Check 38. Mining Frequent and Similar Patterns with Attribute Oriented Induction High Level Emerging.pdf

by Arta Sundjaja

Submission date: 30-Apr-2019 03:47PM (UTC+0700)

Submission ID: 1102510710

File name: tterns_with_Attribute_Oriented_Induction_High_Level_Emerging.pdf (532.11K)

Word count: 7980

Character count: 40358

International Association of Scientific Innovation and Research (IASIR)
(An Association Unifying the Sciences, Engineering, and Applied Research)

ISSN (Print): 2279-0047
ISSN (Online): 2279-0055

International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

www.iasir.net

Mining Frequent and Similar Patterns with Attribute Oriented Induction High Level Emerging Pattern (AOI-HEP) Data Mining Technique

Spits Warnars
Database, Datawarehouse and Data Mining Research Center,
Human Computer Interaction Department, Surya University,
Jl. Boulevard Gading Serpong blok O/1 Summarecon Serpong,
Tangerang, 15810, INDONESIA.

***Abstract**: Attribute Oriented Induction High level Emerging Pattern (AOI-HEP) is a novel idea which is influenced by Attribute Oriented Induction (AOI) and Emerging Pattern (EP). AOI-HEP discovers patterns such as Total Subsumption HEP (TSHEP), Subsumption Overlapping HEP (SOHEP) and Total Overlapping HEP (TOHEP) include frequent and similar patterns. Mining TSHEP, SOHEP, TOHEP, frequent and similar patterns for each dataset is influenced by learning on high level concept in one of chosen attribute. The experiments used four datasets from UCI machine learning repository and most datasets have SOHEP but not TSHEP and TOHEP and the most rarely found were TOHEP. There are total twenty two High level Emerging Pattern (HEP) where four HEP are TSHEP, sixteen HEP are SOHEP and two HEP are TOHEP, and there are five frequent and four similar patterns from the experiments. Moreover, the experiment showed that adult and breast cancer datasets are interested to mine frequent pattern while breast cancer and IPUMS datasets are interested to mine similar pattern. However, census dataset is not interested to be mined for both frequent and similar patterns. AOI-HEP is suitable for dealing with large dataset since can handle million tuples in dataset in one digit seconds.*

***Keywords**: Attribute-oriented induction; Emerging pattern; High Emerging Pattern; Frequent pattern; Similar pattern*

I. Introduction

This paper proposes Attribute Oriented Induction High level Emerging Pattern (AOI-HEP) [19, 20] (as a hybrid approach which is influenced by two data mining techniques i.e. Attribute Oriented Induction (AOI) [4,11] and Emerging Pattern (EP) [6,7,10,17,18]. AOI influences AOI-HEP by using AOI characteristic rule algorithm which was run twice with two input datasets, derived from the same dataset in order to create two rulesets which are then processed with High level Emerging Pattern (HEP) algorithm. EP influences AOI-HEP by extending growth rate equation and propose HEP algorithm which is not influenced by border-based algorithm. EP was proposed earlier by a border-based algorithm and influences most other EP mining algorithms. The border-based algorithm avoids the long process naive algorithms do to get the counts of all itemsets in a large collection of candidates, by manipulating only borders of some two collections and derive all EPs whose support satisfies a minimum support threshold in dataset [6]. The first proposed AOI-HEP was only to mine Total Subsumption HEP (TSHEP) and Subsumption Overlapping HEP (SOHEP) [19], and was extended to mine frequent pattern from both of TSHEP and SOHEP [20]. Meanwhile, this paper proposes extension AOI-HEP with Total Overlapping HEP (TOHEP), mine both of frequent and similar patterns.

Firstly, the HEP algorithm starts with Cartesian product between two rulesets which eliminates rules in rulesets with a metric similarity using the categorization of attribute comparison. Secondly, the output rules between two rulesets from metric similarity are discriminated with growth rate to find ratio of supports between rules from two rulesets. The categorization of attribute comparisons is based on similarity hierarchy level and values which have three options in how they subsume each other. These are Total Subsumption HEP (TSHEP), Subsumption Overlapping HEP (SOHEP) and Total Overlapping HEP (TOHEP). From certain similarity hierarchy level and values, we can mine frequent and similar patterns.

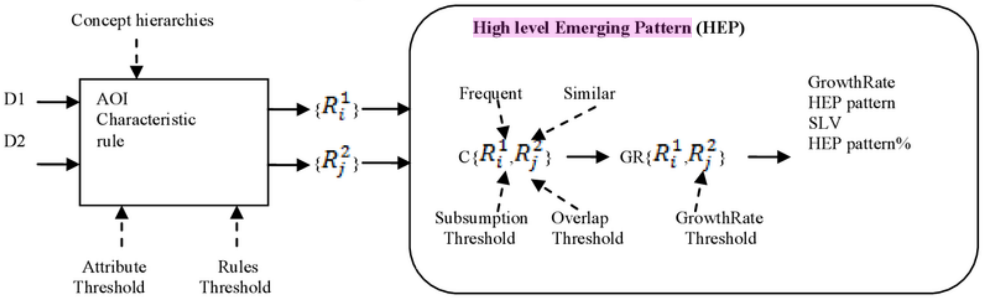
The main purpose or motivation of proposing AOI-HEP which is influenced with AOI and EP is to use its typical strength of extracting important high-level emerging knowledge from data. The typical strength of AOI is using concept hierarchy [2] to produce high-level data, and moreover, AOI is recognized as an important mining technique since has been tested successfully against large relational database and can learn different kinds of rules. Meanwhile, the typical strength of EP is using growth rate as ratio of the supports in one dataset to another dataset. In addition, EP is recognized as a powerful mining technique to discriminate datasets. AOI concerns with high level data whereas EP concerns with low level data. The new framework, AOI-HEP, is able to produces high level emerging patterns which discriminates two datasets. AOI-HEP will be better than AOI since AOI-HEP pattern results are less than AOI pattern results and AOI-HEP will be better than EP since AOI-HEP pattern results are on high level while EP pattern results are on low level.

The paper is organized as follows: Section 2 shows AOI-HEP framework which show combination AOI and HEP. Section 3 describes representation rules and rulesets, TSHEP, SOHEP and TOHEP definitions. Section 4 presents HEP algorithm as implementation part of AOI-HEP framework. Moreover, section 5 defines metric similarity function $C\{R_i^1,R_j^2\}$ and determine concept mining for TSHEP, SOHEP, TOHEP, frequent and similar patterns. Furthermore, section 6 discusses growth rate function $GR\{R_i^1,R_j^2\}$ as used in current EP but have two the same or different high level itemset instead of one low level itemset. Meanwhile, section 7 discusses experiments for four UCI repository datasets with each user defined concept hierarchies, include experimental mining for TSHEP, SOHEP, TOHEP, frequent and similar patterns. Conclusion is given in section 8.

II. AOI-HEP Framework

Figure 1 shows the proposed AOI-HEP framework where traditional AOI characteristic rule algorithm is run twice with two datasets D1 and D2 (horizontal partitions of the dataset). AOI uses concept hierarchy as background knowledge for data generalization. AOI eliminates distinct attributes and tuples until they are less or equal than attribute and rules thresholds respectively [11]. AOI's outputs are rulesets R_i^1 and R_j^2 from datasets D1 and D2 respectively. Rulesets R_i^1 and R_j^2 are inputs for High level Emerging Pattern (HEP) which include two functions i.e. similarity function $C\{R_i^1,R_j^2\}$ and growth rate function $GR\{R_i^1,R_j^2\}$. The $C\{R_i^1,R_j^2\}$ function is a metric similarity function which applies cartesian product between rulesets R_i^1 and R_j^2, and eliminate the cartesian product by determining the type of HEP i.e. either TSHEP, SOHEP or TOHEP [19].

Figure 1. AOI-HEP Framework.



III. HEP Definition

For High level Emerging Patterns (HEP), let D1 and D2 be horizontal partitions of some dataset $D^x = \{A_i, \dots, A_p\}$ with p attributes $1 \le i \le p$and $1 \le x \le 2$. Rulesets $\{R_i^1\}$ and $\{R_j^2\}$ from datasets D1 and D2 are represented as $R^x = \{r_1^x, r_2^x, \dots, r_n^x\}$ in figure 2. In figure 2 each ruleset R^xconsists of n rules where n ≤ rules threshold. Each rule in a ruleset R^x is represented by attributes $r_n^x = \{A_1^x, A_2^x, A_{\dots}^x, A_m^x, |r_n^x|\}$, where $|r_n^x|$ is number of tuples forming the rule and m is the number of attributes in a ruleset as in equation 1. Figure 2 shows the representation of rulesets $R^x = \{r_1^x, r_2^x, \dots, r_n^x\}$ vertically where $r_n^x \in R^x$ and each rule $r_n^x = \{A_1^x, A_2^x, A_{\dots}^x, A_m^x, |r_n^x|\}$ horizontally where $A_m^x \in r_n^x$. For example we have used rule r_1^1in ruleset 1 and rule r_1^2 in ruleset 2. $A_m^1 \in r_1^1$ where all attributes A_m^1 are member of rule r_1^1 in ruleset 1 and $A_m^2 \in r_1^2$ where all attributes $A_m^2$are member of rule r_1^2 in ruleset 2. If there are four attributes (m=4 in equation 1) then rule $r_1^1 = \{A_1^1, A_2^1, A_3^1, A_4^1, |r_1^1|\}$ and rule $r_1^2 = \{A_1^2, A_2^2, A_3^2, A_4^2, |r_1^2|\}$.

Figure 2. Representation rule and rulesets.

$$\begin{aligned} D^x \rightarrow R^x \rightarrow r_1^x &= \{A_1^x, A_2^x, A_{\dots}^x, A_m^x, |r_1^x|\} \\ r_2^x &= \{A_1^x, A_2^x, A_{\dots}^x, A_m^x, |r_2^x|\} \\ r_{\dots}^x &= \{A_1^x, A_2^x, A_{\dots}^x, A_m^x, |r_{\dots}^x|\} \\ r_n^x &= \{A_1^x, A_2^x, A_{\dots}^x, A_m^x, |r_n^x|\} \end{aligned}$$

A. Definition of Total Subsumption HEP (TSHEP)

For Total Subsumption HEP (TSHEP) we say rule r_1^1 is totally subsumed by rule r_1^2 if $A_{[1..m]}^1 \subseteq A_{[1..m]}^2$ then $r_1^1 \subseteq r_1^2$. This means rule r_1^1 is TSHEP by rule r_1^2 ($r_1^1 \subseteq r_1^2$, rule r_1^1 is a subset of rule r_1^2) if each attribute in A_m^1 is subsumed by each attribute in A_m^2 ($A_{[1..m]}^1 \subseteq A_{[1..m]}^2$). Based on example four attributes for rules r_1^1 and r_1^2 if each attribute in A_m^1 is subsumed by each attribute in A_m^2 $\{A_1^1 \subseteq A_1^2, A_2^1 \subseteq A_2^2, A_3^1 \subseteq A_3^2, A_4^1 \subseteq A_4^2\}$ then $r_1^1 \subseteq r_1^2$.

B. Definition of Total Overlapping HEP (TOHEP)

Meanwhile, for Total Overlapping HEP (TOHEP) we say rule r_1^1 totally overlaps with rule r_1^2 if $A_{[1..m]}^1 \cap A_{[1..m]}^2$ then $r_1^1 \cap r_1^2$. This means rule r_1^1 is TOHEP with rule r_1^2 ($r_1^1 \cap r_1^2$, rule r_1^1 is overlap with rule r_1^2) if each attribute in A_m^1 is overlap with each attribute in A_m^2 ($A_{[1..m]}^1 \cap A_{[1..m]}^2$). Based on example four attributes for rules r_1^1 and r_1^2, if each attribute in A_m^1 is overlap with each attribute in A_m^2 $\{A_1^1 \cap A_1^2, A_2^1 \cap A_2^2, A_3^1 \cap A_3^2, A_4^1 \cap A_4^2\}$ then $r_1^1 \cap r_1^2$.

C. *Definition of Subsumption Overlapping HEP (SOHEP)*

Moreover, for Subsumption Overlapping HEP (SOHEP) we say rule r_1^1 is subsumed by and overlaps with rule r_1^2: if $A_{[1..m1]}^1 \subseteq A_{[1..m1]}^2$ and $A_{[m1+1..m]}^1 \cap A_{[m1+1..m]}^2$ then $r_1^1 \subset r_1^2$ and $r_1^1 \cap r_1^2$. This means rule r_1^1 is SOHEP with rule r_1^2 ($r_1^1 \subset r_1^2$,rule r_1^1 is a proper-subset of rule r_1^2) and ($r_1^1 \cap r_1^2$, rule r_1^1 overlaps with rule r_1^2), if some attributes from 1 to m1 in A_m^1 are subsumed by some attributes from 1 to m1 in A_m^2 ($A_{[1..m1]}^1 \subseteq A_{[1..m1]}^2$) and if some attributes from m1+1 to m in A_m^1 are overlap with some attributes from m1+1 to m in A_m^2 ($A_{[m1+1..m]}^1 \cap A_{[m1+1..m]}^2$), where m1 is the number subsumption attribute and m is the number of attributes in a ruleset as in equation 1. Based on example four attributes for rules r_1^1 and r_1^2, if the first two attributes in A_m^1 are subsumed by the first two attributes in A_m^2 and certainly the last two attributes in A_m^1 are overlap with the last two attributes in A_m^2 $\{A_1^1 \subseteq A_1^2, A_2^1 \subseteq A_2^2, A_3^1 \cap A_3^2, A_4^1 \cap A_4^2\}$ then $r_1^1 \subset r_1^2$and $r_1^1 \cap r_1^2$.

IV. HEP Algorithm

Figure 3 shows the HEP algorithm as part of AOI-HEP framework in figure 1. The HEP algorithm has inputs such as rulesets R_i^1 and R_j^2, subs_threshold, overlap_threshold, GR_threshold,num_attr, |D2| , |D1|, Frequent and Similar. The HEP algorithm inputs are in accordance with inputs for HEP in AOI-HEP framework figure 1 where for HEP in figure 1 there are rulesets R_i^1 and R_j^2 inputs, subs_threshold, overlap_threshold, Frequent and Similar for C{ R_i^1 , R_j^2 } function, GR_threshold for GR{ R_i^1 , R_j^2 } function. The three thresholds i.e.: subs_threshold, overlap_threshold and GR_threshold have default value 0 and for subs_threshold and overlap_threshold have maximum value 100. Moreover, num_attr input is the number attributes in rulesets R_i^1 and R_j^2 as m in equation 1. The outputs from HEP algorithm are in accordance with the HEP outputs shown in figure 1 and they are GrowthRate, HEP pattern, SLV and HEP pattern%. The outputs are printed in line 17 in HEP algorithm.

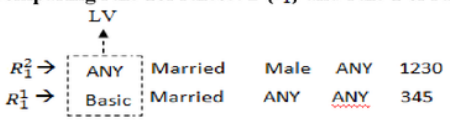
Figure 3. AOI-HEP Algorithm.

```
HEP algorithm
Input:  {$R_i^1$} , {$R_j^2$}, subs_threshold, overlap_Threshold, GR_threshold, num_attr,|D2|,|D1|, Frequent, Similar
Output:  growth rate, HEP pattern, SLV, HEP pattern%
1.  While( noAllANY($R_i^1$))
2.   {While ( noAllANY($R_j^2$))
3.     {SLV=0,    over=0, subs=0, F=0,S=0
4.      for x=1 to num_attr
5.        {if $R_i^1$[x]== $R_j^2$ [x] and $R_i^1$[x]=="ANY"                     SLV=SLV+2.1,   over=over+1,S++
6.         if $R_i^1$[x]== $R_j^2$[x] and $R_i^1$[x]!="ANY"                      SLV=SLV+2,       over=over+1
7.         if $R_i^1$[x]!= $R_j^2$[x] and $R_j^2$[x] subsump by $R_i^1$[x]       SLV=SLV+0.4,     subs=subs+1
8.         if $R_i^1$[x]!= $R_j^2$[x] and $R_i^1$[x] subsump by $R_j^2$[x]       SLV=SLV+0.5,     subs=subs+1,F++}
9.      subs_=subs/num_attr*100
10.     over_=over/num_attr*100
11.     if subs_>subs_threshold and over_>over_threshold
12.        if subs>0   and over==0    HEP pattern="TSHEP", HEP pattern%=subs_
13.        if subs>0   and over>0     HEP pattern="SOHEP", HEP pattern%=subs_+over_
14.        if subs==0  and over>0     HEP pattern="TOHEP", HEP pattern%=over_
15.        growth rate=($R_j^2$[x+1]/|D2|)  /  ($R_i^1$[x+1]/|D1|)
16.         if growth rate > GR_threshold  and/or  (Frequent and F==x or F==x-1) and/or  (Similar and S<x-1)
17.            print   growth rate, HEP pattern,SLV,HEP pattern%    }    }
```

V. Metric Similarity

This section presents the metric similarity function C{R_i^1,R_j^2} between rulesets {R_i^1} and {R_j^2}. As mention in section 2, the C{R_i^1,R_j^2} function is a metric similarity function which apply cartesian product between rulesets R_i^1 and R_j^2, and eliminate the cartesian product by determining type of HEP. The determining type of HEP is applied by summing categorization of attribute comparison value and hierarchy level based on subsumption and overlap thresholds. To derive similarity hierarchy level value (SLV) in the HEP algorithm, firstly, we determine categories of attribute values between the rulesets as shown in figure 4. The categorization is based on similarity hierarchy level and the values shown in equation 1 as LV. Secondly, by summing the attribute categorizations or LV values, we get SLV (equation 1) as the similarity between the two rules. The two steps described above are shown between line numbers 4 and 8 in the HEP algorithm of figure 3.

Figure 4. Comparing rule 1of ruleset 2 $\{r_1^2\}$ and rule 1 of ruleset 1 $\{r_1^1\}$



$$SLV=\sum_{i=1}^{m} LV_i \qquad (1)$$

where:

SLV =similarity value based on the similarity of attributes hierarchy level and values
M = number of attributes in a ruleset, where m > 1
(number of attributes in concept hierarchies - 1)
i = attribute position
LVi = categorization of attributes comparison based on similarity hierarchy level and values, and the options are :

a. If hierarchy level is different and the attribute in rule of ruleset R2 is subsumed by the attribute in rule of ruleset R1, LV=0.4.

b. If hierarchy level is different and the attribute in rule of ruleset R1 is subsumed by the attribute in rule of ruleset R2, LV=0.5.

c. If hierarchy level and values are the same and the attributes values are not ANY, LV=2.

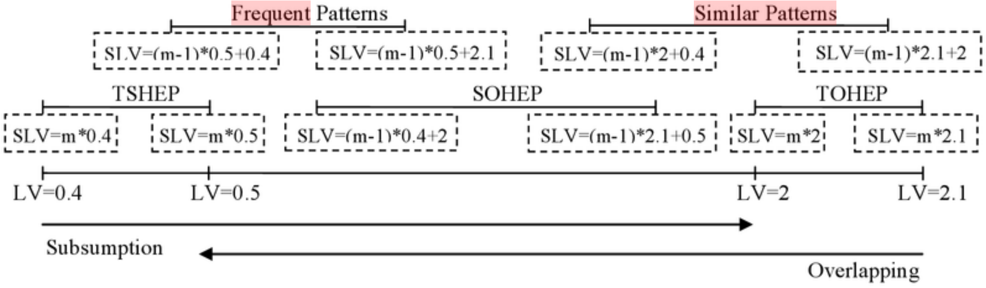
d. If hierarchy level and values are the same and the attributes values are ANY, LV=2.1.

The four categorization of attribute comparisons or LV in equation 1 is based on two main categorizations i.e. subsumption (LV=0.4 or LV=0.5) and overlapping (LV=2 or LV=2.1). For each LV option values 0.4,0.5,2 and 2.1 are user defined number, where option numbers 0.4 and 0.5 as values for subsumption categorization (minimum categorization) and option numbers 2 and 2.1 as values for overlapping categorization (maximum categorization). After the similarity between the two rules (SLV) has been derived, then we can determine type of HEP between TSHEP, SOHEP or TOHEP and mining frequent and similar patterns.

A. Mining TSHEP, SOHEP and TOHEP

Determining type of HEP between TSHEP, SOHEP or TOHEP is shown between line 12 and 14 in figure 3 which is categorized with variables over and subs. Variable over represents the overlapping (LV=2 or LV=2.1) and variable subs represents the subsumption (LV=0.4 or LV=0.5) which are possibly having increment as shown between line number 5 and 6, and number 7 and 8 in figure 3 respectively. The mining between TSHEP, SOHEP or TOHEP can be filtered when the variables over and subs are limited with over_threshold and subs_threshold as inputs HEP algorithm respectively as shown in line number 11 figure 3. TSHEP and TOHEP are composition subsumption (LV=0.4 or LV=0.5) and overlapping (LV=2.0 or LV=2.1) respectively, whilst SOHEP as composition between subsumption (LV=0.4 or LV=0.5) and overlapping (LV=2.0 or LV=2.1) have minimum and maximum SLV values as shown in figure 5.

Figure 5. Composition subsumption and overlapping for mining patterns



The overlapping arrow line shows the influence overlapping from LV=2.1 (maximum value for overlapping categorization) until LV=0.5 (maximum value for subsumption categorization). Whilst subsumption arrow line shows the influence subsumption from LV=0.4 (minimum value for subsumption categorization) until LV=2 (minimum value for overlapping categorization). SLV is categorized as TSHEP when have all subsumption LV values (LV=0.4 or LV=0.5) where minimum and maximum SLV values between m*0.4 and m*0.5. Meanwhile, SLV is categorized as SOHEP when have combination subsumption and overlapping LV values (LV=0.4 or LV=0.5 and LV=2 or LV=2.1) where minimum and maximum SLV values between (m-1)*0.4+2 and (m-1)*2.1+0.5. Moreover, SLV is categorized as TOHEP when have all overlapping LV values (LV=2 or LV=2.1) where minimum and maximum SLV values between m*2 and m*2.1.

B. Mining Frequent pattern

Frequent pattern is a combination of feature patterns that appear in dataset with frequency not less than a user-specified threshold [12,13] and the frequent pattern synonym with large pattern was first proposed for market basket analysis in the form of association rules [1]. With frequent pattern we can have strong/sharp discrimination power where have large growth rate and support in target (D2) dataset and other support in contrasting (D1) dataset is small [6,8,14]. From frequent patterns, we can create a discrimination rule and are interested in mining the frequent pattern with strong/sharp discrimination power. In EP, the strength of discrimination power is expressed by its large growth rate and support in target (D2) dataset [6,8]. This is called an essential Emerging Patterns (eEP) [8]. In AOI-HEP, the strength of discrimination power is expressed by its large growth rate and support in target (D2) dataset and expressed by subsumption LV=0.5 where R2 in target (D2) dataset is superset with large support and R1 in contrasting (D1) dataset is subset with low support.

Since frequent pattern in AOI-HEP are expressed by value LV=0.5 then frequent pattern can be mined from TSHEP or SOHEP as shown in figure 5, where minimum and maximum SLV values between (m-1)*0.5+0.4 and (m-1)*0.5+2.1, showed that not all frequent pattern have the same LV=0.5 values. Frequent pattern without the same LV=0.5 values, have been allocated to percentage value of (m-1)/m*100 and it is accordance where two parts of objects are similar if they are similar in all features (full matching similarity) or if the percentage of similar features is greater than the 80% [5] or if they are similar in at least 90% of the features [15]. Indeed, frequent similarity subsumption LV=0.5 at percentage value of (m-1)/m*100 shows that at least LV values have greater than (m-1)/m*100.

The HEP algorithm in figure 3 shows the process of mining frequent pattern with strong discrimination power, which is executed by giving condition true to input frequent variable. Moreover, variable counter F, will be incremented when have subsumption LV=0.5 as shown in line number 8. In line number 16, if input Frequent variable is true and variable F=x or F=x-1 then the output will be categorized as frequent pattern with strong discrimination power, where x is m in equation 1. F=x represents to TSHEP with full similarity subsumption LV=0.5, while F=x-1 represents to TSHEP or SOHEP with frequent similarity subsumption LV=0.5.

C. Mining Similar pattern

Similar patterns are interesting to mine because similarity pattern between datasets show the equality pattern which can represent similar behavior patterns. There are many examples of the important similar patterns in data mining process. In business, it is important to discover companies with similar patterns such as similar growth patterns, similar product selling patterns and etc. In education, it is important to discover students with similar patterns such as similar student behavior patterns, similar student progress patterns and etc. In banking system, it is important to discover customer with similar patterns such as similar customer behavior patterns, similar customer loan patterns and etc. Searching similar patterns are important and can be used for segmentation or prediction. For example in banking system, banking segmentation and banking prediction with similar banking transaction could help to show banking transaction prediction, with similar customer behavior patterns could help to uncover fraud, and loan prediction [16]. The similarity patterns can be measured with similarity two or more attributes or by calculating distance with euclidean distance or manhattan distance [3].

In AOI-HEP, similar patterns are shown by overlapping LV=2.0 or LV=2.1 and as shown in figure 5, similar pattern are mined from SOHEP or TOHEP, having minimum and maximum SLV values between (m-1)*2+0.4 and (m-1)*2.1+2. As mentioned before, that two parts of objects are similar if they are similar in all features (full matching similarity) or if the percentage of similar features is greater than the 80% [5] or if they are similar in at least 90% of the features [15]. Therefore, AOI-HEP similar pattern are interested to SOHEP or TOHEP with frequent overlapping LV=2.0 or frequent combination overlapping LV=2.0 and LV=2.1 at percentage value of (m-1)/m*100 where m as in equation 1. However, AOI-HEP similar pattern are not interested to SOHEP or TOHEP with frequent overlapping LV=2.1 at percentage value of (m-1)/m*100, where LV=2.1 is ANY and means nothing. Moreover, AOI-HEP similar pattern are not interested to TOHEP with full similarity overlapping LV=2.1 and shown in line number 1 and 2 in HEP algorithm figure 3 which show the exclusion rule with ANY values in all attributes in rulesets. Similar like frequent pattern, discrimination rule can be created from similar pattern.

The HEP algorithm in figure 3 shows the process mining similar pattern is executed by giving condition true to input similar variable. Moreover, variable counter S will be incremented when have overlapping LV=2.1 as shown in line number 5. In line number 16, if input Similar variable is true and variable S<x-1 then the output will be categorized as similar pattern, where x is m in equation 1. S<x-1 represents to SOHEP with frequent similarity overlapping LV=2.1 < x-1 where SOHEP with frequent similarity overlapping LV=2.1 at percentage value of (m-1)/m*100 is not interesting (for instance SOHEP with SLV=2.1+2.1+2.1+0.5).

VI. HEP Growth Rate

Besides eliminating patterns with similarity function $C\{R_i^1,R_j^2\}$, the large number of HEP (Cartesian product between rulesets) is eliminated by the growth rate function $GR\{R_i^1,R_j^2\}$ with given a GrowthRate threshold. Growthrate is a standard function used in Emerging Patterns (EP) [6], and the difference in our approach is discovering high level emerging pattern with the same or different itemset instead of low level pattern with the same itemset. As mentioned in section 3, rulesets are AOI outputs and each of rule in ruleset has $|r_n^x|$ as the number of tuples forming the rule (figure 2). Because of rule in ruleset has $|r_n^x|$ as the number of tuples, then there is no Jumping High level Emerging Patterns (JHEP), where JHEP is related as a term of Jumping EP (JEP). JEP is EP with support is 0 in one dataset and more than 0 in the other dataset or EP as special type of EP which is having infinite growth rate (∞).

Growth rate GR$\{R_i^1, R_j^2\}$ is shown in figure 1 and in line number 15 in the HEP algorithm in figure 3 is used to discriminate between datasets D2 and D1. This growth rate can be calculated using equation 2. We can define that a HEP is a ruleset whose support changes from one ruleset in dataset D1 to another ruleset in dataset D2. In other words, HEP is a ruleset whose strength of high level rule Y of ruleset R1 in dataset D1 changes to high level rule X of ruleset R2 in dataset D2. Conventionally, this is defined as follows:

$$GR(X,Y) = \frac{\text{Support D2}(X)}{\text{Support D1}(Y)} = \frac{\text{Count R2}(X)\ /\ |D2|}{\text{Count R1}(Y)\ /\ |D1|} \qquad (2)$$

where:

X = High level rule of ruleset R2 in dataset D2.
Y = High level rule of ruleset R1 in dataset D1.
D2 = Dataset D2.
D1 = Dataset D1.
|D2| = Total number of instances in dataset D2.
|D1| = Total number of instances in dataset D1.
Count R2(X) = Number of high level rule X of ruleset R2 in dataset D2.
Count R1(Y) = Number of high level rule Y of ruleset R1 in dataset D1.
Support D2(X) = Composition number of high level rule X of ruleset R2 in D2.
Support D1(Y) = Composition number of high level rule Y of ruleset R1 in D1.

VII. Experimental evaluation

The experiments used four datasets from the UCI machine learning repository: adult, breast cancer, census, and IPUMS datasets with the number of instances 48842, 569, 2458285 and 256932 respectively [9]. Each dataset has concept hierarchies built from five chosen attributes with a minimum concept level of three. The attributes in concept hierarchies for adult dataset include workclass, education, marital-status, occupation, and native-country attributes. The attributes in concept hierarchies for the breast cancer dataset contains attributes i.e. clump thickness, cell size, cell shape, bare nuclei and normal nucleoli attributes. Meanwhile, class, marital status, means, relat1 and yearsch attributes, were given to concept hierarchies for the Census dataset. Finally, the attributes in concept hierarchies for the IPUMS dataset consists of relateg, marst, educrec, migrat5g and tranwork attributes.

Each dataset was divided into two sub datasets based on learning the high level concept in one of their attributes. Learning the high level concept in one of their five chosen attributes for concept hierarchies, makes the parameter m in equation 1 has value 4, where value 4 comes from five chosen attributes for concept hierarchies minus 1 and 1 is the attribute for the learning concept. In the adult dataset, we learn by discriminating between the "government" (4289 instances) and "non government" (14 instances) concepts of the "workclass" attribute in datasets D2 and D1 respectively. In the breast cancer dataset, we learn by discriminating between "aboutaverclump" (533 instances) and "aboveaverclump" (289 instances) concepts of the "clump thickness" attribute in datasets D2 and D1 respectively. Meanwhile Census dataset learns "green" (1980 instances) and "no green" (809 instances) concepts of the "means" attribute for datasets D2 and D1 respectively. Finally, the IPUMS dataset learns "unmarried" (140124 instances) and "married" (77453 instances) concepts of the "marst" attribute as datasets D2 and D1 respectively.

Experiments were carried out by a java application as shown in figure 6. The experiments were tested on Intel(R) Atom(TM) CPU N550 (1.50 GHz) with 1.00 GB RAM. The AOI-HEP application has an input dataset and corresponding concept hierarchies in the form of flat files respectively. The AOI-HEP application was run 4 times as the number of experimental datasets and with the attribute and rule thresholds 6 which were chosen based on the preliminary experiments done on all datasets such that to get meaningful numbers of rules. higher threshold is preferable after trial experiments. The AOI algorithm is part of the AOI-HEP application which is combined with the HEP algorithm.

Running AOI-HEP application with input adult, breast cancer, census and IPUMS datasets have running time of approximately 3, 3, 4 and 13 seconds respectively. Incredibly, the extraordinary running time of 13 seconds with the input IPUMS dataset happened because IPUMS has huge instances learning dataset's unmarried and married concepts with 140124 and 77453 instances respectively. In running AOI-HEP application, each dataset has rulesets R2 and R1 based on learning concepts in one chosen of its attribute. Since there is page's limit for paper publishing, then we limit to rulesets R2 and R1 for only adult dataset as shown in tables 1 and 2 respectively. Rulesets R2 and R1 in table 1 and 2 are the result learning from "government" and "non government" concepts from the same "workclass" attribute of adult dataset. Ruleset R2 or R1 as shown in tables 1 or 2 has 6 tuples (rules) include number of instances for each tuple (rule) and has four attributes (m in equation 1) as representation rules and rulesets in figure 2.

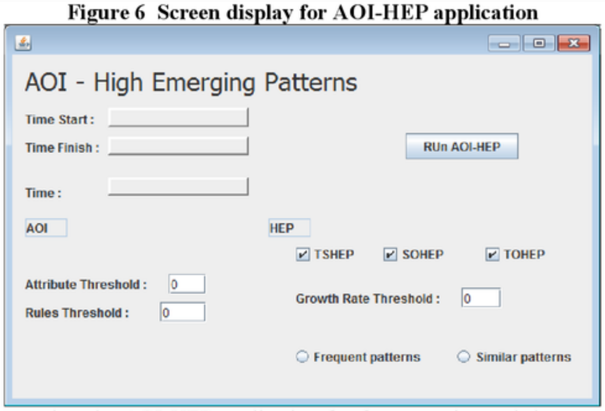
Table I Ruleset R2 for learning government concept from "workclass" attribute of adult dataset

No	Education	Marital	Occupation	Country	Number of instances
0	Intermediate	ANY	ANY	ANY	3454
1	ANY	ANY	ANY	America	786
2	Advanced	ANY	ANY	Asia	30
3	Advanced	ANY	ANY	Europe	17
4	Basic	Married-spouse	Services	Europe	1
5	Advanced	Married-spouse	Services	Antartica	1

Table II Ruleset R1 for learning non government concept from "workclass" attribute of adult dataset

No	Education	Marital	Occupation	Country	Number of instances
0	7th-8th	Widowed	Tools	United-states	1
1	HS-grad	Never-married	ANY	United-states	4
2	HS-grad	Married-civ-spouse	ANY	ANY	5
3	Assoc-adm	Married-civ-spouse	Tools	United-states	1
4	Some-college	Married-civ-spouse	ANY	United-states	2
5	Some-college	Married-spouse-absent	Tools	United-states	1

Figure 6 Screen display for AOI-HEP application



AOI - High Emerging Patterns

Time Start :

Time Finish :

Time :

☑ TSHEP ☑ SOHEP ☑ TOHEP

Attribute Threshold : 0

Rules Threshold : 0

Growth Rate Threshold : 0

○ Frequent patterns ○ Similar patterns

Overall, the results for running the AOI-HEP application for four experimental datasets can be seen in table 5 where the adult dataset has two TSHEP, four SOHEP and no TOHEP, the breast cancer dataset has no TSHEP, two SOHEP and no TOHEP, whilst the census dataset has two TSHEP, six SOHEP and no TOHEP and the IPUMS dataset has no TSHEP, four SOHEP and two TOHEP. Due to page's limit for paper publishing, then we limit only to adult dataset which has two TSHEP and four SOHEP as shown in tables 3 and 4 respectively. Tables 3 and 4 are outputs which are stated in line number 17 HEP algorithm in figure 3. Tables 3 and 4 have number of growth rates grouped either as TSHEP or TOHEP, where growth rate is discrimination between rulesets R2 and R1 as mentioned in equation 2.

Tables 3 and 4 have position rulesets R2(X) and R1(Y), support D2(X), support D1(Y), Growth rate, HEP pattern and HEP%, where parameters X and Y, R2(X), R1(Y), support D2(X) and support D1(Y) refer to equation 2. Position ruleset R2(X) and R1(Y) in tables 3 and 4 refer to position tuple (rule) in tables 1 and 2 respectively since they are from the same dataset (adult dataset), where R2(X) and R1(Y) for learning government and non government concepts respectively from the same "workclass" attribute of adult dataset. Table 5 shows SLV and growth rate values (SLV/growth rate) with equations 1 and 2 whilst figure 7 shows the SLV values composition.

Table III TSHEP from adult dataset

No	R2(X	R1(Y)	Support D2(X)	Support D1(Y)	GR	HEP Pattern	HEP %
1	0	3	3454/4289=0.80532	1/14=0.07143	11.27442	0.5+0.5+0.5+0.5=2	100%
2	0	5	3454/4289=0.80532	1/14=0.07143	11.27442	0.5+0.5+0.5+0.5=2	100%

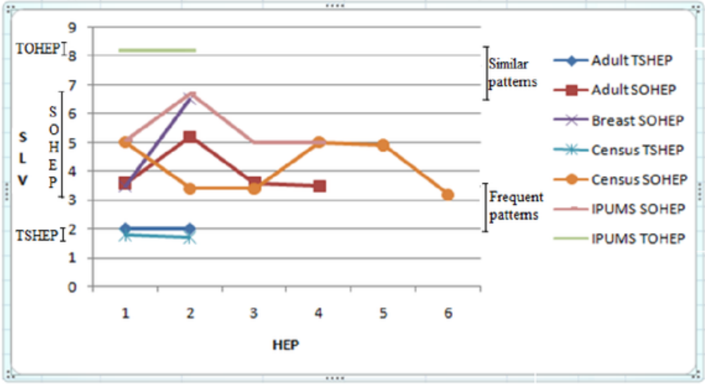
Table IV SOHEP from adult dataset

No	R2(X	R1(Y)	Support D2(X)	Support D1(Y)	GR	HEP Pattern	HEP %
1	0	1	3454/4289=0.80532	4/14=0.28571	2.81861	0.5+0.5+2.1+0.5=3.6	75%+25%
2	0	2	3454/4289=0.80532	5/14=0.35714	2.25488	0.5+0.5+2.1+2.1=5.2	50%+50%
3	0	4	3454/4289=0.80532	2/14=0.14286	5.63721	0.5+0.5+2.1+0.5=3.6	75%+25%
4	1	2	786/4289=0.18326	5/14=0.35714	0.51313	0.5+0.5+2.1+0.4=3.5	75%+25%

Table V Composition SLV values and Growth rate for four experimental datasets

Adult		Breast Cancer	Census		IPUMS	
TSHEP	SOHEP	SOHEP	TSHEP	SOHEP	SOHEP	TOHEP
2/11.274	3.6/2.818	3.5/10.302	1.8/0.697	5/0.088	5.1/0.629	8.2/1.530
2/11.274	5.2/2.255	6.5/0.677	1.7/0.057	3.4/0.275	6.7/3.466	8.2/0.446
0	3.6/5.637	0	0	3.4/0.123	5/0.851	0
0	3.5/0.513	0	0	5/0.379	5/0.261	0
0	0	0	0	4.9/0.272	0	0
0	0	0	0	3.2/0.009	0	0

Figure 7. Composition SLV values for four experimental datasets



A. *Experimental mining TSHEP, SOHEP and TOHEP*

The graph in figure 7 shows the consistency between minimum and maximum SLV values for TSHEP, SOHEP and TOHEP in figure 5, where TSHEP, SOHEP and TOHEP have small, medium and high SLV values respectively. The graph in figure 7 shows the position TSHEP at the bottom of graph (below SLV=2) which indicates that TSHEP have small SLV values. Firstly, TSHEP for adult and census datasets are consistent where minimum and maximum SLV value between m*0.4 (SLV=4*0.4=1.6) and m*0.5 (SLV=4*0.5=2). The SOHEP position in the middle of the graph (between SLV=3 and SLV=7) indicates that SOHEP have medium SLV values and secondly, SOHEP for all four experimental datasets are consistent where minimum and maximum SLV value between (m-1)*0.4+2 (SLV=(4-1)*0.4+2=3.2) and (m-1)*2.1+0.5 (SLV=(4-1)*2.1+ 0.5= 6.8). Lastly, TOHEP position at the upper part of the graph (above SLV=8) indicates that TOHEP have high SLV values and thirdly, TOHEP for IPUMS dataset is consistent where minimum and maximum SLV value between m*2 (SLV= 4*2=8) and m*2.1 (SLV=4*2.1=8.4).

B. *Experimental mining frequent pattern*

The graph in figure 7 shows the consistency between minimum and maximum SLV values for frequent pattern in figure 5, where position frequent pattern between TSHEP and SOHEP. Frequent pattern for all four experimental datasets are consistent where minimum and maximum SLV value between (m-1)*0.5+0.4 (SLV=(4-1)*0.5+0.4=1.9) and (m-1)*0.5+2.1 (SLV=(4-1)*0.5+ 2.1= 3.6). From running results of AOI-HEP application for four experimental datasets in tables 5, there are nine candidate frequent patterns based on minimum and maximum SLV values between 1.9 and 3.6. However, only five frequent patterns as shown in table 6 which fulfilled frequent pattern with strong discrimination power where having large growth rate and support in target (D2) dataset as mentioned in sub section 5.2. For example of frequent pattern which did not fulfil as strong discrimination power is the fourth result in table 4, where support in target (D2) dataset (0.18326) is lower than in contrasting (D1) dataset (0.35714), even it has SLV value=3.5 which fulfilled as frequent pattern and furthermore it has small growth rate (0.513). Tables between 7 and 11 show ruleset relation between rulesets R_i^1 and $R_j^2$for each frequent pattern in table 6.

Table VI Frequent patterns from four experimental datasets

No	Dataset	HEP	= Growth rate	SLV
1	Adult	TSHEP	= 11.2744	0.5+0.5+0.5+0.5=2
2	Adult	TSHEP	= 11.2744	0.5+0.5+0.5+0.5=2
3	Adult	SOHEP	= 2.81861	0.5+0.5+2.1+0.5=3.6
4	Adult	SOHEP	= 5.63721	0.5+0.5+2.1+0.5=3.6
5	Breast cancer	SOHEP	= 10.30286	2.0+0.5+0.5+0.5=3.5

Table VII TSHEP in adult dataset for rulesets R_3^1 to R_0^2 with GR=(3454/4289)/(1/14) = 0.80532/0.07143 = 11.27442

Rulesets	Education	Marital	Occupation	Country	Instances
R_0^2	Intermediate	ANY	ANY	ANY	3454
R_3^1	Assoc-adm	Married-civ-spouse	Tools	United-states	1
LV	0.5	0.5	0.5	0.5	

Table VIII TSHEP in adult dataset for rulesets R_5^1 to R_0^2 with GR=(3454/4289)/(1/14) = 0.80532/0.07143 = 11.27442

Rulesets	Education	Marital	Occupation	Country	Instances
R_0^2	Intermediate	ANY	ANY	ANY	3454
R_5^1	Some-college	Married-spouse-absent	Tools	United-states	1
LV	0.5	0.5	0.5	0.5	

Table IX Frequent subsumptionSOHEP in adult dataset for rulesets R_1^1 to R_0^2 with GR=(3454/4289)/(4/14) = 0.80532/0.28571 = 2.81861

Rulesets	Education	Marital	Occupation	Country	Instances
R_0^2	Intermediate	ANY	ANY	ANY	3454
R_1^1	HS-Grad	Never-married	ANY	United-states	4
LV	0.5	0.5	2.1	0.5	

Table X Frequent subsumptionSOHEP in adult dataset for rulesets R_4^1 to R_0^2 with GR=(3454/4289)/(2/14) = 0.80532/0.14286=5.63721

Rulesets	Education	Marital	Occupation	Country	Instances
R_0^2	Intermediate	ANY	ANY	ANY	3454
R_4^1	Some-college	Married-civ-spouse	ANY	United-states	2
LV	0.5	0.5	2.1	0.5	

Table XI Frequent subsumptionSOHEP in breast cancer dataset for rulesets R_4^1 to R_2^2 with GR=(19/533)/(1/289)= 0.03565/0.00346=10.30206

Rulesets	Cell Size	Cell Shape	Bare Nuclei	Normal Nucleoli	Instances
R_2^2	VeryLargeSize	ANY	ANY	ANY	19
R_4^1	VeryLargeSize	smallShape	MediumNuclei	VeryLargeNucleoli	1
LV	2.0	0.5	0.5	0.5	

Here are listing of discriminant rule for each of the frequent pattern in table 6 which are detailed between tables 7 and 11:

a. There are 11.2744 growth rate for TSHEP adult dataset with 80.53% frequent pattern in government workclass with intermediate education and 7.14% infrequent pattern in non government workclass with assoc-adm education, married-civ-spouse marital status, tools occupation and from the United States.

b. There are 11.2744 growth rates for TSHEP adult dataset with 80.53% frequent pattern in government workclass with an intermediate education and 7.14% infrequent pattern in non government workclass with some college education, married-spouse-absent marital status, tools occupation and from the United States.

c. There are 2.81861 growth rates for SOHEP adult dataset with 80.53% frequent pattern in government workclass with an intermediate education and 28.57% infrequent pattern in non government workclass with HS-Grad education, Never-married marital status and from the United States.

d. There are 5.63721 growth rates for SOHEP adult dataset with 80.53% frequent pattern in government workclass with intermediate education and 14.28% infrequent pattern in non government workclass with some college education, married-civ-spouse marital status and from the United States.

e. There are 10.30206 growth rates for SOHEP breast cancer dataset with 3.56% frequent pattern in AboutAverClump "clump thickness" with VeryLargeSize "Cell Size" and 0.34% infrequent pattern in AboveAverClump "clump thickness" with VeryLargeSize "Cell Size", SmallShape "Cell shape", mediumNuclei "Bare Nuclei" and VeryLargeNucleoli "Normal Nucleoli".

Experimental mining similar pattern

The graph in figure 7 shows the consistency between minimum and maximum SLV values for similar pattern in figure 5, where position similar pattern between SOHEP and TOHEP. Similar pattern for all four experimental datasets are consistent where minimum and maximum SLV value between (m-1)*2+0.4 (SLV=(4-1)*2+0.4=6.4) and (m-1)*2.1+2 (SLV=(4-1)*2.1+ 2=8.3). From running results of AOI-HEP application for four experimental datasets in tables 5, there are four similar patterns which are fulfilled minimum and maximum SLV values between 6.4 and 8.3, as shown in table 12. Tables between 13 and 16 show ruleset relation between rulesets R_i^1 and R_j^2 for each similar pattern in table 12.

Table XII Similar patterns from four experimental datasets

No	Dataset	HEP	= Growth rate	SLV
1	IPUMS	TOHEP	= 1.530	2.1+2.0+2.0+2.1=8.2
2	IPUMS	TOHEP	= 0.446	2.1+2.0+2.0+2.1=8.2
3	Breast cancer	SOHEP	= 0.67777	2.0+2.0+0.4+2.1=6.5
4	IPUMS	SOHEP	= 3.46636	2.1+2.0+0.5+2.1=6.7

Table XIII TOHEP in IPUMS dataset for rulesetsR_3^1 to R_4^2 with GR=(6356/140124)/(2296/77453)= 0.045/0.029= 1.530

Rulesets	Relateg	Educrec	Migrat5g	Tranwork	Instances
R_4^2	ANY	Primary School	Not-known	ANY	6356
R_3^1	ANY	Primary School	Not-known	ANY	2296
LV	2.1	2.0	2.0	2.1	

Table XIV TOHEP in IPUMS dataset for rulesets R_4^1 to R_5^2 with GR=(4603/140124)/(5706/77453) = 0.033/0.074=0.446

Rulesets	Relateg	Educrec	Migrat5g	Tranwork	Instances
R_5^2	ANY	College	Not-known	ANY	4603
R_4^1	ANY	College	Not-known	ANY	5706
LV	2.1	2.0	2.0	2.1	

Table XV Frequent overlapping SOHEP in breast cancer dataset for rulesets R_3^1 to R_5^2 with GR=(5/533)/(4/289) = 0.00938/0.01384 =0.67777

Rulesets	Cell Size	Cell Shape	Bare Nuclei	Normal Nucleoli	Instances
R_5^2	largeSize	VeryLargeShape	VeryLargeNuclei	ANY	5
R_3^1	largeSize	VeryLargeShape	ANY	ANY	4
LV	2.0	2.0	0.4	2.1	

Table 16. Frequent overlapping SOHEP in IPUMS dataset for rulesets R_5^1 to R_1^2 with GR=(7632/140124)/(1217/77453) = 0.05447/0.01571=3.46636

Rulesets	Relateg	Educrec	Migrat5g	Tranwork	Instances
R_1^2	ANY	Secondary School	ANY	ANY	7632
R_5^1	ANY	Secondary School	Not-known	ANY	1217
LV	2.1	2.0	0.5	2.1	

Here are listing of discrimination rule for each of the similar pattern in table 12 which are detailed between tables 13 and 16:

a. There are 1.53 growth rates similar patterns for TOHEP IPUMS dataset with 4.5% unmarried "marital status" and 2.9% Married "marital status" with a similar pattern in the Primary School education and Not-known "Migration status".

b. There are 0.446 growth rates similar patterns for TOHEP IPUMS dataset with 3.3% unmarried "marital status" and 7.4% Married "marital status" with a similar pattern in College education and Not-known "Migration status".

c. There are 0.6777 growth rates similar patterns for SOHEP breast cancer dataset with 0.938% AboutAverClump "clump thickness" and 1.384% AboveAverClump "clump thickness" with similar pattern largeSize "Cell Size" and VeryLargeShape "Cell shape".

d. There are 3.46636 growth rates similar patterns for SOHEP IPUMS dataset with 5.447% unmarried "marital status" and 1.571% Married "marital status" with Not-known "Migration status" and there is a similar pattern in Secondary School education.

VIII. Conclusion

AOI-HEP has been successfully implemented using four large real datasets from UCI machine learning repository and discovered TSHEP, SOHEP, TOHEP, frequent and similar patterns. The experiments showed that there are five frequent and four similar patterns from twenty two HEP, where all frequent patterns and two similar patterns have strong discrimination rules with growth rate values between 1.530 and 11.2744 respectively. Since AOI-HEP can strongly discriminate high-level data, assuredly AOI-HEP can be implemented to discriminate datasets such as finding bad and good customers for banking loan systems or credit card applicants and etc. Moreover, since AOI-HEP can mine similar patterns, certainly AOI-HEP can be implemented to mine similar patterns, for instance, mining similar customer loan patterns and etc.

AOI-HEP can be extended to learn other knowledge patterns such as characteristic, classification, data evolution regularities, association and cluster description. Moreover, AOI-HEP knowledge discovery can be extended to mine disjoint as dissimilar pattern, inverse the discovery learning, learning from more than two datasets and learning multidimensional view. In the future, this AOI-HEP should be compared with current data mining technique in order to improve the performance and patterns which can be mined.

IX. References

[1] Agrawal, R., Imielinski, T. & Swami, A. (1993). Mining association rules between sets of items in large databases. ACM SIGMOD Rec, 22(2), 207-216.

[2] Beneditto, M.E.M.D & Barros, L.N.D. (2004). Using Concept Hierarchies in Knowledge Discovery. Lectures Notes in Computer Science, 3171, 255-265.

[3] Chen, M.S., Han, J. & Yu, P.S. (1996). Data Mining: An Overview from a Database Perspective. IEEE Trans. on Knowledge and Data Engineering, 8(6), 866-883.

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

[5] Danger, R., Ruiz-Shulcloper, J. & Llavori, R.B. (2004). Objectminer: A new approach for Mining Complex objects. In Proceedings of the 6th international conference on Enterprise Information Systems (ICEIS '04), 42-47.

[6] Dong, G. & Li, J. (1999). Efficient mining of emerging patterns: discovering trends and differences. In Proceedings of the 5th ACM SIGKDD international Conference on Knowledge Discovery and Data Mining, 43-52.

[7] Dong, G. & Li, J. (2005). Mining border description of emerging patterns from dataset pairs. Journal of Knowledge and Information Systems, 8(2), 178-202.

[8] Fan, H. & Ramamohanarao, K. (2003). A Bayesian approach to use emerging patterns for classification. In Proceedings of the 14th Australasian database conference (ADC '03), 17, 39-48.

[9] Frank, A. & Asuncion, A. (2010). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[10] Garcia-Borroto, M., Martinez-Trinidad, J.F. & Carrasco-Ochoa, J.A. (2010). New Emerging Pattern Mining Algorithm and Its Application in Supervised Classification. In proceedings of the 14th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD 2010), 150-157.

[11] Han, J., Cai,Y. & Cercone, N. (1992). Knowledge discovery in databases: An attribute-oriented approach. In Proceedings of the 18th International Conference Very Large Data Bases, 547-559.

[12] Han, J., Cheng, H., Xin, D. & Yan, X. (2007). Frequent pattern mining: current status and future directions. Data Mining Knowledge Discovery, 15(1), 55-86.

[13] Han. J., Pei, J., Yin, Y. and Mao, R. (2004). Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach. Data Mining Knowledge Discovery, 8(1), 53-87.

[14] Muyeba, M.K., Crockett, K. & Keane, J.A. (2011). A hybrid interestingness heuristic approach for attribute-oriented mining. In proceedings of the 5th KES International Conference on Agent and multi-agent systems: technologies and applications (KES-AMSTA 2011), 414-424.

[15] Muyeba, M.K, Crockett, K., Wang, w. & Keane, J.A.(2014). A hybrid heuristic approach for attribute-oriented mining. Decision Support Systems, 57,139-149.

[16] Ramamohanarao, K., Bailey, J. & Fan, H. (2005). Efficient Mining of Contrast Patterns and Their Applications to Classification. In Proceedings of the 3rd International Conference on Intelligent Sensing and Information Processing (ICISIP 2005), 39-47.

[17] Rodriguez-Gonzalez, A. Y. , Martinez-Trinidad, J.F., Carrasco-Ochoa, J.A. & Ruiz-Shulcloper, J. (2008). Mining Frequent Similar Patterns on Mixed Data. In Proceedings of the 13th Iberoamerican congress on Pattern Recognition: Progress in Pattern Recognition, Image Analysis and Applications (CIARP '08), 136-144.

[18] Schubert, S. & Lee, T. (2011). Time Series Data Mining with SAS@ Enterprise minerTM. In proceedings of the SAS@ Global Forum.

[19] Sherhod, R., Gillet, V. J., Judson, P. N. & Vessey, J. D. (2012). Automating Knowledge Discovery for Toxicity Prediction Using Jumping Emerging Pattern Mining. J. Chem. Inf. Model, 52(11), 3074−3087.

[20] Sherhod, R., Judson, P.N., Hanser, T., Vessey, J.D., Webb, S.J. &Gillet, V.J. (2014). Emerging Pattern Mining to Aid Toxicological Knowledge Discovery. J. Chem. Inf. Model, 54(7), 1864-1879.

[21] Warnars, S. (2012). Attribute Oriented Induction of High-level Emerging Patterns. In Proceedings of IEEE the International Conference on Granular Computing (GrC 2012), 525-530.

[22] Warnars, S. (2014). Mining Frequent Pattern with Attribute Oriented Induction High level Emerging Pattern (AOI-HEP). In Proceedings of IEEE the 2nd International Conference on Information and Communication Technology (IEEE ICoICT 2014), 144-149.

Check 38. Mining Frequent and Similar Patterns with Attribute Oriented Induction High Level Emerging.pdf

ORIGINALITY REPORT

22%	8%	17%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Spits Warnars. "Attribute Oriented Induction High Level Emerging Pattern (AOI-HEP) future research", Proceedings of International Conference on Information, Communication Technology and System (ICTS) 2014, 2014 Publication	8%
2	Submitted to Universitas Esa Unggul Student Paper	3%
3	e-space.mmu.ac.uk Internet Source	3%
4	Submitted to Florida Polytechnic University Student Paper	2%
5	Harco Leslie Hendric Spits Warnars, Ford Lumban Gaol, Yaya Heryadi, Agung Trisetyarso, Horacio Emilio Perez Sanchez. "Mining Association rule with Attribute Oriented Induction High level Emerging Pattern (AOI-HEP) data mining technique", 2018 Second International Conference on Inventive	1%

	Communication and Computational Technologies (ICICCT), 2018 Publication	
6	Harco Leslie Hendric Spits Warnars, Ford Lumban Gaol, Nesti Fronika Sianipar, Bahtiar Saleh Abbas, Horacio Emilio Perez Sanchez. "Easy understanding for mining discriminant itemset with emerging patterns", 2017 International Conference on Applied Computer and Communication Technologies (ComCom), 2017 Publication	1 %
7	journal.uad.ac.id Internet Source	1 %
8	Submitted to Federal University of Technology Student Paper	1 %
9	ccc.inaoep.mx Internet Source	1 %
10	link.springer.com Internet Source	1 %
11	citeseerx.ist.psu.edu Internet Source	1 %
12	Métivier, Jean-Philippe, Alban Lepailleur, Aleksey Buzmakov, Guillaume Poezevara, Bruno Crémilleux, Sergei Kuznetsov, Jérémie Le Goff, Amédéo Napoli, Ronan Bureau, and	1 %

Bertrand Cuissart. "Discovering structural alerts for mutagenicity using stable emerging molecular patterns", Journal of Chemical Information and Modeling, 2015.

Publication

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		