



**2019 IEEE 10th International Conference on Awareness Science and
Technology (iCAST)
October 23-25, Morioka, Japan**

Proceedings

Sponsor:

Japan Chapter of IEEE Systems, Man and Cybernetics Society (SMCS).

Supporters:

Iwate Prefectural University, Japan,

IEEE SMCS Technical Committee on Awareness Computing,

IEEE CIS Task Force on Awareness Computing,

IEEE Sendai Section,

IEEE Society on Social Implications of Technology (SSIT).



Table of Contents

Keynote Speakers

Brain Machine Interface Systems - Translating Thoughts to Actions	i
<i>Vinod A Prasad</i>	
Signal Representation with Non-Linear Networks	ii
<i>Wen-Liang Hwang</i>	

Invited Papers

Re-identifying people from anonymous histories of their activities	1
<i>Hiroshi Yoshiura</i>	
Awareness of Privacy and Intellectual Property Rights under the Economic Partnership Agreement between EU and Japan	6
<i>Peter Burgstaller</i>	
Confront Phishing Attacks — from a Perspective of Security Education	10
<i>Toyoo Takata and Kanayo Ogura</i>	
Security and Privacy Awareness: for Software Creators and Users	14
<i>Glenn Mansfield Keeni and Hiroshi Tsunoda</i>	

General Session

Improve the generalization of the cross-task emotion classifier using EEG based on feature selection and SVR	18
<i>Liu Shuang, Wu Wenyi, Zhai Siyu, Liu Xiaoya, Ke Yufeng, An Xingwei and Ming Dong</i>	
A Study on Multi-objective Chaotic Evolution Algorithms Using Multiple Chaotic Systems	22
<i>Zitong Wang and Yan Pei</i>	
Classification of Sleep Disorder from Single Lead Non-overlapping of ECG-apnea based Non-Linear Analysis using Ensemble Approach	28
<i>Iman Fahruzi, I Ketut Eddy Purnama, Hideya Takahashi and Mauridhi Hery Purnomo</i>	
Parallel Implementation of Chaos Neural Networks for an Embedded GPU	34
<i>Zhongda Liu, Takeshi Murakami, Satoshi Kawamura and Hitoaki Yoshida</i>	
Computer-Aided Diagnosis for Lung Lesion in Companion Animals from X-ray Images Using Deep Learning Techniques	40
<i>Pattaramanee Arsomngern, Nichakorn Numcharoenpinij, Jitpinun Piriyataravet, Wutiwong Teerapan, Woranich Hinthong and Phond Phunchongharn</i>	
Using Smart Insoles and RGB Camera for Identifying Stationary Human Targets	46
<i>Sevendi Eldrige Rifki Poluan and Yan-Ann Chen</i>	
Acoustic-to-Articulatory Inversion of a Three-dimensional Theoretical Vocal Tract Model using Deep Learning-based Model	52
<i>Thanat Laphawn and Santitham Prom-On</i>	

Bridge Sub Structure Defect Inspection Assistance by using Deep Learning.....	57
<i>Pravee Kruachottikul, Nagul Cooharajanone, Gridsada Phanomchoeng, Thira Chavarnakul, Kittikul Kovitangoon, Donnaphat Trakulwaranont and Kanokwan Atchariyachanvanich</i>	
A Selective Modular Neural Network Framework	63
<i>Intisar Chowdhury and Qiangfu Zhao</i>	
Indoor Location Estimation of a Moving Subject Based on a Simple Motion Sensor Array	69
<i>Ryo Ota, Tatsuya Hanyu and Qiangfu Zhao</i>	
Misinformation Containment in OSNs leveraging Community Structure.....	75
<i>Arnab Kumar Ghoshal, Nabanita Das and Soham Das</i>	
A Proposal for Shape Aware Feature Extraction for Time Series Classification.....	81
<i>Hidetoshi Ito and Basabi Chakraborty</i>	
A Development of Ad-hoc Smart Room Based on User-oriented Context-aware System Architecture	87
<i>Kohei Suzuki, Haruki Igarashi, Yusuke Manabe and Kenji Sugawara</i>	
Liver Cancer Prediction using Synthetic Minority based on Probabilistic Distribution (SyMProD) Oversampling Technique.....	93
<i>Intouch Kunakornnum, Woranich Hinthong, Sumet Amonyinchareon and Phond Phunchongharn</i>	
A Deep Learning Model for Dimensional Valence-Arousal Intensity Prediction in Stock Market.....	99
<i>Jheng-Long Wu, Chi-Sheng Yang, Kai-Hsuan Liu and Min-Tzu Huang</i>	
Detecting Anomalous Events on Distributed Systems Using Convolutional Neural Networks	105
<i>Purimpat Cheansunan and Phond Phunchongharn</i>	
Risk-Sensitive Portfolio Management by using Distributional Reinforcement Learning.....	110
<i>Thammasorn Harnpadungkij, Warasinee Chaisangmongkon and Phond Phunchongharn</i>	
Static Analysis Method on Portable Executable Files for REMNIX based Malware Identification.....	116
<i>Muhammad Salman, Diyanatul Husna and Nindya Viani</i>	
A Novel Maximum-Likelihood Detection for the Binary MIMO System Using DC Programming	122
<i>Benying Tan, Xiang Li, Shuxue Ding, Yujie Li, Shotaro Akaho and Hideki Asoh</i>	
Knowledge Acquisition with Deep Fuzzy Inference Model and Its Application to a Medical Diagnosis.....	128
<i>Yuki Mori, Hirosato Seki and Masahiro Inuiguchi</i>	
Ambiguous Utterance Aware: Communication Learning System for Pragmatic Language Disorders Using a Humanoid Robot	134
<i>Keigo Yabuki and Kaoru Sumi</i>	
Non-reference Quality Assessment Model using Deep learning for Omnidirectional Images.	140
<i>Tung Q. Truong, Huyen Thi Thanh Tran and Truong Cong Thang</i>	

Data Acquisition Framework for Cloud Robotics	145
<i>Yutaka Watanobe, Yuichi Yaguchi, Toshimune Miyaji, Ryuhei Yamada and Keitaro Naruse</i>	
A Framework for Usability Design to Promote Awareness of Information Disseminated via Mobile Government Applications	152
<i>Pinnaree Kureerung and Lachana Ramingwong</i>	
Investigating Parallelization of Cross-language Plagiarism Detection System Using the Winnowing Algorithm in Cloud Based Implementation	158
<i>Anak Agung Putri Ratna, Fransiskus Astha Ekadiyanto, Ihsan Ibrahim, Diyanatul Husna and Fathimah Rahimullah</i>	
Latent Semantic Analysis and Winnowing Algorithm Based Automatic Japanese Short Essay Answer Grading System Comparative Performance.....	165
<i>Anak Agung Putri Ratna, Lea Santiar, Ihsan Ibrahim, Prima Dewi Purnamasari, Dyah Lalita Luhurkinanti and Adisa Larasati</i>	
Mobile EEG Based Drowsiness Detection using K-Nearest Neighbor	172
<i>Prima Dewi Purnamasari, Pratiwi Yustiana, Anak Agung Putri Ratna and Dodi Sudiana</i>	
Modeling and Prediction of Time-Series - A Case Study with Forex Data	177
<i>Yi Chen Shiao, Goutam Chakraborty, Shin Fu Chen, Li Hua Li and Rung Ching Chen</i>	
Dehazed Image Enhancement by a Gamma Correction with Global Limits.....	182
<i>Cheng-Hsiung Hsieh</i>	
Binary Owl Search Algorithm for Feature Subset Selection	186
<i>Ashis Kumar Mandal, Rikta Sen and Basabi Chakraborty</i>	
A Comparative Study of the Stability of Filter based Feature Selection Algorithms	192
<i>Rikta Sen, Ashis Kumar Mandal, Saptarsi Goswami and Basabi Chakraborty</i>	
Object Dynamics from Video Clips using YOLO Framework.....	198
<i>Wei-Hsiang Chung, Goutam Chakraborty, Rung-Ching Chen and Cédric Bornand</i>	
A Study on the Tourism Features Extraction from Photos in a Tourism Website by Image Analysis.....	204
<i>Shuang Li and Jun Sasaki</i>	
Definition and Goal of Graph Clustering - Motivation to Explore a New Algorithm	209
<i>Yuching Lu and Goutam Chakraborty</i>	
Selection of Effective Features for BGP Anomaly Detection.....	215
<i>Tastuya Arai, Kotaro Nakano and Basabi Chakraborty</i>	
<hr/>	
Special Session: Awareness Technology for Economic and Social Data Analysis	
A Re-Identification Strategy Using Machine Learning that Exploits Better Side Data	221
<i>Eina Hashimoto, Masatsugu Ichino and Hiroshi Yoshiura</i>	
Pattern Classification of Disasters Impact on Companies Stock Price	229
<i>Kenji Yamaguchi and Yukari Shirota</i>	

Sifting for Deeper Insights from Public Opinion: Towards Crowdsourcing and Big Data for Project Improvement	235
<i>Jean Marie Tshimula, Mary M. Njuguna, Thierry Roger Bayala, Mbuyi Mukendi Didier, Achraf Essemlali, Hugues Kanda and Numfor Solange Ayuni</i>	
Time Series Electricity Consumption Analysis using Non-negative Matrix Factorization ...	240
<i>Akira Kusaba, Tetsuji Kuboyama and Takako Hashimoto</i>	
Effective feature extraction from driving data for detection of danger awareness.....	246
<i>Kotaro Nakano and Basabi Chakraborty</i>	
<hr/>	
Special Session: The Data Analysis and Intelligent Computing for Awareness Applications	
<hr/>	
Finding Longest (s, t)-paths of O-shaped Supergrid Graphs in Linear Time.....	252
<i>Ruo-Wei Hung, Fatemeh Keshavarz-Kohjerdi, Yuh-Min Tseng and Guo-Hao Qiu</i>	
Quality Enhancement of P2P using Interested Clusters in the P4P Architecture	260
<i>Shu-Ching Wang, Hsiu-Wei Hsu, Hung-Hsuan Wang, Hsin-Hung Pan and Po-Chien Lin</i>	
Use Product Segmentation to Enhance the Competitiveness of Enterprises in the IoT.....	266
<i>Shu-Ching Wang, Hsiu-Wei Hsu, Chan-Grong Dai, Chin-Ling Ho and Fang-Yu Zhang</i>	
The Data Analysis for the Influence of Cross-cultural Training on the Adaptability of Foreign Workers.....	272
<i>Shu-Ching Wang, Wei-Shu Hsiung, Yi-Syuan Lin, Rong-Chung Chen and Hsin-Hung Pan</i>	
Alterations in brainwaves caused by different Music Genres	278
<i>Shao Kuo Tai and Yu-jen Kuo</i>	
Similar Music Instrument Detection via Deep Convolution YOLO-Generative Adversarial Network	282
<i>Christine Dewi, Rung-Ching Chen, Hendry Hong and Yan-Ting Liu</i>	
Estimation of the Kansei Information obtained from Musical Scores via Machine Learning Algorithms	288
<i>Satoshi Kawamura, Zhongda Liu and Hitoaki Yoshida</i>	
A Study of Social Media Reviews Effects on the Success of Crowdfunding Projects.....	293
<i>Long-Sheng Chen and Ying-Jung Chuang</i>	
A Study of Influences of Social Network Awareness on MOOC Learner Behaviors: Case of Chulalongkorn University Free MOOC.....	298
<i>Nagul Cooharajanone, Orawun Moolpun, Pitchapa Pawong, Jidapa Dilokpabhaphat, Thanaporn Rimnong-Ang, Manutsaya Choosuan, Pattamon Bunram and Suporn Pongnumbkul</i>	
Productivity-based Features from Article Metadata for Fuzzy Rules to Classify Academic Expert	304
<i>Diana Purwitasari, Chastine Fatichah, Adri Gabriel Sooi, Surya Sumpeno and Mauridhi Hery Purnomo</i>	
A Short Answer Grading System in Chinese by CNN	310
<i>Shih-Hung Wu and Chun-Yu Yeh</i>	

Route-Based Ship Classification	315
<i>Shoichi Ichimura and Qiangfu Zhao</i>	

Special Session: Technologies for Women – The Good, the bad and the ugly

Breast Cancer Detection Based on Medio-Lateral ObliqueView and Cranio-Caudal View Mammograms: An Overview	321
<i>Sasikala.S, Bharathi M, Ezhilarasi M and Arunkumar S</i>	
Diagnosis of Diabetes using an Extreme Learning Machine Algorithm based Model	327
<i>Shanthi M, Ramalatha M Marimuthu, Shivappriya S.N. and Navaneethakrishnan R</i>	
A Survey about WSN and IoT Based Health Care Applications and ADPLL Contribution for Health Care Systems	331
<i>Dinesh Ramachandran and Dr.Ramalatha Marimuthu</i>	
PERFORMANCE COMPARISON OF MACHINE LEARNING ALGORITHMS TO PREDICT LABOR COMPLICATIONS AND BIRTH DEFECTS BASED ON STRESS ..	339
<i>Madhusri V, Kesav Krishnan G, Ramalatha Marimuthu and Sathyanarayanan Ramesh</i>	
INFLUENCE OF ILLITERACY ON MATERNAL AND CHILD HEALTH AMONG WOMEN IN INDIA.....	343
<i>Shivappriya S.N, Shanthi M, Kesav Krishnan G, Sathyanarayanan Ramesh and Ramalatha Marimuthu</i>	
Maternal ehealth Monitoring System using LoRa Technology	349
<i>Alagumeenaakshi Muthiah, Kavitha K, Bharathi M, Ajitha S, Monisha Thangam K S, Viveka Vikram K and Ramalatha Marimuthu</i>	
SMART SHOE FOR WOMEN SAFETY	353
<i>Vishesh Sharma, Yati Tomar and Vydeki D</i>	

Special Session: Intuitive Human-System Interaction

Equilibrium Selective Role Coordination for Autonomous Driving	357
<i>Naoto Iwahashi</i>	
Interactive Evolutionary Computation System Using Multiple Users' Gaze Information Considering User's Partial Evaluation Participation	365
<i>Hiroshi Takenouchi and Masataka Tokumaru</i>	
Implementation of multi-modal interface for VR application	370
<i>Yushi Machidori, Ko Takayama and Kaoru Sugita</i>	
Method for left atrial appendage segmentation using heart CT images	374
<i>Itaru Takayashiki, Akio Doi, Tour Kato, Hiroki Takahashi, Syoto Sekimura, Maiko Hozawa and Yoshihiro Morino</i>	
Physics Projection	379
<i>Naoto Iwahashi</i>	

Special Session: Text Mining and Analytics

Imbalanced Twitter Sentiment Analysis using Minority Oversampling	384
<i>Kushankur Ghosh, Arghasree Banerjee, Sankhadeep Chatterjee and Soumya Sen</i>	

Cyberbullying Detection on Twitter using Multiple Textual Features	389
<i>Jianwei Zhang, Taiga Otomo, Lin Li and Shinsuke Nakajima</i>	
Method to Present Situation Aware Suitable Reply Candidate on Mobile Terminal	395
<i>Norifumi Watanabe, Masafumi Matsuhara, Goutam Chakraborty and Hiroshi Mabuchi</i>	
Improving Awareness of Emotional Meaning of Emoticon by Representing as Numerical Vectors	401
<i>Yuki Okude, Masafumi Matsuhara, Goutam Chakraborty and Hiroshi Mabuchi</i>	
An Approach for Automatic Aspect Extraction by Latent Dirichlet Allocation	407
<i>Subha Jyoti Das and Basabi Chakraborty</i>	
Study of recurrent neural networks and Utilization of Oversampling for multiclass sentiment analysis	413
<i>Anirban Mukherjee, Sabyasachi Mukhopadhyay, Prasanta K. Panigrahi and Saptarsi Goswami</i>	
Comparison of Sentiment Analysis of Review Comments by Unsupervised Clustering of Features Using LSA and LDA	419
<i>Shu-Cih Tseng, Yu-Ching Lu, Goutam Chakraborty and Long-Sheng Chen</i>	
<hr/> Special Session: Disaster Awareness Communication Networks, Systems, and Applications <hr/>	
Implementation of a Quality Measurement Software for Virtual Reality Content on a head mount display	425
<i>Ko Takayama, Yushi Machidori and Kaoru Sugita</i>	
Proposal on the Disaster Information System based on QoS Controls of Heterogenous Radio Networks and Radio Atmosphere	429
<i>Noriki Uchida, Kanji Nakano, Chiho Takegawa, Tomoyuki Ishida and Yoshitaka Shibata</i>	
Evaluation of Performance on LPWA Network Realizes for Multi-wavelength Cognitive V2X Wireless System	434
<i>Akira Sakuraba, Yoshitaka Shibata and Toshihiro Tamura</i>	
<hr/> Special Session: Security and Privacy in Data-Aware Systems <hr/>	
Second Level Steganalysis - Secret Data Position Prediction Based on Machine Learning ..	441
<i>Takumi Saito, Zhao Qiangfu and Hiroshi Naito</i>	
Reversible Data Hiding Scheme based on Difference Expansion using Shiftable Block Strategy for Enhancing Image Fidelity	447
<i>Chin-Feng Lee, Jau-Ji Shen, Yi-Jhen Wu and Somya Agrawal</i>	
Effective Dual-images based Reversible Information Hiding Scheme based on Complexity Analysis and Thresholds Controlling	453
<i>Tzu-Chuen Lu, Jau-Ji Shen and Ting-Chi Chang</i>	
Hierarchical MQTT with Edge Computation	457
<i>Hung-Yu Chien, Guo-Hao Qiu, Ruo-Wei Hung, An-Tong Shih and Chunhua Su</i>	
Image Steganographic Scheme Based on Double-layer Magic Matrix	462
<i>Jau-Ji Shen, Chin-Feng Lee, Yen-Hsi Li and Somya Agrawal</i>	

Exploration of Fast Edible Oil Classification Using Infrared Spectrum, Machine Learning, and Chemometrics	468
<i>Hung-Yu Chien, An-Tong Shih and Y. M. Tseng</i>	
Fault-Tolerant Visual Secret Sharing Scheme Using Meaningful Shares	474
<i>Yu-Chun Chung, Jia-Hao Ou and Justie Su-Tzu Juan</i>	
A Study on Security Trend based on News Analysis	480
<i>Chia-Mei Chen, Dan-Wei Marian Wen, Jun-Jie Fang, Gu-Hsin Lai and Yi-Hung Liu</i>	
Device-Free Localization via Sparse Coding with Log-Regularizer	484
<i>Zhaoyang Han, Chunhua Su, Shuxue Ding, Huakun Huang and Lingjun Zhao</i>	
Applying Convolutional Neural Network for Malware Detection	490
<i>Chia-Mei Chen, Shi-Hao Wang, Dan-Wei Wen, Gu-Hsin Lai and Ming-Kung Sun</i>	
A New Steganography Method Based on Generative Adversarial Networks	495
<i>Hiroshi Naito and Qiangfu Zhao</i>	

Special Session: Advances on Biomedical Informatics and Engineering for Healthcare

Recognition of Nerve and Dura Mater in Spinal Endoscopic Images Using Deep Learning .	501
<i>Peng Cui, Zhe Guo, Jianbo Xu, Tianhui Li and Wenxi Chen</i>	
Improvement of ECG based Personal Identification Performance In Different Bathtub Water Temperature By CNN	506
<i>Jianbo Xu, Tianhui Li, Peng Cui and Wenxi Chen</i>	
Utilizing Subject-Specific Discriminative EEG Features for Classification of Motor Imagery Directions	511
<i>Kavitha Thomas, Neethu Robinson and Vinod A P</i>	
Continuous Cuffless Blood Pressure Estimation Based on Pulse Transit Time: A New Evaluation with Invasive BP Reference	516
<i>Zunyi Tang, Ying Chen, Shing-Hong Liu, Toshio Kobayashi and Wenxi Chen</i>	
Sleep Stage Classification Based on EEG, EOG, and CNN-GRU Deep Learning Model....	521
<i>S.M. Isuru Niroshana, Xin Zhu, Ying Chen and Wenxi Chen</i>	
A New Approach to Classify Knee Osteoarthritis Severity from Radiographic Images based on CNN-LSTM Method	528
<i>Rima Tri Wahyuningrum, Lilik Anifah, I Ketut Eddy Purnama and Mauridhi Hery Purnomo</i>	
The Effect of Stress on Optimal Bathing Time	534
<i>Tianhui Li, Jianbo Xu, Peng Cui and Wenxi Chen</i>	

Special Session: Medical Data Analytics

Deep Learning for Automatic Identification of Nodule Morphology Features and Prediction of Lung Cancer	539
<i>Weilun Wang and Goutam Chakraborty</i>	

Engineering for psychiatry: a case study.....	545
<i>Cedric Bornand, Angelika Güsewell, Alexia Stantzos, Emilie Bovet, Gilles Bangerter and Matthieu Thomas</i>	
A machine-learning approach for predicting success in smoking cessation intervention.....	551
<i>Khishigsuren Davagdorj and Keun Ho Ryu</i>	
Identification of DNA Methylation Signatures for Diagnosis of Lung Adenocarcinoma.....	557
<i>Yongjun Piao, Kwang Ho Park, Keun Ho Ryu and Rong Xiang</i>	
NOVEL FEATURES FOR DIAGNOSIS OF PARKINSON'S DISEASE FROM OFF-LINE ARCHIMEDEAN SPIRAL IMAGES.....	562
<i>Jija Dasgupta and Bhabatosh Chanda</i>	
<hr/> Appendix <hr/>	
Keyword Index.....	568
Author Index.....	578

Author Index

A P, Vinod	511
Agrawal, Somya	447, 462
Akaho, Shotaro	122
Amonyingchareon, Sumet	93
Anifah, Lilik	528
Arai, Tastuya	215
Arsomngern, Pattaramanee	40
Asoh, Hideki	122
Atchariyachanvanich, Kanokwan	57
Banerjee, Arghasree	384
Bangerter, Gilles	545
Bayala, Thierry Roger	235
Bornand, Cedric	545
Bornand, Cédric	198
Bovet, Emilie	545
Bunram, Pattamon	298
Burgstaller, Peter	6
Chaisangmongkon, Warasinee	110
Chakraborty, Basabi	81, 186, 192, 215, 246, 407
Chakraborty, Goutam	177, 198, 209, 395, 401, 419, 539
Chanda, Bhabatosh	562
Chang, Ting-Chi	453
Chatterjee, Sankhadeep	384
Chavarnakul, Thira	57
Cheansunan, Purimpat	105
Chen, Chia-Mei	480, 490
Chen, Long-Sheng	293, 419
Chen, Rong-Chung	272
Chen, Rung Ching	177
Chen, Rung-Ching	198, 282
Chen, Shin Fu	177
Chen, Wenxi	501, 506, 516, 521, 534
Chen, Yan-Ann	46
Chen, Ying	516, 521
Chien, Hung-Yu	457, 468
Choosuwan, Manutsaya	298
Chowdhury, Intisar	63
Chuang, Ying-Jung	293
Chung, Wei-Hsiang	198
Chung, Yu-Chun	474
Cooharojananone, Nagul	57, 298
Cui, Peng	501, 506, 534

D, Vydeki	353
Dai, Chan-Grong	266
Das, Nabanita	75
Das, Soham	75
Das, Subha Jyoti	407
Dasgupta, Jija	562
Davagdorj, Khishigsuren	551
Dewi, Christine	282
Dilokpabhaphat, Jidapa	298
Dinesh Ramachandran,	331
Ding, Shuxue	122, 484
Doi, Akio	374
Dong, Ming	18
Dr.Ramalatha Marimuthu,	331
Ekadiyanto, Fransiskus Astha	158
Essemlali, Achraf	235
Fahruzi, Iman	28
Fang, Jun-Jie	480
Fatichah, Chastine	304
G, Kesav Krishnan	339, 343
Ghosh, Kushankur	384
Ghoshal, Arnab Kumar	75
Goswami, Saptarsi	192, 413
Guo, Zhe	501
Güsewell, Angelika	545
Han, Zhaoyang	484
Hanyu, Tatsuya	69
Harnpadungkij, Thammasorn	110
Hashimoto, Eina	221
Hashimoto, Takako	240
Hinthong, Woranich	40, 93
Ho, Chin-Ling	266
Hong, Hendry	282
Hozawa, Maiko	374
Hsieh, Cheng-Hsiung	182
Hsiung, Wei-Shu	272
Hsu, Hsiu-Wei	260, 266
Huang, Huakun	484
Huang, Min-Tzu	99
Hung, Ruo-Wei	252, 457
Husna, Diyanatul	116, 158
Ibrahim, Ihsan	158, 165

Ichimura, Shoichi	315
Ichino, Masatsugu	221
Igarashi, Haruki	87
Inuiguchi, Masahiro	128
Ishida, Tomoyuki	429
Ito, Hidetoshi	81
Iwahashi, Naoto	357, 379
Juan, Justie Su-Tzu	474
K S, Monisha Thangam	349
K, Kavitha	349
K, Viveka Vikram	349
Kanda, Hugues	235
Kato, Tour	374
Kawamura, Satoshi	34, 288
Keeni, Glenn Mansfield	14
Keshavarz-Kohjerdi, Fatemeh	252
Kobayashi, Toshio	516
Kovitangoon, Kittikul	57
Kruachottikul, Pravee	57
Kuboyama, Tetsuji	240
Kunakorntum, Intouch	93
Kuo, Yu-jen	278
Kureerung, Pinnaree	152
Kusaba, Akira	240
Lai, Gu-Hsin	480, 490
Laphawn, Thanat	52
Larasati, Adisa	165
Lee, Chin-Feng	447, 462
Li, Li Hua	177
Li, Lin	389
Li, Shuang	204
Li, Tianhui	501, 506, 534
Li, Xiang	122
Li, Yen-Hsi	462
Li, Yujie	122
Lin, Po-Chien	260
Lin, Yi-Syuan	272
Liu, Kai-Hsuan	99
Liu, Shing-Hong	516
Liu, Yan-Ting	282
Liu, Yi-Hung	480
Liu, Zhongda	34, 288
Lu, Tzu-Chuen	453
Lu, Yu-Ching	419
Lu, Yuching	209

Luhurkinanti, Dyah Lalita	165
M, Bharathi	321, 349
M, Ezhilarasi	321
M, Shanthi	327, 343
Mabuchi, Hiroshi	395, 401
Machidori, Yushi	370, 425
Manabe, Yusuke	87
Mandal, Ashis Kumar	186, 192
Marimuthu, Ramalatha	339, 343, 349
Marimuthu, Ramalatha M	327
Matsuhara, Masafumi	395, 401
Mbuyi Mukendi Didier,	235
Miyaji, Toshimune	145
Moolpun, Orawun	298
Mori, Yuki	128
Morino, Yoshihiro	374
Mukherjee, Anirban	413
Mukhopadhyay, Sabyasachi	413
Murakami, Takeshi	34
Muthiah, Alagumeenaakshi	349
Naito, Hiroshi	441, 495
Nakajima, Shinsuke	389
Nakano, Kanji	429
Nakano, Kotaro	215, 246
Naruse, Keitaro	145
Niroshana, S.M. Isuru	521
Njuguna, Mary M.	235
Numcharoenpinij, Nichakorn	40
Numfor Solange Ayuni,	235
Ogura, Kanayo	10
Okude, Yuki	401
Ota, Ryo	69
Otomo, Taiga	389
Ou, Jia-Hao	474
Pan, Hsin-Hung	260, 272
Panigrahi, Prasanta K.	413
Park, Kwang Ho	557
Pawong, Pitchapa	298
Pei, Yan	22
Phanomchoeng, Gridsada	57
Phunchongharn, Phond	40, 93, 105, 110
Piao, Yongjun	557
Piriyataravet, Jitpinun	40
Poluan, Sevend Eldrige Rifki	46

Pongnumbkul, Suporn	298
Prom-On, Santitham	52
Purnama, I Ketut Eddy	28, 528
Purnamasari, Prima Dewi	165, 172
Purnomo, Mauridhi Hery	28, 304, 528
Purwitasari, Diana	304
Qiangfu, Zhao	441
Qiu, Guo-Hao	252, 457
R, Navaneethakrishnan	327
Rahimullah, Fathimah	158
Ramesh, Sathyanarayanan	339, 343
Ramingwong, Lachana	152
Ratna, Anak Agung Putri	158, 165, 172
Rimnong-Ang, Thanaporn	298
Robinson, Neethu	511
Ryu, Keun Ho	551, 557
S, Ajitha	349
S, Arunkumar	321
S.N, Shivappriya	343
S.N., Shivappriya	327
Saito, Takumi	441
Sakuraba, Akira	434
Salman, Muhammad	116
Santiar, Lea	165
Sasaki, Jun	204
Sasikala.S,	321
Seki, Hirosato	128
Sekimura, Syoto	374
Sen, Rikta	186, 192
Sen, Soumya	384
Sharma, Vishesh	353
Shen, Jau-Ji	447, 453, 462
Shiao, Yi Chen	177
Shibata, Yoshitaka	429, 434
Shih, An-Tong	457, 468
Shirota, Yukari	229
Shuang, Liu	18
Siyu, Zhai	18
Sooai, Adri Gabriel	304
Stantzios, Alexia	545
Su, Chunhua	457, 484
Sudiana, Dodi	172
Sugawara, Kenji	87
Sugita, Kaoru	370, 425
Sumi, Kaoru	134

Sumpeno, Surya	304
Sun, Ming-Kung	490
Suzuki, Kohei	87
Tai, Shao Kuo	278
Takahashi, Hideya	28
Takahashi, Hiroki	374
Takata, Toyoo	10
Takayama, Ko	370, 425
Takayashiki, Itaru	374
Takegawa, Chiho	429
Takenouchi, Hiroshi	365
Tamura, Toshihiro	434
Tan, Benying	122
Tang, Zunyi	516
Teerapan, Wutiwong	40
Thang, Truong Cong	140
Thi Thanh Tran, Huyen	140
Thomas, Kavitha	511
Thomas, Matthieu	545
Tokumar, Masataka	365
Tomar, Yati	353
Trakulwaranont, Donnaphat	57
Truong, Tung Q.	140
Tseng, Shu-Cih	419
Tseng, Y. M.	468
Tseng, Yuh-Min	252
Tshimula, Jean Marie	235
Tsunoda, Hiroshi	14
Uchida, Noriki	429
V, Madhusri	339
Viani, Nindya	116
Wahyuningrum, Rima Tri	528
Wang, Hung-Hsuan	260
Wang, Shi-Hao	490
Wang, Shu-Ching	260, 266, 272
Wang, Weilun	539
Wang, Zitong	22
Watanabe, Norifumi	395
Watanobe, Yutaka	145
Wen, Dan-Wei	490
Wen, Dan-Wei Marian	480
Wu, Jheng-Long	99
Wu, Shih-Hung	310
Wu, Yi-Jhen	447

Xiang, Rong	557
Xiaoya, Liu	18
Xingwei, An	18
Xu, Jianbo	501, 506, 534
Yabuki, Keigo	134
Yaguchi, Yuichi	145
Yamada, Ryuhei	145
Yamaguchi, Kenji	229
Yang, Chi-Sheng	99
Yeh, Chun-Yu	310
Yoshida, Hitoaki	34, 288
Yoshiura, Hiroshi	1, 221
Yufeng, Ke	18
Yustiana, Pratiwi	172
Zhang, Fang-Yu	266
Zhang, Jianwei	389
Zhao, Lingjun	484
Zhao, Qiangfu	63, 69, 315, 495
Zhu, Xin	521

Productivity-based Features from Article Metadata for Fuzzy Rules to Classify Academic Expert

Diana Purwitasari^{1,2}, Chastine Fatichah², Adri Gabriel Sooi^{1,3}, Surya Sumpeno^{1,4}, Mauridhi Hery Purnomo^{1,4}

¹Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

²Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

³Department of Computer Science, Universitas Katolik Widya Mandira Kupang, Indonesia

⁴Department of Computer Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

diana@if.its.ac.id, chastine@cs.its.ac.id, adrigabriel@ieee.org, surya@ee.its.ac.id, hery@ee.its.ac.id

Abstract—Since modeling expertise is necessary in an expert recommendation system, this paper addressed the issue to obtain researcher expertise in the academic field on certain topic interest. The profile considers productivity and dynamicity of an expert. The productivity of research activities through published articles as research output determine expertise that changes over time to indicate the dynamicity aspect. Here, the resulted expertise status on certain topic interest augments the expert profile. However, the expertise status is unavailable in the expert finder dataset. This paper discussed on approaches to classify the status from features of productivity and dynamicity in the form of fuzzy rules, which can be applied later in the expert recommendation system. Then, the approaches include of determining topics, mapping expertise candidates, extracting features, and labeling expertise status for training to generate fuzzy rules. Because of unavailable expertise status, to get better labels, the results of linear model and clustering were compared. Based on the empirical experiments, rules trained from scaled data with expertise labels from fuzzy clustering gave better results. After simplifying the rules, if-then forms with two features were representable enough for identifying the status of specialist or thriving experts on a topic interest.

Keywords—expert profile, productivity and dynamicity features, expertise status on topics, generating fuzzy rules

I. INTRODUCTION

In the academic field, an expert finder recommends experts as collaboration candidates [1] like being academic supervisors [2] regardless of similar or cross-domain research interests [3][4]. Experts or researchers in universities are lectures who often have various interests that not always restricted to the customary fields within departments or faculties. One of the major steps in the recommendation process is to model experts and identified tacit knowledge as their expertise by extracting information from document contents concerned with them [5]. The expertise is extracted from metadata texts of published articles like title or abstract. Because recommendations in real-world scenarios have scarce information of expertise, the unsupervised clustering experts based on metadata texts with [4] or without [3] word embedding for keyword weighting became a common approach. Then, the system recommends experts in one cluster with the same interest relatedness to the finder.

Instead of keywords, experts also can be clustered through the dynamic performance of research activities after monitored over time [6]. The performance is influenced by the number of papers published as scientific articles defined by productivity term and the research impact of citations. Both productivity and impact were also considered with regards to co-authors as collaboration aspect. Then those indicators were monitored over

time to obtain the dynamic performance of researchers. But researchers can change their interest [7] such that the expertise degree in certain domain becomes dynamically varied. General models in expert recommendation system are expertise-oriented and topic-oriented [5]. The first model generates expert profiles and ranks them based on the user query, while the second model searches related documents to the query and finds the experts. Recommendations in the academic field are accustomed to topical representations called as meta topics which are extracted from titles and abstracts of articles as parts of text mining related applications [8]. Meta topics are used to reduce the dimensions of a document vector representation, thus make a faster process in the recommendation system. Word embedding for weighting terms is also used in extracting the topics. Here, the topics are further applied in mapping the research domains. This paper used both models by creating expert profile from productivity and dynamicity aspects for expertise-oriented, then meta topics from documents are mapped to the experts for topic-oriented.

The expert profile is complemented with expertise status on topics of research interests which can be used later in an expert recommendation system. Scarce information with real-world scenarios includes the absence of topics and expertise status on topics in commonly used expert finder dataset of Association for Computational Linguistics (ACL) [9] or AMiner (Academic Network Miner) [10]. Thus, to provide the expertise status, here we focused on utilizing the productivity-dynamicity based features to classify researchers without topic information. The expertise status labels the experts as specialist or thriving ones based on research activities that causes dynamic aspect in expert productivities. Unavailable (meta) topics are obtained from text mining of article metadata as tacit knowledge engineering of the experts. In this paper, we proposed feature extractions for productivity and dynamicity aspects of research activities. The expertise status on topic interests defined from comparing the results of linear model and clustering on the features also became another significance. Both features and expertise status as target were used to generate fuzzy rules to classify the status.

An expert recommendation system with awareness context can suggest suitable collaborators according to the needs and conditions of the collaboration seekers, like awareness in the context of mutual interest for finding collaborators between university and industry [11]. Here, we focused on the conditions that may easily change because of productivity and dynamicity research activities. To avoid rigid settings in discretize values of productivity and dynamicity features, fuzzy rules is preferable in applying recommendation process. For instance, one expert

who continues to annually publish 1-2 articles of a topic in some periods and another one who publishes 2-3 articles with one year gap can be considered to have the same medium level in features related to publication numbers. Fuzzy rules generated from the expert features that set in the recommendation system is going to make the system has awareness context in finding the experts.

After presenting problem background and our contributions, following discussions are about related works in expert profiles that considering research domains in the profiles. Although lists of experts based on domains are provided in standard datasets, there is no information about expertise on topics of research interests. Meta topics were used in the experiments to represent the research interests, consequently the later section is about data processing tailored to create dataset for our problem. After describing the proposed steps to classify experts, we discussed the results in empirical settings of selected research domains.

II. RELATED WORKS

Recommendation process with awareness to the context of research interest is prevalent on the expert finder system in the academic field [2][3][4][11]. Similarities between experts for recommendation consider the expert profiles to represent their expertise. The similarities can be obtained without text mining of title-abstract but using intersection-union of authors [2]. It was common for mining the unstructured texts using clustering to explore interest similarities [3][4]. Thorough calculations investigated keyword relevance in the context of topic interest with considering author networks [11].

Monitoring research activities through the published articles over time detects expert potential. Recent approaches exploit time span in profiling the experts [6][7][12] like to recommend advisors for supervising tasks [6]. In pursuing a scientific career, rising stars is detected from productivity-dynamicity in research activities. However, previous work on detecting the rising starts did not emphasize on domains [6]. Monitoring activities also considered time span that may influence topic drifts or interest changes [7]. In finding experts, the time concept to validate expertise through question and answer networks was explored as well [12]. The used assumption in the questionnaire was that experts had been encouraged to share knowledge by responding to the questions related to their expertise. Here, productivity-dynamicity based features which used in rising star detection have been modified [6]. The modification adapted topic domains which is unavailable information as the issue to classify experts.

Expert recommendations of real-world scenarios had better models to some extent by applying fuzziness, such as reviewer assignment to appraise proposals [13]. Without fuzzy approach, the problem of reviewer and reviewee in the academic field used document keyword similarities [14]. Before setting reviewers, self-evaluation was conducted to get matching degrees between reviewers and proposals according to the reviewers themselves. Then, the values of matching degrees were used for generating the fuzzy model.

Based on previous works, the approaches with clustering and time span are often used in recommending process. Time span is associated with weighing down expertise or giving some penalty values of the candidates because of interest changes. There is no information of expertise on topic interest with real-world data in this paper. Accordingly, our approaches handle the

incomplete data problem through extracting topics that is called as meta topics, mapping the topics as candidate expertise though article topic labeling, obtaining productivity-dynamicity based features, and generating fuzzy rules to classify experts for the expertise status on topic interest as classes. Then the results are going to be utilized later by an expert recommendation system which is not discussed in this paper (Fig. 1).

III. DATASET PREPARATION

This paper used an article collection with bibliographic data of authors, titles, abstracts, and references. Most researches on bibliographic data work on datasets of ACL (Association for Computational Linguistics) [9] or AMiner [10] which have research domains. The expert lists of some domains have been provided [15]. The experiments with selected AMiner experts were using two related domains of Natural Language Processing (NLP) and Information Extraction (IE). Some NLP-IE experts who indicated author name disambiguation and only had less than 20 articles were removed. Because of the first removal rule, the results were 70 experts with their $\pm 4,800$ articles (Fig. 2).

The second removal rule ensured that selected experts had collaboration experience by only including researchers who at least has seven articles within AMiner database. They should co-author with 70 experts in the NLP-IE list. Then, the results of 51 experts were semi-manually labeled as a specialist or thriving expert in certain topic interests following some rules. Labeling process with both approaches of linear model and clustering are discussed later as well as their result comparisons.

There is no mapping between experts and articles to the topic interests since AMiner dataset does not provide research topics. Thus, after selecting experts, dataset preparation was about meta topics with clustering approach. We clustered only the title texts. Text preprocessing on article titles of 70 experts removed words with three characters or less, stop words, and words appeared in fewer than ten articles. The results of text preprocessing were 3,500 candidate keywords. Based on best practices, next process was Word2Vec word embedding with 200 dimensions [16] and ignored words with document frequency less than ten. The embedded results also took out words with an average embedded weight value less than 0.04 and generated a 3,500 x 200 feature matrix of remaining keywords. The dimension means that each word has 200 values as the relation weights. A negative value in certain dimension reveals the negative sentiment of the word.

The next process was clustering words in the embedding result with KMeans++ and ensuring that each cluster has at least 100 words. With Silhouette indicators to measure the cluster goodness [17], there were K=30 clusters of words from 3,500 keywords. Meta topics or clusters of words resulted from the clustering process were used to map the expertise of researchers.

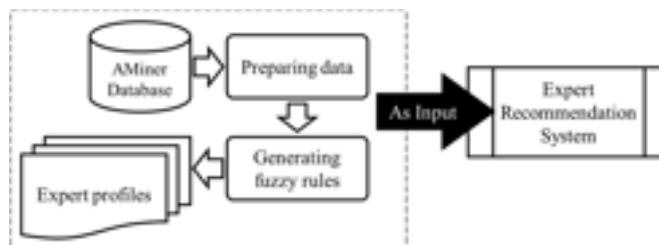


Fig. 1 General processes in the proposed approach

Meta topics are shortened as the topics for the rest of this paper. And so, the preparation results as shown in Fig. 2 are Data-1 (experts), Data-2 (articles), and Data-3 (topics). After preparing data, the next process from Fig. 1 is generating fuzzy rules as described in the following section.

IV. PROPOSED MODEL TO GENERATE FUZZY RULES FOR CLASSIFYING EXPERTS WITH PRODUCTIVITY-BASED FEATURES

The proposed model for generating fuzzy rules to classify experts needed expertise information as shown in Fig. 3. First, each article from 51 experts was labeled with clustered topics by calculating Cosine similarities of keywords between the article vector and the topic vector. Keywords in a topic are the clustered words. An article is transformed with the embedded matrix such that the article vector at most only has 3500 word items. The value of an item is averaged from associated values in the 200 dimensions. The topic label for each article was set with similarities ≥ 0.90 which representing a small angle or the closer degree between vectors of the article and the topic. An article has some topic labels as shown in Data-2a for the collection of labeled articles. Afterward, the distinct topics from the published articles of one academic expert are mapped to the expert as candidate expertise of topic interests.

The six productivity features in this paper principally come from number of articles and citations which used to detect an expert who has a rising star potential [6]. The first three features are number of articles, the cumulative article number without and with penalty weights.

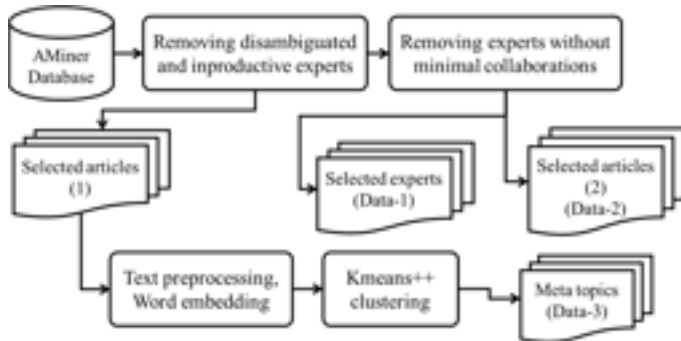


Fig. 2 Dataset preparation using AMiner

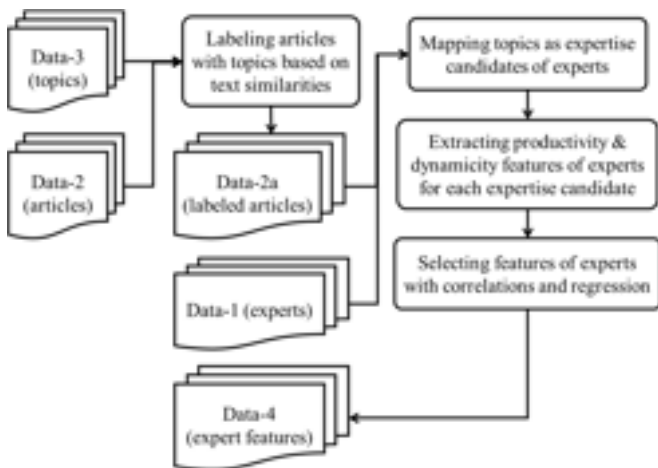


Fig. 3 Generating expert features before fuzzy rules for classification

The penalty is a period denominator between two observed years. Thus, feature values from recent observation are larger than values obtained from much longer period. The other three productivity features are derived from number of cited articles, cumulative cited article number without and with penalty weights. The dynamic features were expected to describe tenacity behavior of the experts from the changes in productivity such as minimum, maximum, last, total and all representation [6]. But here, our modification contributed to the productivity features for each topic of an expert during the observed years. Consequently, the dynamic features from the productivity are also accordingly adjusted with topic information.

There are five additional dynamic based features for each productivity feature. For example, an expert with four topics, T1-T4, shown in Fig. 4 with $J=4$. As candidate expertise being observed in N years, the expert is going to have four set of data instance for each topic. After extracting productivity-dynamic based features, there are $6 \times 5 = 30$ features for each topic. Because of too many dimensions, the final process in Fig. 3 is selecting the 30 features. Functions in detail to obtain 30 feature values without our modified cluster parameter are shown in [6], but the functions with topic information are described in here. To make equivalent features, at first, all values are column normalized in range 0-1. A candidate expertise on certain topic is assumed valid if the expertise status = 1. However, there is no such status in AMiner. Therefore, as the second step, the expertise status for certain topic of an expert is set to 1 if the sum of 30 feature values ≥ 5.5 , that indicated a linear model. The threshold value was specified by assuming that at least there are ten dynamic aspects of an expert with weights > 0.5 . Notes that the expertise status is binary. Before selecting features, we made correlations for all combination pairs of 30 features. Then, the last step was joining all features that have correlations < 0.5 or have weak relations into regressions with expertise status as the target.

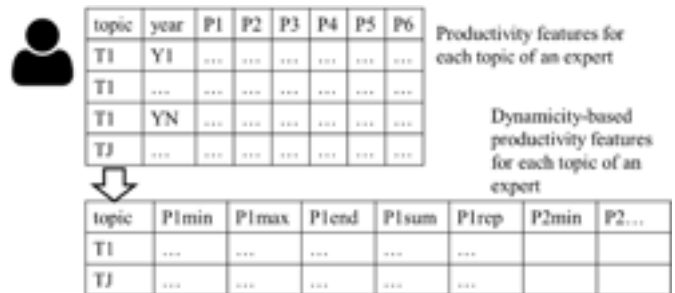


Fig. 4 Sample of features in the extraction process

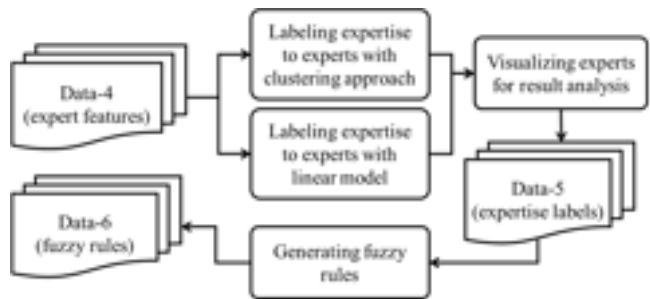


Fig. 5 Expertise model used in generating fuzzy rules for classification

Detail results of selecting six features are described later, but the functions to calculate feature values as follows. Notes that the selected six features construct better regression model with higher values of accuracy and sensitivity.

- a. **P1min** (the minimum difference of article number between two consecutive years) (2)

Calculating P1min needs $P_1(a_x, c_i, t_n)$ as the number of articles published by expert a_x which are labeled as topic c_i in year t_n . The observed period are some years set between t_a, t_z .

$$P_{1.chg}(a_x, c_i, t_{y-1}, t_y) = |P_1(a_x, c_i, t_y) - P_1(a_x, c_i, t_{y-1})| \quad (1)$$

$$P_{1.min}(a_x, c_i, t_a, t_z) = \min_{t_{y-1} < t_y; t_{y-1}, t_y \in t_a \dots t_z} P_{1.chg}(a_x, c_i, t_{y-1}, t_y) \quad (2)$$

- b. **P1end** (the difference of article number from the last year in monitoring periods) (3)

The function is essentially (1) but only for the last year t_z .

$$P_{1.end}(a_x, c_i, t_{z-1}, t_z) = |P_1(a_x, c_i, t_z) - P_1(a_x, c_i, t_{z-1})| \quad (3)$$

- c. **P2rep** (average-like estimation for the cumulative of article number within the observed period) (4)

$$P_{2.rep}(a_x, c_i, t_a, t_z) = \frac{P_1(a_x, c_i, t_{z-1}) + P_1(a_x, c_i, t_z)}{t_a - t_z + 1} \quad (4)$$

- d. **P3sum** (sum of differences for cumulative article number with penalty weights) (7)

$$P_3(a_x, c_i, t_m, t_n) = \sum_{t_y \in t_m \dots t_n} \frac{P_1(a_x, c_i, t_y)}{t_y - t_m + 1} \quad (5)$$

$$P_{3.chg}(a_x, c_i, t_a, t_{y-1}, t_y) = P_3(a_x, c_i, t_a, t_y) - P_3(a_x, c_i, t_a, t_{y-1}) \quad (6)$$

$$P_{3.sum}(a_x, c_i, t_a, t_z) = \sum_{t_{y-1} < t_y; t_{y-1}, t_y \in t_a \dots t_z} P_{3.chg}(a_x, c_i, t_a, t_{y-1}, t_y) \quad (7)$$

- e. **P4min** (the minimum difference of cited article number between two consecutive years, similar to P1min) (10)

For calculating P4min, use function $ncite(d_k, t_n)$ to return the total citation number of an article d_k which is published by expert a_x and labeled as topic c_i . Then use (2) but change $P_1(a_x, c_i, t_n)$ to $P_4(a_x, c_i, t_n)$.

$$P_4(a_x, c_i, t_n) = \sum_{d_k \in c_i} ncite(d_k, t_n) \quad (8)$$

$$P_{4.chg}(a_x, c_i, t_{y-1}, t_y) = P_4(a_x, c_i, t_y) - P_4(a_x, c_i, t_{y-1}) \quad (9)$$

$$P_{4.min}(a_x, c_i, t_a, t_z) = \min_{t_{y-1} < t_y; t_{y-1}, t_y \in t_a \dots t_z} P_{4.chg}(a_x, c_i, t_{y-1}, t_y) \quad (10)$$

- f. **P5end** (the cumulative difference of cited article number from the last year) (12)

$$P_5(a_x, c_i, t_m, t_n) = \sum_{t_y \in t_m \dots t_n} P_4(a_x, c_i, t_y) \quad (11)$$

$$P_{5.end}(a_x, c_i, t_{z-1}, t_z) = |P_5(a_x, c_i, t_a, t_z) - P_5(a_x, c_i, t_a, t_{z-1})| \quad (12)$$

To determine expertise status for each topic, we used the linear model as mentioned before, and compared it with the clustering of six features computed with (2) (3) (4) (7) (10) (12) as shown in Fig. 5. Clustering based on fuzzy membership like Fuzzy C-Means (FCM) becomes basic approach to generate potential fuzzy rules from existing data [18]. Here, we used FCM on the six chosen features and observed the results through visualization with data transformation using t-SNE [19].

The processes in Fig. 5 give the results of Data-4 (expert features) and Data-5 (expertise labels). With only a few number of features, fuzzy rules are easily generated and interpreted [20], and then embedded later into an expert recommendation system. Initially, the six selected features have wide-ranging values. We transformed the values to have same scales of 1-10 according to rules in Table I after analyzing by creating histograms for each feature. For example, an expert in a topic has P1min=7. It means that, in two consecutive years, the expert has an increasing or decreasing productivity with the difference of seven articles. P1min = 0 is often interpreted as the research consistency in publishing, because the expert has the same article number for consecutive years. Rules for P4min and P5end which related to citations have much larger range values than P1min. It is possible because an article can received ± 40 citation times in just one year, such as the 2012 citation of an article ‘‘Open language learning for information extraction’’ by Stephen Soderland.

V. RESULTS AND DISCUSSIONS

The experiments used Python libraries of Natural Language Toolkit, Scikit-learn, Gensim, and Orange Toolkit, in addition to R package of frbs. In the beginning, there are 30 productivity-dynamicity based features. Next step was iterating all possible combinations with several correlation values of 0.5-0.7 to filter out sets of features. Started with 475,020 combinations of six features, and then there were 16 sets having weaker correlations of 0.5. Those six selected features and their extracted functions were described before. There were 813 data instances of nine columns along with features of ExpertID, TopicID and Expertise from 51 experts, 30 topics, and binary values for expertise status. After using the linear model for the status as mentioned before, there were $\pm 19.7\%$ data instance which auto-labeled as positive (status = 1 or the expert is a specialist on the topic). Data instance with status = 0 represented that he or she is a thriving expert in the topic. The instance with status = 0 indicates that the expert has just few articles related to a topic as well, which means the expert features are non-null values.

Table II showed comparisons of classifiers on the 813 data instance with and without scaling process. The target labels from linear model and FCM were compared as well such that training and testing processes applied on the same data. The results showed four outstanding features of P2rep, P3sum, P4min, and P5end that as good as the six selected features.

TABLE I RULES FOR TRANSFORMING DATA INTO SCALED VALUE

Scale	P1min	P1end	P2rep	P3sum	P4min	P5end
1	0	-4	0.1	<-1.0	0	0
2	1	-3	0.2	-1.0	3	3
3	2	-2	0.3	-0.5	5	5
4	3	-1	0.4	0.0	10	10
5	4	0	0.5	0.5	20	20
6	5	1	0.6	1.0	30	30
7	6	2	0.7	1.5	40	40
8	7	3	0.8	2.0	50	50
9	8	4	0.9	2.5	70	70
10	9	5	≥ 1.0	> 3	≥ 100	≥ 100
Avg.	1.92	5.06	3.49	4.20	2.64	2.47
Std.	0.81	0.75	2.96	1.87	2.12	2.18

TABLE II CLASSIFICATION ACCURACIES WITH VARIOUS SELECTED FEATURES

Methods (Orange Toolkit)	P3sum + P4min + P5end			P2rep + P3sum + P4min + P5end			All six features		
	Lin. non	Lin. scaled	FCM scaled	Lin. non	Lin. scaled	FCM scaled	Lin. non	Lin. scaled	FCM scaled
Logistic Regression	86.8%	87.2%	87.7%	86.8%	87.2%	98.4%	88.3%	88.2%	97.5%
Random Forest	95.4%	91.4%	92.0%	96.2%	93.5%	99.5%	96.4%	94.6%	99.4%
SVM	86.2%	85.2%	89.1%	87.7%	87.5%	99.6%	88.9%	90.4%	99.1%

Even though, there was fairly high accuracy on non-scaled data but the visualization by t-SNE displayed mixed labels (Fig. 6). Thus, all feature values of the data instance were scaled into 1-10. Next analysis was about the quality of expertise labels from both methods of linear model (Fig. 7) and FCM (Fig. 8) on scaled data. FCM used $C=2$ to comply with binary values of expertise status. The visualizations after t-SNE transformations were plots in two dimensional. Transformed and scaled data with the labels from linear model showed muddled points of class-0 and class-1 at the fourth quadrant of Fig. 7, although the scaled data plot was still preferable than the non scaled data plot. In contrast, same data points in Fig. 8 with the labels from FCM exhibited more distinctive partitions. Zoom-out fragments at the fourth quadrant in both plots highlighted the mixed points.

The next experiments were generating fuzzy rules based on data with R package frbs using several suggested methods of FRBCS.W, FRBCS.CHI, GFS.GCCL, and FH.GBML [20]. The first two methods are based on space partition while the rests are based on genetic algorithms. Training and testing on the non-scaled data of six features with linear model labels presented accuracies of 20%, 30%, 80.6%, and 80.9% for FRBCS.W to FH.GBML in that order as baselines. After manual analysis, the rules generated from FRBCS.CHI had simple interpretation. Therefore, in the next generating process with the scaled data, FRBCS.CHI was the selected method.

The fuzzy rules learned with labels from linear model had accuracies of 85% for P3sum-P4min-P5end and 61% for the features of P2rep-P3sum-P4min-P5end. The rules from three features with the labels of linear model are listed in Table III. Then, we used four features with FCM labels to generate better fuzzy results with the accuracy of 93.7% as shown in Table IV. As mentioned before, both sets of rules with certainty factor ≥ 0.95 were generated from FRBCS.CHI on the scaled dataset. Label analysis with t-SNE visualizations or by performing fuzzy rules in classification revealed not contradictory results (Fig. 7 and Fig. 8).

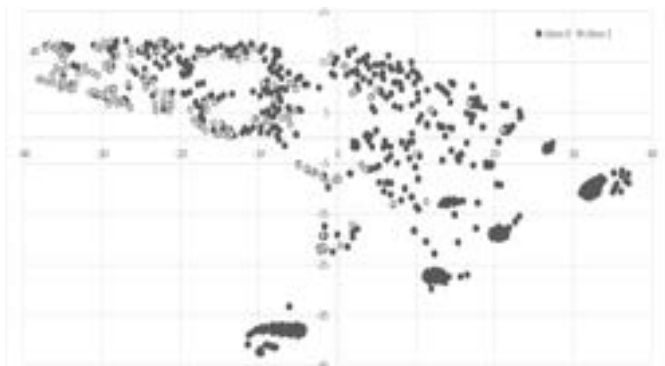


Fig. 6 t-SNE visualization of non-scaled data with labels from linear model

The ambiguity of Expertise Status for Rule-A...Rule-C by having both values of class-0 and class-1 for the same fuzzy rules in Table III confirmed the label quality from linear model (mixed points of Fig. 7).

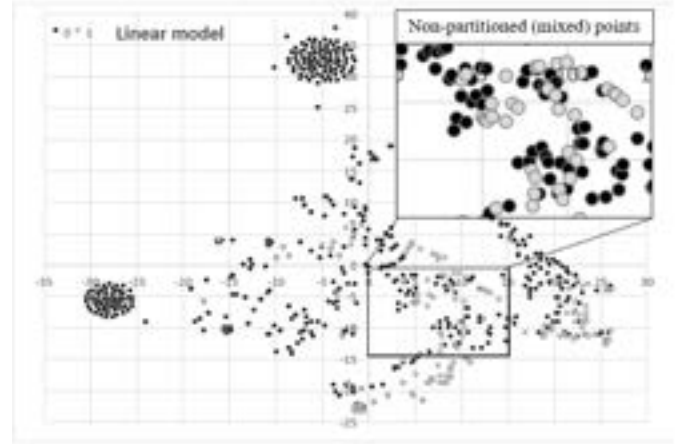


Fig. 7 t-SNE visualization of scaled data with labels from linear model

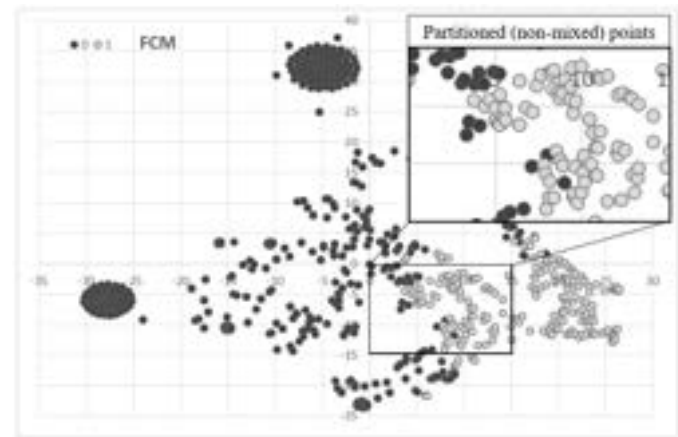


Fig. 8 t-SNE visualization of scaled data with labels from FCM approach

TABLE III AND-CONNECTOR FUZZY RULES GENERATED WITH LABELS FROM LINEAR MODEL ON SCALED DATA

Rule	P3sum	P4min	P5end	Expertise Status
1	large	large	large	
2	large	medium	large	
3	medium	small	large	1-specialist expert
4	medium	small	medium	
5	small	large	large	
6	small	small	small	0-thriving expert
A	large	medium	medium	0-thriving expert
B	medium	medium	medium	1-specialist expert
C	medium	small	small	

TABLE IV AND-CONNECTOR FUZZY RULES GENERATED WITH LABELS FROM FCM ON SCALED DATA

Rule	P2rep	P3sum	P4min	P5end	Expertise Status
1	large	large	large	large	
2	large	large	medium	large	
3	large	large	medium	medium	1-specialist
4	large	medium	medium	medium	
5	large	medium	small	small	
6	medium	medium	medium	medium	1-specialist
7	medium	medium	small	small	0-thriving
8	small	medium	small	large	
9	small	medium	small	medium	0-thriving
10	small	medium	small	small	

Finally, after preceding investigations, unambiguous fuzzy rules were generated from the scaled dataset with FCM labels (Table IV). Gaussian membership functions for all four features were small $f(x, \sigma_{0.175}, \mu_{0.0})$, medium $f(x, \sigma_{0.175}, \mu_{0.5})$, and large $f(x, \sigma_{0.175}, \mu_{1.0})$. To sum up from the fuzzy rules in Table IV after one decade observation, an expert with large value of **P2rep** (Rule-1...Rule-5), estimated by (4), or at least has 2 articles annually published on certain topic can be stated as a specialist or becomes the topic expertise. If that criterion has not met, at least the expert with medium degree receives 5-10 citations on the particular topic in a year (**P4min** in Rule-6 estimated by (10)). However, if the expert is not being cited at all for the particular topic in one observed year, then it means that he or she is still in thriving state (small **P4min** in Rule-7...Rule-10). Those uncomplicated rules are easy to be executed in an expert recommendation system with rather straightforward if-then clauses.

VI. CONCLUSIONS

We have described fuzzy rules to classify expertise of an academic researcher on certain topic based on the productivity performance in publishing articles and receiving citations after period. Since the rules are generated from incomplete expertise status, some preprocessing steps were necessary. Those steps included determining topics, mapping expertise candidates, extracting productivity-dynamicity based feature, and labeling expertise status for training a complete dataset to generate the rules. After some preceding analysis, the generating process of fuzzy rules only needed four representative features of **P2rep**, **P3sum**, **P4min**, and **P5end**. Improved results were influenced by the selected features and the labels from FCM on scaled data. Then, classification of the expertise status on topic interest with fuzzy rules showed comparable results with commonly used classifiers. Though penalty weights from the year difference as the denominator may have some influence in some features, finally, the simple fuzzy rules with just features of **P2rep** and **P4min** were enough. Next works are focusing on implementing the fuzzy rules for recommending experts.

ACKNOWLEDGMENT

This work was supported by the Indonesia Endowment Fund for Education, or in Indonesian called as *Lembaga Pengelola Dana Pendidikan*, LPDP, under Indonesian Education Scholarship for Master and Doctoral Programs with the grant

number PRJ-4228/LPDP.3/2016 of the LPDP Doctoral Scholarship Programme fiscal year 2017-2020.

REFERENCES

- [1] F. Xia, Z. Chen, W. Wang, J. Li, and L. T. Yang, "MVCWalker: Random Walk-Based Most Valuable Collaborators Recommendation Exploiting Academic Factors," *IEEE Trans. Emerg. Top. Comput.*, vol. 2, no. 3, pp. 364–375, Sep. 2014.
- [2] F. Alarfaj, U. Kruschwitz, D. Hunter, and C. Fox, "Finding the right supervisor: expert-finding in a university domain," *Proc. 2012 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. Student Res. Work.*, pp. 1–6, 2012.
- [3] S. Anongnart, "Building Fexpert: System for searching experts in research university using K-MEANS algorithms," in *2012 IEEE Symposium on Computers Informatics (ISCI)*, 2012, pp. 176–179.
- [4] D. Purwitasari, C. Fatichah, I. K. E. Purnama, S. Sumpeno, and M. H. Purnomo, "Inter-departmental research collaboration recommender system based on content filtering in a cold start problem," in *2017 IEEE 10th International Workshop on Computational Intelligence and Applications (IWCI/A)*, 2017, pp. 177–184.
- [5] N. Nikzad-Khasmaki, M. A. Balafar, and M. R. Feizi-Derakhshi, "The state-of-the-art in expert recommendation systems," *Eng. Appl. Artif. Intell.*, vol. 82, pp. 126–147, 2019.
- [6] G. Panagopoulos, G. Tsatsaronis, and I. Varlamis, "Detecting rising stars in dynamic collaborative networks," *J. Informetr.*, vol. 11, no. 1, pp. 198–222, 2017.
- [7] T. Amjad, A. Daud, and M. Song, "Measuring the Impact of Topic Drift in Scholarly Networks," in *Companion Proceedings of the The Web Conference 2018*, 2018, pp. 373–378.
- [8] I. Mendonça, A. Trouvé, A. Fukuda, and K. Murakami, "Exploring a Topical Representation of Documents for Recommendation Systems," *2018 9th Int. Conf. Aware. Sci. Technol. iCAST 2018*, pp. 73–78, 2018.
- [9] D. R. Radev, P. Muthukrishnan, V. Qazvinian, and A. Abu-Jbara, "The ACL anthology network corpus," *Lang. Resour. Eval.*, vol. 47, no. 4, pp. 919–944, Dec. 2013.
- [10] J. Tang, "AMiner: Toward Understanding Big Scholar Data," in *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, 2016, p. 467.
- [11] Q. Wang, J. Ma, X. Liao, and W. Du, "A context-aware researcher recommendation system for university-industry collaboration on R&D projects," *Decis. Support Syst.*, vol. 103, pp. 46–57, 2017.
- [12] M. Dehghan, M. Biabani, and A. A. Abin, "Temporal expert profiling: With an application to T-shaped expert finding," *Inf. Process. Manag.*, vol. 56, no. 3, pp. 1067–1079, 2019.
- [13] G. S. Daş and T. Göçken, "A fuzzy approach for the reviewer assignment problem," *Comput. Ind. Eng.*, vol. 72, pp. 50–57, 2014.
- [14] J. Protasiewicz *et al.*, "A recommender system of reviewers and experts in reviewing problems," *Knowledge-Based Syst.*, vol. 106, pp. 164–178, 2016.
- [15] H. Deng, J. Han, M. R. Lyu, and I. King, "Modeling and Exploiting Heterogeneous Bibliographic Networks for Expertise Ranking," *Proc. ACM/IEEE Jt. Conf. Digit. Libr.*, pp. 71–80, 2012.
- [16] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and Their Compositionality," in *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, 2013, pp. 3111–3119.
- [17] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [18] M.-Y. Chen and D. A. Linkens, "Rule-base self-generation and simplification for data-driven fuzzy models," *Fuzzy Sets Syst.*, vol. 142, no. 2, pp. 243–265, 2004.
- [19] L. van der Maaten and G. Hinton, "Visualizing Data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.
- [20] L. Riza, C. Bergmeir, F. Herrera, and J. Benitez, "frbs: Fuzzy Rule-Based Systems for Classification and Regression in R," *J. Stat. Software, Artic.*, vol. 65, no. 6, pp. 1–30, 2015.