

ICT Research

JOURNAL ON INFORMATION TECHNOLOGIES & COMMUNICATIONS

BỘ THÔNG TIN VÀ TRUYỀN THÔNG - MINISTRY OF INFORMATION AND COMMUNICATIONS

ISSN 1859-3534

Research and Development on Information & Communication Technology

Volume
2024,
Number 1,
March

Các công trình
nghiên cứu, phát triển
và ứng dụng công nghệ
thông tin và truyền thông

Journal on Information Technologies & Communications

Deputy Director & Editor-in-Chief
Tran Anh Tu

Technical Editor-in-Chief
Le Hoai Bac
University of Science,
Vietnam National University Ho Chi Minh City

Associate Technical Editor-in-Chief
Tran Xuan Nam
Le Quy Don University of Technology, Vietnam
Nguyen Khanh Van
Hanoi University of Science and Technology

Editor
Nguyen Thanh Thuy,
University of Engineering and Technology,
Vietnam National University Hanoi
Nguyen Xuan Huy
Institute of Information Technology,
Vietnam Academy of Science and Technology
Tu Minh Phuong
Posts and Telecommunications Institute of Technology, Vietnam
Tzung-Pei Hong
National University of Kaohsiung, Taiwan
Nguyen Xuan Huan
Middlesex University, London, England
Giuseppe D'Aniello
University of Salerno, Italia

Secretary
Nguyen Quang Hung
E-mail: nguyenquanghung@mic.gov.vn

Journal Address
115 Tran Duy Hung Rd., Cau Giay District, Hanoi
Tel: (84)-24-37737136, Fax: (84)-24-37737130,
Website <https://ictmag.vn>
Email: chuyensanbcvt@mic.gov.vn
Journal Office Branch
27 Nguyen Binh Khiem Street, District 1, Ho Chi Minh City,
Vietnam
Tel/Fax: (84)-28-38992898

Journal marketing and distribution
Bui Hong Duong
Email: bhduong@mic.gov.vn, Tel: 84- 24. 37737136

Phó Tổng biên tập phụ trách
Trần Anh Tú

Thường trực Hội đồng biên tập
Chủ tịch
GS. TS. Lê Hoài Bắc,
Trường Đại học Khoa học Tự nhiên,
Đại học Quốc gia TP. Hồ Chí Minh

Phó Chủ tịch
GS. TS. Trần Xuân Nam,
Trường Đại học Kỹ thuật Lê Quý Đôn
PGS. TS. Nguyễn Khanh Văn,
Trường Đại học Bách khoa Hà Nội

Thành viên
GS. TS. Nguyễn Thanh Thủy,
Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội
PGS. TSKH. Nguyễn Xuân Huy,
Viện Công nghệ thông tin, Viện Hàn Lâm KHCN Việt Nam

GS. TS. Từ Minh Phương,
Học viện Công nghệ Bưu chính Viễn thông
GS. TS. Tzung-Pei Hong,
National University of Kaohsiung, Đài Loan
GS. TS. Nguyễn Xuân Huấn,
Middlesex University, Anh
TS. Giuseppe D'Aniello,
University of Salerno, Italia

Thư ký
TS. Nguyễn Quang Hưng
E-mail: nguyenquanghung@mic.gov.vn

License No 365/GP-BTTTT, dated 19th December 2014;
Modified license No 233/GP-BTTTT dated 23rd May 2017.
Printed by Cong ty TNHH MTV In Tap chi Cong san.
Printed and delivered in March 2023.

Price: 40,000 VND

Journal on
Information Technologies & Communications

<https://www.ictmag.vn>

Volume 2024, No.1

March

TABLE OF CONTENTS

Network Traffic Attention Features Fusion for Unauthorized IoT Devices Detection	<i>Duong Van Dan, Ta Minh Thanh</i>	1
Vietnamese Scene Text Detection Method based on Deep Learning	<i>Huynh Van Huy, Nguyen Thi Thanh Tan, Ngo Quoc Tao</i>	11
A Method for Change Impact Analysis of C# Projects	<i>Hoang-Viet Tran, Nguyen Thi Mai Loan, Doan Duc Kien, Nguyen Ha Trang, Le Van Huy, Pham Ngoc Hung</i>	24
A Recommendation Method for an Online Programming Portal	<i>Nguyen Duy Anh, Nguyen Duy Phuong, Nguyen Van Tien</i>	39
A Meta-Heuristic Approach for Enhancing Performance of Associative Classification		47
Exploring the Feasibility of Meeting the MDS Criterion in Random Matrices and Searching for Efficient Random Involutory Hadamard MDS Matrices	<i>Tran Thi Luong</i>	54
Data Power Optimization Algorithm for Downlink Multi-ARS Small Cell Communication Systems	<i>Bui Anh Duc, Tran Manh Hoang, Nguyen Thu Phuong, Xuan Nam Tran, Pham Thanh Hiep</i>	62

A Meta-Heuristic Approach for Enhancing Performance of Associative Classification

Communication: received 30 Nov 2023, revised 12 Jan 2024, accepted 22 Jan 2024
DOI: 10.32913/mic-ict-research.v2024.n1.1246

Abstract: Associative Classification is an interesting approach in data mining to create more accurate and easily interpretable predictive systems. This approach is often built on both association rule mining and classification techniques, to find a set of rules called association rules for classification (CAR) of label attributes. There are many kinds of associative classification such as CPAR, CBA, CMAR but the accuracy is still low on large datasets, and the running time is not reasonable as well. This paper proposes a heuristic approach to significantly enhance the performance of Associative Classification algorithms in running time, reducing the rule set, and accuracy on large data. Experimental results show that heuristic searching the optimal data set makes the associative classification more useful on big data, and reasonable in practice.

Keywords: *Associative Classification, Heuristic search, CAR, Feature Selection*

I. INTRODUCTION

The association classifier is a supervised learning model that uses association rules to assign labels. The model uses association rules to label new data. Thus, the model can be seen as a list of “if-then” clauses: if a new data meets the left-hand attributes of the rule, then the data will be classified according to the right-hand value of the rule [1, 7, 8, 9]. Association classifiers without requiring prior knowledge of feature significance is very effective in heterogeneous high-dimensional data [12]. Most classifiers combine sequential scanning of the set of rules and labeling new data by matching the attributes on the left-hand side of rules. Association classifier rules inherit several metrics from association rules, such as Support or Confidence, that can be used to rank or filter the rules in the model and evaluate their quality. There are many different associative classification methods such as CBA, CMAR, and CPAR.

CBA uses techniques in association rules to classify data, this method has higher performance than traditional classification techniques. The limitation of this method is

that the number of rules generated is too large in case the support threshold is low. CMAR applies FP-tree efficiently, consuming less memory and space than the CBA method. However, FP-tree will meet limitations if the data has a large number of attributes, and the memory capacity is not enough to fit. CPAR is a type of association classifier based on predicted association rules. This method combines the advantages of association classification and traditional rule-based classification by generating rules directly during the training phase to avoid missing important rules.

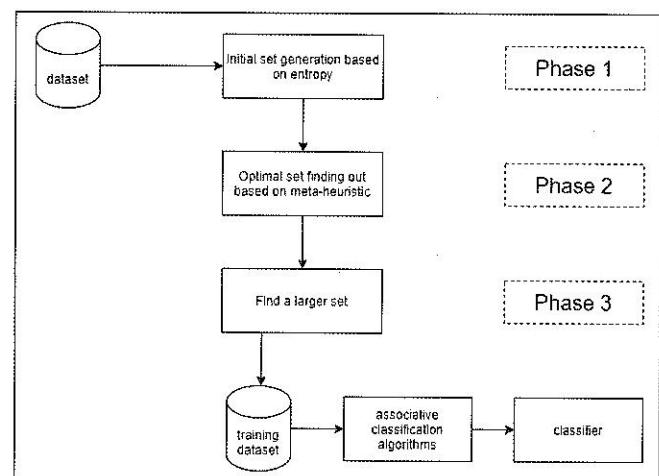


Figure 1. Three phases of heuristic approach

However, association classifier still faces many limitations in terms of time and accuracy rate of classification in case of the attributes is large [5, 6, 9, 13]. Table 1 shows the accuracy of classification, running time, as well as the number of rules of CAR algorithms in SPMF [13] are not reasonable at all. SPMF is an open-source software and data mining library written in Java, the latest version is 2.56.

This paper proposes a heuristic approach to remove unimportant attributes to enhance the associative classification method. Our approach works in three stages as

TABLE I
LIMITATIONS OF CAR IN SPMF[13]

	CPAR	CBA	CMAR
Accuracy	0.5429	0.9828	0.8012
Rules	55	22	453
Times(ms)	1120	945343	1500

shown in Figure 1. This improvement helps decrease running time while significantly increasing accuracy rate of classification.

Phase 1 selects the initial attribute set according to the Entropy. At this stage, attributes are eliminated based on the Entropy relationship of that attribute with the class attribute. The Entropy represents a correlation to check whether the attributes play an important role with the class attribute in classification and decide whether or not to remove those attributes. At this phase, we temporarily select a candidate set that will be the initial set in phase 2.

Phase 2 finds the optimal set of attributes derived from the initial set in phase 1. The original attribute set is divided into subsets and a associative classification (AC) algorithm is run based on these subsets of data. Based on the accuracy of the model, we can add or remove randomly some attributes on the current set to obtain a new subset, and run AC repeatedly until getting the best subset of attributes. This method create subsets by using a greedy approach and evaluates the goodness of all subsets derived from the initial set in phase 1 instead of testing all combinations of possible attributes.

Phase 3 expands the optimal attribute set to increase classification accuracy. Expand the optimal set of attributes in phase 2 by adding each remaining attribute to find a new optimal set with higher accuracy.

The main contributions focus on part II, part III. The rest content is arranged as follows: part II describes our method in three phases of feature selection, and three algorithms correspondingly. The experimental results and evidences of effectiveness are described in part III, and the conclusion is presented in part IV.

II. OUR PROPOSED METHOD

With large data sets and many attributes, the cost of selecting an optimal attribute set is huge. The proposed method combines the Entropy measure and the heuristic approach to select features with three phases. Phase 1 is mainly filtering unimportant attributes according to the Entropy measure (the Entropy value between an attribute and the class attribute) and feature selection method in the form of Hill Climbing. The accuracy of the model after removing attributes was tested again using the cross-

validation method. The selected attribute set will be the optimal attribute set for classification.

Algorithm 1 Phase1(FSAC)

Input: minSup, minConf, minAllConf; D: dataset (D_{train} , D_{test});
Output: D_e (all best attributes)

```

1: e := entropy threshold ;
2:  $D_e :=$  attribute  $a_i$  from  $D_{train}$  where  $\text{entropy}(a_i, \text{label}) \leq e$ ;
3: neighbor=[e,e ± Δ];
4:  $e' := \text{random}(\text{neighbor})$ ;
5:  $D_{e'} :=$  attribute  $a_i$  from  $D_{train}$  where  $\text{entropy}(a_i, \text{label}) \leq e'$ ;
6: repeat
7:    $\text{model}_e := \text{AC-RG}(D_e)$ ;
8:    $\text{accuracy}_e := \text{model}_e.\text{predict}(D_{test})$ ;
9:   neighbor=[e,e ± Δ];
10:   $e' := \text{random}(\text{neighbor})$ ;
11:   $\text{model}_{e'} := \text{AC-RG}(D_{e'})$ ;
12:   $\text{accuracy}_{e'} := \text{model}_{e'}.\text{predict}(D_{test})$ ;
13:  if ( $\text{accuracy}_e \leq \text{accuracy}_{e'}$ ) then
14:     $e := e'$ 
15:  end if
16: until  $\text{accuracy}_e \leq \text{accuracy}_{e'}$ 
17: return  $D_e$ 

```

The Entropy value is in the range [0,1], attributes with large Entropy values that will not contribute much information to the classification should be eliminated. In data sets with a few attributes, we can calculate all the entropies of the attributes and arrange them from highest to lowest. Then, each attribute is checked repeatedly for removing it from high to low entropy and evaluate again the accuracy of the remaining set of attributes. If the accuracy is equal to or greater than the old value (without removing the attribute), then the removed attribute should really be removed. Because this is a bottom-up approach, the cost of eliminating the initial large data set is not feasible. We proposes an initial Entropy threshold, instead of testing Entropy in order from high to low as in the Phase 1 algorithm.

The FSAC [11] is the algorithm for phase 1. With the initial entropy threshold, the FSAC algorithm finds the optimal entropy value that eliminates all redundant attributes. The remaining attribute set is a relatively good attribute set. Line 7 is AC-RG(D_e) algorithm that generates a large set of rules, Associative classification is a special case of association rule discovery in which only the class attribute is considered on the rule's right-hand side. Line 8 calculates the accuracy of the model obtained from line 7.

The code from lines 6 to 16 repeats the improvement until the model's accuracy reaches its peak. In each iteration, based on the current entropy neighbor (lines 11, 12), and the entropy level e' , the algorithm determines the attribute set D_e including the attributes whose entropy are less than e' . In other words, in this iteration the algorithm tried to remove attributes which are redundant from the training set. Line 13 checks whether the removal is correct or not, if so, updates the better value. After exiting the loop, the algorithm finds a relatively good set of attributes like line number 16. The dataset D is divided into 2 groups Dtrain and Dtest according to the ratio 80/20. Based on Dtrain to find a set of classification rules, and Dtest is used to test the accuracy of the set of classification rules.

Algorithm 2 Phase2.1(Hill Climbing)

Input: The initAttr set, the subAttr set, initValid, delta
Output: bestAttr(all best attributes)

```

1: for ( ; ; ) do
2:   newAttr := GenAttributes(initAttr, subAttr, delta);
3:   validation := AC(newAttr);
4:   if validation > initValid then
5:     initValid := validation;
6:     bestAttr := newAttr
7:     subAttr := allAttr - bestAttr
8:   end if
9: end for
10: return bestAttr

```

After implementing Phase 1, we obtain an initial set of attributes that are relatively good but not really optimal because of the limitations of the Filter method for each attribute. Phase 2 bases on the heuristic approach, the starting solution is the initial set of attributes obtained from phase 1. At this phase, attributes are divided into two attribute groups (bestAttr and subAttr). The bestAttr attribute set contains the attributes obtained from phase 1, temporarily considered a good solution.

This task iterates until finding the best set with the same number of attributes compared to the attribute set obtained from phase 1. In each iteration, the new attribute set generated from the GenAttributes function in line 3 is tried, based on the delta value (neighborhood parameter) to give a delta number of any random attribute in the bestAttr set, and add the delta number of random attributes from the subAttr set to the bestAttr set. If the new attribute set has better results than the current bestAttr attribute set, update it again. This process stops when after a finite number of iterations a better set of attributes is still not found. The subAttr set contains the remaining attributes. The heuristic algorithms in Phase 2 (Hill-Climbing or Simulated

Annealing) find out whether there is any set that is better than the current bestAttr set. If so, the bestAttr set will be updated. This task is from lines 2 to 9 in the algorithm 2 (Hill-Climbing), and is from lines 2 to 16 in the algorithm 3 (Simulated Annealing).

After completing phase 2, we have the optimal attribute set with the same number of attributes as in phase 1. Phase 3 is the expanding the number of attribute sets to find a better value than the set of attributes obtained in phase 2. Lines 2 to 7 in the Phase 3 algorithm is a task that tries to find a more optimal extended set by combining the set obtained in phase 3. with each remaining attribute in the subAttr set.

Algorithm 3 Phase2.2 (Simulated Annealing)

Input: The initAttr set, the subAttr set, initValid, delta, T, k

Output: bestAttr

```

1: for ( ; ; ) do
2:   newAttr := GenAttributes(initAttr, subAttr, delta);
3:   validation := AC(newAttr);
4:   if validation > initValid then
5:     initValid := validation;
6:     bestAttr := newAttr
7:     subAttr := allAttr - bestAttr
8:   else
9:      $\Delta f = validation - initValid$ 
10:    r := random r(0,1)
11:    if r > exp(- $\Delta f/kT$ ) then
12:      initValid := validation;
13:      bestAttr := newAttr
14:      subAttr := allAttr - bestAttr
15:    end if
16:   end if
17: end for
18: return bestAttr

```

III. EXPERIMENTS

The experimental environment is processed centrally on a computer configured with Intel(R), Core(TM) i7-6820HQ CPU @ 2.70GHz (8 CPUs), 2.7GHz and Windows 10 operating system. In this paper, the experiment has 3 phases as follows:

- Phase 1: Choose the entropy threshold and initial attribute set
- Phase 2: Find the optimal set of attributes by using HC and SA algorithms
- Phase 3: Find a larger attribute set with better results than the original attribute set

Algorithm 4 Phase 3

Input: The initAttr set, the subAttr set, initValid
Output: bestAttr

```

1: bestAttr := initAttr;
2: for  $i \leftarrow 0$ ;  $i \leq subAttr.length$ ;  $i++$  do
3:   newAttr := initAttr + subAttr [i];
4:   validation := AC(newAttr);
5:   if validation > initValid then
6:     initValid := validation;
7:     bestAttr := newAttr;
8:   end if
9: end for
10: return bestAttr

```

1. Simple Dataset

The first experiment is implemented on Mushroom data [2]. Because this data is used by the original ACAC [1, 10] algorithm, it has 8125 rows and 23 columns. In the data, the first attribute is a target attribute named 'class' to classify mushrooms as poisonous or non-poisonous (edible=e, poisonous=p). The thresholds are minSupp = 0.1, minConf = 0.8, minAllConf = 0.5.

In Figure 2, vertical axis describes the accuracy of model, the horizontal axis describes the entropy of each removed attribute. The graph peaks at accuracy of 99.51% correspond to the entropy score of 0.757, with 8 attributes. However, the attributes are removed continuously with entropy values (0.745, 0.727, 0.714), and the accuracy still keeps at 99.51%. The classification performance will go down if one more attribute is removed. So the the entropy value of 0.714 is the best value, and gains the highest accuracy of 99.51%.

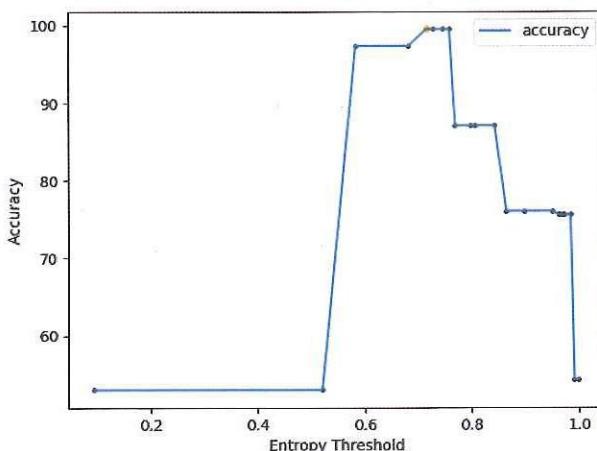


Figure 2. The accuracy of classification

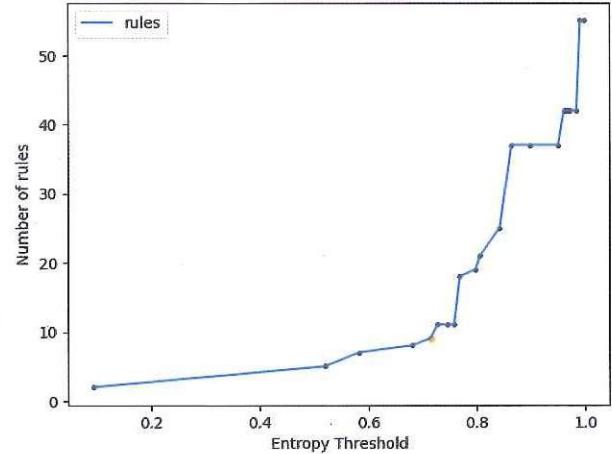


Figure 3. The optimal CAR set

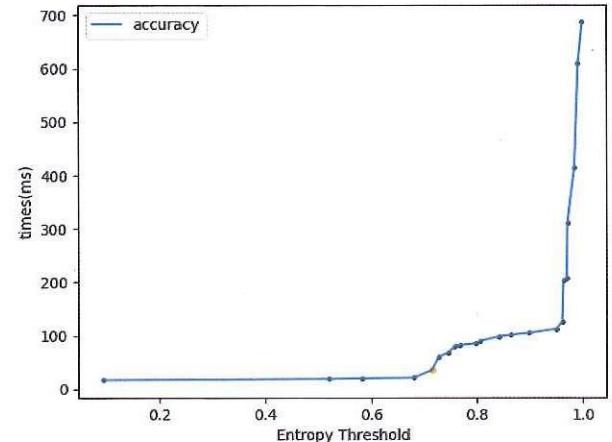


Figure 4. Running time of processing

After phase 1, the prediction model achieves very high accuracy in small datasets (we can see the performance of CPAR, CBA, CMAR on 5 attributes of Mushroom in Table 2), so the phase 2 and phase 3 is not necessary to implement more. Our research focuses on the challenge of associative classification problem, the large dataset with many attributes, so we do experiments on more challenge dataset, SpamEmail.

TABLE II
 REASONABLE VALUES ON MUSHROOM

	CPAR	CBA	CMAR
Accuracy	0.9951	0.9397	0.8876
Rules	9	8	14
Times(ms)	169	146	200

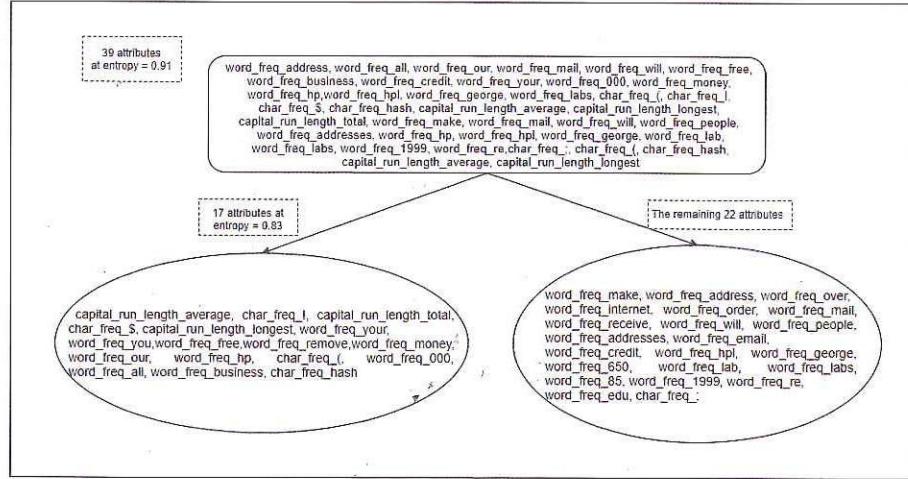


Figure 5. Divide 2 subsets from output of phase 1

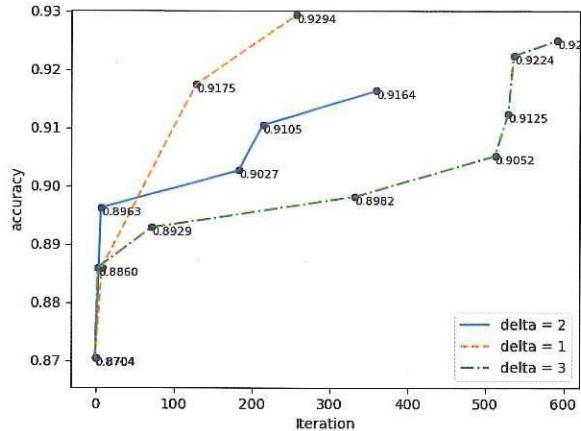


Figure 6. Wrapper with delta neighbor by SA

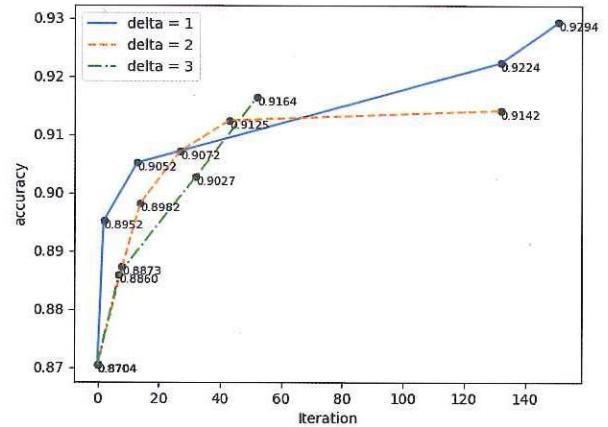


Figure 7. Wrapper with delta neighbor by HC

2. Large Dataset

The Spam Emails Dataset [3] is selected for experiment, it is a rather complex data set, with 4602 rows and 58 columns. In particular, the last attribute is a label attribute called 'spam' to classify emails as spam or spam (spam=1, spam=0). This data is also used in paper[11]. The thresholds are minSupp = 0.1, minConf = 0.9, minAllConf = 0.5.

a) Phase 1: Selecting the Entropy threshold, the initial set

Among the remaining 57 attributes, there are many attributes that have no value in classification and need to be eliminated. To eliminate, we uses entropy to determine the ability of each attribute to contribute to the classification. Any attribute with a high entropy value will be removed from the data set, the new data set will be tried again using the ACAC method for classification. This model will be tested using cross-validation method (80% - 5000 lines of data used for training, 20% - 1001 lines of data used for

testing). If the precision increases then the attribute removal is correct, and this is done until the precision reaches its highest value. In other words, work stops when accuracy is degraded.

TABLE III
OUTPUT OF PHASE 1 ON SPAMEMAIL

Entropy	Attributes	rules	times(ms)	Accuracy(%)
0.91	39	?	?	?
0.885	33	3108	7.783E+13	94.2
0.88	31	2239	3.5E+11	94.4
0.875	29	1612	11E+9	94.9
0.87	28	1254	261000000	94.06
0.865	27	989	51000000	93.04
0.86	26	692	10000000	92.4
0.85	24	338	2200000	90.99
0.84	22	109	35170	88.93
0.83	17	24	343	87.04
0.81	14	17	151	81.11
0.79	13	13	145	80.61

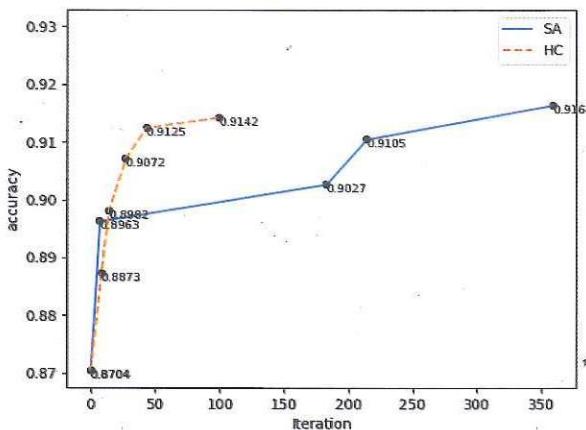


Figure 8. Comparison between HC and SA

TABLE IV
OUTPUT OF PHASE 3

Entropy	Attributes	times(ms)	rules	Accur(%)
0.308	capital_run_length_a	7862	170	9294
0.589	capital_run_length_t	7165	170	0.9294
0.658	capital_run_length_l	6891	170	0.9294
0.786	word_freq_hp	18390	170	0.9294
0.789	char_freq_(7029	169	0.9294
0.815	word_freq_all	17124	219	0.9294
0.820	word_freq_george	19007	170	0.9294
0.835	word_freq_mail	19316	332	0.9305
0.836	word_freq_address	18231	225	0.9294
0.836	word_freq_hpl	18023	170	0.9294
0.838	word_freq_will	12262	267	0.9294
0.846	word_freq_email	13710	244	0.9294
0.866	word_freq_credit	13292	278	0.9305
0.871	word_freq_re	12716	194	0.9294
0.877	word_freq_people	17010	185	0.9294
0.878	word_freq_1999	18207	170	0.9294
0.884	char_freq_;	17159	327	0.9305
0.887	word_freq_addresses	17589	204	0.9305
0.892	word_freq_edu	18668	211	0.9294
0.898	word_freq_labs	17062	170	0.9294
0.898	word_freq_85	17108	170	0.9294
0.901	word_freq_650	17400	171	0.9294

In this case choosing word_freq_credit because of its performance (Table 4, column Attributes, accuracy = 0.9305). After going through 3 phases, we have the best attributes set so we can use Associative classification algorithms to build prediction model.

TABLE V
HEURISTIC OUTPUT OF SPAMEMAIL

	CPAR	CBA	CMAR
Accuracy	0.9305	0.848	0.6916
Rules	278	184	158
Times(ms)	13292	2257889	3496

Based on Table 3, the entropy threshold = 0.91 is chosen to reduce the number of attributes from 57 to 39, thereby reducing space, processing time and improving evaluation

indicators. Although the entropy threshold = 0.91 helps improve the limitation score compared to the original data set, the number of 39 attributes is still too large for one processing. Therefore, we propose an approach that is to divide the set of 39 attributes into 2 subsets. The first subset (original data set) includes 17 attributes at entropy threshold = 0.81, and the second subset includes the remaining 22 attributes. The reason for choosing the entropy threshold = 0.81 as the initial attribute set for experimentation is because at this threshold, although the performance is not the best, it is not bad to be improved and expanded more.

b) Phase 2: Find out the optimal set

After selecting the initial attribute set, there are 17 attributes at the entropy threshold = 0.81. The next step is to find a set of 17 attributes that are more optimal than the original 17 attributes. To do that, the experimental part chooses delta values of 1, 2 and 3 respectively, each different delta value will give different results. On this experimental, we select HC and SA to implement in heuristic search.

In Figure 6, it shows the peak changes (accuracy) of the data set over iterations of the program. At delta = 1, the number of peak changes when finding a set with higher accuracy than the current set is 7 times, this number at delta = 3 is 3 times and delta = 2 is 4 times. The following data sets tend to have higher accuracy than the original (0.8704) with each new peak being found, and the highest peak achievable at delta=3 across iterations is 0.9294. For delta = 3, finding the peak with accuracy = 0.9294 takes less time and also has higher accuracy than delta = 2 or delta = 3.

SA (Figure 7) is considered an improved approach to improve the disadvantages of Hill Climbing, we can refer charts comparing the indicators of each approach on the same data set (17 attributes) at different delta values.

c) Phase 3: Find a larger attribute set that has better result than the original attribute set

From phase 2, we have selected a set of rules including 17 attributes with accuracy = 0.9294 (highest peak) and the remaining 22 attributes. In phase 3, we will find a set of 18 attributes with accuracy higher than 0.9294.

To concretize the above goal, this experimental part will combine the set of 17 attributes and each of the remaining 22 attributes. Each set of 18 attributes will give different possible results.

IV. CONCLUSION

The main advantage of AC is that the set of "If-Then" rules has the ability to infer new knowledge that other classification methods can not do. Another important advantage

of AC is the ease of interpretation of labeling for new data. AC algorithms have limitations that often occur when minsupp is set to a very small value.

Experimental results show that AC algorithms on small and simple data such as Mushroom (23 attributes) are very effective. On large and complex data like SpamEmail (58 attributes), the algorithms show limitations (as shown in Table 3).

The heuristic approach still maintains almost the same classification accuracy, but running time and rule pruning are more efficient (as shown in Table 5 compared to Table 3). The comparison on the same classification accuracy (0.9305 vs 0.9304), processing time (13292 vs 51000000) is 3750 times faster, the number of rules is much reduced (278 vs 989). The running time is more important than the number of rules, that why we select 278 rules instead of 204 rules.

REFERENCES

- [1] Z. Huang, Z. Zhou, T. He, and X. Wang, "ACAC: Associative Classification Based on All-Confidence," *IEEE International Conference on Granular Computing*, pp. 289–293, 2011.
- [2] Mushroom Classification Dataset, [Online]. Available: <https://www.kaggle.com/dataset/suclm/mushroom-classification>.
- [3] Spam Emails Dataset, [Online]. Available: <https://www.kaggle.com/datasets/yasserspamemaildataset>
- [4] H. F. Ong, C. Y. M. Neoh, V. K. Vijayaraj, Y. X. Low, "Information-Based Rule Ranking for Associative Classification," *ISPACS*, 2022.
- [5] M. Abrar, A. Tze and S. Abbas, "Associative Classification using Automata with Structure based Merging," *IJACSA*, vol. 10, 2019.
- [6] D. L. Olson and G. Lauhoff, "Market Basket Analysis" in Descriptive Data Mining," *Springer Singapore*, 2019.
- [7] K. D. Rajab, "New Associative Classification Method Based on Rule Pruning for Classification of Datasets," *IEEE Access*, vol. 7, pp. 157783-157795, 2019.
- [8] H. F. Ong, N. Mustapha, H. Hamdan, R. Rosli and A. Mustapha, "Informative top-k class associative rule for cancer biomarker discovery on microarray data," *Expert Systems with Applications*, vol. 146, 2020.
- [9] Majid Seyf, Yue Xu, Richi Nayak, "DAC: Discriminative Associative Classification", *SN Computer Science*, 2023.
- [10] E.R.Omiecinski, "Alternative interest measures for mining associations in databases", *IEEE Transactions on Knowledge and Data Engineering*, vol 15, pp. 57-69, 2003.
- [11] N. Q. Huy, T. A. Tuan, N. T. N. Thanh, "An efficient algorithm that optimizes the classification association rule set", *VNICT* 26, pp. 13-19, 2023.
- [12] K.H.T Dam,T.G Wilson,A Legay,R Veroneze, "Packer classification based on association rule mining", *Applied Soft Computing*, Vol 127, Issue C, 2022.
- [13] Philippe Fournier-Viger, "SPMF - An Open-Source Data Mining Library", <https://www.philippe-fournier-viger.com/spmf/index.php>, Version 2.59, 2022.



HỘI THẢO QUỐC GIA: MỘT SỐ VẤN ĐỀ CHỌN LỌC CỦA CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

Viện Công nghệ thông tin & Trường Đại học Nha Trang

Nha Trang 11 - 12 / 10 / 2024. Website: <https://vnict.vn>



THÔNG BÁO SỐ 1

Hội thảo Quốc gia lần thứ XXVII "Một số vấn đề chọn lọc của Công nghệ thông tin và Truyền thông" (VNICT 2024) do Viện Công nghệ thông tin – Viện Hàn lâm Khoa học và Công nghệ Việt Nam và Trường Đại học Nha Trang đồng tổ chức tại Nha Trang vào các ngày 11-12 tháng 10 năm 2024, và có sự tham gia phối hợp của Câu lạc bộ các Khoa-Trường-Viện CNTT – TT Việt Nam (FISU). Hội thảo là diễn đàn thường niên để những người làm công tác nghiên cứu, triển khai, giảng dạy và quản lý trong lĩnh vực công nghệ thông tin và truyền thông trao đổi học thuật, chia sẻ kinh nghiệm ứng dụng và tìm kiếm sự hợp tác. Đặc biệt hội thảo khuyến khích các nhà khoa học trẻ, nghiên cứu sinh, học viên cao học tham gia báo cáo trao đổi kết quả nghiên cứu và học tập của mình. Hội thảo khuyến khích trưng bày giới thiệu các sản phẩm khoa học công nghệ và những kết quả nghiên cứu ứng dụng vào thực tiễn. Chủ đề chính của hội thảo năm nay: **Khoa học dữ liệu và ứng dụng**

Một số thời hạn

Toàn văn báo cáo
30 / 6 / 2024

Chấp nhận đăng ký yêu
31 / 8 / 2024

Xuất bản ký yếu
30 / 9 / 2024

Đăng ký tham dự
30 / 9 - 7 / 10 / 2024

BAN TỔ CHỨC

Đồng Trưởng Ban

PGS. TS Nguyễn Trường Thắng, Viện Công nghệ thông tin
TS Khổng Trung Thắng, Trường Đại học Nha Trang

Thành viên

PGS. TS Phạm Quốc Hùng, Trường Đại học Nha Trang
TS Phạm Thị Thu Thủy, Trường Đại học Nha Trang
TS Nguyễn Thu Anh, Viện Công nghệ thông tin
TS Nguyễn Như Sơn, Viện Công nghệ thông tin
PGS. TS Bùi Thu Lâm, CLB các Khoa-Trường-Viện CNTT – TT VN (FISU Việt Nam)

CÁC CHỦ ĐỀ CHÍNH

- Trí tuệ nhân tạo
- An toàn thông tin
- Học máy
- Xử lý ảnh và thực tại ảo
- Xử lý ngôn ngữ tự nhiên
- Các hệ thống thông minh
- Tin-sinh học
- Công nghệ da phương tiện và mô phỏng
- Truyền thông và mạng máy tính
- Kỹ thuật điều khiển, tự động hóa
- Cơ sở toán học cho tin học

BAN CHƯƠNG TRÌNH

Thông tin liên hệ

TS Ngô Hải Anh
Viện Công nghệ thông tin
✉ ngoahaianh@oit.ac.vn
☎ 098 859 6582

TS Phạm Thị Thu Thủy
Trường DH Nha Trang
✉ thuthuy@ntu.edu.vn
☎ 090 190 5679

TÀI TRỢ

Viện Hàn lâm Khoa học và Công nghệ
Việt Nam



Viện Công nghệ thông tin



Trường Đại học Nha Trang



Câu lạc bộ các Khoa-Trường-Viện
CNTT – TT Việt Nam



Học viện Khoa học và Công nghệ



Tạp chí Thông tin và Truyền thông

THÔNG TIN TRUYỀN THÔNG

Tạp chí An toàn thông tin

AN TOÀN
Thông tin

an toan cuu tru

Trưởng Ban
PGS. TS Nguyễn Long Giang, Viện Công nghệ thông tin

Thành viên

► GS. TS Đăng Quang Á, Viện Hàn lâm Khoa học và Công nghệ Việt Nam
► PGS. TS Nguyễn Việt Anh, Viện Công nghệ thông tin
► PGS. TS Huỳnh Thị Thành Bình, Đại học Bách Khoa Hà Nội
► PGS. TS Nguyễn Thành Bình, Trường Đại học Bách Khoa Đà Nẵng
► GS. TS Nguyễn Đường, Viện Công nghệ thông tin
► TS Nguyễn Văn Cẩn, Trường Đại học Kỹ thuật-Hỗn cản CAND
► TS Nguyễn Mạnh Cường, Trường Đại học Nha Trang
► TS Nguyễn Ngọc Cường, Bộ Công an
► TS Nguyễn Khắc Cường, Trường Đại học Nha Trang
► PGS. TS Trịnh Viết Cường, Trường Đại học Hồng Đức
► TS Hà Mạnh Dào, Trường Đại học Công nghiệp Hà Nội
► TS Nguyễn Huân Đức, Trường Đại học Thủ Đức
► PGS. TS Đặng Văn Đức, Viện Công nghệ thông tin
► PGS. TS Lê Bá Dũng, Viện Công nghệ thông tin
► PGS. TS Nguyễn Đức Dũng, Viện Công nghệ thông tin
► PGS. TS Ngô Quốc Dũng, Học viện Công nghệ Bưu chính Viễn thông
► PGS. TS Lương Thế Dũng, Học viện Kỹ thuật Mật mã
► PGS. TS Pham Thanh Giang, Viện Công nghệ thông tin
► GS. TS Huỳnh Xuân Hiệp, Trường Đại học Cần Thơ
► TS Nguyễn Hữu Hòa, Trường Đại học Cần Thơ
► PGS. TS Nguyễn Ngọc Hóa, Trường Đại học Công nghệ – DHQG Hà Nội
► TS Đặng Trọng Hùng, Trường Đại học Công nghiệp Hà Nội
► PGS. TSKH Nguyễn Xuân Huy, Viện Công nghệ thông tin
► PGS. TS Trần Đăng Hưng, Trường Đại học Sư phạm Hà Nội
► TS Nguyễn Đình Hưng, Trường Đại học Nha Trang
► TS Nguyễn Quang Hưng, Tạp chí Thông tin Truyền thông và Truyền thông
► TS Nguyễn Minh Hương, Trường Đại học Khoa học và Công nghệ Hà Nội
► TS Hồ Văn Hương, Cục Quản lý dân số và Kinh định sản phẩm mầm mống
► PGS. TS Bùi Thu Lâm, Học viện Kỹ thuật Mật mã
► PGS. TS Trần Văn Lãng, Viện Hàn lâm Khoa học và Công nghệ Việt Nam
► PGS. TS Ngô Thành Long, Học viện Kỹ thuật Quân sự
► PGS. TS Hoàng Việt Long, Trường Đại học Kỹ thuật-Hỗn cản CAND
► TS Nguyễn Thị Lương, Trường Đại học Đà Lạt
► PGS. TS Lương Chí Mai, Viện Công nghệ thông tin
► PGS. TS Nguyễn Thị Hồng Minh, Trường Đại học KHTN – DHQG Hà Nội
► TS Nguyễn Hải Minh, Trường Đại học Công nghệ thông tin và Truyền thông
► TS Pham Ngoc Minh, Viện Công nghệ thông tin

► TS Lê Quang Minh, Đại học Quốc gia Hà Nội
► TS Phạm Văn Nam, Trường Đại học Nha Trang
► PGS. TS Trần Thị Ngân, Trường Quốc tế – DHQG Hà Nội
► TS Trần Đức Nghĩa, Viện Công nghệ thông tin
► PGS. TS Phùng Trung Nghĩa, Trường DH CNTT và Truyền thông
► PGS. TS Lê Anh Phương, Trường Đại học Sư phạm Huế
► GS. TS Từ Minh Phương, Học viện Công nghệ Bưu chính Viễn thông
► TS Nguyễn Duy Phương, Học viện Công nghệ Bưu chính Viễn thông
► TS Bùi Thị Thành Quyên, Viện Công nghệ thông tin
► PGS. TS Nguyễn Hữu Quỳnh, Trường Đại học Thủ Đức
► PGS. TS Nguyễn Quốc Tạo, Viện Công nghệ thông tin
► TS Vũ Đức Thái, Trường DH Công nghệ thông tin và Truyền thông
► TS Hoàng Đức Thắng, Viện Công nghệ thông tin
► PGS. TS Huỳnh Quyết Thắng, Đại học Bách Khoa Hà Nội
► PGS. TS Nguyễn Trường Thắng, Viện Công nghệ thông tin
► TS Trần Thiên Thành, Trường Đại học Quy Nhơn
► GS. TS Vũ Đức Thi, Đại học Quốc gia Hà Nội
► TS Nguyễn Văn Thiên, Trường Đại học Công nghiệp Hà Nội
► TS Hoàng Đức Tho, Học viện Kỹ thuật Mật mã
► GS. TS Nguyễn Thành Thúy, Đại học Quốc gia Hà Nội
► TS Phạm Thị Thu Thủy, Trường Đại học Nha Trang
► PGS. TS Đỗ Năng Toàn, Viện Công nghệ thông tin
► PGS. TS Trần Minh Triết, Trường DHKH Tự nhiên TPHCM
► TS Lê Anh Tú, Trường Đại học Hạ Long
► TS Trần Mạnh Tuấn, Trường Đại học Thủ Đức
► TS Nguyễn Như Tuấn, Tạp chí An toàn thông tin
► TS Hoàng Dồ Thành Tùng, Viện Công nghệ thông tin
► PGS. TS Thái Quang Vinh, Viện Công nghệ thông tin
► TS Nguyễn Trần Quốc Vinh, Trường Đại học Sư phạm Đà Nẵng
► PGS. TS Lê Sỹ Vinh, Trường Đại học Công nghệ – DHQG Hà Nội
► PGS. TS Lê Trọng Vinh, Trường Đại học KHTN – DHQG Hà Nội

BAN THƯ KÝ

TS Ngô Hải Anh, Viện Công nghệ thông tin
TS Bùi Chí Thành, Trường Đại học Nha Trang
TS Trần Thị Thành Vân, Trường Đại học Nha Trang
TS Hoàng Thị Thu Hằng, Tạp chí An toàn thông tin

KÝ YÊU HỘI THẢO & TUYỂN CHỌN BÀI ĐĂNG TẠP CHÍ

► Các bài gửi tham gia Hội thảo (qua hệ thống EasyChair: <https://easychair.org/conferences/?conf=vnict2024>) có chất lượng tốt sau khi qua vòng phản biện sẽ được lựa chọn đăng trên ký yếu có mã số ISBN, được xuất bản bởi NXB Khoa học và Kỹ thuật (Bộ Khoa học và Công nghệ). Một số bài xuất sắc sẽ được chọn đăng trên Tạp chí Tin học và Điều khiển học (Viện Hàn lâm Khoa học và Công nghệ Việt Nam), và các chuyên san tạp chí: Các công trình nghiên cứu, phát triển và ứng dụng Công nghệ Thông tin và Truyền thông (Tạp chí Thông tin và Truyền thông – Bộ Thông tin và Truyền thông) và Nghiên cứu khoa học và Công nghệ trong lĩnh vực An toàn thông tin (Tạp chí An toàn thông tin - Ban Cơ yếu Chính phủ).

► Các bài được chọn đăng Ký yếu hoặc Tạp chí phải được ít nhất một tác giả trình bày slide tại hội thảo.

Phí tham dự hội thảo: 1.000.000 (tác giả báo cáo), 500.000 (khách tham dự không báo cáo).

Website: <https://vnict.vn>

Địa điểm diễn ra hội thảo: Trường Đại học Nha Trang, 02 Nguyễn Dinh Chiểu, phường Vĩnh Thọ, Nha Trang, Khánh Hòa.

INFORMATION FOR AUTHORS

Journal on Information Technologies & Communications or *ICT Research* in short – is a scientific publication of the Journal of Information and Communication, published by Vietnam Ministry of Information and Communications (MIC) since 1999. *ICT Research* is peer-reviewed, open-access, and published four times a year: two issues in English (ISSN: 1859-3534), and two issues in Vietnamese (ISSN: 1859-3526) (Chuyên san Các công trình nghiên cứu và ứng dụng phát triển Công nghệ Thông tin và Truyền thông).

The purpose of *ICT Research* is to provide a forum for researchers and professionals to disseminate original and innovative ideas in the fields of information technology, communications, and electronics in Vietnam and worldwide. Its Editorial Board is composed of renowned scientists from research institutions throughout Vietnam and other countries.

AUTHOR GUIDELINES

Authors are encouraged to follow good practice of technical paper writing, such as the guidelines on “How to write for technical periodicals & conferences” of the Institute of Electrical and Electronics Engineers.

Submissions must be made online on the website of *ICT Research* [<https://ictmag.vn>]. Authors should format the manuscript as closely as possible to the *ICT Research* manuscript template.

By submitting a manuscript, authors are required to ensure that the submission has not been previously published, nor is currently submitted for publication elsewhere. Submission also implies that the corresponding author has the consent of all authors. The submission file must be in PDF format. To facilitate double-blind review, authors are requested to submit two versions of the manuscript, one with full information about all authors, the other without.

Submissions of revised manuscripts should follow the same guidelines as for the initial manuscript, plus with a document entitled “Response to reviewers’ comments” which explains the modifications according to the comments of the anonymous reviewers.

Upon formal acceptance of a manuscript for publication, the authors are required to prepare the final manuscript in LaTeX, using the IEEE Transactions style.

PUBLICATION ETHICS

ICT Research is committed to following the rules, standards, and guidelines set forth by the Committee on Publication Ethics (COPE) in ensuring publication integrity and in handling cases of misconduct. All involved parties (Editors, Authors, Reviewers, and Publisher) are expected to be aware of and to conform to the respective rules, standards, and guidelines, available on COPE website.

The Editorial Board of *ICT Research* will immediately screen all articles submitted for publication. All submissions we receive are checked for plagiarism by using available online tools. Any type of plagiarism is unacceptable and is considered unethical publishing behavior. Such manuscripts will be rejected.

PEER REVIEW PROCESS AND PUBLICATION

An area editor of *ICT Research* (to be considered as the Editor from the point of view of the Authors) is assigned, by the Technical Editor-in-Chief, to each submission to coordinate the review process. The Editor pre-screens the submission and determines whether it is appropriate to the interests of readers of *ICT Research*. If appropriate, the Editor then initiates the review process. Each submission is then reviewed by at least two Reviewers. The review process is double blind, that is, identities of authors and reviewers are not revealed to each other.

Review decisions are: (A) *Accept submission* (i.e., no changes required), (B1) *Revision required* (i.e., minor revision), (B2) *Resubmit for review* (i.e., major revision, the submission needs to be re-worked with major changes and is then subject to a second round of review), and (C) *Decline submission* (i.e., reject). Each major round of review may take approximately 1 month. Should authors be requested by the Editor to revise the manuscript, the revised version should be submitted within 4 weeks for a major revision or 2 weeks for a minor revision. No more than 1 major revision and 2 minor revisions are allowed.

ICT Research is published with two issues a year in its *English Edition* (ISSN: 1859-3534). An accepted article will be immediately published online, with a DOI number, having been copyedited and laid out in compliance with *ICT Research* editing rules.

COPYRIGHT NOTICE

Upon acceptance for publication, transfer of copyright will be made to the Publisher of *ICT Research*, who guarantees that full content of the published article is freely distributed on the website of *ICT Research*. The copyright transfer gives the Publisher of *ICT Research* full authority to resolve any complaints of misuse or abuse (such as infringement or plagiarism) of the published article. The author has the freedom to redistribute and reuse the published article in any medium or format for any purpose, provided the original published article is properly cited. An article submission implies author agreement with this policy.

Call For Papers

July 19-20, 2024 - Da Nang and Hoi An, Vietnam

THE 13th CONFERENCE ON INFORMATION TECHNOLOGY AND ITS APPLICATIONS

<https://cita.vku.udn.vn/2024>

CITA (Conference on Information Technology and its Applications) is an annual scientific conference on information technology and its applications in all fields.

The 13th CITA, 2024, will be hosted by the Vietnam - Korea University of Information and Communication Technology (VKU), a member of University of Danang, with the supports of other researching and training institutions belonging to ASEAN Consortium for Innovation and Research (ACIR) as well as Vietnam ICT Association of Faculties-Institutes-School-Universities (FISU Vietnam). The conference will take place on July 19-20, 2024 in Da Nang City and Hoi An City, which are the beautiful cities and the most livable cities in Vietnam.

Papers for the CITA 2024 must be written in English and will be peer-reviewed. Accepted papers will be published in one of the following proceedings:

- Volume 1: Lecture Notes in Network and Systems (LNNS) series (Springer) indexed in, among others, DBLP, Scopus, Web of Science. The fee for a paper accepted to publish in volume of LNNS is only \$200. For Vietnamese students, who are the first authors, the fee is only \$100.

- Volume 2: CITA 2024 Proceedings (with ISBN) with the fee is only \$100.

* CITA 2024 will also have an Online session (with conference the fee is \$100) exclusively for authors, delegates who are outside of Vietnam and could not attend the conference in person.

As the post-conference publication, a number of selected papers will be invited to be extended for publication in special issues of the following journals:

1. Journal of Information and Telecommunication, Taylor and Francis Publisher (<https://www.tandfonline.com/toc/jiti20/current>) indexed in Scopus and eSCI

2. Vietnam Journal of Computer Science, World Scientific Publishing (<https://www.worldscientific.com/page/vjcs/abstracted-indexed>) indexed in Scopus and eSCI.

3. The special issue "Research and Development on Information and Communication Technology" (ISSN 1859-3526), Journal of Information and Communication, Vietnam Information and Communications Publisher.



Wroclaw University
of Science and Technology



University College Dublin
Ireland's Global University



INTERESTED TOPICS (BUT NOT LIMITED TO)

- 1. Data Science & AI
- 2. Image and Natural Language Processing
- 3. Software Engineering and Information System
- 4. Network and Communications
- 5. Digital Economy
- 6. Cyber Security and Cryptography

SUBMISSION GUIDELINES

All submitted papers in CITA 2024 must conform to the Springer LNCS / LNAI Formatting Guidelines (<https://www.springer.com/gp-computer-science/lncs/conference-proceedings-guidelines>). When posting, the authors need to agree to the following terms:
- Submissions must be original papers, never be posted/published before;
- Use the template specified by the conference, do not change the font style, format, header/footer, no page numbering;
- Language: English;
- Length: no more than 12 pages.

- November 10th, 2023: Special Session proposal
- January 15th, 2024: Submission deadline
- March 31th, 2024: Notification of acceptance
- April 20th, 2024: Camera ready and Registration
- July 19th-20th, 2024: Conference

<https://easychair.org/conferences/?conf=cita2024>

ORGANIZING COMMITTEE

➤ General Chair

Assoc. Prof. Dr. Huynh Cong Phap – Rector of Vietnam-Korea University of Information and Communication Technology (VKU), Vietnam

➤ Program Chair

Prof. D.Sc. Ngoc Thanh Nguyen - Chair of IEEE SMC Technical Committee on Computational Collective Intelligence; Wroclaw University of Science and Technology, Poland.

➤ Program Co-chairs

Prof. Dr. Nguyen Thanh Thuy - Chairman of FISU Vietnam, University of Engineering and Technology, Vietnam National University
Assoc. Prof. Dr. Le Kha Nhan An - University College Dublin (UCD), Ireland

➤ Steering Committee

Prof. D.Sc. Ngoc Thanh Nguyen - Chair of IEEE SMC Technical Committee on Computational Collective Intelligence; Wroclaw University of Science and Technology, Poland.
Assoc. Prof. Dr. Huynh Cong Phap – Rector of Vietnam-Korea University of Information and Communication Technology (VKU), Vietnam.
Prof. Dr. Nguyen Thanh Thuy - Chairman of FISU Vietnam, University of Engineering and Technology, Vietnam National University
Assoc. Prof. Dr. Le Kha Nhan An - University College Dublin (UCD), Ireland
Prof. Dr. Chia- Nan Wang - National Kaohsiung University of Science and Technology, Taiwan

Prof. Dr. Tzong-Pei Hong - National University of Kaohsiung, Taiwan
Prof. Dr. Le Minh Hoa - Northumbria University, Newcastle, UK
Prof. Dr. David Camacho - Universidad Autonoma de Madrid, Spain
Prof. Dr. Nitish Chawla - University of Notre Dame, USA
Prof. Dr. Kang-Hyun Jo - University of Ulsan, South Korea
Prof. Dr. Lipo Wang - Nanyang Technological University, Singapore
Prof. Dr. Franck Leprevost - University of Luxembourg, Luxembourg
Prof. Dr. Thang LE DINH - Université du Québec à Trois-Rivières, Québec, Canada

➤ Special Session Chair

Dr. Nguyen Quang Vu - Vietnam-Korea University of Information and Communication Technology (VKU), Vietnam

➤ Student Track Chair

Dr. Nguyen Huu Nhat Minh - Vietnam-Korea University of Information and Communication Technology (VKU), Vietnam



Language
Understanding
 $\lambda(\mu)$ Laboratory

