Jurnal
Teknologi

Full Paper

MINING SIMILAR PATTERN WITH ATTRIBUTE ORIENTED INDUCTION HIGH LEVEL EMERGING PATTERN (AOI-HEP) DATA MINING TECHNIQUE

Harco Leslie Hendric Spits Warnars[a*], Nizirwan Anwar[b], Richard Randriatoamanana[c], Horacio Emilio Perez Sanchez[d]

[a]Department of Computer Science, Binus Graduate Program, Bina Nusantara University, Jakarta 11530, Indonesia
[b]Faculty of Computer Science, Esa Unggul University, Jakarta 11510, Indonesia
[c]Institut de Calcul Intensif, Ecole Centrale de Nantes, Nantes 44321, France
[d]Structural Bioinformatics and High Performance Computing Research Group (BIO-HPC), Universidad Católica de Murcia (UCAM), Guadalupe 30107, Spain

Article history
Received
21 March 2016
Received in revised form
28 September 2017
Accepted
21 June 2017

*Corresponding author
shendric@binus.edu

Abstract

AOI-HEP (Attribute Oriented Induction High Emerging Pattern) as new data mining technique has been success to mine frequent pattern and is extended to mine similar patterns. AOI-HEP is success to mine 3 and 1 similar patterns from IPUMS and breast cancer UCI machine learning datasets respectively. Meanwhile, the experiments showed that there was no finding similar patterns on adult and census UCI machine learning datasets. The experiments showed that finding AOI-HEP similar pattern in dataset is influenced by learning on chosen high level concept attribute in concept hierarchy and it is applied to AOI-HEP frequent pattern in previous research as well. The experiments chosed high level concept attributes such as workclass, clump thickness, means and marts for adult, breast cancer, census and IPUMS datasets respectively. In order to proof that the chosen high level concept attribute will influences the AOI-HEP similar pattern in dataset, then extended experiments were carried on and the finding were census dataset which had been none AOI-HEP similar pattern, had AOI-HEP similar pattern when learned on high level concept in marital attribute. Meanwhile, Breast cancer which had been had 1 AOI-HEP similar pattern, had none AOI-HEP similar pattern when learned on high level concept in attributes such as cell size, cell shape and bare nuclei. The 2 of 3 finding Similar patterns in IPUMS dataset have strong discriminant rule since having large growth rates such as 1.53% and 3.47%, and having large supports in target dataset such as 4.54% and 5.45 respectively. Moreover, there have small supports in contrasting dataset such as 2.96% and 1.57% respectively.

Keywords: Similar pattern, Data Mining, AOI-HEP, High Level Emerging Pattern, Attribute Oriented Induction, discriminant rule

1.0 INTRODUCTION

Attribute Oriented Induction High Level Emerging Pattern (AOI-HEP) is combination of Emerging Pattern (EP) algorithms [1, 13, 14, 17] and Attribute Oriented Induction (AOI) characteristic rule [2, 15, 16]. AOI-HEP as a new coming data mining technique has opportunity to be explored in order to find such as inversed discovery learning, learning more than 2 datasets, multidimensional views, learning other knowledge rules and so on [3]. AOI-HEP has been success to mine frequent pattern where the target dataset subsumed to contrasting dataset or target dataset has superset rule/frequent pattern and contrasting dataset has subset rule/infrequent pattern [4, 5, 6]. From this frequent pattern we can create strong/sharp discrimination rules where having large growth rate and support in target dataset, include small support in contrasting dataset [9, 10, 11, 12].

The similar patterns are interested to be explored since with the similarity we can find the equality pattern which present similar pattern behavior. Similar pattern can be used for prediction purposes in many fields such as banking, education and so on. In banking system, similar pattern can be used to combat fraud and find similar pattern for new credit card user if they can have ability as customer without delaying or fail to pay their credit card installment. In education field, similar pattern can be used to predict which one suitable study program for student candidate, what the student behaviour and probability to finish their study in suitable time and with high score.

This AOI-HEP similar pattern will also refer to AOI-HEP frequent pattern where having strong/sharp discrimination rules with huge growth rate and support in target dataset, including small support in contrasting dataset [7, 8, 9, 10]. AOI-HEP similar pattern will be mined from dataset where the number of attributes similarity are full or dominant/frequent and the number of similarity with ANY values are infrequent. It has been known that the ANY value has been known as no meaning value and should be neglected.

2.0 METHODOLOGY

AOI-HEP algorithm is consist of 2 sub algorithms such as AOI characteristic rule[4] and HEP similar pattern algorithms as seen on Figures 1 and 2 respectively. The process of AOI-HEP algorithm will be executed with 4 steps and they are:

1. Input dataset and concept hierachy file.
2. Split dataset into sub datasets as many as number of high level concepts in chosen concept hierarchy's attribute.
3. Running AOI characteristic rule algorithm for each of sub dataset.
4. Running HEP similar pattern algorithm for all sub datasets.

The steps 1 and 2 are complement which prepare the input dataset and concept hierarchy files and split dataset into number of sub datasets based on chosen concept hierarchy's attribute. Number of attributes in sub dataset will less than number of attributes in each dataset since the chosen attribute will not be included.

This AOI Characteristic rule algorithm in Figure 1 is executed for each sub dataset which were splitted from input dataset file as explained in previous section. If there are 2 sub datasets then this algorithm will be run in 2 times and if there are 5 sub datasets then this algorithm will be run in 5 times. This AOI characteristic rule algorithm will need input such as attribute and rule thresholds where attribute threshold as limit control of number of distinct value for each attribute while rule threshold as limit control of number of final rules respectively. Attribute and rule thresholds are shown in AOI characteristic rule algorithm on Figure 1, in line number 2 and 9 respectively.

Input: dataset, concept hierarchies, attribute_threshold, rule_threshold
Output: characteristic rule of learning task, $\{R_i^1\}$, $\{R_j^2\}$, num_attr, |D2|,|D1|

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 attr Ai
4. { 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. }

Figure 1 AOI Characteristic rule algorithm

AOI characteristic rule algorithm will have outputs such as characteristic rule result either as $\{R_i^1\}$ or $\{R_j^2\}$ and total number of each sub dataset either as |D2| or |D1| for just in case if there are 2 sub datasets. If there are 3 sub datasets then it will have 1 more characteristic rule result such as $\{R_k^3\}$ and total number of sub dataset such as |D3|.

This AOI characteristic rule algorithm will looping as many as number of attributes in sub datasets as shown in line number 1 in order to check number distinct value in every attribute until <= attribute threshold as shown in line number 2. While number distinct value in every attribute > attribute threshold, then process between line number 3 and 7 are executed. The line number 3 will check if each attribute sub dataset has higher level concept in concept hierarchy file and if there is no higher-level concept then the attribute will be removed as will execute line number 4. Meanwhile, if the attribute has higher-level concept then the value of attribute will be changed with the value of corresponding higher-level concept in concept hierarchy file. Line number 6 will merge some identical tuples and if it is so then probably the number of tuples will be reduced.

The line number between 9 and 12 will check if number tuple of characteristic rule result are > rule threshold as shown in line number 9. If number tuple of characteristic rule result are > rule threshold, then line number between 10 and 11 are executed. Line number 10 will execute specific algorithm which reduce number tuple of characteristic rule result and it has been discussed in previous research [8]. The line number 11 is identical with line number 6 which will merge some identical tuples and if it is so then probably the number of tuples will be reduced until <= rule threshold as seen at line number 9.

Input :{R_i^1} , {R_j^2}, num_attr, |D2| , |D1| , GR_threshold
Output: R_j^2, |R_j^2| , (|R_j^2|/|D2|),R_i^1, |R_i^1|, (|R_i^1|/|D1|), HEP_GR

1. {While(noAllANY(R_i^1))
2. {While(noAllANY(R_j^2))
3. {Sim=0, any=0
4. for x=1 to num_attr
5. {If(R_i^1[x]== R_j^2[x] and R_i^1[x] == "ANY")
6. {Sim++; any++;}
7. If(R_i^1[x]== R_j^2[x] and R_i^1[x]!="ANY")
8. Sim++;
9. }
10. If (Sim>=num_attr-1 and any<num_attr-1)
11. {HEP_GR=(|R_j^2|/|D2|)/(|R_i^1|/|D1|)
12. If HEP_GR > GR_threshold
13. Print R_j^2, |R_j^2|,(|R_j^2|/|D2|), R_i^1, |R_i^1|, (|R_i^1|/|D1|),HEP_GR
14. }
15. }
16. }
17. }

Figure 2 HEP Similar pattern algorithm

Meanwhile, HEP similar pattern algorithm which is shown in Figure 2 will implement similar pattern cartesian product between characteristic rule result of sub datasets where each rule in characteristic rule result in one sub dataset will be checked their similarity with each rule in characteristic rule result from another sub dataset. The implementation of similar pattern cartesian product is shown in HEP similar pattern algorithm, Figure 2 at line number 4, 5, 6, 7, 8 and 9, where the similarity will be examined attribute by attribute. The process of similar pattern cartesian product only interest with rare ANY value as shown in Figure 2 at line number 7, where ANY value is identical with unmeaning value. Indeed, the similar pattern with all ANY value is uninteresting.

Furthermore, line number between 10 and 14 in Figure 2 will be executed if the number of attributes similarity are dominant/frequent (sim>=num_attr-1) and the number of similarity with ANY values are infrequent (any<num_attr-1). Line number 11 will executed High emerging pattern (HEP) growth rate where support of target/positive sub dataset is divided with support of background/negative/contrasting sub dataset. Moreover, as shown in line number 12 and 13, the similar pattern will be printed if HEP > growth rate threshold and the similar pattern will show both rules in characteristic rule result from one and another sub dataset.

3.0 RESULTS AND DISCUSSION

For AOI-HEP algorithm experiments will use dataset files which taken from UCI machine learning repository as free public dataset and they are 4 datasets like adult, breast cancer, census and IPUMS datasets with 48,842; 569; 2,458,285 and 256,932 instances respectively [7]. In the experiments, the AOI characteristic rule algorithm in Figure 1 will divided each of dataset into 2 sub datasets which based on learning the high-level concept in 1 of their chosen attributes which only discriminate only 2 concepts in these experiments. These 4 Datasets have each chosen attribute like worklass for adult dataset, clump thickness for breast cancer dataset, means for census dataset and marst for IPUMS dataset.

AOI characteristic rule algorithm were executed with given number attribute and rule thresholds such as 6. The number 6 was selected based on preliminary experiments, in order to running the experiments and finding the suitable number for attribute and rule threshold. The experiments found that similar pattern is numerous when did with attribute thresholds 4,5 and 6, and rules thresholds 5,6,7,8,9 and 10.

Each dataset creates a concept hierarchy which is built from 5 chosen attributes with a minimum concept level of 3. Adult dataset creates concept hierarchy which is built from 5 chosen attributes like workclass, education, marital-status, occupation, and native-country. Breast cancer dataset creates concept hierarchy which is built from 5 chosen attributes such as clump thickness, cell size, cell shape, bare nuclei and normal nucleoli. Census dataset creates concept hierarchy which is built from 5 chosen attributes like class, marital status, means, relat1 and yearsch. IPUMS dataset creates concept hierarchy which is built from 5 chosen attributes such as relateg, marst, educrec, migrat5g and tranwork.

Table 1 ruleset R2 of learning unmarried concept from "marst" attribute of ipums dataset with total number of tuples 140124

No	Relateg	Educrec	Migrat5g	tranwork	Number of tuples /Support
0	ANY	ANY	ANY	ANY	108026 /77.09%
1	ANY	Secondary School	ANY	ANY	7632 / 5.45%
2	ANY	Primary School	Other-state	ANY	10332 / 7.37%
3	ANY	Reception School	Other-state	ANY	3175 / 2.27%
4	ANY	Primary School	Not-Known	ANY	6356 / 4.54%
5	ANY	College	Not-Known	ANY	4603 / 3.28%

Table 2 ruleset R1 of learning married concept from "marst" attribute of ipums dataset with total number of tuples 77453

No	Relateg	Educrec	Migrat5g	tranwork	Number of tuples /Support
0	ANY	ANY	ANY	ANY	56087 / 72.41%
1	ANY	Basic	Moved	ANY	6707 / 8.66%
2	ANY	Academy	Not-Known	ANY	5440 / 7.02%
3	ANY	Primary School	Not-Known	ANY	2296 / 2.96%
4	ANY	College	Not-Known	ANY	5706 / 7.37%
5	ANY	Secondary School	Not-Known	ANY	1217 / 1.57%

Table 3 Ruleset R2 for learning AboutAverClump concept from "clump thickness" attribute of breast cancer dataset with total number of tuples 533

No	Cell Size	Cell Shape	Bare Nuclei	Normal Nucleoli	Number of tuples /Support
0	ANY	ANY	ANY	ANY	496 /93.06%
1	mediumSize	smallShape	ANY	aboutAverNucleoli	3/ 0.56%
2	VeryLargeSi:	ANY	ANY	ANY	19 / 3.56%
3	mediumSize	largeShape	aboveAverNuclei	ANY	7 / 1.31%
4	VeryLargeSi:	mediumShape	ANY	VeryLargeNucleoli	3 / 0.56%
5	LargeSize	VeryLargeShape	VeryLargeNuclei	ANY	5 / 0.94%

Table 4 Ruleset R1 for learning AboveAverClump concept from "clump thickness" attribute of breast cancer dataset with total number of tuples 289

No	Cell Size	Cell Shape	Bare Nuclei	Normal Nucleoli	Number of tuples /Support
0	ANY	ANY	ANY	ANY	277 /95.85%
1	smallSize	largeShape	VeryLargeNuclei	VeryLargeNucleoli	1/ 0.35%
2	mediumSize	ANY	ANY	AboveAverNucleoli	5 / 1.73%
3	largeSize	VeryLargeShape	ANY	ANY	4 / 1.38%
4	VeryLargeSiz	VeryLargeShape	mediumNuclei	VeryLargeNucleoli	1 / 0.35%
5	LargeSize	smallShape	mediumNuclei	largeNucleoli	1 / 0.35%

Each of dataset was divided into 2 sub datasets based on learning the high level concept in 1 of their attributes. Learning the high- level concept in 1 of their 5 chosen attributes will split each dataset become 2 sub datasets such as:

1. Adult dataset was learned on workclass attribute, which differentiate between the "government" (4289 instances) and "non government" (14 instances) concepts.
2. Breast cancer dataset was learned on "clump thickness" attribute, which differentiate between "aboutaverclump" (533 instances) and "aboveaverclump" (289 instances) concepts.
3. Census dataset was learned on means attribute, which differentiate between "green" (1980 instances) and "no green" (809 instances) concepts.
4. IPUMS dataset was learned on marst attribute which differentiate between "unmarried" (140124 instances) and "married" (77453 instances) concepts. Concept hierarchy of marst attribute which content concepts "unmarried" and "married" can be seen in Figure 4.

Finally, the AOI characteristic rule algorithm in Figure 1 produced characteristic rules for each of sub dataset with 6 tuples and since there are 4 datasets which divided into 8 sub datasets then there were 8 characteristic rules. After running this AOI characteristic rule algorithm, then only IPUMS and breast cancer datasets which have 3 and 1 finding similar patterns respectively. Tables 1 and 2 show the characteristic rules from IPUMS dataset whilst Tables 3 and 4 show the characteristic rules from breast cancer dataset. Other 4 tables as learning from Adult and Census datasets are not shown due to limitation of publication page and there is no finding similar patterns in these 2 datasets.

Table 1 shows characteristic rule for learning unmarried concept on attribute "marst" in IPUMS dataset with total number of tuples 140,124, whilst Table 2 shows characteristic rule for learning Married on attribute "marst" in IPUMS dataset with 77,453 tuples. Tables 1 and 2 have 6 characteristic rules with each number of tuples and support percentage where support=total number of tuples divide with number each characteristic rule. For example, first characteristic rule in Table 1 has support = 108,026 / 140,124=77.09 and first characteristic rule in Table 2 has support =56,087 / 77,453=72.41. Meanwhile, concept hierarchy in Figure 4 shows that unmarried and married concepts as high level concepts in concept hierarchy for attribute marst. Moreover, Table 3 shows characteristic rule for learning AboutAverClump concept on attribute "Clump Thickness" in breast cancer dataset, whilst Table 4 shows characteristic rule for learning AboveAverClump concept on attribute "Clump Thickness" in breast cancer dataset.

Similar like Tables 1and 2, then Tables 3 and 4 have 6 characteristic rules with each number of tuples and support percentage where support=total number of tuples divide with number each characteristic rule. For example, first characteristic rule in Table 3 has support = 496 / 533=93.06 and first characteristic rule in Table 4 has support =277 / 289=95.85.

Meanwhile, as mentioned before where ANY value is uniteresting, then the tuples with all ANY value is uninteresting as shown in first row in all Tables 1, 2, 3 and 4. Moreover, because of page limitation too and the result of experiment where only IPUMS and breast cancer datasets which have 3 and 1 similar patterns, then only IPUMS dataset's concept hierarchies will be shown between Figures 3 and 7, where each figure as concept hierarchy for IPUMS dataset chosen attributes such as relateg, marst, educrec, migrat5g and tranwork as shown in Figure 3, 4, 5, 6 and 7 respectively.

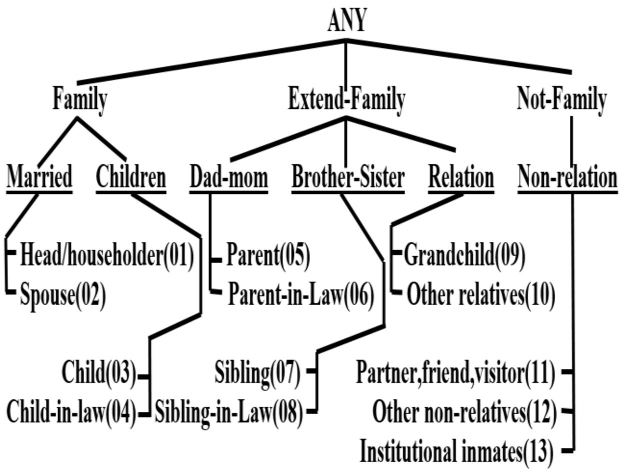

Figure 3 Concept hierarchy of relateg attribute of IPUMS dataset

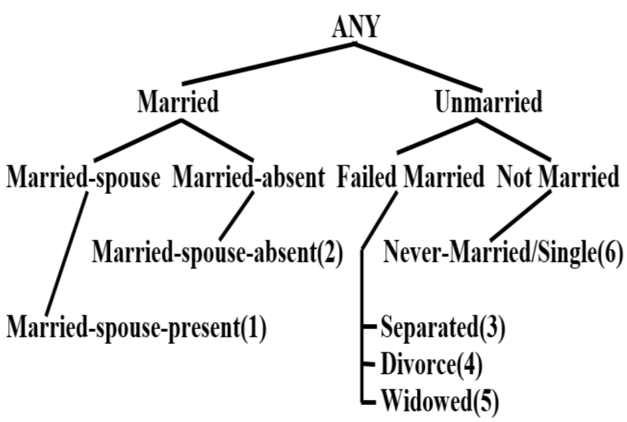

Figure 4 Concept hierarchy of marst attribute of IPUMS dataset

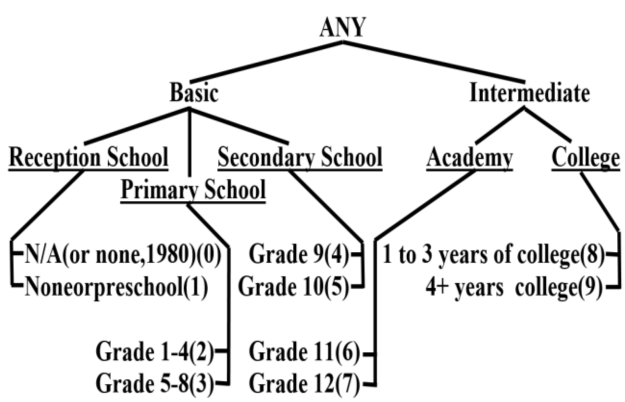

Figure 5 Concept hierarchy of educrec attribute of IPUMS dataset

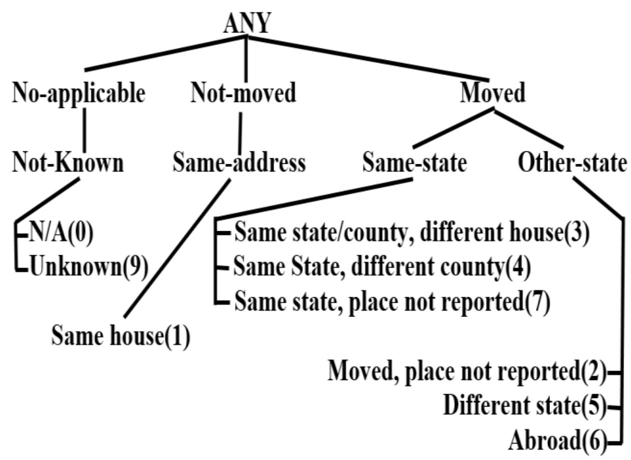

Figure 6 Concept hierarchy of migrat5g attribute of IPUMS dataset

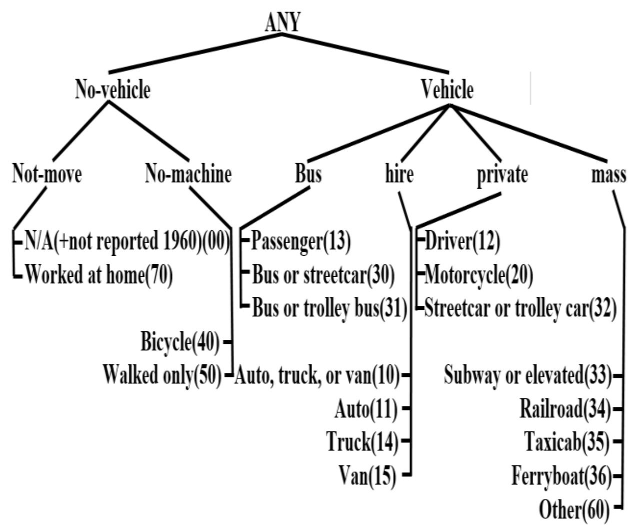

Figure 7 Concept hierarchy of Tranwork attribute of IPUMS dataset

Table 5 similar pattern of ipums dataset with rulesets R_3^1 to R_4^2 with Growth rate = (6356/140124)/ (2296/77453) = 0.045/0.0296 = 1.53

Rulesets	Relateg	Educrec	Migrat5g	Tranwork	tuples	Support
R_4^2	ANY	Primary School	Not-Known	ANY	6356	4.54%
R_3^1	ANY	Primary School	Not-Known	ANY	2296	2.96%
Growth Rate (GR) =						**1.53%**

Table 6 similar pattern of ipums dataset with rulesets R_4^1 to R_5^2 with Growth rate =(4603/140124)/ (5706/77453) = 0.0328/0.0737 = 0.45

Rulesets	Relateg	Educrec	Migrat5g	Tranwork	tuples	Support
R_5^2	ANY	College	Not-Known	ANY	4603	3.28%
R_4^1	ANY	College	Not-Known	ANY	5706	7.37%
Growth Rate (GR) =						**0.45%**

Table 7 similar pattern of ipums dataset with rulesets R_5^1 to R_1^2 with Growth rate =(7632/140124)/ (1217/77453) = 0.0545/0.0157 = 3.47

Rulesets	Relateg	Educrec	Migrat5g	Tranwork	tuples	Support
R_1^2	ANY	Secondary School	ANY	ANY	7632	5.45%
R_5^1	ANY	Secondary School	Not-Known	ANY	1217	1.57%
Growth Rate (GR) =						3.47%

Table 8 similar pattern of breast cancer dataset with rulesets R_3^1 to R_5^2 with Growth rate= (5/533) / (4/289) = 0.0094 / 0.0138 =0.68

Rulesets	Cell Size	Cell Shape	Bare Nuclei	Normal Nucleoli	tuples	Support
R_5^2	LargeSize	VeryLarge Shape	VeryLarg eNuclei	ANY	5	0.94%
R_3^1	LargeSize	VeryLarge Shape	ANY	ANY	4	1.38%
Growth Rate (GR) =						0.68%

After the running of AOI characteristic rule algorithm as shown in Figure 1, then HEP similar pattern algorithm as shown in Figure 2 will be executed by finding similar patter as cartesian product between characteristic rule result of sub datasets as shown Between Tables 1 and 4, where Table 1 will be compared with Table 2, and Table 3 will be compared with Table 4. Interestingly, there is no finding similar patterns when running this HEP Similar pattern algorithm on adult and census datasets. On other hand, there are 3 and 1 finding similar patterns when running the HEP Similar pattern algorithm on IPUMS and breast cancer datasets as shown between Tables 5 to 7 and Table 8 respectively.

Table 5 shows there is similarity for pattern ANY, Primary School, Not-known, ANY between ruleset number 4 in Table 1 and ruleset number 3 in Table 2. Table 6 shows there is similarity for pattern ANY, College, Not-known, ANY between ruleset number 5 in Table 1 and ruleset number 4 in Table 2. Moreover, Table 7 shows there is similarity for pattern ANY, Secondary School, ANY, ANY between ruleset number 1 in Table 1 and ruleset number 5 in Table 2. Table 8 shows there is similarity for pattern LargeSize, VeryLargeShape, ANY, ANY between ruleset number 5 in Table 3 and ruleset number 3 in Table 4.

Based on similar patterns between Tables 5 and 8, next are discrimination rule of each similar pattern:

1. Table 5 shows that IPUMS dataset has 1.53 growth rates similar patterns between unmarried and Married "marital status" with support 4.54% and 2.96% respectively for similarity pattern in the Primary School education and Not-known "Migration status".
2. Table 6 shows that IPUMS dataset has 0.45 growth rates similar patterns between unmarried and Married "marital status" with support 3.28% and 7.37% respectively for similarity pattern in the College education and Not-known "Migration status".
3. Table 7 shows that IPUMS dataset has 3.47 growth rates similar patterns between unmarried and Married "marital status" with support 5.45% and 1.57% of Not-Known "Migration status" respectively for similarity pattern in the SecondarySchool education.
4. Table 8 shows that Breast cancer dataset has 0.68 growth rates similar patterns for breast cancer dataset between AboutAverClump and AboveAverClump "clump thickness" with support 0.94% of VeryLargeNuclei "Bare Nuclei" and 1.38% respectively for similarity pattern in largeSize "Cell Size" and VeryLargeShape "Cell shape".

Discriminant rules number 1 and 3 are strong discriminant rule which is showed as strong discrimination power when they have large growth rates (1.53% and 3.47%) and supports in target (D2) datasets (4.54% and 5.45%). Furthermore, they have small supports in contrasting (D1) datasets (2.96% and 1.57%) where each of supports in the contrasting (D1) dataset is < than support in the target (D2) dataset. Meanwhile, discriminant rules numbers 2 and 4 are weak discriminant rules since have small growth rates (0.45% and 0.68%) and supports in target (D2) datasets (3.28% and 0.94%). Furthermore, they have large supports in contrasting (D1) datasets (7.37% and 1.38%) where each of supports in contrasting (D1) dataset is > than the support in target(D2) dataset.

The experimental upon 4 UCI machine learning public dataset like Adult, breast cancer, census and IPUMS datasets, show that IPUMS and breast cancer datasets are interested for AOI-HEP similar pattern since having AOI-HEP similar pattern finding. The AOI-HEP mining interest similar pattern for each dataset, is influenced by learning on high-level concept in one of chosen attribute. Adult, breast,cancer, census and IPUMS datasets learn on high level concept in "workclass", "clump thickness", "means" and "marst" attributes respectively.

Extended experiment upon census dataset which have no AOI-HEP mining interest for similar patterns shows that census dataset have AOI-HEP mining interest for similar patterns when learn on high level concept in "marital" attribute. Another extended experiment upon breast cancer dataset which have AOI-HEP mining interest for similar patterns shows that no AOI-HEP mining interest for similar patterns when learn on high level concept in "cell size" or "cell shape" or "bare nuclei" attribute respectively.

4.0 CONLUSION

In current research experiments, each of dataset was divided into 2 sub datasets based on learning the high level concept in one of their attributes, where the chosen of high level concept only discriminate 2 concepts. For example, datasets such as adult, breast cancer, census and IPUMS have high level concept with only discriminate 2 concepts in attributes such as workclass, clump thickness, means and marst respectively.

In future research, using other attributes with learning more than 2 concepts should be extended to such as adult dataset with attributes like education and native country, census dataset with attributes like class, relat1 and yearsch, IPUMS dataset with attributes like relateg and migrat5g.

Acknowledgement

This research 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] Warnars, H. L. H. S. 2012. Attribute Oriented Induction of High-level Emerging Patterns. *IEEE International Symposium on Foundations and Frontiers of Data Mining in conjunction with IEEE International Conference on Granular Computing (IEEE GrC2012), Hangzhou, China, 11-13 August 2012.*

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

[3] Warnars, H. L. H. S. 2014. Attribute Oriented Induction High Level Emerging Pattern (AOI-HEP) Future Research. *IEEE the 8th International Conference on Information & Communication Technology and Systems (ICTS), Surabaya, Indonesia, 24-25 September 2014*. 13-18.

[4] Warnars, H. L. H. S. 2015. Mining Patterns with Attribute Oriented Induction, the International Conference on Database, Data Warehouse, Data Mining and Big Data (DDDMBD2015), Tangerang, Indonesia, 10-12 September 2015.

[5] Warnars, H. L. H. S. 2014. 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): 266-276.

[6] Warnars, H. L. H. S. 2016. Using Attribute Oriented Induction High level Emerging Pattern (AOI-HEP) to mine Frequent Pattern. *International Journal of Electrical and Computer Engineering (IJECE).* Desember 2016. 6(6).

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

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

[9] Ramamohanarao, K., Bailey, J. and Fan, H. 2005. Efficient Mining of Contrast Patterns and Their Applications to Classification. *Proceedings of the 3rd International Conference on Intelligent Sensing and Information Processing (ICISIP '05), IEEE Computer Society*. 39-47.

[10] Fan, H. and Ramamohanarao, K. 2003. A Bayesian Approach to Use Emerging Patterns for Classification. In Proceedings Of The 14th Australasian Database Conference (ADC '03). 39-48.

[11] Dong, G. and Li, J. 1999. Efficient Mining of Emerging Patterns: Discovering Trends and Differences. *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 43-52.

[12] Podraza, R. and Tomaszewski, K. 2005. KTDA: Emerging Patterns Based Data Analysis System. *Proceedings of XXI Fall Meeting of Polish Information Processing Society*. 213-221.

[13] Dong, G. and Li, J. 2005. Mining Border Description of Emerging Patterns from Dataset Pairs. *Journal of Knowledge and Information Systems.* 8(2): 178-202.

[14] Dong, G., Zhang, X., Wong, L. and Li, J. 1999. CAEP: Classification by Aggregating Emerging Patterns. *Proceeding of the 2nd international Conference on Discovery Science S. Arikawa and K. Furukawa, Eds. Lecture Notes in Computer Science*. Springer-Verlag, London. 1721: 30-42.

[15] Han, J., Cai, Y. and Cercone, N. 1992. Knowledge Discovery in Databases: An Attributed Approach. *Proceeding of the 18th International Conference on Very Large Data Bases*. 547-559.

[16] Cai, Y., Cercone, N. and Han, J. 1990. An Attribute-Oriented Approach for Learning Classification Rules From Relational Databases. *Proceedings of 6th International Conference on Data Engineering*. 281-288.

[17] Li, J., Dong, G. and Ramamohanarao, K. 2000. Instance-based Classification by Emerging Patterns. *Proceeding of the 4th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD'00)*. 191-200.