Topic Discovery and Future Trend Prediction In Scholarly Networks

Qiwen Shen
515030910600
F1503025
Shanghai, China
t_nine@sjtu.edu.cn

***Abstract*—Accurate forecasting for the trend of future research is required while allocating the resources of research. Judgmental or numerical analysis are two of the frequently used methods. Bibliometric analysis is a quantitative method to determine the trend of research area by counting the frequency of certain keywords using journal publication or patents. This paper reports a combination method which can select proper forecasting method for the dataset. The experimental result shows that the performance of the new methods is better than the fix combination of predictors.**

***Index Terms*—forecasting, research topics, topic forecast, data analytics**

I. Introduction

Deriving topics from a large set of document corpus and predicting their popularity is a critical requirement in personalized applications like recommender systems, e-commerce, e-learning etc. As per recent statistics, thousands of documents are being continuously indexed on the Web each minute of the day, which is a testimony to its ever-increasing popularity. The problem of effectively processing, storing and retrieving documents as per a defined user need is further compounded due to the wide diversity of the corpus, which means incremental approaches that are capable of dealing with this are needed. Managing these documents for making best use of them across domains becomes a critical problem for Information Retrieval/Extraction (IR/IE) based applications like scholarly article search, recommender systems etc. Incorporating techniques for automatic topic extraction and categorization can be a major help in streamlining temporal document management in such applications.

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. The volume of scholarly articles published every year has grown exponentially over the years. With these growths in both core and interdisciplinary areas of research, analyzing interesting research trends can be helpful for new researchers and organizations geared towards collaborative work. Efficient and fast topic discovery models and future trend forecasters can be helpful in building intelligent applications like recommender systems for scholarly articles.

II. Related works

The ways to predict the future topics of research, in general can be categorized into judgmental and quantitative analysis. 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 [1] 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. Small [2] using co-citation clustering area for research in the field of science, while Rahayu and Hasibuan [3] and Zhu et al., [4] used co-word analysis. To determine the categories of research topics that are growing, a certain percentage limit was used in [3], such as between one to three percent of the total number of patents or scientific publications. Those researchers generally 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.

Meanwhile, Woon et al., [5] conducted a study on technology trends using bibliometrics, namely the number of records generated from a query against the database online. From the titles and abstracts obtained, derived keywords (terms) are then arranged in a taxonomy by using the Normalized Google Distance. Technology growth scores are calculated based on the average year of publication (average publication year), the term frequency multiplied by the year divided by the frequency of the term during the observation period. Thus, the last year will provide a greater impact on the size of the growth. Similarly, Kobayashi et al., [6] made predictions of specific technology by using online information. Terms in the dictionary are used to filter the keywords obtained from the search results. In [7], similarity measures are used to combine similar keywords, to determine the hierarchy between the keywords and to determine the importance of a keywords.

However, single best method is not always applicable for different kind of data. Thus, model selection is advocated to

overcome this shortcoming. For example, by training many networks and then picking the one that guarantees the best prediction on out-of-sample (verification) data, as done in [8]. A more general approach is to take into account some best prediction results, and then combine them into an ensemble system to get the final forecast result as suggested by Huang et al., [9] and Armstrong et al., [10]. Poncela et al., [11] and Siwek et al., [8].

Previous literatures calculate the weight of the predictors at once using all training data. In this study, quite different approach is proposed in which 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. In addition, based on the prediction result, we identify the emerging topics in Springer for the next periods.

The most common method to find the time series similarity is computing their distances. These distances can measure by Euclidean distance or Dynamic Time Warping. Others used likelihood to find similarity, such as Hassan [12], who used Hidden Markov Model to identify similar pattern including time series. He 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.

The rest of this paper is organized as follows. In Section 3 I present the theoretical background of forecast combination and emerging topics based on bibliometrics, in Section 4 I describe the experimental setup including our proposed model selection, the steps taken and the dataset, and Section 5 is the conclusion.

III. Theoretical Background

A. Forecast Combinations

The forecast combination problem 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., [6] 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

$$f_t = w_t y_{t+1|t}$$

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 .

Various integration methods 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

$$x = \frac{1}{M}\sum_{1}^{M} x_i$$

This process of averaging 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

$$x = \sum_{1}^{M} w_i x_i$$

where w_i is weight associated with each predictor. One way to determine the values of the weights (i=1, 2, ..., 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, where the weight is specified by

$$w_i = \frac{1/MSE_i}{\sum_{i=1}^{M} 1/MSE_i}$$

In this weighted average, the high performance predictor will be given larger weight and vice versa. Some large constant is sometimes needed when the value of inverse MSE (Mean Squared Error) is close to infinity, which could happen when MSE is or close to zero.

B. Model Selection

Franses [9] stated that the prediction methods that need to be combined are those which contribute significantly to the increased accuracy of prediction. The selection of prediction models in the ensemble is usually done by calculating the performance of each model toward the hold-out sample. He also proposed an encompassing mechanism where a model is selected in the final combination if the combination with that model yields more forecast accuracy than a combination without that model. This method is quite close to forward feature selection. Seeking the combination of best model, however, is not trivial since there are many possibility to form the combination.

In addition, Andrawis et al., [13] use 9 best models out of 140 models to combine. The combination method used in their study is simple average. Previously, Armstrong [10] states that only five or six best models are needed to get better prediction result. Therefore, it is obvious that not all models should be selected for the combination. The next task is finding

the optimal number of models to select, which will be depend upon the characteristics of the data as well as the predictors.

C. Time Series Similarity

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

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

This methods is quite easy to compute and take complexity of O(n).

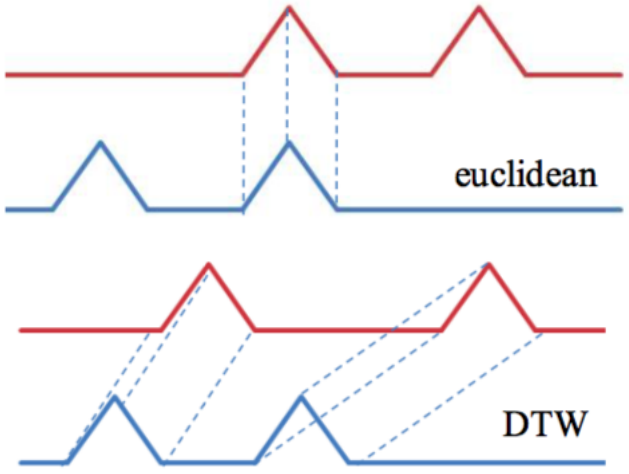

Fig. 1. Two Time Series to Compare

Meanwhile, Dynamic Time Warping allows acceleration-deceleration of signals along the time dimension [24]. For two series $X = x_1, x_2, ..., x_n$, and $Y = y_1, y_2, ..., y_n$, each sequence may be extended by repeating elements such that the Euclidean distance can calculated between the extended sequences X' and Y'. For example, for two time series in Figure 1, it is exactly the same for DTW, whereas it is not for Euclidean.

D. Growth Rate

In [14], several alternatives are devised to calculate the growth rate of a research topic, 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

$$AverageYearOfPublication = \frac{\sum_i y_i x h_i}{\sum_i h_i}$$

Thus, the publication last year will have a weight higher than previous years. The first and second methods do not take the cumulative amount of frequency into account, whereas the year as the weighting factor in third method is somewhat arbitrary. Hence, the Average Year of Publication can be modified by substituting the year with a score from 1 to n, where n is the number of year time series dataset, such as

$$GrowthRate = \frac{\sum_i w_i x h_i}{\sum_i h_i}$$

For instance, a time series having 21 year period started from 1991 until 2011 will be weighted as 1 to 21, consecutively. Each voter (predictor) rank orders the candidates (selected predictors). If there are N candidates, the first-place candidate receives N-1 votes, the second-place candidate receives N-2, with the candidate in i^{th} place receiving N-i votes. The candidate ranked last receives 0 votes.

IV. Experiments

A. Dataset

The dataset is derived from research report compiled by the Springer site. The total number of topics within the S&T category is 14, with consideration of the availability of data with a minimum time span of 25 years. The frequency of topics each year is shown in Table 1.

The number of samples to be used as training and testing is determined by the length of time series. If there are k values to predict, the ytest vector will contain k values, and xtest matrix will consist of $m \times k$, where m is the sliding window. Thus, having 3 values to predict, the vector ytest consists of 3 values, and the matrix xtest consists of $m \times 3$ series, where m is the sliding window. The value of m is determined while constructing the training dataset, namely the xtrain and ytrain, whose matrix's size are $m \times n$ and n. The shorter the value of m the larger the dataset (which is n) that can be constructed, and vice versa. The resulting input matrix is scaled and centered so that its standard deviation is equal to one. This scale is then applied to the output vector. The result of the prediction is later converted back to the actual scale.

B. Performance Evaluation

The Mean Squared Error (MSE) of an estimator is one of many ways to quantify the difference between values implied by an estimator and the true values of the quantity being estimated. Let $X = x_1, x_2, ..x_T$ be a random sample of points in the domain of f, and suppose that the value of $Y = y_1, y_2, ..y_T$ is known for all x in X. Then, for all N samples, the error is computed as

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

TABLE I
FREQUENCY OF S&T TOPICS PER YEAR FROM SPRINGER

	Medicine & Public Health	Life Sciences	Engineering	Chemistry	Biomedicine	Mathmatics	Physics
1993	18332	11980	2787	7745	6464	5147	7158
1994	18861	11942	2810	8210	6503	5165	7927
1995	19834	12149	2945	8595	6179	5172	7547
1996	21703	12665	3061	8466	6444	5539	9033
1997	23963	13163	3776	9401	6361	5863	9194
1998	25561	13902	3990	8828	5829	6071	10199
1999	27665	13427	3954	8614	5999	6103	10422
2000	30007	14798	4367	10463	6811	6600	11745
2001	31810	15779	4792	11691	6783	6969	11779
2002	32792	15908	5565	11469	6914	6831	12122
2003	34607	18002	6018	11993	5508	6879	11939
2004	35889	18165	6313	12154	6339	7134	12694
2005	41902	19216	6981	13401	7982	7458	12455
2006	46580	19667	7807	15283	10194	8483	12777
2007	50721	21379	9544	17697	11579	8764	14257
2008	59133	23941	11382	18232	13483	9502	15791
2009	58206	24373	11405	17880	14548	9411	15709
2010	67820	26294	12563	18839	15027	9897	16257
2011	74770	28262	13868	20798	16848	10383	18204
2012	83278	29300	15388	21228	19413	11258	18836
2013	87635	29779	16929	23599	19877	12684	19549
2014	91034	31559	18363	22955	22051	12574	20090
2015	97773	30927	18924	23720	21912	12979	20270
2016	91041	29488	20781	22892	20737	13533	20782
2017	100577	30879	27681	24737	20761	16458	20076

	Materials Science	Earth Science	Computer Science	Environment	Dentistry	Pharmacy	Energy
1993	4190	2097	1187	1862	62	3781	109
1994	4260	2207	1279	1690	69	3646	118
1995	4276	2269	1419	2484	58	3480	109
1996	4884	2621	1518	2005	78	3259	64
1997	3771	2525	1659	2531	109	3291	91
1998	3770	2704	1789	2204	136	3507	128
1999	4014	2752	2034	2224	142	3458	120
2000	4188	3005	2128	2281	112	3569	106
2001	4116	3123	2120	2806	120	3648	116
2002	4286	3182	2273	2669	115	3602	76
2003	4172	3247	2390	2573	127	3616	73
2004	4653	3509	2577	2320	125	3871	101
2005	4094	3754	2693	2642	123	4144	134
2006	4362	5707	3381	2744	190	4166	156
2007	4524	4641	3864	3103	217	4704	157
2008	6202	6751	4174	3340	270	4268	201
2009	7270	7668	4368	3137	249	853	218
2010	7304	8080	4628	3699	486	777	208
2011	9064	9170	5066	4394	439	891	237
2012	9548	9473	5692	5102	565	889	244
2013	10697	10525	6268	5595	1055	839	37
2014	11276	11246	7079	6592	1384	857	121
2015	12058	12571	6850	7670	1625	888	153
2016	12958	13139	8067	8190	1655	840	413
2017	14318	14864	12086	9276	1780	839	474

An MSE of zero means that the estimator predicts observations with perfect accuracy, which is the ideal. Two or more statistical models may be compared using their MSEs as a measure of how well they explain a given set of observations. In addition, to measure the ranking performance of the most emerging trends, the average precision is used. In the field of information retrieval, precision is the fraction of retrieved documents that are relevant to the search:

$$precision = \frac{|RelevantAndRetrievedDocument|}{RetrievedDocument}$$

C. Comparison among Individual Predictor

The first experiment in this study is to compare the performance of each predictor. There are 2 predictors used, namely (1) Neural Network 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 varying the parameters of those predictors.

The patterns are rather fluctuating but in moderate scale as the trends are generally smooth initially but then moving

upward or downward quite abruptly at the end. Thus time series derived from bibliometric is quite different from economic time series which mostly has seasonal component and repeatable pattern. The MSE is quite low for Neural Network with number of hidden nodes equal 2 or 3. SVR with kernel RBF having smaller sigma and SVR with kernel Polynomial having higher degree also gives quite better result. Recall that smaller sigma value in SVR implies smaller variance which fits the data tighter. Thus, the SME on training data is smaller, but SME on testing data may get bigger as the model tends to overfit. Similar behavior is observed for polynomial kernel of higher degree, which tends to yield poor generalization error.

Table 2 also shows that most predictors have good prediction for some time series but not for the others. For example, SVR polynomial of degree 2 is very good for the first time series but performs badly for fourth time series. Likewise, NN with hidden node 10 yields fair result for fifth dataset but deteriorates for the first dataset. Hence, the ensemble of predictors shall choose the best of predictors for a given pattern in a time series, assuming that the pattern in testing data can be found in the validation data. In this experiment, the number of cross validation folds is three times the number of points to forecast. More records for cross validation is desirable but the feature available for sliding window will become smaller.

D. Combination of Models using Similarity Measure

The second experiment in this study 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 either Euclidean Distance or DTW. The performance of all possible numbers of best models is shown in Table 3 for DTW similarity, Euclidean 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 only by the training sample similar to the testing data. Simple average and ranking by performance method are used to combine the forecasting results.

Table 2 also indicates that the first and second best models 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 decreases up until about half of the total number of model. Figure 2 further shows 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. The MSEs of all methods decrease as more models are added, but they increase after the ninth model is added. Among those three model selections, Euclidean similarity is the one that may yield the lowest MSE.

Lastly, the most often used models as the best models are depicted in Figure 3. To find out the total rank of usage of each model, we assign a certain value for each model based on its performance ranking on all time series such that for n^{th} ranked model will get value of 1/n. For example, if a model is ranked

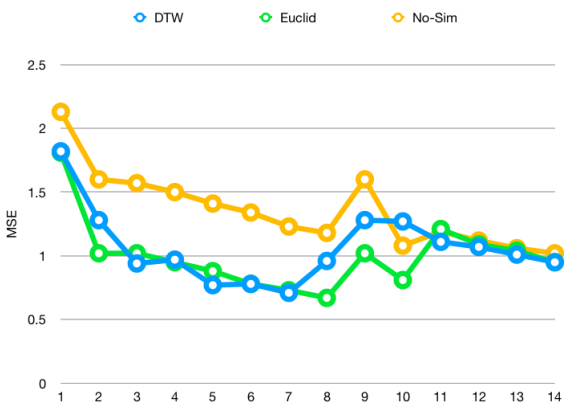

Fig. 2. Average Performance of Forecast Combination using Models Selected by Euclidean and DTW Similarity Compared to the one using best and all Models without Similarity Measure

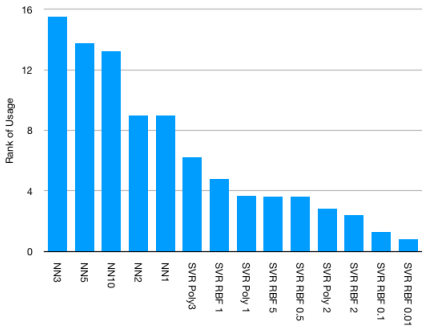

Fig. 3. The Most often Selected Models for the First Eight Models of all Series

first twice, ranked second once and ranked fourth three times then the total rank of usage is $1 \times 2 + \frac{1}{2} \times 2 + \frac{1}{4} \times 3$. Thus, the last ranked model will get a negligible value. It turns out that Neural Networks are chosen more often as best models than the SVRs. Among the NNs, the moderate number of hidden node, such as 3 and 5, are more preferable while among the SVRs, the polynomial kernel of degree 3 and RBF kernel of width 1, which are more suitable for fluctuating pattern, are closely following the NNs.

V. Conclusions

The experimental result shows that the combination of methods selected based on the similarity between training and testing data may perform better compared to the combination of all methods. 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.

For future works, this method shall be tested against many other time series data, especially in the domain of research topics, to confirm its feasibility. There are also many possibilities of employing different predictors other than NN and

TABLE II
FORECASTING PERFORMANCE USING MSE AMONG INDIVIDUAL MODELS ON 14 TIME SERIES

No	Methods	ts1	ts2	ts3	ts4	ts5	ts6	ts7	ts8	ts9	ts10	ts11	ts12	ts13	ts14	Avg
1	NN hn = 1	0.06	0.45	7.19	0.86	0.11	0.07	3.74	0.14	1.82	0.61	2.21	2.53	7.16	0.10	1.92
2	NN hn = 2	0.17	0.75	0.16	0.23	0.06	0.06	1.36	0.40	0.80	0.45	1.09	1.92	0.25	1.33	0.65
3	NN hn = 3	0.41	0.08	0.11	1.53	0.04	0.08	0.86	0.40	2.79	0.67	1.01	0.79	0.05	0.37	0.63
4	NN hn = 5	0.82	2.49	0.59	0.72	0.05	0.06	3.21	2.03	2.46	0.42	0.36	2.47	0.10	2.44	1.30
5	NN hn = 10	19.45	13.87	2.35	0.15	0.04	0.22	6.69	0.39	0.52	0.25	11.26	2.66	6.72	0.28	4.64
6	SVR RBF 0.01	0.33	0.14	1.12	0.07	0.70	0.07	0.93	0.15	0.71	0.44	0.46	0.14	0.13	0.42	0.41
7	SVR RBF 0.1	0.36	0.14	1.04	0.09	0.63	0.10	0.70	0.15	0.76	0.24	0.38	0.15	0.73	0.14	0.41
8	SVR RBF 0.5	0.0	0.21	1.31	0.00	0.59	0.08	3.26	0.00	4.70	0.43	0.04	0.23	0.40	0.90	0.90
9	SVR RBF 1	0.07	0.24	5.00	0.17	1.13	0.13	1.78	0.42	3.68	0.33	0.51	1.48	0.52	0.04	1.11
10	SVR RBF 2	0.00	2.31	0.12	0.72	5.27	0.86	1.41	1.27	2.35	0.39	0.28	1.94	1.97	2.33	1.36
11	SVR RBF 5	0.00	0.65	0.16	0.00	0.14	0.44	0.63	0.38	1.38	0.41	0.63	1.37	0.51	0.14	0.48
12	SVR Poly 1	0.00	10.49	0.13	0.31	0.08	1.34	0.54	2.73	2.39	0.43	88.45	1.57	0.00	0.34	7.77
13	SVR Poly 2	0.21	0.32	116.73	377.74	194.59	2.26	4.20	1.21	1.15	0.33	3.77	1.68	0.33	2.23	50.48
14	SVR Poly 3	0.04	67.59	65.92	18.78	13.87	1.05	2.79	1.11	1.88	0.35	0.82	1.54	65.89	13.89	18.25
	Average	1.57	7.12	14.42	28.73	15.52	0.45	2.29	0.77	1.96	0.41	7.95	1.46	6.05	1.78	6.46

TABLE III
MSE ON COMBINATION OF METHODS USING EUCLIDEAN, DTW AND WITHOUT SIMILARITY

Number of combination	Average			Inverse-MSE			Raank		
	DTW	Euclid	No-Sim	DTW	Euclid	No-Sim	DTW	Euclid	No-Sim
1 best model	1.84	1.77	2.17	1.82	1.81	2.13	1.85	1.79	2.14
2 best models	1.17	1.05	0.97	1.28	1.02	1.60	1.37	1.19	1.25
3 best models	4.05	1.02	1.07	0.94	1.02	1.57	1.95	1.06	1.28
4 best models	2.54	0.90	1.01	0.97	0.95	1.50	1.29	1.03	1.20
5 best models	1.72	0.81	0.96	0.77	0.88	1.41	1.04	0.95	1.15
6 best models	1.40	1.75	1.36	0.78	0.78	1.34	0.92	0.92	1.14
7 best models	1.17	0.67	1.07	0.71	0.73	1.23	0.85	0.82	1.05
8 best models	1.31	0.63	1.05	0.96	0.67	1.18	0.97	0.69	1.03
9 best models	1.17	1.05	0.97	1.28	1.02	1.60	1.37	1.19	1.25
10 best models	1.42	0.90	1.14	1.27	0.81	1.08	0.97	0.68	0.97
11 best models	1.37	1.39	1.31	1.11	1.21	1.19	0.94	0.92	1.01
12 best models	1.27	1.45	1.43	1.07	1.09	1.12	0.82	0.90	0.99
13 best models	1.14	1.21	1.15	1.01	1.04	1.06	0.93	0.86	0.90
14 best models	1.01	1.01	1.01	0.95	0.95	1.01	0.78	0.75	0.87

SVR, such as kernel learning approach to automatically select the kernel. In addition, we need to explore other similarity measures besides the Euclidean and DTW that suit better for comparing testing and training of time series dataset.

REFERENCES

[1] M. Bengisu and R. Nekhili, "Forecasting emerging technologies with the aid of science and technology databases", Technological Forecasting and Social Change, vol. 73, no. 7, (2006) September, pp. 835-844.

[2] H. Small, "Tracking and predicting growth areas in science", Scientometrics, vol. 68, no. 3, (2006), pp. 595-610.

[3] E. Sri and R. Rahayu, "Identification of Technology Trend on Indonesian Patent Documents and Research Reports on Chemistry and Metallurgy Fields", Asia-Pacific Conference on Library and Information Education & Practice, (2006), pp. 581-586.

[4] D. Zhu, A. Porter, S. Cunningham, J. Carlisie and Anustup Nayak, "A process for mining science & technology documents databases, illustrated for the case of knowledge discovery and data mining", Artigos, (1998).

[5] W. L. Woon, "A Framework for Technology Forecasting and Visualization", (2009) October.

[6] S.-ichi Kobayashi, Y. Shirai, K. Hiyane and F. Kumeno, "Technology Trends Analysis", (2005), pp. 820-825.

[7] G.0 P. Zhang, B. E. Patuwo and M. Y. Hu, "A simulation study of artificial neural networks for nonlinear time-series forecasting", vol. 28, (2001).

[8] K. Siwek, S. Osowski and R. Szupiluk, "Ensemble Neural Network Approach for Accurate Load Forecasting in a Power System", International Journal of Applied Mathematics and Computer Science, vol. 19, no. 2, (2009) June, pp. 303-315.

[9] C. Huang, D. Yang and Y. Chuang, "Application of wrapper approach and composite classifier to the stock trend prediction", Expert Systems with Applications, vol. 34, no. 4, (2008) May, pp. 2870-2878.

[10] J. S. Armstrong, "Combining forecasts, in Principles of forecasting: a handbook for researchers and practitioners", Kluwer Academic Publishing, (2001), pp. 417-439.

[11] P. Poncela, J. Rodrguez, R. Snchez-Mangas and E. Senra, "Forecast combination through dimension reduction techniques", International Journal of Forecasting, vol. 27, no. 2, (2011) April, pp. 224-237.

[12] M. R. Hassan, "Neurocomputing A combination of hidden Markov model and fuzzy model for stock market forecasting", Neurocomputing, vol. 72, no. 16-18, (2009), pp. 3439-3446.

[13] R. R. Andrawis, A. F. Atiya and H. El-Shishiny, "Forecast combinations of computational intelligence and linear models for the NN5 time series forecasting competition", International Journal of Forecasting, vol. 27, no. 3, (2011) July, pp. 672-688.

[14] B. Ziegler, "Methods for Bibliometric Analysis of Research : Renewable Energy Case Study by", (2009) September.