

VOLUME 1

2011 INTERNATIONAL CONFERENCE ON
**UNCERTAINTY REASONING
AND
KNOWLEDGE ENGINEERING**

4-7 AUGUST 2011
BALI, INDONESIA

2011 INTERNATIONAL CONFERENCE ON UNCERTAINTY REASONING AND KNOWLEDGE ENGINEERING



IEEE Conference Publications Management Group
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08854

IEEE Catalog Number: CFP1115N-PRT
ISBN: 978-1-4244-9983-0



Proceedings of the International Conference on
Uncertainty Reasoning and
Knowledge Engineering

4-7 August 2011, Bali, Indonesia

URKE 2011

All right reserved.

Print Version

Copyright and Reprint Permission: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Operations Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved. Copyright 2011 by the Institute of Electrical and Electronics Engineers

URKE 2011
Table of Contents
Volume – 1

Preface – Volume 1.....	VI
Conference Organization – Volume 1	VII
Alternative implementation techniques of parallelized differential evolution for multi-core processors.....	1
<i>Kiyaharu Tagawa and Takashi Ishimizu</i>	
Designing of expert system for troubleshooting diagnosis on gas chromatography gc-2010 by means of inference method.....	5
<i>Suyanto M. Abduh Pancaputra, and Ruri Agung Wahyuono</i>	
Selection of districts for data entry sites in aceh household census.....	9
<i>Bonivasius Prasetya, Stefanie Intan Christienova, and Idaman</i>	
Collaborative internet content filtering on the internet infrastructure in Malaysia.....	12
<i>Fajri Achmad Maulana, Sahidan Abdulmana, and Fauzan Alfariti</i>	
Reinforcement mechanism and engineering applications of jointed rock mass.....	16
<i>Jingsui Zhang, Fei Wang, Shaojun Fu, and Tingting Liao</i>	
Uncertainty, protocol behavior prediction through chaos theory?.....	20
<i>Prabhu Shankar Kaliappan</i>	
Soaking with uncertainty for problem solving.....	24
<i>Wei Hui and Fu Xixu</i>	
A model for information systems evaluation.....	28
<i>Liang-Chuan Wu</i>	
Flexible decision making in uncertain R&D projects.....	32
<i>Juite Wang</i>	
Semantic technology classification - a defence and security case study.....	36
<i>Dirk Thorleuchter and Dirk Van den Poel</i>	
Companies website optimising concerning consumer’s searching for new products.....	40
<i>Dirk Thorleuchter and Dirk Van den Poel</i>	
A framework for vision-based swimmer tracking.....	44
<i>Wen-Hui Chen, Po-Chuan Cho, Ping-Lin Fan, and Yi-Wen Yang</i>	
Initiating layers architecture design for software product line.....	48
<i>I Made Murwantara</i>	

Research on remote intelligent monitoring system.....	52
<i>Wang Xiao-li, Gao Xiao-hong, and Li Jian-wei</i>	
Empirical study on selection and evaluation of TPL based on CS.....	56
<i>Wei Xu, Songzheng Zhao, and Liangdong Lu</i>	
Research on the application of neural network PID to wipe film motor of molecular distillation.....	60
<i>Guan Changjun and Li Hui</i>	
Design of a pet drinking vessel based on TRIZ & growth design.....	64
<i>Hongwu Chen, Xuan Liu, and Can Wang</i>	
Knowledge sharing platform for project team based on web feeds.....	67
<i>I-Ching Hsu, Lee Jang Yang, and Der-Chen Huang</i>	
New switched filtering method for recurrent neural networks.....	71
<i>Choon Ki Ahn</i>	
Research on China's studying abroad situation and countermeasures.....	75
<i>Chen Aijuan and Zhang Qian</i>	
Classifier combination for telegraphese restoration.....	79
<i>Leo Willyanto Santoso</i>	
Design and implementation of the soft-sensing system for fermentation process.....	83
<i>Wu Zi-yue, Xu Zhe, and Jinfeng Geng</i>	
The study of mobile education development based on 3G technique and cloud computing.....	86
<i>Huang Bin</i>	
The analysis of human judgment accuracy using decision tree models.....	90
<i>Hsi-Peng Lu, Yi-Wen Yang, and Wen-Hui Chen</i>	
A weighted fuzzy time series model based on adoptive OWA operators.....	94
<i>Jing-Rong Chang, and Yu-Jie Huang</i>	
Research on randomized greedy algorithm for k-median problem.....	98
<i>Wang Shouqiang</i>	
An algorithm for splitting an orthogonal polyhedron with a polyplane.....	102
<i>Jefri Marzal, Hong Xie, and Chun Che Fung</i>	
Implementing user-oriented interfaces: from user analysis to framework's components.....	107
<i>Sergio Inzunza, Andrés Mejía, Reyes Juárez-Ramírez, and Manuel Gómez-Ruelas</i>	
Towards a formal model for knowledge transference: An initial proposal to support the adoption of process models.....	111
<i>Violeta ocegueda-miramontes, Reyes Juárez-Ramírez, Rafael Pimienta-Romo, and Manuel Gómez-Ruelas</i>	
A formal approach for measuring the lexical ambiguity degree in Natural language requirement specification.....	115
<i>Carlos Huertas, Reyes Juárez-Ramírez, Manuel Gómez-Ruelas, and Héctor Plata</i>	
Towards improving user interfaces: a proposal for integrating functionality and usability since early phases.....	119

<i>Reyes Juárez-Ramírez, Alan A. Gutiérrez, Manuel Gómez-Ruelas, and Pavel Negrete</i>	
Using iterative learning control methods for 2-D systems.....	124
<i>Zahra Maghsoodzade Sarvestani, Ahmadreza Argha, and Mehdi Roopaei</i>	
Direct torque control of brushless dc motor drives with reduced starting current using fuzzy logic controller.....	129
<i>N. Parhizkar, M. Shafiei, and M. Bahrami Kouhshahi</i>	
Application of topic based vector space model with wordnet.....	133
<i>Adi Wibowo and Andreas Handojo</i>	
Processing comparisons and evaluations in business intelligence: a question answering system.....	137
<i>Kenston Choi, Rosalyn Margret Pacaña, Adrian Lester Tan, Jonathan Yiu, and Nathalie Rose Lim</i>	
A study of applying arima and svm model to software reliability prediction.....	141
<i>Jung-Hua Lo</i>	
Incorporating information quality criteria in blog search engine.....	145
<i>Fatemeh Azimzadeh, Abd. R. Ramli, Borhanoddin. M. Ali, and Hamidah Ibrahim</i>	
Study on the development trend of the digital printing.....	149
<i>Qian Hua</i>	
Using particle swarm optimization algorithm for image enhancement.....	154
<i>Hanxing Chen and Jun Tian</i>	
Chaos-based gray image watermarking algorithm.....	158
<i>Junmei Guo, Yong Yang, and Ning Wang</i>	
Special equipment safety analysis and countermeasures.....	161
<i>Jun Tian, and Hanxing Chen</i>	
A study on home information network.....	165
<i>Yong Yang and Junmei Guo</i>	
Non-sampled contourlets and gray level co-occurrence matrix based images segmentation.....	168
<i>Zhang Jian and Chen Xiaowei</i>	
Preparation for structural path analysis from plain text input output matrix.....	171
<i>Nova Nurviana, Rizchi Wahyuni, and Idaman</i>	
A study on the design of half pipe.....	175
<i>Zhao Yang and Li zhaonan</i>	
Evaluation of building materials logistics provider based on FAHP—TOPSIS.....	180
<i>Sun Shusheng, Zhang Qin, and Zhang Min</i>	
An approach for multimodal biometric fusion under the missing data scenario.....	185
<i>Quang Duc Tran, Panos Liatsis, Bing Zhu, and Changzheng He</i>	
Hybrid-dimension association rules for diseases track record analysis at Dr. Soetomo general hospital.....	189
<i>Silvia Rostianingsih, Gregorius Satia Budhi, and Ni Wayan Yessy Dwijayanti</i>	
The forecast for the engine wear trend based on grey gm(1,1) and markov chain model.....	192
<i>Zhang Yong, Chen Liang, and Gu Xiaoxing</i>	

Data mining market basket analysis' using hybrid-dimension association rules, case study in minimarket X.....	196
<i>Djoni Haryadi Setiabudi, Gregorius Satia Budhi, I Wayan Jatu Purnama, and Agustinus Noertjahyana</i>	
Product cost estimation using supper efficiency data envelopment analysis.....	200
<i>Fantahun M. Defersha, Adil Salam, and Nadia Bhuiyan</i>	
Prioritization of information quality criteria in the blog context.....	205
<i>Mohammad Javad Kargar</i>	
Towrads dedicated information quality frameworks in social network products.....	209
<i>Mohammad Javad Kargar</i>	
Study on the fast image registration based on genetic algorithm.....	213
<i>Shuai Feng, Yufeng Li, and Weiping Zhao</i>	
Compression and realization of power quality disturbance data based on wavelet analysis.....	217
<i>Zhao Hongtu and Xi Dongmei</i>	
Research on face detection algorithm in instant message robot.....	220
<i>Yufeng Li, Haijuan Zhang, and Yanan Zhang</i>	
Three-dimensional visual geographic information system.....	224
<i>Bo Hang</i>	
Runtime monitoring web services implemented in BPEL.....	228
<i>Zhang Haiteng, Shao Zhiqing, and Zheng Hong</i>	
Based on stm32f103 implement profibus-dp slave with high-speed transmission.....	232
<i>Xu Hongyan, Wu Guichu, and Chen Chong</i>	
LVMM: a lightweight virtual machine memory management Architecture for virtual computing environment.....	235
<i>Weizhe Zhang, Tao Cheng, Hui He, and Albert M.K. Cheng</i>	
Primary signal detection over rayleigh fading channel for cognitive radio.....	239
<i>Yufeng Li, Yanan Zhang, and Haijuan Zhang</i>	
A spt-routing key management scheme for heterogeneous wireless sensor networks based on ECC.....	243
<i>Gang Han, Rui Zhou, and Hua Yang</i>	
Design of video surveillance system based on 3G wireless network in underground coal mine.....	248
<i>Wei Zhou</i>	
A hybrid key management scheme for heterogeneous wireless sensor networks based on ecc and trivariate symmetric polynomial.....	251
<i>Rui Zhou and Hua Yang</i>	
Comprehensive evaluation and selection model of supplier under E-business.....	256
<i>Wang Mian</i>	

Voice conversion application (VOCAL).....	259
<i>Liliana, Resmana Lim, and Elizabeth Kwan</i>	
Clustering search engine at petra christian university library using suffix tree clustering.....	263
<i>Andreas Handojo, Adi Wibowo, and Jemmy Lay Santo</i>	
Author Index – Volume 1	267

Preface

These proceedings record the papers presented at the 2011 International Conference on Uncertainty Reasoning and Knowledge Engineering (URKE 2011), held in the city of Bali, Indonesia on 4-7 August, 2011. The conference proceeding is published by IEEE Press.

The International Conference on Uncertainty Reasoning and Knowledge Engineering (URKE'11), technical co-sponsored by IEEE Indonesia Section and Bina Nusantara University, sponsored by IAMSIE , which will provides a forum for the exchange of ideas and results among researchers, developers, and practitioners working on all aspects of uncertainty reasoning and knowledge engineering and their applications. The program of URKE'11 will consist of invited lectures, tutorials, refereed research papers, and tool demonstrations. Research contributions can report new results as well as experimental evaluations and comparisons of existing techniques.

All the submitted papers in these proceedings have been peer reviewed by at least two reviewers drawn from the chairs of committees depending on the subject matter of the paper. Reviewing and initial selection were undertaken electronically. A joint committee meeting was held to resolve the final paper selection and a draft programme for the conference. The conference provides three day's focus on the science and technology that are the basis for the uncertainty reasoning and knowledge engineering topics including uncertainty reasoning, data modelling, machine learning, web application and digital society and information (organized by workshop). It is featuring a range of contributions by distinguished invited speakers drawn from both industry and academia. The invited speakers address significant recent industrial applications of formal methods, as well as important academic advances serving to enhance their potency and widen their applicability.

We hope that all participants and other interested readers benefit scientifically from these proceedings and also find it stimulating in the process.

General Chairs of URKE 2011
5 July 2011

Committees

Conference Chair

Dr. Ford Lumban Gaol, Bina Nusantara University, Indonesia

Program Co-chair

Dr. Mehdi Roopaei, Shiraz University, Shiraz, Iran

Dr. Kelly Robinson, Electrostatic Answers LLC, USA

Program Members

Prof. Ivan Jelínek, Czech Technical University In Prague, CZ

Dr. Prabhat K. Mahanti, University of New Brunswick, CA

Dr. Kunal Patel, Ingenuity Systems, USA

Dr. Lefteris Gortzis, University of Patras, Greece

Prof. Po-Tsun Liu, National Chiao Tung University

Prof. Vladimir O. Safonov, St. Petersburg University, Russia

Prof. S. Ablameyko, Belarusian State University, Belarus

Prof. Tsung-Chih Lin, Feng Chia University, Taiwan

Dr. Jivika Govil, Carnegie Mellon University, USA

Prof. Genoveffa Tortora, University of Salerno, Italy

Prof. Santo Banerjee, Politecnico di Torino, Italy

Prof. Sikh Namh C., University of Delhi, India

Prof. Haiyin Sun, Xi'an Jiaotong University

Prof. Mo.Jamishi, Montreal University, Canada

Prof. P. Tjeerd, Technische Universiteit Delft, Netherlands

Prof. J.Upadhyaya, New York University, USA

Prof. Chang-Ho Lee, New York University, USA

Prof. Anthony F.J, Technische Universiteit Delft, Netherlands

Dr. Ruiping Xie, Xiamen University, China

Dr. Jin Zhao, Xi'an University of Technology, China

Prof. Yanlong Cao, University of Huddersfield, UK

Prof. M. Jamishi, Montreal University, Canada

Dr. Y. F. Ly, Zhejiang University, China

Prof. Svetlana, P., Kazan State University, Russia

Dr. B. Craciun, University of Bucharest, Romania

Prof. G. X. Liu, Huazhong Normal University, China

Prof. L. Shen, Boston Graduate School of Psychoanalysis, USA

Prof. S. Komithe, Malaysia University of Science and Technology, Malaysia

Prof. Jasmko Tochiny, University Malaysia Sabah, Malaysia

Dr. Deng Sun, Tsinghua University, China
Prof. Macos, B., Federal University of Rio de Janeiro, Brazil
Dr. S. S. Tang, Lenoir-Rhyne University, USA
Dr. P. Q. Wu, Guizhou University, China
Dr. Muslim, K. L., University of Pune, India
Dr. Y. Liu, Shanghai Jiaotong University, China
Dr. W. Pan, Carroll University, USA
Dr. Ming Xu, Shenzhen University, China
Dr. Xiaolin Zheng, Zhejiang University, China
Dr. Min He, Hunan University, China
Dr. Ray Klefstad, University of California, Irvine, USA
Dr. Sharad Mehrotra, University of California, Irvine, USA
Prof. Kenji Saito, University of Tokushima, Japan

Classifier Combination for Telegraphese Restoration

Leo Willyanto Santoso
 Department of Computer Science
 Petra Christian University
 Siwalankerto 121-131 Surabaya, Indonesia
 e-mail: leow@petra.ac.id

Abstract—This paper presents a classifier combination to solve telegraphese restoration problem. By implementing more than one classifier, it can support other classifier, and finally it can improve the performance. Using supplied development data, training data and testing data, the best model had an accuracy $F = 79\%$.

Keywords: Classifier combination, telegraphese restoration, Penn Treebank tagset, chunk parsing

I. INTRODUCTION

Telegraphese restoration is an interesting topic in machine learning. It is used to restore “telegraphese” to its original form; focusing on case and punctuation. For example, given the following text:

pawang or medicine man johari albert 78 said he had a dream thursday

The correct text as follows:

Pawang, or medicine man, Johari Albert, 78, said he had a dream Thursday

In this research, Penn Treebank Tagset, Chunk and Claws 7 in 5 different taggings, across 5 positions in the window are used. All data will be made available in ARFF format, with a single instance per word token in the source text. For each word token, an instance index, the word, and the class label are provided. The following is a data sample:

```
245,pawang,cap1+comma
246,or,nochange
247,medicine,nochange
248,man,nochange+comma
249,johari,cap1
250,albert,cap1+comma
251,78,nochange+comma
252,said,nochange
253,he,nochange
254,had,nochange
255,a,nochange
256,dream,nochange
257,thursday,cap1
```

The classes describe first necessary changes to the capitalisation of the word, and then any insertions of punctuation to the end of that word. Note that a given word may require multiple changes to its capitalisation, and also

potentially multiple punctuation insertions, and that the atomic class labels are additive. The detailed information about class and description can be seen in Table I.

TABLE I. CLASS AND DESCRIPTION

Class	Description	Example
nochange	No changed required	Nochange(medicine) → medicine
allcaps	Convert to all caps	allcaps(unesco) → UNESCO
capN	Convert the letter at position N to upper case	cap1(thursday) → Thursday
+comma	Insert a comma at the end of the word	nochange+comma(78) → 78,
+fullstop	Insert a full stop at the end of the word	nochange+fullstop(popularity) → popularity
+colon	Insert a colon at the end of the word	nochange+colon(it) → it:
+semicolon	Insert a semicolon at the end of the word	nochange+semicolon(10) → 10;
+exclmark	Insert a exclamation mark at the end of the word	nochange+exclmark(proof) → proof!
+questmark	Insert a question mark at the end of the word	nochange+questmark(man) → man?

The research will take the form of a shared task: pre-classified training and development data are provided, to use in feature engineering and the classifier development. After that, unannotated test data will be provided. The output of this research is final classifiers over that data.

In this paper, we present a potential approach for improving the performance of telegraphese restoration by using classifier combination techniques such as bagging and boosting. To the best of our knowledge, this is the first effort that utilizes classifier combination for improving telegraphese restoration.

Combination methods have been applied to many problems in natural-language processing (NLP). For examples: ROVER system for speech recognition [3], the Multi-Engine Machine Translation (MEMT) system [7], and improving lexical disambiguation [1]. Most of these techniques have shown a considerable improvement over the performance of a single classifier and therefore, considering implementation such a multiple classifier system for telegraphese restoration as well decision.

Using classifier combination techniques one can potentially achieve a classification accuracy that is superior to that of the single best classifier. This is based on the assumption that the errors made by each of the classifiers are not identical, and therefore if we intelligently combine multiple classifier outputs, we may be able to correct some of these errors

The remaining part of this paper is organised as follows. Section 2 presents an overview of current proposals for dealing with natural language processing. Section 3 depicts the approach that we have delineated to solve proposed problems. Section 4 discusses the performance of proposed method. Finally, section 5 concludes the paper

II. BACKGROUND AND RELATED WORK

In this section, the previous work of Part of Speech (POS) tagging, Penn Treebank tagset, chunk parsing are presented.

A. Part of Speech (POS) Tagging

Part of Speech (POS) Tagging is the process that classifies word into several categories based on its definition and context. POS tagging is now done in the context of computational linguistics; using algorithms which associate discrete terms, as well as hidden parts of speech, in accordance with a set of descriptive tags.

Part-of-speech tagging is harder than just having a list of words and their parts of speech, because some words can represent more than one part of speech at different times [6].

B. Penn Treebank Tagset

Penn Treebank Tagset is developed by researchers of Computer Science and Information Department at the University of Pennsylvania. It annotates naturally; occurring text for linguistic structure. Table II shows the sample of Treebank Tagset.

TABLE II. SAMPLE TREEBANK TAGSET

Tagset	Description
CC	Coordinating conjunction. e.g. and, or, but
CD	Cardinal number
DT	Determiner
EX	Existential <i>there</i>
FW	Foreign word
...	...

The complete tagset can be browsed from this URL¹.

C. Chunk Parsing

Chunk Parsing or Shallow Parsing is an important step to extract word into linguistic fragment [4]. Compared to full parsers that would fail to deliver any (even partial) linguistic information if the whole utterance cannot be completely analysed in accordance with some competence

model of the particular language, this parsing method is robust (since it always delivers some linguistic information).

A chunk parser attempts to model human parsing by breaking the text up into small pieces; each parsed separately. There are some advantages of chunk parsing, such as: better modeling human behaviour. Moreover, chunk parsing is fast because it only deals with small part without recursion process

III. CLASSIFIER COMBINATION

Running this experiment, let's go through steps as follows. The first task is formatting input and output of classifier; this task was done by developing simple software in Java. Next, the final classifier is performed to classify formatted input.

The final classifier was developed using Weka. The Weka workbench itself is a collection of modern machine learning algorithm and data pre processing tools. It includes most of state-of-the-art algorithms for doing classification, including preparing the input data, evaluating learning schemes statistically, and visualizing the input data and the result of learning.

A. Feature Engineering

Successful machine learning method involves not only selecting a learning algorithm, but transformations engineering between input and output is also important [2, 5]. In this experiment, the author performed data transformation, feature selection, cleansing data and detecting outliers. Let consider the following fragment of data set (development.data, train.data, test.data) and feature set (development.features, train.features, test.features) as can be seen in Table III and Table IV respectively.

TABLE III. SAMPLE DEVELOPMENT DATA

ID	Word	Class
1	are	cap1
2	we	nochange
3	going	nochange
4	to	nochange
5	remember	nochange

TABLE IV. SAMPLE DEVELOPMENT FEATURES

ID	Word	Tagging
1	be	VBP, _ , _ , _ , O, _ , _ , _ , _ , VBR
2	we	PRP, PRP, _ , _ , B-NP, B-NP, _ , _ , _ , _ , PPIS2, PPIS2
3	go	VBG, VBG, VBG, _ , _ , B-VP, B-VP, B-VP, _ , _ , _ , VVGK, VVGK, VVGK
4	to	TO, TO, TO, TO, _ , B-VP, I-VP, I-VP, I-VP, _ , _ , TO, TO, TO
5	remember	VB, VB, VB, VB, VB, B-VP, I-VP, I-VP, I-VP, I-VP, VV0, VV0, VV0, VV0, VV0

With respect to the data condition above, joining both of them are necessary. By developing simple java program,

¹ <http://www.computing.dcu.ie/~acahill/tagset.html>

both files are joined together. The result of this process is described in Table V.

TABLE V. OUTPUT OF JOIN FILE

ID	Word	Tagging	Class
1	be	VBP, _, _, _, O, _, _, _, _, VBR	cap1
2	we	PRP, PRP, _, _, B-NP, B-NP, _, _, PPIS2, PPIS2	nochange
3	go	VBG, VBG, VBG, _, B-VP, B-VP, B-VP, _, VVGK, VVGK, VVGK	nochange
4	to	TO, TO, TO, TO, B-VP, I-VP, I-VP, I-VP, _, TO, TO, TO	nochange
5	remember	VB, VB, VB, VB, VB, B-VP, I-VP, I-VP, I-VP, VV0, VV0, VV0, VV0, VV0	nochange

By applying Part of Speech (POS) tagging approach, Tagging column in Table 4 need to be split to better understanding of each attributes that has correspondents to class. The result of this task is depicted in Table VI as follows.

TABLE VI. OUTPUT OF JOIN FILE

Instance 1		Instance 2	
Attribute	Value	Attribute	Value
ID	1	ID	2
Word	be	Word	we
Penn-1	VBP	Penn-1	PRP
Penn-2	_	Penn-2	PRP
Penn-3	_	Penn-3	_
Penn-4	_	Penn-4	_
Penn-5	_	Penn-5	_
Chunk-1	O	Chunk-1	B-NP
Chunk-2	_	Chunk-2	B-NP
Chunk-3	_	Chunk-3	_
Chunk-4	_	Chunk-4	_
Chunk-5	_	Chunk-5	_
Claw-1	_	Claw-1	_
Claw-2	_	Claw-2	_
Claw-3	_	Claw-3	_
Claw-4	_	Claw-4	PPIS2
Claw-5	VBR	Claw-5	PPIS2
Class	nochange	Class	nochange

Using Weka tool, data with missing value are easily deleted. After running function deleteWithMissingClass() using Java+Weka, the quality of input data was gotten.

B. ARFF Format

The standard format in representing data set in Weka is ARFF file. To produce ARFF file, the author uses generate-arff script, was developed in perl language.

C. Feature Selection

Most of learning algorithms are designed to learn which are the most appropriate attributes to use for making best decisions. The negative effect of irrelevant attributes on the system is it will eliminate and make ambiguity of other attributes which could be the appropriate one [2]. The best way to select relevant attributes is manually, based on the understanding of the problem context and what the attributes mean.

In this problem, when we used all attributes (ID, word, all Tagging, and class), the system looks like give up to produce the output. Attribute ID is not relevant to the machine learning system, and attribute word is harmful, because if we use it as an input, our machine learning may to remember the word only. For example, if in the data set we have 10 words "we", 7 is categorised as nochange class, 2 as cap1 and 1 as nochange+comma, the built system will produce output for word we as the member of nochange class.

After eliminating 2 attributes, further investigating to other attributes are still crucial. Attribute selection could be done by machine learning algorithm [2]. First, the author tries to apply decision tree algorithm to the full data set, and then select only those attributes that are actually used in tree. Unfortunately, because of big data set (65,000 instances), the system always goes to not responding and out of memory.

Another way to doing selection is by using instanced-based learning method. The author takes sample of instances randomly from data set, then doing "near hits" and "near misses" analysis.

The result is, 4 attributes can be eliminated, 2 from chunks attribute and 1 from claw attribute and 1 from penn attributes.

D. Classifier Architecture

In this experiment, author implements classifier combination that consists of two classifiers and 1 meta classifiers to replace vote system in stacking design.

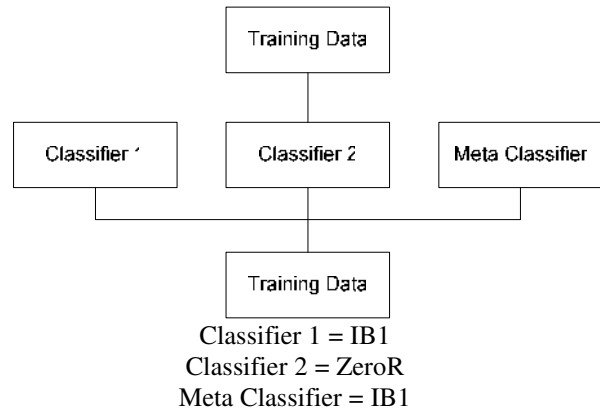


Figure 1. Design of Classifier Combination

To combine outputs, we use meta classifier rather than voting, because sometimes the majority prediction from several classifier is incorrect. Actually, voting is good if we have one classifier as an "expert", that we can trust.

Actually, decision tree is quite better classifier, unfortunately, author can't implement tree classifiers, because of big data sets, and so the memory is going to low.

E. Output

The performance of classifier combination within supplied development and training data set is described in the Table VII.

TABLE VII. RESULT OF DEVELOPMENT

Classes	F
allcaps	0.81
allcaps+fullstop	0.88
cap1	0.70
cap1+colon	0.72
cap1+fullstop	0.89
cap1-3	0.73
nochange	0.81
nochange+colon	0.64
nochange+comma	0.76

Overall Accuracy: F = 0.79

IV. EXPERIMENTS

From the result, classifier combination is quite good to solve telegraphese problem. Actually, after doing comparison with the same input data but using single classifier (IB1), overall accuracy increases significantly. IB1 has F score around 34% and ZeroR has F score around 42%.

Improvements could be done by focusing on the R score. Applying other machine learning algorithm could assist this problem. To get better performance, we have spent more time at pre-processing task. We can do like manipulating training set, manipulating learning algorithm, manipulating input features and class as well. In manipulating the training set, we can do sampling from original data, and then we develop classifier for each training set.

In manipulating input features and class can be performed. In this approach, we can focus on the weak performance result of class. For example, in this experiment, we split the class into two groups. The first group contains all class that have performance $F > 0$, and the second class who have F score = 0. Then, investigating for each attributes

are necessary to reduce redundancy and ambiguity. By applying symmetric uncertainty formula, redundant attributes could be detected. In manipulating learning algorithm, we can perform the experiment several times with the same data and algorithm.

V. CONCLUSION

Overall, classifier combination performs well than single classifier. Each classifier has advantage and disadvantage, by combining more than one classifier algorithm, better result will we get.

As was argued in experiment part, feature engineering is very important task in machine learning. There are number of improvements which could be done to these models such as re-feature selection to reduce over training and manipulating learning algorithm. These tasks can be expected to give modest boost to the performance of the best model.

ACKNOWLEDGMENT

The author would like to acknowledge Timothy Baldwin, The University of Melbourne, Australia, for his guidance and support in conducting the research presented in this paper.

REFERENCES

- [1] E. Brill and J. Wu, "Classifier Combination for Improved Lexical Disambiguation", Proc. of 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics – Volume 1 (COLING '98), 1998, pp 191-195, doi: 10.3115/980451.980876.
- [2] I. Witten, E. Frank and M. Hall, "Data mining: Practical Machine Learning, Tools and Technique", 3rd ed. Morgan Kaufmann Publishers, 2011.
- [3] J. Fiscus. A Post-Processing System to Yield Reduced Word Error Rates: Recognizer Output Voting Error Reduction (ROVER). Proc. of Automatic Speech Recognition and Understanding (ASRU 1997), Dec. 1997, pp. 347-354, doi: 10.1109/ASRU.1997.659110.
- [4] P. Brooks, "SCP: A Simple Chunk Parser", Artificial Intelligence Center, The University of Georgia. Athens, Georgia, 2003.
- [5] P. Tan, M. Steinbach and V. Kumar, "Introduction to Data Mining", Adison Wesley, 2006.
- [6] R. Navigli, "Word Sense Disambiguation: A Survey", ACM Computing Surveys (CSUR), vol. 41, issue. 2, Feb. 2009, pp. 1-69, doi: 10.1145/1459352.1459355.
- [7] S. Jayaraman and A. Lavie. "Multi-engine Machine Translation Guided by Explicit Word Matching", In Proc. of the ACL Interactive Poster and Demonstration Sessions, June 2005, pp. 101-104, doi: 10.3115/1225753.1225779.

CERTIFICATE

OF APPRECIATION

The Certificate is presented to

Leo Willyanto Santoso

for paper presentation at the 2011 International Conference on
Uncertainty Reasoning and Knowledge Engineering (URKE)
4 - 7 August 2011, Bali, INDONESIA





Proceedings of the IEEE

COUNTRY

United States



Universities and research institutions in United States

SUBJECT AREA AND CATEGORY

Computer Science
Computer Science (miscellaneous)
Engineering
Electrical and Electronic Engineering

Ad closed by Google

PUBLISHER

Institute of Electrical and Electronics Engineers Inc.

H-IND

27

PUBLICATION TYPE

Journals

ISSN

00189219, 15582256

COVERAGE

1927, 1963-2020


INFORMATION

Home
How to publish this journal
proceedings@ieee.org




Ad closed by Google

SCOPE

Proceedings of the IEEE provides in-depth review, survey, and tutorial coverage of the technical developments in electronics, electrical engineering, and computer science. Consistently ranked as one of the top journals by Impact Factor, it serves as a trusted resource for engineers around the world. The Proceedings publishes approximately ten special issues and two regular paper issues per year. -Special Issues are led by distinguished Guest Editor teams and contain articles, typically surveys, reviews or tutorials, from leading experts in the technology area being covered. They serve as a guide to state-of-the-art and are highly valued by the core research community, as well as specialists in other areas, looking to quickly come up to speed on the latest and more promising advances in areas outside of their own expertise. -Regular Issues consist of three to four papers on more focused topics, giving readers background and insight into emerging areas. Papers published in the Proceedings are usually reviews, surveys, or tutorials. -Reviews critically examine a technology, tracing its progress from its inception to the present—and perhaps into the future. -Surveys comprehensively view a technology—its applications, issues, ramifications, and potential. -Tutorial papers explain a technology and may give practical information for implementing it. These papers are written for the purpose of informing non-specialist engineers about a particular technology.

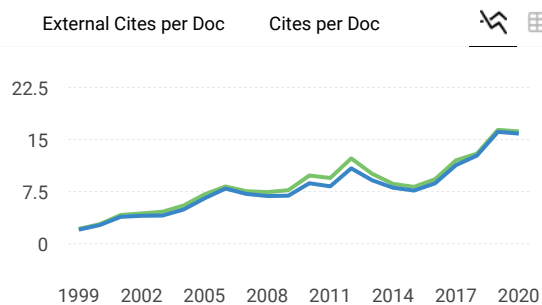
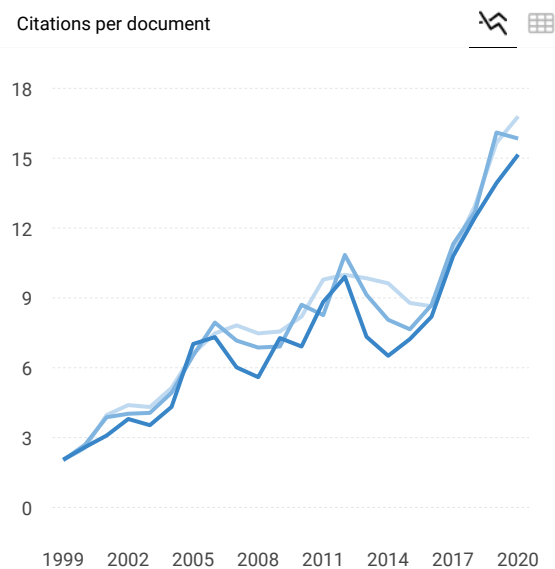
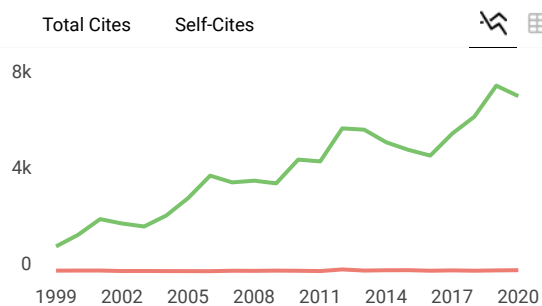
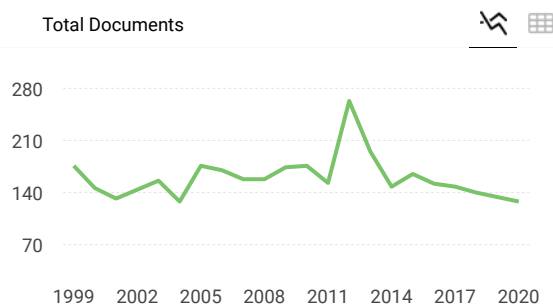
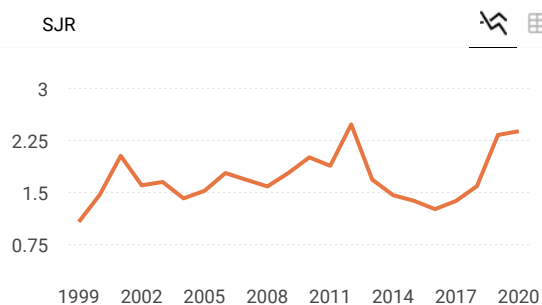
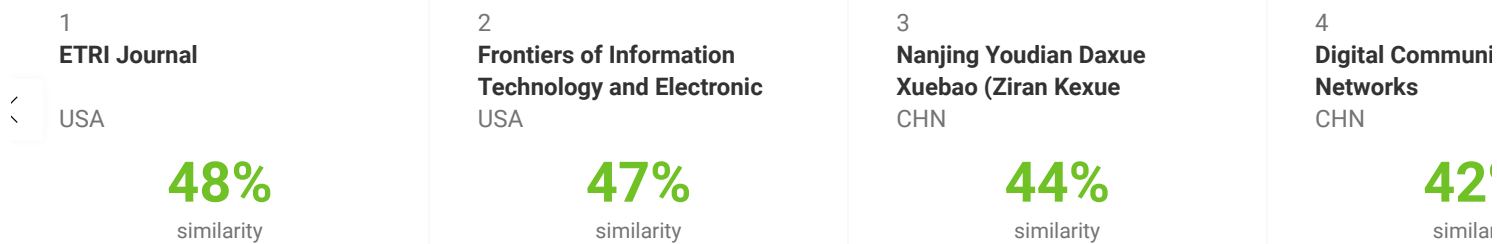
 Join the conversation about this journal

Ad closed by Google

 Quartiles



FIND SIMILAR JOURNALS 

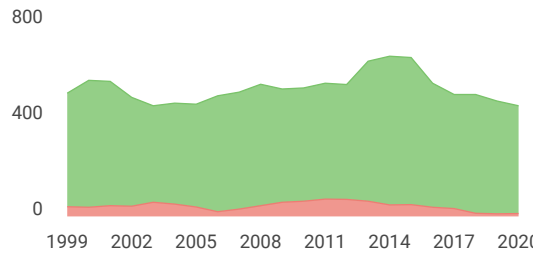
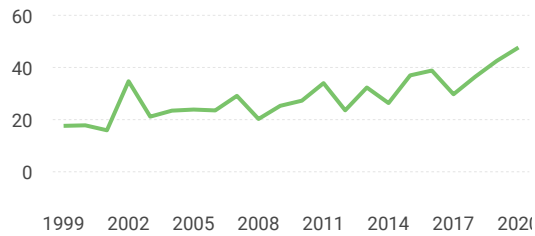
Ad closed by Google



% International Collaboration

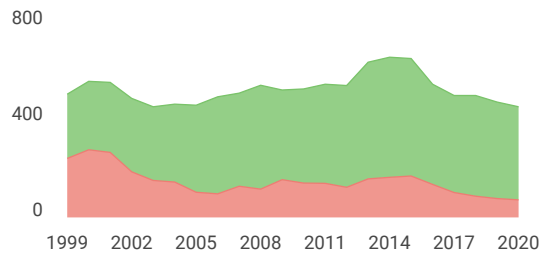
Citable documents

Non-citable documents



Cited documents

Uncited documents



Proceedings of the IEEE

Q1 Computer Science (miscellaneous) best quartile

SJR 2020 2.38

powered by scimagojr.com

← Show this widget in your own website

Just copy the code below and paste within your html code:

```
<a href="https://www.scimagojr.com" data-bbox="611 281 757 294">
```

SCImago Graphica

Explore, visually communicate and make sense of data with our new free tool.

Get it



Metrics based on Scopus® data as of April 2021

S **Sultan** 1 year ago

SJR shows this journal in Q1. What is the actual rank of this journal?

reply



Melanie Ortiz 1 year ago

SCImago Team

Dear Sultan, thank you for contacting us.

For every journal, the annual value of the SJR is integrated into the distribution of SJR values of all the thematic categories to which the journal belongs. There are more than 300 thematic categories. The position of each journal is different in any category and depends on the performance of the category, in general, and the journal, in particular .

Best Regards, SCImago Team



Melanie Ortiz 2 years ago

Dear Kason,
thank you for contacting us.
Sorry to tell you that SCImago Journal & Country Rank is not a journal. SJR is a portal with scientometric indicators of journals indexed in Elsevier/Scopus.
Unfortunately, we cannot help you with your request, we suggest you to visit the journal's homepage (see Scope) or contact the journal's editorial staff , so they could inform you more deeply. You can see the updated journal's information just above .
Best Regards, SCImago Team

Leave a comment

Name

Email

(will not be published)

I'm not a robot reCAPTCHA
Privacy - Terms

Submit

The users of Scimago Journal & Country Rank have the possibility to dialogue through comments linked to a specific journal. The purpose is to have a forum in which general doubts about the processes of publication in the journal, experiences and other issues derived from the publication of papers are resolved. For topics on particular articles, maintain the dialogue through the usual channels with your editor.

Developed by:



Powered by:



Follow us on @ScimagoJR

Scimago Lab, Copyright 2007-2020. Data Source: Scopus®

