

Confidence of AOI-HEP Mining Pattern

Harco Leslie Hendric Spits Warnars¹, Agung Trisetarso², Richard Randriatoamanana³

^{1,2}Computer Science Department, BINUS Graduate Program-Doctor of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480

³Institut de Calcul Intensif, Ecole Centrale de Nantes, Nantes 44321, France

*Corresponding author, e-mail: spits.hendric@binus.ac.id¹, trisetarso@binus.ac.id², richard.randriatoamanana@ec-nantes.fr³

Abstract

Attribute Oriented Induction High level Emerging Pattern (AOI-HEP) has been proven can mine frequent and similar patterns and the finding AOI-HEP patterns will be underlined with confidence mining pattern for each AOI-HEP pattern either frequent or similar pattern, and each dataset as confidence AOI-HEP pattern between frequent and similar patterns. Confidence per AOI-HEP pattern will show how interested each of AOI-HEP pattern, whilst confidence per dataset will show how interested each dataset between frequent and similar patterns. The experiments for finding confidence of each AOI-HEP pattern showed that AOI-HEP pattern with growth rate under and above 1 will be recognized as uninterested and interested AOI-HEP mining pattern since having confidence AOI-HEP mining pattern under and above 50% respectively. Furthermore, the uninterested AOI-HEP mining pattern which usually found in AOI-HEP similar pattern, can be switched to interested AOI-HEP mining pattern by switching their support positive and negative value scores.

Keywords: data mining, AOI-HEP mining pattern, confidence similar pattern, confidence frequent pattern, confidence pattern.

Copyright © 2018 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction

Attribute Oriented Induction High level Emerging Pattern (AOI-HEP)[1] has been proven as data mining algorithm which can mine AOI-HEP frequent and similar patterns [2],[4]. AOI-HEP frequent pattern is recognized when have maximum subsumption target (superset) into contrasting (subset) datasets (contrasting \subset target) and having large High Emerging Pattern (HEP) growth rate and support in target dataset [2]. Whilst AOI-HEP similar pattern will be mined from dataset where number of attributes similarity are full or dominant/frequent and the number of similarity with ANY values are infrequent [4]. AOI-HEP as data mining technique has opportunity to be more explored such as inverse discovery learning, learning more than 2 datasets, multidimensional view, learning other knowledge rules and so on [3]. The current finding AOI-HEP patterns cannot be measured in term of confidence the finding AOI-HEP patterns either for each AOI-HEP pattern or dataset. The exploration of confidence AOI-HEP patterns will be explored in order to give confidence mining pattern for each AOI-HEP pattern either for frequent or similar pattern, and each dataset as confidence AOI-HEP pattern between frequent and similar patterns.

2. Confidence of Finding Pattern with Emerging Pattern Data Mining Algorithm

In Emerging Pattern (EP), the confidence of finding pattern is formulated with equation (1) or (2), where $GR(x)$ in equation (1) is GrowthRate which formulated in equation (3) as growth rate of x itemset between positive and negative sub datasets [7]. Moreover, $Sup(x)_{pos}$ and $sup(x)_{neg}$ in equation (3) are dividend and divisor where each of them is support $sup Dy(x)$ in equation (4). Dividend support or $sup(x)_{pos}$ is support of $sup Dy(x)$ which is support for target dataset or recognized as positive class, whilst divisor support or $sup(x)_{neg}$ is support of $sup Dy(x)$ which is support for background dataset or recognized as negative class [5].

Meanwhile, $\text{sup}(x)$ pos and $\text{sup}(x)$ neg are shown in equation (2) as well and formulated in equation (4) with $\text{sup Dy}(x)$ as division of number of x itemset in Dy sub dataset between target/positive and background/negative ($\text{count_Dy}(x)$), with total instances of Dy sub dataset, either as target/positive or background/negative too ($|Dy|$). Growth rate is recognized as EP which is growth rate of x itemset from $\text{sup}(x)\text{neg}$ to $\text{sup}(x)\text{pos}$ or EP of $\text{sup}(x)\text{pos}$ [6]. Its means number of x itemset in $\text{sup}(x)\text{pos}$ is GR or $\text{sup}(x)\text{pos}/\text{sup}(x)\text{neg}$ times the number of x itemset in $\text{sup}(x)\text{neg}$.

$$\text{Conf}(x) = \frac{GR(x)}{GR(x)+1} \quad (1)$$

$$\text{Conf}(x) = \frac{\text{sup}(x)\text{pos}}{\text{sup}(x)\text{pos}+\text{sup}(x)\text{neg}} \quad (2)$$

$$GR(x) = \frac{\text{sup}(x)\text{pos}}{\text{sup}(x)\text{neg}} \quad (3)$$

$$\text{supDy}(x) = \frac{\text{count_Dy}(x)}{|Dy|} \quad (4)$$

where

D=Dataset.

y=option between positive and negative.

Dy=option between positive and negative sub dataset.

|Dy|=total instances of Dy sub dataset.

x=itemset or pattern.

count_Dy(x)=number of x itemset in Dy sub dataset

sup Dy(x)=number of support of x itemset in Dy sub dataset.

GR(x)=Growth rate of x itemset between positive and negative sub datasets.

Sup(x) pos=support (sup Dy(x)) for target/positive dataset.

Sup (x) neg=support (sup Dy(x)) for background/negative dataset.

Conf(x)=confidence pattern of x itemset between positive and negative sub datasets.

For example is itemset {(outlook,sunny)} with equation (4) has $\text{sup}(x)\text{pos}=3/5=0.6$ and $\text{sup}(x)\text{neg}=2/9=0.22$ and $x=\{(outlook,sunny)\}$. Based on equation (4), $\text{sup}(x)\text{pos}$ has 3 records of {(outlook,sunny)} itemset in positive sub dataset with total record of 5. Meanwhile, $\text{sup}(x)\text{neg}$ has 2 records of {(outlook,sunny)} itemset in negative sub dataset with total record of 9. Growth rate of {(outlook,sunny)} itemset can be counted with equation (3) where $GR(x)=\text{sup}(x)\text{pos}/\text{sup}(x)\text{neg}=0.6/0.22=2.7$. The EP confidentiality for {(outlook,sunny)} itemset can be counted with equation (1) or (2), where with equation (1) $\text{conf}(x)=GR(x)/(GR(x)+1)=2.7/(2.7+1)=2.7/3.7=0.729$ or with equation (2) $\text{conf}(x)=\text{sup}(x)\text{pos}/(\text{sup}(x)\text{pos}+\text{sup}(x)\text{neg})=0.6/(0.6+0.22)=0.6/0.82=0.729$.

Meanwhile, if we change the number $\text{sup}(x)\text{pos}$ to $\text{sup}(x)\text{neg}$ and vice versa, and for example of {(outlook,sunny)} itemset with equation (4), then $\text{sup}(x)\text{pos}=2/9=0.22$ and $\text{sup}(x)\text{neg}=3/5=0.6$. The same like previous composition, then based on equation (4), $\text{sup}(x)\text{pos}$ has 2 records of {(outlook,sunny)} itemset in positive sub dataset with total record of 9. On other hand, $\text{sup}(x)\text{neg}$ has 3 records of {(outlook,sunny)} itemset in negative sub dataset with total record of 5. Growth rate of {(outlook,sunny)} itemset can be counted with equation (3) where $GR(x)=\text{sup}(x)\text{pos}/\text{sup}(x)\text{neg}=0.22/0.6=0.367$. The EP confidentiality for {(outlook,sunny)} itemset can be counted with equation (1) or (2), where with equation (1) $\text{conf}(x)=GR(x)/(GR(x)+1)=0.367/(0.367+1)=0.367/1.367=0.268$ or with equation (2) $\text{conf}(x)=\text{sup}(x)\text{pos}/(\text{sup}(x)\text{pos}+\text{sup}(x)\text{neg})=0.22/(0.22+0.6)=0.22/0.82=0.268$.

The explanation from previous 2 paragraphs is summarized in Table 1 where equations (1) to (4) are applied and Table 1 shows that itemset {(outlook,sunny)} with $\text{sup}(x)\text{pos}=3/5=0.6$ and $\text{sup}(x)\text{neg}=2/9=0.22$ is interested since having $GR(x)=2.7$ and moreover having confidence pattern 0.729 or 72.9%. Meanwhile, when itemset {(outlook,sunny)} with

$\text{sup}(x)\text{pos}=2/9=0.22$ and $\text{sup}(x)\text{neg}=3/5=0.6$ is uninterested since having $\text{GR}(x)=0.367$ and moreover having uninterested confidence pattern 0.268 or 26.8%. The number of growth rate less than 1 is uninterested and moreover will have confidence pattern less than 50% which means also as uninterested.

Table 1. Interested and uninterested of example itemset {(Outlook,Sunny)}

Sup\Eq	Sup (4)	Conf (1)	Conf (2)	GR (3)
Sup(x)pos 3/5	3/5=0.6	$2.7/(2.7+1)=2.7/3.7=0.729$	$0.6/(0.6+0.22)=0.6/0.82=0.729$	$\text{GR}=0.6/0.22=2.7$
Sup(x)neg 2/9	2/9=0.22			
Sup(x)pos 2/9	2/9=0.22	$0.367/(0.367+1)=0.367/1.367=0.268$	$0.22/(0.22+0.6)=0.22/0.82=0.268$	$\text{GR}=0.22/0.6=0.367$
Sup(x)neg 3/5	3/5=0.6	0.268	0.22/0.82 = 0.268	

3. Finding AOI-HEP Frequent and Similar Patterns

AOI-HEP has been success to mine frequent and similar patterns from UCI machine learning dataset with experiment dataset such as adult, breast cancer, census and IPUMS datasets with number of record 48842, 569, 2458285 and 256932 respectively [8]. Each of experiment dataset was limited to 5 chosen attributes where each of attribute will have concept hierarchy as can be seen in [9]. The 5 chosen attributes for adult dataset are workclass, education, marital-status, occupation, and native-country, and the 5 attributes for breast cancer dataset are attributes i.e. clump thickness, cell size, cell shape, bare nuclei and normal nucleoli. Meanwhile, class, marital status, means, relat 1 and yearsch attributes, were given to census dataset and the 5 attributes for the IPUMS dataset consists of relateg, marst, educrec, migrat5g and tranwork attributes.

Moreover, each of dataset was divided into two sub datasets based on learning the high level concept in one of their attributes. Finding AOI-HEP patterns are influenced by learning on high level concept in one of chosen attribute and extended experiment upon adult dataset where learn on marital-status attribute showed that there is no finding AOI-HEP frequent pattern. Other extended experiments for finding AOI-HEP similar pattern were carried on and the finding on census dataset which had been none AOI-HEP similar pattern, had AOI-HEP similar pattern when learned on high level concept in marital attribute. Moreover, breast cancer dataset 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. Learning the high level concept in one of their five chosen attributes for concept hierarchy will split each dataset become 2 sub datasets such as:

- Adult dataset was learned on workclass attribute and discriminates between the "government" (4289 instances) and "non government" (14 instances).
- Breast cancer dataset was learned on clumpthickness attribute and discriminates between "aboutaverclump" (533 instances) and "aboveaverclump" (289 instances).
- Census dataset was learned on means attribute and discriminates between "green" (1980 instances) and "no green" (809 instances).
- IPUMS dataset was learned on marst attribute and discriminates between "unmarried" (140124 instances) and "married" (77453 instances).

Experiments upon these 4 datasets show that there are 9 frequent pattern from adult and breast cancer datasets which can be seen in Table 2 and 3 respectively. Meanwhile, 4 similar pattern are found from IPUMS and breast cancer datasets and can be seen in Table 4 and 5 respectively. The experiments showed that adult, breast cancer and IPUMS datasets are interested whilst census dataset is uninterested since there is no finding pattern. Table 2 till 5 show the frequent or similar pattern resulted from learning of the high level concept in one of their five chosen attributes. Moreover, each table includes number of record and support where number of support was got it from each number of record which divided with total number of record from each of learning. Support number in those tables is implementation of equation (4) which applied in AOI-HEP algorithm.

Table 2 shows 8 frequent patterns which content 4 attributes from 5 chosen adult dataset attributes which was not elected as the learning of high level concept of one of their attributes (workclass attribute) and they are education, marital-status, occupation, and native-country. Each frequent pattern contents 2 line rulesets from result of learning of workclass

attribute where 1st and 2nd lines from learning government and non government with ruleset number of record either 3454 or 786 and either 1 or 2 where ruleset total number of record 4289 and 14 respectively. The number of support for each ruleset in frequent pattern is division of ruleset number of record and ruleset total number record of learning between government and non government, for example in 1st frequent pattern, the 1st and 2nd rulesets have support number $3454/4289=0.8053$ and $1/14=0.0714$ respectively. Frequent patterns in Table 2 have similarity in number of record and support such as frequent patterns number 1 and 2, frequent patterns number 3,4 and 5. Moreover, there is no similarity number of record and support in other AOI-HEP pattern in Table 3,4 and 5.

Table 2. AOI-HEP Frequent Pattern in Adult Dataset

No	Frequent pattern	Rec	support
1	Intermediate,ANY,ANY,ANY	3454	0.8053
	Assoc-adm, Married-civ-spouse,Tools, United-States	1	0.0714
2	Intermediate,ANY,ANY,ANY	3454	0.8053
	Some-college, Married-spouse-absent,Tools, United-States	1	0.0714
3	ANY,ANY,ANY, America	786	0.1833
	7 th -8 th , w idowed,Tools, United-States	1	0.0714
4	ANY,ANY,ANY, America	786	0.1833
	Assoc-adm, Married-civ-spouse,Tools, United-States	1	0.0714
5	ANY,ANY,ANY, America	786	0.1833
	Some-college, Married-spouse-absent,Tools, United-States	1	0.0714
6	Intermediate,ANY,ANY,ANY	3454	0.8053
	HS-Grad, Never-married,ANY, United-States	4	0.2857
7	Intermediate,ANY,ANY,ANY	3454	0.8053
	Some-college, Married-civ-spouse,ANY, United-States	2	0.1429
8	ANY,ANY,ANY, America	786	0.1833
	Some-college, Married-civ-spouse,ANY, United-States	2	0.1429

Table 3. AOI-HEP Frequent Pattern in Breast Cancer Dataset

No	Frequent pattern	Rec	support
1	VeryLargeSize, ANY, ANY, ANY	19	0.0356
	VeryLargeSize, SmallShape, MediumNuclei, VeryLargeNucleoli	1	0.0035

Table 4 has 3 similar patterns which content 4 attributes from 5 chosen IPUMS dataset attributes which was not elected as the learning of high level concept of one of their attributes (marst attribute) and they are relateg, educrec, migrat5g and tranwork. Each similar pattern contents 2 line rulesets from result of learning of marst attribute where 1st and 2nd lines from learning unmarried and married with ruleset number of record 6356,4603,7632 and 2296,5706,1217 where ruleset total number of record 140124 and 77453 respectively. The number of support for each ruleset in similar pattern is division of ruleset number of record and ruleset total number record of learning between unmarried and married, for example in 1st similar pattern, the 1st and 2nd rulesets have support number $6356/140124=0.0454$ and $2296/77453=0.0296$ respectively.

Meanwhile, Table 3 and 5 show 1 frequent and similar pattern respectively which content 4 attributes from 5 chosen breast cancer dataset attributes which was not elected as the learning of high level concept of one of their attributes (clumpthickness attribute) and they are cell size, cell shape, bare nuclei and normal nucleoli. Each frequent and similar patterns content 2 line rulesets from result of learning of clumpthickness attribute where 1st and 2nd lines from learning aboutaverclump and aboveaverclump with ruleset number of record either 19 or 5 and either 1 or 4 where ruleset total number of record 533 and 289 respectively. The number of support for each ruleset in frequent and similar patterns are division of ruleset number of record and ruleset total number record of learning between aboutaverclump and aboveaverclump. Frequent pattern in Table 3 shows that the 1st and 2nd rulesets have support number $19/533=0.0356$ and $1/289=0.0035$ respectively. Moreover, similar pattern in Table 5 shows that the 1st and 2nd rulesets have support number $5/533=0.0094$ and $4/289=0.0138$ respectively.

Table 4. AOI-HEP Similar Pattern in IPUMS Dataset

No	Similar pattern	Rec	support
1	ANY, Primary school, Not-Known, ANY	6356	0.0454
	ANY, Primary school, Not-Known, ANY	2296	0.0296
2	ANY, College, Not-Known, ANY	4603	0.0328
	ANY, College, Not-Known, ANY	5706	0.0737
3	ANY, Secondary school, ANY, ANY	7632	0.0545
	ANY, Secondary school, Not-Known, ANY	1217	0.0157

Table 5. AOI-HEP Similar Pattern in Breast Cancer Dataset

No	Similar pattern	Rec	support
1	LargeSize, VeryLargeShape, VeryLargeNuclei, ANY	5	0.0094
	LargeSize, VeryLargeShape, ANY, ANY	4	0.0138

4. Confidence of AOI-HEP Pattern

The confidence of AOI-HEP pattern will be explored in 2 ways and they are confidence of each AOI-HEP pattern and confidence of AOI-HEP pattern in each dataset.

a. Confidence of each AOI-HEP pattern.

Confidence of AOI-HEP pattern which score each AOI-HEP pattern with confidence equation Emerging Pattern (EP) as shown in equation (1) or (2). Finding each AOI-HEP pattern can be justified between uninterested and interested AOI-HEP pattern with confidence between under and above 50% respectively. However, threshold percentage can be applied in order to find interested AOI-HEP pattern and for example, threshold 60% can be applied as minimum confidence of interested AOI-HEP pattern.

b. Confidence of AOI-HEP pattern in each dataset.

The confidence of AOI-HEP pattern in each dataset will assess the confidence of dataset between AOI-HEP frequent and similar pattern. The confidence number will show how confidence its AOI-HEP pattern both frequent and similar patterns. Confidence of AOI-HEP pattern in each dataset will be executed with equation (5) where x1 is total number of AOI-HEP frequent or similar pattern in dataset and x2 is total number of AOI-HEP frequent and similar patterns per dataset.

$$\text{conf} = x1/x2 \quad (5)$$

where:

x1= total number of AOI-HEP pattern in dataset either frequent or similar pattern.

x2= total number of AOI-HEP pattern in dataset both frequent and similar patterns.

4.1. Confidence of each AOI-HEP pattern.

Support number which showed between Tables 2 to 5 are implementation of equation (4) and in order to find confidence finding pattern in AOI-HEP pattern, then the equation (1) or (2) will be implemented in AOI-HEP algorithm. Meanwhile, growthrate number which is showed with equation (3) has been implemented in AOI-HEP algorithm [4]. However, in order to implement equation (1) and (2) in finding frequent and similar patterns between Table 2 and 5, then growth rate in equation (3) will be assessed together with equation (1) and (2) as shown between Tables 6 and 9.

Table 6 is implementation of equation (1), (2) and (3) upon frequent pattern in adult dataset as shown in Table 2 and since there are similar number of record and support in Table 2 such as frequent pattern number 1 and 2, frequent pattern number 3,4 and 5 then automatically their finding growth rate and confidence numbers will similar between of them. For example, Growth rate and confidence numbers for frequent pattern number 1 is similar to number 2 as shown in table 6. Moreover, Growth rate and confidence numbers for frequent pattern number 3 is similar to number 4 and 5 as shown in table 6 too. Table 6 shows that finding AOI-HEP frequent patterns in adult dataset are interested since having maximum and minimum confidence 0.92 (92%) and 0.56 (56%) respectively. The maximum confidence 92% is trully confidence since having growth rate 11.27 which mean there are 11.27 times more.

Meanwhile, the minimum confidence 56% is still more than 50% which show as win by a nose, and other than that the minimum confidence 56% has minimum growthrate 1.28.

Table 6. Growth rate and confidence of AOI-HEP frequent pattern in adult dataset

No	Growth Rate (eq(3))	Confidence (eq(2))	Confidence (eq(1))
1	$(3454/4289)/(1/14)=0.8053/0.0714$ = 11.27	$0.8053/(0.8053+0.0714)=0.8053/0.8767$ = 0.92	$11.27/(11.27+1)=11.27/12.27=$ 0.92
2	$(3454/4289)/(1/14)=0.8053/0.0714$ = 11.27	$0.8053/(0.8053+0.0714)=0.8053/0.8767$ = 0.92	$11.27/(11.27+1)=11.27/12.27=$ 0.92
3	$(786/4289)/(1/14)=0.1833/0.0714$ = 2.56	$0.1833/(0.1833+0.0714)=0.1833/0.2547$ = 0.72	$2.56/(2.56+1)=2.56/3.56=$ 0.72
4	$(786/4289)/(1/14)=0.1833/0.0714$ = 2.56	$0.1833/(0.1833+0.0714)=0.1833/0.2547$ = 0.72	$2.56/(2.56+1)=2.56/3.56=$ 0.72
5	$(786/4289)/(1/14)=0.1833/0.0714$ = 2.56	$0.1833/(0.1833+0.0714)=0.1833/0.2547$ = 0.72	$2.56/(2.56+1)=2.56/3.56=$ 0.72
6	$(3454/4289)/(4/14)=0.8053/0.2857$ = 2.82	$0.8053/(0.8053+0.2857)=0.8053/1.0910$ = 0.74	$2.82/(2.82+1)=2.82/3.82=$ 0.74
7	$(3454/4289)/(2/14)=0.8053/0.1429$ = 5.64	$0.8053/(0.8053+0.1429)=0.8053/0.9482$ = 0.85	$0.8053/(0.8053+0.1429)=0.8053/0.9482=$ 0.85
8	$(786/4289)/(2/14)=0.1833/0.1429$ = 1.28	$0.1833/(0.1833+0.1429)=0.1833/0.3261$ = 0.56	$1.28/(1.28+1)=1.28/2.28=$ 0.56

Table 7. Growth rate and confidence of AOI-HEP frequent pattern in breast cancer dataset

No	Growth Rate (eq(3))	Confidence (eq(2))	Confidence (eq(1))
1	$(19/533)/(1/289)=0.0356/0.0035$ = 10.30	$0.0356/(0.0356+0.0035)=0.0356/0.0391=$ 0.91	$10.30/(10.30+1)=10.30/11.30$ = 0.91

Moreover, growthrate and confidence of AOI-HEP frequent pattern of breast cancer dataset as shown in Table 7 which related to AOI-HEP frequent pattern in breast cancer dataset as shown in Table 3. Table 7 shows that finding AOI-HEP frequent patterns in breast cancer dataset is interested with confidence 0.91(91%) and growth rate 10.30. Furthermore, growthrate and confidence of AOI-HEP similar pattern of IPUMS dataset as shown in Table 8 which related to AOI-HEP similar pattern in IPUMS dataset as shown in Table 4. Table 8 shows that finding AOI-HEP similar patterns in IPUMS dataset are interested for AOI-HEP similar patterns number 1 and 3 where they have confidence 0.60(60%) and 0.78 (78%).

The confidences of 60% and 78% are inline with their growth rate 1.53 and 3.47 respectively, and particularly for confidence 60% is a narrow won. AOI-HEP similar pattern number 2 has uninterested confidence 0.31(31%) and moreover has low growthrate such as 0.45. In order to increase the confidence number, the position support positive (4603/140124) will be changed to support negative (5706/77453) and vice versa as shown in AOI-HEP similar pattern extension number 2 in Table 8. As result, AOI-HEP similar pattern number 2 has interested confidence 0.69 (69%) with number growthrate 2.24. Furthermore, the ruleset for AOI-HEP similar pattern number 2 in Table 4 such as {ANY, College, Not-Known, ANY} are similar then there is no problem to change the position.

Finally, growthrate and confidence of AOI-HEP similar pattern of breast cancer dataset as shown in Table 9 which related to AOI-HEP similar pattern in breast cancer dataset as shown in Table 5. Table 9 shows that finding AOI-HEP similar patterns in breast cancer dataset is uninterested with confidence 0.40(40%) and moreover has unsatisfied growth rate 0.68 which is under score 1. In order to increase the confidence number, the position support positive (5/533) will be changed to support negative (4/289) and vice versa as shown in AOI-HEP similar pattern extension in Table 9. As result, AOI-HEP similar pattern become interested confidence 0.59 (59%) with number growthrate 1.47. Furthermore, the ruleset for AOI-HEP similar pattern in Table 5 such as ruleset for support positive like {LargeSize, VeryLargeShape, VeryLargeNuclei, ANY} is similar with ruleset for support negative like {LargeSize, VeryLargeShape, ANY, ANY}, then there is no problem to change the position.

Table 8. Growth rate and confidence of AOI-HEP similar pattern in IPUMS dataset

No	Growth Rate (eq(3))	Confidence (eq(2))	Confidence (eq(1))
1	$(6356/140124)/(2296/77453)=0.0454/0.0296 =$ 1.53	$0.0454/(0.0454+0.0296)=0.0454/0.0750$ = 0.60	$1.53/(1.53+1)=1.53/2.53$ = 0.60
2	$(4603/140124)/(5706/77453)=0.0328/0.0737=$ 0.45	$0.0328/(0.0328+0.0737)=0.0328/0.1065$ = 0.31	$0.45/(0.45+1)=0.45/1.45$ = 0.31
	$(5706/77453)/(4603/140124)=0.0737/0.0328=$ 2.24	$0.0737/(0.0737+0.0328)=0.0737/0.1065$ = 0.69	$2.24/(2.24+1)=2.24/3.24$ = 0.69
3	$(7632/140124)/(1217/77453)=0.0545/0.0157=$ 3.47	$0.0545/(0.0545+0.0157)=0.0545/0.0702$ = 0.78	$3.47/(3.47+1)=3.47/4.47$ = 0.78

Table 9. Growth rate and confidence of AOI-HEP similar pattern in breast cancer dataset

No	Growth Rate (eq(3))	Confidence (eq(2))	Confidence (eq(1))
1	$(5/533)/(4/289)=0.0094/0.0138=$ 0.68	$0.0094/(0.0094+0.0138)=0.0094/0.0232=$ 0.40	$0.68/(0.68+1)=0.68/1.68$ = 0.40
	$=(4/289)/(5/533)=0.0138/0.0094=$ 1.47	$0.0138/(0.0138+0.0094)=0.0138/0.0232=$ 0.59	$1.47/(1.47+1)=1.47/2.47=$ 0.59

4.2. Confidence of AOI-HEP pattern in each dataset.

After we have confidence for each AOI-HEP pattern which show how each AOI-HEP pattern has confidence AOI-HEP pattern between uninterested and interested with confidence score under and above 50% respectively. Meanwhile, this section will explore the confidence of AOI-HEP pattern in each dataset between AOI-HEP frequent and similar patterns with equation (5). Obviously, confidence per AOI-HEP pattern will show how interested each of AOI-HEP pattern, whilst confidence per dataset will show how interested each dataset between frequent and similar patterns.

Table 10. Confidence of dataset between AOI-HEP frequent And Similar Patterns

Dataset	Frequent pattern	Similar pattern
Adult	$x1/x2=8/8=1$	$x1/x2=0/8=0$
Breast cancer	$x1/x2=1/2=0.5$	$x1/x2=1/2=0.5$
census	$x1/x2=0/0=0$	$x1/x2=0/0=0$
IPUMS	$x1/x2=0/3=0$	$x1/x2=3/3=1$

Table 10 shows the confidence of each dataset between AOI-HEP frequent and similar patterns. The experiments using 4 datasets such as adult, breast cancer, census and IPUMS from UCI machine learning dataset[8], and each dataset will be assessed with equation (5). Based on AOI-HEP patterns' result upon these 4 datasets, which are shown between Tables 2 and 9 and next are the detail:

- a. Tables 2 and 6 show 8 AOI-HEP frequent patterns in adult dataset.
- b. Tables 3 and 7 show 1 AOI-HEP frequent pattern in breast cancer dataset.
- c. Tables 4 and 8 show 3 AOI-HEP similar patterns in IPUMS dataset.
- d. Tables 5 and 9 show 1 AOI-HEP similar pattern in breast cancer dataset

Finally, the experiments upon these 4 datasets will be summarized and the confidence of each dataset will be assessed with equation (5) and next are the detail:

- a. In adult dataset, there are total 8 AOI-HEP patterns where consist of 8 and 0 AOI-HEP frequent and similar patterns respectively. Based on equation (5) then variable $x2=8$ for adult dataset as total number of AOI-HEP pattern in adult dataset both frequent and similar patterns. Lastly, total number of AOI-HEP frequent and similar patterns in adult dataset are $x1=8$ and $x1=0$ with confidence value= $x1/x2=8/8=1$ and $=x1/x2=0/8=0$ respectively as shown in 1st line of table 10. This is show that Adult dataset is confidence 100% for mining AOI-HEP frequent pattern and 0% for mining AOI-HEP similar pattern.

- b. In breast cancer dataset, there are total 2 AOI-HEP patterns where each of 1 is AOI-HEP frequent and similar patterns respectively. Since there are total 2 of AOI-HEP both frequent and similar patterns in breast cancer dataset then variable $x_2=2$. Lastly, total number of AOI-HEP frequent and similar patterns in breast cancer are $x_1=1$ each with the same confidence value $=x_1/x_2=1/2=0.5$ as shown in 2nd line of table 10. This is show that breast cancer dataset is equally confidence 50% for mining both AOI-HEP frequent and similar patterns.
- c. In census dataset, there is none of AOI-HEP frequent and similar patterns. Then $x_1=0$ and $x_2=0$ and automatically both of AOI-HEP frequent and similar patterns in census dataset have similar confidence value $=x_1/x_2=0/0=0$ as shown in 3rd line of table 10. This is show that census dataset is equally none or 0% confidence for mining both AOI-HEP frequent and similar patterns.
- d. In IPUMS dataset, there are total 3 AOI-HEP patterns where consist of 0 and 3 AOI-HEP frequent and similar patterns respectively. Based on equation (5) then variable $x_2=3$ as total number of AOI-HEP pattern in IPUMS dataset both frequent and similar patterns. Lastly, total number of AOI-HEP frequent and similar patterns in IPUMS dataset are $x_1=0$ and $x_1=3$ with confidence value $=x_1/x_2=0/3=0$ and $=x_1/x_2=3/3=1$ respectively as shown in 4th line of table 10. This is show that IPUMS dataset is confidence 0% for mining AOI-HEP frequent pattern and 100% for mining AOI-HEP similar pattern.

Finally, based on above explanation, adult and IPUMS datasets have confidence 100% for mining AOI-HEP frequent and similar patterns respectively. Meanwhile, breast cancer dataset has equally confidence 50% for mining both AOI-HEP frequent and similar patterns, and census dataset is equally confidence none or 0% for mining both AOI-HEP frequent and similar patterns.

5. Conclusion

The confidence of AOI-HEP pattern will be explored in 2 ways and they are confidence of each AOI-HEP pattern and confidence of AOI-HEP pattern in each dataset. Confidence per AOI-HEP pattern will show how interested each of AOI-HEP pattern, whilst confidence per dataset will show how interested each dataset between frequent and similar patterns. Confidence of each AOI-HEP pattern will be scored with equations (1) and (2) whilst confidence of AOI-HEP pattern in each dataset will be scored with equation (5). The experiments showed that using confidence equations from Emerging Pattern as shown in equation (1) or (2) will increase the confidence of AOI-HEP mining patterns either frequent or similar pattern.

The experiments showed that AOI-HEP pattern with growthrate under 1 will recognized as uninterested since having confidence AOI-HEP mining pattern under 50%. AOI-HEP similar pattern in Table 8 (number 2) and Table 9 show in IPUMS and breast cancer datasets respectively, are example of uninterested AOI-HEP pattern where they have growthrate under 1 like 0.45 and 0.68 respectively.

In order to change the uninterested AOI-HEP pattern into interested AOI-HEP pattern then the position support positive and negative will be switched. The experiments showed that uninterested change into interested AOI-HEP mining pattern since its changed their growthrate score from 0.45 and 0.68 to 2.24 and 1.47 as shown in Table 8 (number 2) and Table 9, with result of interested confidence score such as 0.69(69%) and 0.59(59%) respectively.

Meanwhile, the experiments showed that AOI-HEP pattern with growthrate above 1 will be recognized as interested since having confidence AOI-HEP mining pattern above 50%. The minimum growth rate above 1 for interested AOI-HEP frequent pattern is in Table 6 (number 8) with growthrate 1.28 which above 1 and has interested confidence such as 0.56(56%) which above 50%. The maximum growth rate above 1 for interested AOI-HEP frequent pattern is in Table 6 (number 1 and 2) with the same growthrate 11.27 which above 1 and have the same interested confidence such as 0.92(92%) which is above 50%. Moreover, the minimum growth rate above 1 for interested AOI-HEP similar pattern is in Table 9 extension with growthrate 1.47 which above 1 and has interested confidence such as 0.59(59%) which above 50%. The maximum growth rate above 1 for interested AOI-HEP similar pattern is in

table 8 (number 3) with growthrate 3.47 which above 1 and has interested confidence such as 0.78(78%) which is above 50% as well.

For a meanwhile, there is no finding uninterested confidence pattern in AOI-HEP frequent pattern and unlike AO-HEP similar pattern which found with uninterested confidence pattern. However, since these experiments only applied in 4 datasets such as adult, breast cancer, census and IPUMS datasets from UCI machine learning dataset, then future experiments should be carried on.

The experiments showed that adult and IPUMS datasets have confidence 100% for mining AOI-HEP frequent and similar patterns respectively. Meanwhile, breast cancer dataset has equally confidence 50% for mining both AOI-HEP frequent and similar patterns, and census dataset is equally confidence none or 0% for mining both AOI-HEP frequent and similar patterns.

Using other confidence equation such as confidence in association rule where confidence antecedent itemset shows how often items consequent itemset appear in antecedent itemset have meaning such a rule is that transactions of database which contain antecedent itemset tend to contain consequent itemset [10]. The strength of association rule is measured with support and confidence, and confidence is probability of transactions which contain both antecedent and consequent itemsets in antecedent itemset transactions [11].

References

- [1] Warnars S. *Attribute Oriented Induction of High-level Emerging Patterns*. Proceeding of the IEEE International Conference on Granular Computing (IEEE GrC), Hangzhou. 2012: 525–530.
- [2] Warnars S. *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 IColCT 2014), Bandung. 2014: 144-149.
- [3] Warnars S. *Attribute Oriented Induction High Level Emerging Pattern (AOI-HEP) future research*. Proceeding of IEEE the 8th International Conference on Information & Communication Technology and Systems (ICTS), Surabaya. 2014: 13-18.
- [4] Warnars S. 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)*. 2014; 3(11): 266-276.
- [5] Dong G, Li J. Mining border description of emerging patterns from dataset pairs. *Journal of Knowledge and Information Systems*. 2005; 8(2): 178-202.
- [6] Dong G, Li J. *Efficient mining of emerging patterns: discovering trends and differences*. Proceeding of the Fifth ACM SIGKDD international Conference on Knowledge Discovery and Data Mining (San Diego, California, United States, August 15 - 18, 1999). KDD '99. ACM, New York, NY. 1999: 43-52.
- [7] Dong G, Zhang X, Wong L, Li J. *CAEP: Classification by Aggregating Emerging Patterns*. Proceeding of the Second international Conference on Discovery Science S. Arikawa and K. Furukawa, Eds. Lecture Notes In Computer Science, Springer-Verlag, London, 1721. 1999: 30-42.
- [8] Frank A, Asuncion A. *UCI Machine Learning Repository* [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science. 2010.
- [9] Warnars S. *Attribute Oriented Induction High level Emerging Pattern (AOI-HEP)*. Ph.D. thesis, Manchester Metropolitan University. 2013.
- [10] Agrawal R, Srikant R. *Fast algorithms for mining association rules*. Proceeding of 20th International Conference on VLDB. 1994: 487-499.
- [11] Srikant R, Agrawal R. *Mining quantitative association rules in large relational tables*. Proceeding of the ACM SIGMOD international conference on Management of data (SIGMOD '96). 1996: 1-12.