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

Special Session: Awareness Technology for Economic and Social Data Analysis

Special Session: The Data Analysis and Intelligent Computing for Awareness Applications

Special Session: Advances on Biomedical Informatics and Engineering for Healthcare

Special Session: Medical Data Analytics

Appendix

Author Index

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

Diana Purwitasari[1,2], Chastine Fatichah[2], Adri Gabriel Sooai[1,3], Surya Sumpeno[1,4], Mauridhi Hery Purnomo[1,4]
[1]Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2]Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[3]Department of Computer Science, Universitas Katolik Widya Mandira Kupang, Indonesia
[4]Department 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 ±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.

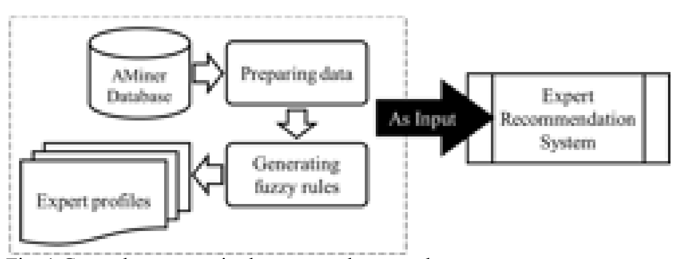

Fig. 1 General processes in the proposed approach

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.

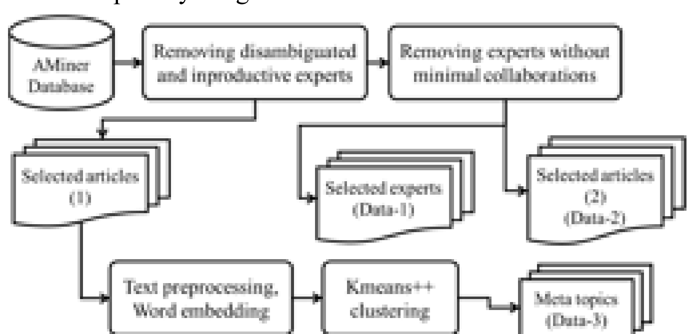

Fig. 2 Dataset preparation using AMiner

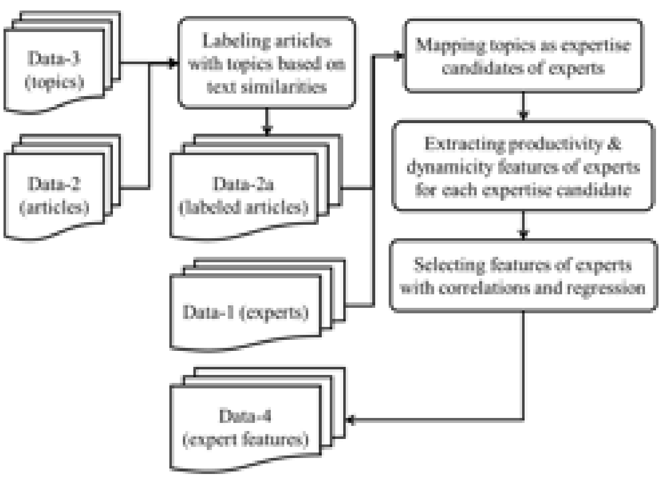

Fig. 3 Generating expert features before fuzzy rules for classification

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 6x5=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.

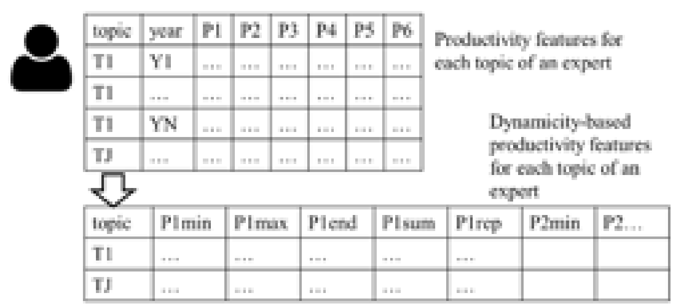

Fig. 4 Sample of features in the extraction process

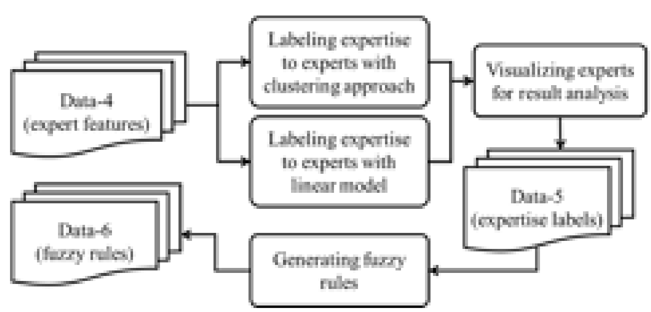

Fig. 5 Expertise model used in generating fuzzy rules for classification

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.

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

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

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)

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

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

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

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 $P_4(a_x, c_i, t_n)$.

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

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

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

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

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

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

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 ±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	P3sum + P4min + P5end			P2rep + P3sum + P4min + P5end			All six features		
(Orange Toolkit)	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).

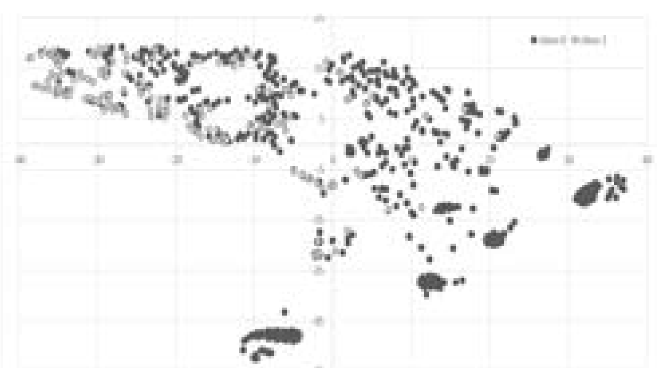

Fig. 6 t-SNE visualization of non-scaled data with labels from linear model

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).

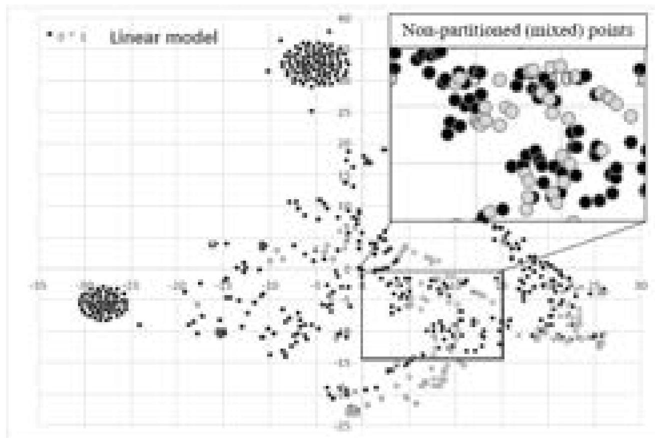

Fig. 7 t-SNE visualization of scaled data with labels from linear model

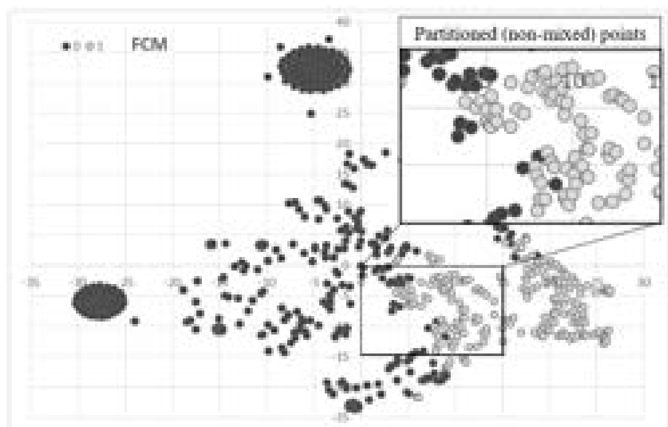

Fig. 8 t-SNE visualization of scaled data with labels from FCM approach

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	1-specialist expert
2	large	medium	large	
3	medium	small	large	
4	medium	small	medium	
5	small	large	large	
6	small	small	small	0-thriving expert
A	large	medium	medium	*0-thriving expert* *1-specialist expert*
B	medium	medium	medium	
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 (IWCIA)*, 2017, pp. 177–184.

[5] N. Nikzad–Khasmakhi, 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. Benítez, "frbs: Fuzzy Rule-Based Systems for Classification and Regression in R," *J. Stat. Software, Artic.*, vol. 65, no. 6, pp. 1–30, 2015.