottawa	3.9
	Graduate	Science	Bombay	3.3
	Undergraduate	Science	Richmond	2.7
	Undergraduate	Science	Victoria	3.5
	Graduate	Science	Shanghai	3.4
	Undergraduate	Art	Burnaby	3.0
	Graduate	Science	Victoria	3.8
	Graduate	Science	Nanjing	3.2
	Undergraduate	Art	Toronto	3.9

Process on fourth attribute or attribute birthplace :
Next, the second line of AOI characteristic rule algorithm is executed with i=4 to check number of distinct value in fourth attribute or attribute birthplace and table 4 shows there are 11 distinct values such as {Vancouver, Calgary, Edmonton, Ottawa, Bombay, Richmond, Victoria, Shanghai, Burnaby, Nanjing, Toronto}. Because distinct value in attribute birthplace=11 and greater than threshold=2, then the third line of AOI characteristic rule algorithm will be run in order to check if attribute birthplace has higher concept in concept hierarchy. Because the fourth attribute or attribute birthplace has higher concept, then line number 5 of AOI characteristic rule algorithm is run in order to substitute the value of attribute birthplace with corresponding minimal generalized concept. The minimal corresponding generalized concept for attribute birthplace can be seen on concept tree for birthplace in figure 3 on previous research[3] and shown in figure 4.

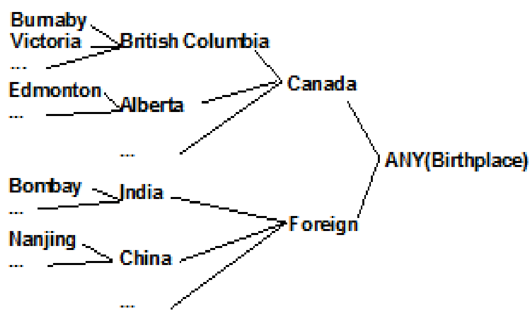

Figure 4: Concept tree for birthplace

Table 5 shows student dataset after attribute birthplace first generalized concept and afterward, line number 6 of AOI characteristic rule will be executed in order to merge identical tuples. However, there is no identical tuples in current student dataset in table 5. Thereafter, the process will loop and the second line of AOI characteristic rule algorithm is executed with still

Table 5: First generalization on attribute birthplace

Name	Category	Major	Birthplace	GPA
	Graduate	Art	British Columbia	3.5
	Undergraduate	Science	Alberta	3.7
	Undergraduate	Art	Alberta	2.6
	Graduate	Science	Ontario	3.9
	Graduate	Science	India	3.3
	Undergraduate	Science	British Columbia	2.7
	Undergraduate	Science	British Columbia	3.5
	Graduate	Science	China	3.4
	Undergraduate	Art	British Columbia	3.0
	Graduate	Science	British Columbia	3.8
	Graduate	Science	China	3.2
	Undergraduate	Art	Ontario	3.9

Table 6: Second generalization on attribute birthplace

Name	Category	Major	Birthplace	GPA	Num
	Graduate	Art	Canada	3.5	1
	Undergraduate	Science	Canada	3.7	1
	Undergraduate	Art	Canada	2.6	1
	Graduate	Science	Canada	3.9	1
	Graduate	Science	Foreign	3.3	1
	Undergraduate	Science	Canada	2.7	1
	Undergraduate	Science	Canada	3.5	1
	Graduate	Science	Foreign	3.4	1
	Undergraduate	Art	Canada	3.0	1
	Graduate	Science	Canada	3.8	1
	Graduate	Science	Foreign	3.2	1
	Undergraduate	Art	Canada	3.9	1

i=4 to check number of distinct value in fourth attribute or attribute birthplace and table 5 shows there are 5 distinct values such as {British Columbia, Alberta, Ontario, India, China}. Because distinct value in attribute birthplace=5 and greater than 2 as threshold, then the third line of AOI characteristic rule algorithm will be run in order to check if attribute birthplace has higher concept in concept hierarchy. Because the fourth attribute or attribute birthplace has higher concept, then line number 5 of AOI characteristic rule algorithm is run in order to substitute the value of attribute birthplace with corresponding minimal generalized concept. The minimal corresponding generalized concept for attribute birthplace can be seen on concept tree for birthplace in figure 4.

Table 6 shows student dataset after attribute birthplace second generalized concept and afterward, line number 6 of AOI characteristic rule will be executed in order to merge identical tuples. However, there is no identical tuples in current student dataset in table 6. Thereafter, the process will loop and executed line number 2 of AOI characteristic rule algorithm with still i=4 which means to check number of distinct value in 4th attribute in table 6, where attribute birthplace in table 6 has 2 distinct values such as {Canada, Foreign}. Because line number 2 has return false value for 2>2 , where 2 for distinct value in attribute birthplace and 2 for threshold, then the process will loop to line number 1 of AOI characteristic rule algorithm with next i=5 and attribute GPA is the fifth attribute in student dataset as shown in table 6.

Table 7: First generalization on attribute GPA

Name	Category	Major	Birthplace	GPA	Num
	Graduate	Art	Canada	Excellent	1
	Undergraduate	Science	Canada	Excellent	1
	Undergraduate	Art	Canada	Average	1
	Graduate	Science	Canada	Excellent	1
	Graduate	Science	Foreign	Good	1
	Undergraduate	Science	Canada	Average	1
	Undergraduate	Science	Canada	Excellent	1
	Graduate	Science	Foreign	Good	1
	Undergraduate	Art	Canada	Good	1
	Graduate	Science	Canada	Excellent	1
	Graduate	Science	Foreign	Good	1
	Undergraduate	Art	Canada	Excellent	1

Table 8: First merging identical tuples

Name	Category	Major	Birthplace	GPA	Num
	Graduate	Art	Canada	Excellent	1
	Undergraduate	Science	Canada	Excellent	2
	Undergraduate	Art	Canada	Average	1
	Graduate	Science	Canada	Excellent	2
	Graduate	Science	Foreign	Good	3
	Undergraduate	Science	Canada	Average	1
	Undergraduate	Art	Canada	Good	1
	Undergraduate	Art	Canada	Excellent	1

Process on fifth attribute or attribute GPA :

Next, the second line of AOI characteristic rule algorithm is executed with i=5 to check number of distinct value in the fifth attribute or attribute GPA and table 6 shows there are 10 distinct values such as {3.5, 3.7, 2.6, 3.9, 3.3, 2.7, 3.4, 3.0, 3.8, 3.2}. Because distinct value in attribute GPA=10 and greater than threshold=2, then the third line of AOI characteristic rule algorithm will be run in order to check if attribute GPA has higher concept in concept hierarchy. Because the fifth attribute or attribute GPA has higher concept, then line number 5 of AOI characteristic rule algorithm is run in order to substitute the value of attribute GPA with corresponding minimal generalized concept. The minimal corresponding generalized concept for attribute GPA can be seen on concept tree for GPA in figure 4 on previous research[3] and shown in figure 5.

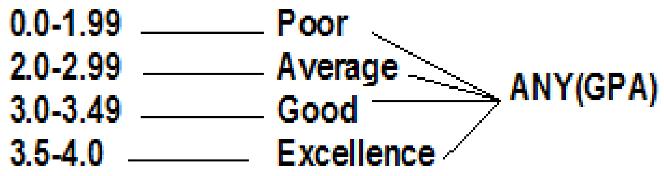

Figure 5: Concept tree for GPA

Table 7 shows student dataset after attribute GPA first generalized concept and afterward, line number 6 of AOI characteristic rule will be executed in order to merge identical tuples. In table 7 there are 7 tuples with 3 different identical tuples, where 2nd tuple is merged with 7th tuple, 4th tuple is merged with 10Th tuple and 5th tuple with 8th and 11th tuples. Table 8 shows the result of first merging identical tuples with only 8 tuples left.

Thereafter, the process will loop to second line of AOI characteristic rule algorithm with still i=5 to check distinct value in fifth attribute or attribute GPA and table 8 shows there are 3 distinct values such as {Excellent, Average, Good}. Because distinct value in attribute GPA=3 and greater than threshold=2, then the third line of AOI characteristic rule algorithm will be run in order to check if attribute GPA has higher concept in concept hierarchy. Because the fifth attribute or attribute GPA has higher concept, then line number 5 of AOI characteristic rule algorithm is run in order to substitute the value of attribute GPA with corresponding minimal generalized concept as can be seen on concept tree for GPA in figure 4.

Table 9: Second generalization on attribute GPA

Name	Category	Major	Birthplace	GPA	Num
	Graduate	Art	Canada	ANY	1
	Undergraduate	Science	Canada	ANY	2
	Undergraduate	Art	Canada	ANY	1
	Graduate	Science	Canada	ANY	2
	Graduate	Science	Foreign	ANY	3
	Undergraduate	Science	Canada	ANY	1
	Undergraduate	Art	Canada	ANY	1
	Undergraduate	Art	Canada	ANY	1

Table 9 shows student dataset after attribute GPA second generalized concept and afterward, line number 6 of AOI characteristic rule will be executed in order to merge identical tuples. In table 9 there are 5 tuples with 2 different identical tuples, where 2nd tuple is merged with 6th tuple and 3th tuple is merged with 7Th and 8th tuples. Table 10 shows the result of second merging identical tuples with only 5 tuples left.

Extended Experiment :
Extended experiment shows that if AOI characteristic rule algorithm is run with threshold more than 2 will have 8 tuples as shown in table 8. The number 2 in "threshold more than 2" is getting from the maximum number of distinct values in dataset as shown in equation 1. In equation 1, the maximum number of threshold (mt) will depend on the maximum number of distinct values in dataset (md).

$$mt=md \qquad (1)$$

where :
mt = maximum number of threshold,
md=maximum number of distinct values in dataset

In this experiment, the maximum number of threshold (mt) is 3 as the maximum number of finding distinct values in dataset is 3 which can find in attribute GPA such as {Excellent, Average, Good}. That it is why for AOI characteristic rule algorithm, particularly in this experiment have the same number 8 tuples as shown in table 8 for threshold 3 and more. Nonetheless, the maximum higher level concept for concept hierarchy GPA which is shown in figure 4 shows there are 4 higher level concepts and they are {Excellent, Average, Good,Poor}.

Finally, the number of tuple result such as 5 or 8 will not be interesting since we just only interested with exact finding number of tuples depend on the number tuples in dataset. Finding the interested characteristic rule will be continued in AOI characteristic rule algorithm between line number 9 and 12, where can be seen on previous research [3].

Table 10: Second merging identical tuples

Name	Category	Major	Birthplace	GPA	Num
	Graduate	Art	Canada	ANY	1
	Undergraduate	Science	Canada	ANY	3
	Undergraduate	Art	Canada	ANY	3
	Graduate	Science	Canada	ANY	2
	Graduate	Science	Foreign	ANY	3

Conclusions

In implementation AOI characteristic rule algorithm between line number of 1 and 8 shows that maximum number of threshold will depend on maximum number of distinct values in dataset as shown in equation 1. In future proposed research, threshold control in AOI which has been known as tedious and consuming work[4], is resolved with equation 1 where at least maximum number of threshold is found.
In this paper, the explanation step by step for each line number between line number 1 and 8 in AOI characteristic rule algorithm and using only 12 tuples student dataset, will easy to understand the steps for mining data with AOI characteristic rule algorithm. At the end, people will think that data mining is not hard as their thought.

For future development, this AOI characteristic rule algorithm should be implemented with server programming language such as PHP (Personal Home Pages) and share to public. The choosing PHP as implementation has motivations such as :

- Prevent of infringement of intellectual property rights, where public can run their dataset but cannot have or download the algorithm, moreover public can submit their dataset into website and get their AOI characteristic rule algorithm result running. However, internet online time connection will become issue.
- Online 24 hours access with browser, access anytime, anywhere and anyplace.

Acknowledgments

This work is supported under Program of research incentive of national innovation system (SINAS) from Ministry of Research, Technology and Higher Education of the Republic of Indonesia, decree number 147/M/Kp/IV/2015, Research code: RD-2015-0020.

References

[1] J. Han, Y. Cai, N. Cercone, Knowledge discovery in databases: An attribute-oriented approach. Proceeding of 18th International Conference Very Large Data Bases, Vancouver, British Columbia, (1992) 547-559.

[2] S.Warnars, Mining Patterns with Attribute Oriented Induction. Proceeding of International Conference on Database, Data Warehouse, Data Mining and Big Data (DDDMBD), Tangerang, Indonesia, (2015) 11-21.

[3] S. Warnars, Measuring Interesting rules in characteristic rule. Proceeding of the 2nd International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT), Bali, Indonesia, (2010) 152-156.

[4] Y. Wu, Y. Chen, R. Chang, Generalized knowledge discovery from relational databases. International Journal of Computer Science and Network, 9(6)(2009) 148-153.

[5] J. Han, Y. Cai, N. Cercone, Data-driven discovery of quantitative rules in relational databases. IEEE Transction on Knowledge and Data Engineering, 5(1)(1993) 29-40.

[6] J. Han, Y. Fu, Exploration of the power of attribute-oriented induction in data mining. in U. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, eds. Advances in Knowledge Discovery and Data Mining, (1995) 399-421.

[7] Y. Cai, Attribute-oriented induction in relational databases. Master thesis, Simon Fraser University, (1989).

[8] D.W. Cheung, H.Y. Hwang, A.W. Fu, J. Han, Efficient rule-based attribute-oriented induction for data mining. Journal of Intelligent Information Systems, 15(2)(2000) 175-200.

[9] S. Warnars, Mining Frequent Pattern with Attribute Oriented Induction High level Emerging Pattern (AOI-HEP). Proceeding of IEEE the 2nd International Conference on Information and Communication Technology (IEEE ICoICT 2014), 28-30 May 2014, Bandung, Indonesia, (2014), 144-149.

[10] S. Warnars, Mining Frequent and Similar Patterns with Attribute Oriented Induction High Level Emerging Pattern (AOI-HEP) Data Mining Technique. International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS), 3(11)(2014), 266-276.