

Prediction of Research Topics Using Ensemble of Best Predictors from Similar Dataset

Indra Budi, Rizal Fathoni Aji, Agus Widodo

Abstract—Prediction of future research topics by using time series analysis either statistical or machine learning has been conducted previously by several researchers. Several methods have been proposed to combine the forecasting results into single forecast. These methods use fixed combination of individual forecast to get the final forecast result. In this paper, quite different approach is employed to select the forecasting methods, in which every point to forecast is calculated by using the best methods used by similar validation dataset. The dataset used in the experiment is time series derived from research report in Garuda, which is an online sites belongs to the Ministry of Education in Indonesia, over the past 20 years. The experimental result demonstrates that the proposed method may perform better compared to the fix combination of predictors. In addition, based on the prediction result, we can forecast emerging research topics for the next few years.

Keywords—Combination, emerging topics, ensemble, forecasting, machine learning, prediction, research topics, similarity measure, time series.

I. INTRODUCTION

RESEARCHERS and policy makers need to understand the current and future state of research, and be able to identify areas of research that has great potential. Meanwhile, the information sources of research today have grown rapidly along with the advancement of the internet. Meanwhile, the ways to predict the future topics of research, in general can be categorized into judgmental and quantitative analysis [19]. Predictions based on numerical data extrapolate historical data through a specific function, whereas the judgmental forecasting can also be based on projections from the past, but the sources of information in the model depend on the subjective judgments of experts. It is stated in [20] stated that the forecasting analysis through Delphi study by panel of experts is partially incompatible with the results of numerical analysis, since the representation of experts in the panel, which cannot always be proportional, would impact the prediction accuracy.

Trend analysis of research topic by using a numerical approach based on scientific publications and / or patents have been done in some previous researchers, for example, Small [21], Rahayu and Hasibuan [22], Daim [23], Woon, Hensche and Madnick [24], Ziegler [25], as well as Vidican et al. [26].

I. Budi is with the University of Indonesia, Depok, West Java 16424, Indonesia (Tel: +62 21 786 3419; Fax: +62 21 786 3415; e-mail: indra@cs.ui.ac.id).

RF. Aji is also with the University of Indonesia, Depok, West Java 16424, Indonesia (e-mail: rizal@cs.ui.ac.id).

A. Widodo is currently PhD student at the University of Indonesia, Depok, West Java 16424, Indonesia (e-mail: agus.widodo@ui.ac.id).

Those researchers calculate the same word in a document to group the documents into a certain category, and calculate the frequency of words to determine its trends.

Analysis of time series of scientific publications and/or patent has been made by several researchers with diverse techniques. For example, Bengisu and Nekhili [20] use a logistic curve on 'machine and materials', while Jun and Uhm [27] use statistical approaches and machine learning techniques to the data of a patent on bio-technology. In our previous research [28], we use time series analysis to make predictions on arrhythmia data from PubMed, and find out that machine learning techniques can provide better performance than statistical approaches. Meanwhile, the ensemble between the logistic curve and statistical techniques (ARIMA) is done by Christodoulos et al. [29] to make predictions with limited data in the field of telecommunication.

Meanwhile, several prediction methods on time series data have been studied and used in practice. The most common ones are linear methods based on autoregressive models of time series, as stated by Romera et al. [17] and Makridakis et al. [8]. More advanced approaches apply nonlinear models based mainly on artificial neural networks (NNs), support vector machine (SVM), and other machine learning methods as studied by Siwek et al. [10], Crone and Kourentzes [3], Huang et al. [7], and Zang et al. [15].

However, some prediction techniques usually yield good performance on certain types of time series. Hence, model selection is considered by several authors. Siwek et al. [10] train many networks and then pick the one that guarantees the best prediction on out-of-sample (verification) data. Huang et al. [7] and Armstrong et al. [1] take into account some best prediction results, and then combine them into an ensemble system to get the final forecast result. In addition, Poncela et al. [9] combine several dimensional reduction methods for prediction and then use ordinary least squares for combination, while Siwek et al. [10] combine prediction results from neural networks using dimensional reduction techniques.

However, previous literatures calculate the weight of the predictors at once using all training data. In this study, every future point is predicted by the best predictors used by similar training dataset. In other words, every point may be predicted by different predictors.

Thus, this paper aims to explore the use of similarity measure as a method for selecting predictors that would be used for forecast combination. Our hypothesis is that the best methods used in training and validation will be suitable for similar time series used in testing.

II. LITERATURE REVIEW

Several combination methods are described by Timmerman [11], such as by least squares estimators of the weights, relative performance weight, minimization of loss function, non-parametric combination, and pooling several best predictors. Time-varying method is also discussed where the combination weight may change over time.

Recently, Poncela et al. [9] combine several dimensional reduction methods, such as principal component analysis, factor analysis, partial least squares and sliced inverse regression, for prediction, using ordinary least squares. The dataset comes from the Survey of Professional Forecasters, which provides forecasts for the main US macroeconomic aggregates. The forecasting results show that partial least squares, principal component regression and factor analysis have similar performances, and better than the usual benchmark models. Mixed result is found for sliced inverse regression which shows an extreme behavior.

Meanwhile, Siwek et al. [10] combine prediction results from neural networks using dimensional reduction techniques, namely principal component analysis and blind source separation. In this paper, all of the predictors are used to form the final outcome. The ensemble of neural predictors is composed of three individual neural networks. The prediction data generated by each component of the ensemble are combined together to form one forecasted pattern of electricity power for 24 hours ahead. The best results have been obtained with the application of the blind source separation method by decomposing the data into streams of statistically independent components and reconstructing the noise-omitted time series.

On the other hand, time series similarity has been widely employed in several fields, namely the gene expression, medical sequences, image, among others. The most common method to find the time series similarity is computing their distances. These distances are usually measured by Euclidean distance. Vlancos [12] describes several variation of this distance computations exist to accommodate the similarity of some parts of the series, namely the Dynamic Time Warping, and Longest Common Subsequence. Besides the Euclidean distance, similarity measure based on Cosine Similarity is also employed by Widodo and Budi [13] for text comparison.

Others used likelihood to find similarity, such as Hassan [6] who uses Hidden Markov Model to identify similar pattern including time series. It is suggested that the forecast value can be obtained by calculating the difference between the current and next value of the most similar training series, and add that differences to the current value of the series to forecast. However, in this paper, the similarity measure is not used to directly compute the next value, but to select the most suitable predictors to compute that value.

III. THEORETICAL BACKGROUND

A. Forecast Combinations

Timmerman [11] indicates that there are several reasons to combine the forecasts. First argument is due to diversification since one model is often suited to one kind of data. Thus, the

higher degree of overlap in the information set, the less useful a combination of forecasts is likely to be. In addition, individual forecasts may be very differently affected by structural breaks in time series. Another related reason is that individual forecasting models may be subject to misspecification bias of unknown form. Lastly, the argument for combination of forecasts is that the underlying forecasts may be based on different loss functions. A forecast model with a more symmetric loss function could find a combination of the two forecasts better than the individual ones.

Combining the forecast results generally seeks an aggregator that reduces the information in a potentially high-dimensional vector of forecasts to a lower dimensional summary measure. Poncela et al. [9] denotes that one point forecast combination is to produce a single combined 1-step-ahead forecast f_t at time t , with information up to time t , from the N initial forecasts; that is

$$f_t = w_t' y_{t+1|t} \quad (1)$$

where w_t is the weighting vector of the combined forecast, $y_{t+1|t}$ is N dimensional vector of forecasts at time t . A constant could also be added to the previous combining scheme to correct for a possible bias in the combined forecast. The main aim is to reduce the dimension of the problem from N forecasts to just a single one, f_t .

1. Methods to Combine

There are various integration methods that may be applied in practice. In this paper, we will compare methods based on the averaging, both simple and weighted on predictor's performance.

In the averaging schema, the final forecast is defined as the average of the results produced by all different predictors. The simplest one is the ordinary mean of the partial results. The final prediction of vector x from M predictors is defined by

$$x = \frac{1}{M} \sum_{i=1}^M x_i \quad (2)$$

This averaging method may reduce the final error of forecasting if all predictive networks are of comparable accuracy. Otherwise, weighted averaging shall be used.

The accuracy of weighted averaging method can be measured on the basis of particular predictor performance on the data from the past. The most reliable predictor should be considered with the highest weight, and the least accurate one with the least weight. The estimated prediction is calculated as

$$x = \sum_{i=1}^M w_i x_i \quad (3)$$

where w_i is weight associated with each predictor. One way to determine the values of the weights ($i=1, 2, \dots, M$) is to solve the set of linear equations corresponding to the learning data, for example, by using ordinary least squares. Another way is using relative performance of each predictor [11], where the weight is specified by

$$w_i = \frac{1/MSE_i}{\sum_{i=1}^M 1/MSE_i} \quad (4)$$

By using the weighted average of relative performances, the high performance predictor will be given larger weight and vice versa.

2. Model Selection

Some methods may not good enough to combine. Franses [18] states that the prediction methods that need to be combined are those which contribute significantly to the increased accuracy of prediction. The prediction models in the ensemble can be selected by calculating the performance of each model using the hold-out sample.

For example, Andrawis et al. [16] use only 9 best models out of 140 models to combine. The combination method used in their study is simple average. Previously, Armstrong [1] states that only five or six best models are needed to get better prediction result. Our previous study [14] on the use of Neural Network for forecast combination also suggests that selecting few best models are crucial for improving the forecasting result.

B. Time Series Similarity

The distance between time series can be measured by calculating the difference between each point of the series. The Euclidean Distance between two time series $Q = \{q_1, q_2, \dots, q_n\}$ and $S = \{s_1, s_2, \dots, s_n\}$ is defined as

$$D(Q, S) = \sqrt{\sum_{i=1}^n (q_i - s_i)^2} \quad (5)$$

This methods is quite easy to compute, and take complexity of $O(n)$. On the other hand, Dynamic Time Warping (DTW) [2] allows acceleration-deceleration of signals along the time dimension. For two series $X = x_1, x_2, \dots, x_n$, and $Y = y_1, y_2, \dots, y_n$, each sequence may be extended by repeating elements such that the Euclidean distance can calculated between the extended sequences X' and Y' .

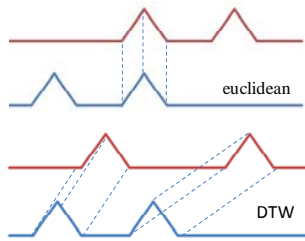


Fig. 1 Two time series to compare

Fig. 1 indicates that those two time series are considered the same for DTW, whereas they are different for Euclidean.

C. Emerging topics

In this paper, the most emerging topics are calculated using the growth rate. Several alternatives are devised to calculate the growth rate of a research topic [25], namely (1) the difference between the frequencies in the last year and early years, (2) the ratio between the frequency in the last year and

early years, (3) the fitting of an exponential curve, and (4) the average year of publication. To provide a more balanced result, then the frequency of certain terms can be normalized by dividing these frequencies by the total number of publications in a given year. Fitting of an exponential curve will result in the form of $a \times e^r$, where r is a measure of growth rate. While the average of publication year is calculated by adding up years of the publication of results between years and the multiplication in the frequency divided by the total number of frequencies, such as

$$avg \text{ year of pub} = \sum_i (y_i \times h_i) / \sum_i h_i \quad (6)$$

Thus, the publication last year will have a weight higher than previous years.

D. Mean Squared Error

The mean squared error (MSE) is measure to quantify the difference between the predicted values and the true values. This measure indicates how well the predictions explain a given set of observations. Let $X = \{x_1, x_2, \dots, x_T\}$ be a random sample of points in the domain of f , and suppose that the value of $Y = \{y_1, y_2, \dots, y_T\}$ is known for all x in X . Then, for all N samples, the error is computed as

$$MSE = \frac{1}{N} \sum_{i=1}^N (f(x_i) - y_i)^2 \quad (7)$$

Hence, an MSE of zero means that the estimator predicts observations with perfect accuracy. In other words, the larger the MSE, the worse the prediction results are.

IV. EXPERIMENTAL SETUP

A. Methodology

The methodology to achieve the goals of this paper follows the quantitative research methods, namely obtaining the dataset, preparing the tools, and evaluating the results. The dataset is obtained by selecting the research topics and constructing the time series. Based on this time series, the matrices for training and testing are built. The prediction algorithms used in this experiment are Neural Network and Support Vector Regressions having various parameters. The similarity of time series, measured by Euclidean and DTW, determines the best model to select among those prediction methods. Forecasting results of the best selected predictors are combined using simple average and weighted average of predictor's performance. Lastly, the performances of the individual predictors and the combination of predictors are measured using the mean squared error.

The steps of (1) comparing time series, (2) selecting best models (3) applying those methods and (4) combining the forecasts are illustrated in Fig. 2.

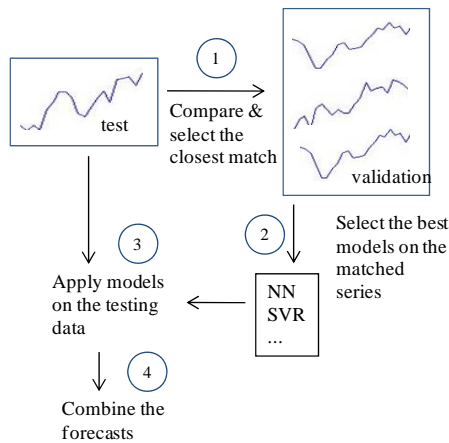


Fig. 2 The proposed method to forecast using the combinations of selected models

B. Datasets

The time series dataset is constructed by compiling the frequencies of each topic per year from the Garuda (Digital Reference Portal) site which is managed by the Directorate General of Higher Education, Indonesian Ministry of Education. The topic is defined by the subject heading codified by Library of Congress Subject Heading (LCSH). The number of topics studied is 12, with consideration of the availability of data with a minimum time span of 16 years. Christodoulos [29] states that the minimum length of training data is 16-20 points. The compiled frequencies of each topic each year is shown in Table I.

TABLE I
FREQUENCY OF TOPICS PER YEAR FROM GARUDA

No	Subject	1991	1992	...	2010	2011
1	Agriculture	182	166		2718	2354
2	Fine Arts	52	27		477	594
3	Geography, Anthrp., & Recrtn.				69	22
4	Language and Literature	16	17		520	357
5	Law	114	105		1863	1459
6	Medicine	211	225		3279	3769
7	Naval Science	1			20	44
8	Philosophy, Psychology, and Religion	1			63	40
9	Political Science	5	2		291	253
10	Science	195	161		2585	2123
11	Social Sciences	65	80		3443	3740
12	Technology	89	84		2474	3628

The number of samples to be used as a training and testing is influenced by the length of time series. If there are k values to predict, the y_{test} vector will contain k values, and x_{test} matrix will consist of $m \times k$, where m is the sliding window. Thus, having 2 values to predict, the vector y_{test} consists of 2 values, and the matrix x_{test} consists of $m \times 2$ series, where m is the sliding window (Fig. 3). The value of m is determined while constructing the training dataset, namely the x_{train} and y_{train} , whose matrix's size are $m \times n$ and n . The shorter the value

of m , the larger the dataset (which is n) can be constructed, and vice versa.

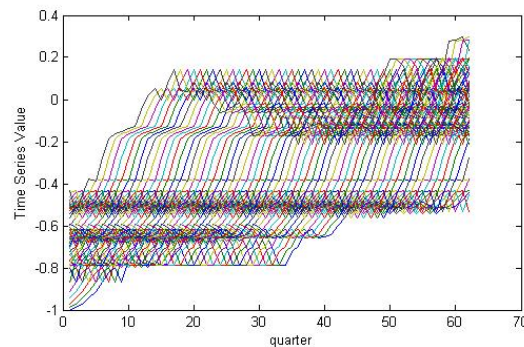


Fig. 3 Example of sliding window of training dataset

C. Performance Evaluation

Mean squared error (MSE) is mainly used to evaluate the prediction's performance on the out-of-sample dataset, which consists of the testing and validation parts. It is also employed to evaluate the forecasting combination results using the simple average as well as the weighted average on individual performance.

D. Hardware and Tools

The hardware used in this experiment is PC with Pentium processor Core i3 and memory of 2GB. The main software used is Matlab version 2008b. The Matlab's command used to perform the Neural Network is the 'newff'. The data is normalized into the range of -1 to 1 using 'mapminmax' command. In addition, Gun [5] provides the toolbox for Support Vector Regression, whereas Felty [4] makes the DWT toolbox available for time series similarity measure.

V. RESULT AND DISCUSSION

A. Forecasting Results of Each Predictor

There are two main predictors used in this experiment, namely (1) Neural Network (NN) having its hidden node set to 1, 2, 3, 5 and 10, (2) Support Vector Regression (SVR) using kernel radial basis function (RBF) of sigma's width of 0.01, 0.1, 0.5, 1, 2, and 5, kernel polynomial of degree 1, 2 and 3. Hence, there are totally 14 models by differentiating the parameters of some predictors. A smaller sigma value in SVR implies smaller variance which fits the data tighter. Likewise, a higher number of hidden nodes in NN would predict the training data more accurately but may overfit for the testing data.

The forecasting result in Table II indicates that the pattern of time series is rather fluctuating as the MSE is quite low for SVR with kernel RBF having smaller sigma, SVR with kernel Polynomial having higher degree, and Neural Network with larger number of hidden nodes. The SME on training is smaller than that of bigger sigma, but SME on testing tends to be bigger as the model tends to overfit. Similarly, using polynomial as kernel of higher degree tends to overfit, hence yield poor generalisation error. Kernel polynomial of degree 2

or 3 is chosen as degree 1 means linear regression and degree higher than 4 tends to overfit.

TABLE II
FORECASTING PERFORMANCE USING MSE AMONG INDIVIDUAL MODELS ON 12 TIME SERIES

No	Predictor	ts1	ts2	...	ts11	ts12	Avg
1	NN 1	0.09	96.32		5.14	5.96	9.17
2	NN 2	2.29	2.09		1.08	0.98	27.97
3	NN 3	4.86	0.17		1.14	0.54	3.05
4	NN 5	3.02	2.07		1.16	0.03	1.71
5	NN 10	3.78	10.52		4.67	1.27	5.85
6	SVR RBF 0.01	0.08	0.33		0.64	0.70	0.36
7	SVR RBF 0.1	0.08	0.33		0.64	0.70	0.36
8	SVR RBF 0.5	0.30	0.29		0.63	0.66	0.35
9	SVR RBF 1	3.23	0.00		0.39	0.24	0.52
10	SVR RBF 2	7.55	0.31		3.18	0.55	2.40
11	SVR RBF 5	3.49	1.46		4.19	0.78	10.11
12	SVR Poly 1	2.92	2.15		4.20	0.36	45.22
13	SVR Poly 2	83.39	40.97		5.12	0.77	15.56
14	SVR Poly 3	14.49	46.82		2.04	0.44	7.04
	Avg	9.26	14.56		2.44	1.00	9.26

There is no single predictor that is best on every dataset. Table II shows that most predictors yield good prediction for some time series but not for the others. Hence, the ensemble of predictors is viable option to choose the best of predictors for a certain time series pattern, assuming that the pattern in testing data can be found in the validation data. Fig. 4 illustrates the result of four validation data and two testing data in which the validation result may not accurately capture the rising pattern of the actual data as the training data of the validation part is rather smooth and flat at the right end.

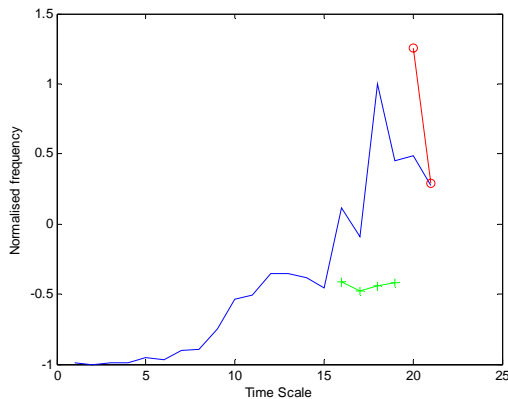


Fig. 4 One of the time series of topic 'Agriculture' The green line with '+' sign is the validation result, and the red line on the right with dotted ('o') sign is the testing result.

B. The Combination of Models

This experiment is to select the predictors that perform best on training time series similar to testing time series to be predicted. The similarity between those series is calculated using Euclidean Distance and DTW. The performance of all possible number of best models is shown in Table III for Euclidean similarity, DTW similarity and without similarity, respectively. By selecting the best models without similarity,

the best models are determined by all training samples. By contrast, using similarity measure, the best models are determined by the training sample similar to the testing data. Simple average method is used to combine the forecasting result.

TABLE III
MSE ON COMBINATION OF METHODS USING EUCLIDEAN, DTW, AND WITHOUT SIMILARITY
CHOOSING HALF OF THE NUMBER OF MODELS YIELDS OPTIMUM PERFORMANCE

Number of best models	Average			Inverse-MSE		
	DTW	Euclid	No-Sim	DTW	Euclid	No-Sim
1 best model	2.38	6.51	10.11	2.38	6.51	10.11
2 best models	3.40	2.21	4.40	3.63	2.50	8.15
3 best models	1.67	1.60	1.94	1.94	1.42	3.36
4 best models	1.36	1.42	1.28	1.94	1.24	2.20
5 best models	0.98	1.06	3.98	1.40	1.17	2.22
6 best models	1.01	0.99	3.30	1.45	1.06	2.14
7 best models	0.97	0.87	2.85	1.32	0.85	2.14
8 best models	1.91	0.91	2.18	39.98	0.90	2.13
9 best models	1.74	0.76	1.72	39.95	0.90	2.12
10 best models	1.61	1.55	1.58	40.16	40.71	2.10
11 best models	1.47	1.41	1.41	40.20	41.34	2.10
12 best models	1.31	1.34	1.99	40.12	41.29	2.10
13 best models	1.95	1.95	1.93	40.12	41.23	2.10
14 best models	2.03	2.03	2.03	40.12	40.98	2.03

The experimental result in Table III indicates that the first and second best model are not good enough as their MSE are quite high. Thus, the best model in validation does not necessarily always imply the best model in testing. However, as the number of model is increased the MSE decreased up until about half of the total number of model.

Furthermore, Fig. 5 illustrates that using combination of methods selected based on the similarity between training and testing data may lead into better prediction result compared to the combination of all methods. Table II presents the detail of performance of the combination of those methods, which actually perform fairly well compared to the individual forecast.

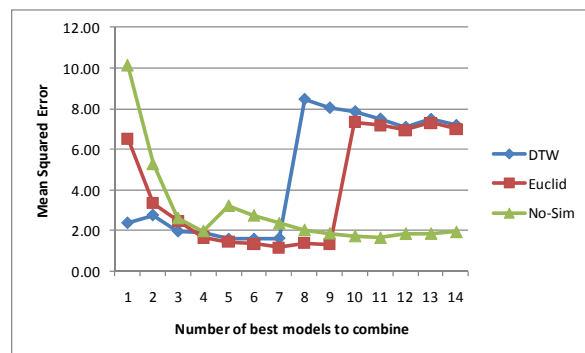


Fig. 5 Average performance of Forecast Combination using models selected by Euclidean and DTW similarity compared to the one using best and all models without similarity measure

Therefore, it can be concluded that the combination of selected methods using similarity measure generally performs better than the best methods without similarity measure when the number of methods combined is greater than three but less than eight for both Euclidean and DTW. According to this experiment, the optimum number of models to combine turns out to be about 50% of all models.

C. The Emerging Topics of Research

The emerging topics of research are calculated using (6) which puts more weight on current frequency. Table III shows the first five of the most emerging topics out of 14 topics from the prediction and from the actual time series. Using the average precision, the ranking performance of the emerging topics is 0.89. The precision of the first until last rank can be elaborated as follows: 1, 0.50, 1, 1, 0.80, 0.83, 0.86, 1, 0.89, 0.90, 0.91, and 1. The first precision is 1 since the predicted topic is the same as the actual one. The second precision is 0.5 since out of 2 predicted topics, there is 1 topic included in actual topics, and so on. The average year of publications of the predicted topics in Table IV are 2001.65, 2001.44, 2001.37, 1998.17, and 1996.70 consecutively, whereas those of the actual topics are 2002.36, 2001.64, 2001.54, 1997.45, and 1995.20.

TABLE IV
THE FIRST FIVE MOST EMERGING TOPICS

Rank	Predicted	Actual
1	'Political Science'	'Political Science'
2	'Philosophy, Psychology, and Religion'	'Naval Science'
3	'Naval Science'	'Philosophy, Psychology, and Religion'
4	'Law'	'Law'
5	'Language and Literature'	'Science'

In addition, we also try to predict the emerging topics for the next two years based on the assumption that its accuracy is about equal as the previous dataset. It turns out that the most emerging research topics for the next two years are 'Social Sciences', 'Agriculture' and 'Political Science'. Thus, the emerging topics for the next two years cannot be assumed to be the same as the last years. Hence, prediction is necessary prior to determining the emerging topics in the future.

VI. CONCLUSION

From the experimental result, it can be concluded that the combination of methods selected based on the similarity between training and testing data may perform better than the combination of all methods. Our experimental result indicates that the optimum number of models to combine is about fifty percent of the number of models. Smaller number of models to combine may not provide enough diversification of method's capabilities whereas greater number of models may select poor performing models. In addition, we may predict the emerging topics for some years ahead by taking account the prediction result and then using growth rate measure such as average year of publication.

For future works, to confirm its feasibility of this model selection, we will try other time series data especially in the domain of research topics. There are also many possibilities of employing different predictors other than NN and SVR, such as ARCH type techniques for volatile time series. In addition, we need to explore other similarity measures besides the Euclidean and DTW such as global time series characteristics that have fixed number of time series features to compare.

ACKNOWLEDGMENT

This work is supported by Research Grant No. DRPM/R/-414/RMUI/2012 from Research and Public Service Directorate, University of Indonesia, 2012.

REFERENCES

- [1] J. S. Armstrong, "Combining Forecasts", *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Kluwer Academic Publishers, 2001.
- [2] D. Berndt, & J. Clifford, J., "Using dynamic time warping to find patterns in time series", *AAAI-94 Workshop on Knowledge Discovery in Databases (KDD-94)*, Seattle, Washington, 1994.
- [3] S. F. Crone, & N. Kourentzes, "Forecasting Seasonal Time Series with Multilayer Perceptrons – an Empirical Evaluation of Input Vector Specifications for Deterministic Seasonality", Lancaster University Management School, Lancaster, UK, 2007.
- [4] T. Felty, "Dynamic Time Warping", toolbox available at Mathlab Central, 2005.
- [5] S. R. Gunn, "Support Vector Machines for Classification and Regression", *Technical Report*, University Of Southampton, 10 May 1998.
- [6] Hassan, M. R., 2007. "Hybrid HMM and Soft Computing modeling with applications to time series analysis", PhD thesis, The University of Melbourne, 2007.
- [7] C. Huang, D. Yang, Y. Chuang, Y., "Application of wrapper approach and composite classifier to the stock trend prediction", *Elsevier, Expert Systems with Applications* 34 (2008) 2870–2878.
- [8] S. Makridakis, S. C. Wheelwright, V. E. McGee, "Forecasting: Methods and Applications", 2nd Ed, John Wiley & Sons, 1983.
- [9] P. Poncela, J. Rodriguez, R. Sanchez-Mangasa, E. Senra, "Forecast combination through dimension reduction techniques", *International Journal of Forecasting* 27 (2011) 224–237.
- [10] K. Siwek, S. Osowski, S., R. Szupiluk, R., "Ensemble Neural Network Approach For Accurate Load Forecasting In A Power System", *Int. J. Appl. Math. Comput. Sci.*, 2009, Vol. 19, No. 2, 303–315
- [11] A. Timmerman, "Forecast Combinations", UCSD, 2005.
- [12] M. Vlachos, "A practical Time-Series Tutorial with MATLAB", *ECML PKDD*, Porto, Portugal, 2005.
- [13] A. Widodo, and I. Budi, "Clustering Patent Document in the Field of ICT", *International Conference on Semantic Technology and Information Retrieval*, Malaysia, 28-29 June, 2011.
- [14] A. Widodo, and I. Budi, "Combination of Time Series Forecasts using Neural Network", *Conference on Electrical Engineering and Informatics*, ITB-Bandung, 18-19 July, 2011.
- [15] G. Q. Zang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural network: The state of the art", *International Journal of Forecasting*, vol. 14, no 1, pp 35062, March 1998.
- [16] R. R. Andrawis, A. F. Atiya, H. El-Shishiny, "Forecast combinations of computational intelligence and linear models for the NN5 time series forecasting competition", *International Journal of Forecasting*, 2011.
- [17] E. Gonzalez-Romera, M. A. Jaramillo-Moran, and D. Carmona-Fernandez, "Monthly electric energy demand forecasting based on trend extraction", *IEEE Transactions on Power Systems* 21(4): 1946–1953, 2006.
- [18] P. H. Franses, "Model selection for forecast combination", *Econometric Institute Research Papers*, Erasmus University Rotterdam, June 2008.
- [19] JR. Meredith, SJ. Mantel, "Technological Forecasting", John Wiley & Sons, Inc., 1995.
- [20] M. Bengisu, R. Nekhili, "Forecasting emerging technologies with the aid of science and technology databases". *Technological Forecasting &*

- Social Change* 2006; 73: 835–844.
- [21] H. Small, "Tracking and predicting growth areas in science". *Scientometrics*, 2006, 68(3):595–610.
 - [22] E.R. Rahayu, and Z.A. Hasibuan, "Identification of technology trend on Indonesian patent documents and research reports on chemistry and metallurgy fields". *Proceeding Asia Pacific Conf.*, Singapore, 2006.
 - [23] TU. Daim, G Rueda, H Martin, and P Gerdri, "Forecasting emerging technologies: Use of bibliometrics and patent analysis". *Technological Forecasting and Social Change*, October 2006, 73(8):981–1012.
 - [24] W.L. Woon, A. Hensche, and S. Madnick, "A Framework For Technology Forecasting And Visualization". *Working Paper Series*, ESD-WP-2009-16, October 2009.
 - [25] B.E. Ziegler. "Methods for Bibliometric Analysis of Research: Renewable Energy Case Study". *Working Paper CISL#2009-10*, September 2009.
 - [26] G. Vidican, W.L. Woon, S. Madnick, "Measuring Innovation Using Bibliometric Techniques: The Case of Solar Photovoltaic Industry", *Working Paper CISL# 2009-05*, 2009.
 - [27] S. Jun, D. Uhm. "Technology Forecasting Using Frequency Time Series Model: Bio-Technology Patent Analysis". *Journal of Modern Mathematics & Statistics*, 2010, 4(3):101-104.
 - [28] A. Widodo., M.I. Fanany, I. Budi, "Technology Forecasting in the Field of Apnea from Online Publications: Time Series Analysis on Latent Semantic". *International Conference on Digital Information Management*, 26-28 September 2011, Melbourne, Australia.
 - [29] C. Christodoulos, C. Michalakelis, D. Varoutas. "Forecasting with limited data: Combining ARIMA and diffusion models". *Technological Forecasting & Social Change* 2010, 77: 558–565.