

A Prediction Model Based on Relevance Vector Machine and Granularity Analysis

Young Im Cho

Department of Computer Engineering, Gachon University, Seongnam, Korea



Abstract

In this paper, a yield prediction model based on relevance vector machine (RVM) and a granular computing model (quotient space theory) is presented. With a granular computing model, massive and complex meteorological data can be analyzed at different layers of different grain sizes, and new meteorological feature data sets can be formed in this way. In order to forecast the crop yield, a grey model is introduced to label the training sample data sets, which also can be used for computing the tendency yield. An RVM algorithm is introduced as the classification model for meteorological data mining. Experiments on data sets from the real world using this model show an advantage in terms of yield prediction compared with other models.

Keywords: Quotient space theory, Granular computing, RVM, Grey model

1. Introduction

The forecast of yield by analyzing meteorological data sets usually can be accomplished with a statistical model (such as regression model, grey model, or Markov model), a neural network, or another method. The disadvantage of these models is that the fluctuation of the yield caused by the meteorological factors cannot be calculated accurately. Thus, a new kind of model based on the quotient space model of granular computing theory is put forward in this paper. Under the framework of quotient space, a grey model [1] combined with a relevance vector machine (RVM) model [2], which can be called a GM-RVM model, is constructed as a forecast model based on meteorological data analysis.

The aim of adopting quotient space granular computing model is that it can investigate the conversion, interaction and dependency with quotient spaces at different granularities in macrocosmic analysis: 1) knowledge discovery at different granularities 2) reduction of the computational complexity of problem solving. Additionally, the data mining of quotient spaces with different granular structure can be dealt with in microcosmic analysis by adopting the RVM model. On the other hand, grey model can analyze the long term tendency of trend yield, and RVM classification model can reflect the fluctuation of yield caused by the meteorological factors more actually. So the analysis of meteorological time series and forecast of yield can be better done with the combination of these two models.

2. The Framework of Quotient Space Theory

In [1] a quotient space model of problem solving is presented. The model was intended to

Received: Jun. 9, 2016
Revised : Sep. 9, 2016
Accepted: Sep. 10, 2016

Correspondence to: Young Im Cho
(yicho@gachon.ac.kr)
©The Korean Institute of Intelligent Systems

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

describe the universe with different grain-sizes easily and be used for analyzing problem hierarchically [3].

2.1 World Presentations at Different Granularities

Suppose that a triple (X, f, T) describes a problem solving space, where X denotes the problem domain, $f(x)$ indicates the attributes (or features) of domain X or is denoted by a function, $f(x) : X \rightarrow Y$, and T is the structure of domain X [4].

Suppose that X represents the problem domain at the finest grain size. When domain X is viewed at a coarser grain size, $[X]$ is a coarse-grained domain. A coarser universe ($[X]$) can be defined by an equivalence relation (R) on X . That is, an element in $[X]$ is equivalent to a set of elements and an equivalence relation (R) on X . Thus, $[X]$ consists of all of the equivalence classes obtained by R . $([X], [f], [T])$ is called the quotient space of (X, f, T) .

2.2 The Relation between Different Grain Size Worlds

Definition 2.1. Under the relation “ $<$,” R is a complete semi-order lattice.

From this definition, a relation is obtained: $R_0 < R_1 < \dots < R_n$, where R_n is the finest relation in R , and R_0 is the coarsest relation in R . Thus, the quotient space from R can be viewed as a hierarchical structure.

2.3 Attribute Preserving

- 1) In the coarse-grained domain, if there does not exist a solution, there is no solution in the fine-grained domain.
- 2) Assuming there is a solution in semi-order quotient space (X_1, f_1, T_1) , (X_2, f_2, T_2) , there will be a solution in their combined quotient space (X_3, f_3, T_3) .

2.4 Fuzzy Quotient Space

In [5], the quotient space model is extended to the fuzzy granular world, and two main conclusions are given. First, the following four statements are equivalent: 1) a fuzzy equivalence relation given in universe X , 2) a normalized isosceles distance given in quotient space $[X]$, 3) a hierarchical structure $\{X(\lambda)\}$ given in X , 4) a fuzzy knowledge base given in X . The third statement is the most essential of the four statements. Consequently, knowledge of the granular structure can be acquired once the hierarchy is given. Second, a whole world with different fuzzy granularities composes a complete semi-order

lattice. The results provide a powerful mathematical model and tool for granular computing.

3. RVM Algorithm

In [2], RVM is introduced, which is a probabilistic sparse kernel model identical in functional form to the SVM model. SVM has been employed for solving classification and forecasting problem like [6]. RVM model also can be applied for solving classification problem successfully such as [7, 8]. The most compelling feature of the RVM is that, while capable of generalization performance comparable to an equivalent SVM, it typically utilizes dramatically fewer kernel functions. Furthermore, the RVM suffers from none of the other limitations of SVM, such as the kernel function having to satisfy Mercer’s condition and the necessity of estimating the error/margin trade-off parameter “ C .” Because of the advantages of RVM, it was chosen as the classification algorithm in our model in this study. The RVM model is described as follows:

Given a training data set of input-target pairs $\{x_n, t_n\}_{n=1}^N$, $x_n \in R^d$, $t_n \in R$, the output of RVM is:

$$y(x) = \sum_{i=1}^N w_i K(x, x_i) + w_0, \quad (1)$$

where $K(x, x_i)$ is a kernel function, and w_i is the weight value for the model.

Assume that $p(t|x)$ is a Gaussian $N(t|y(x), \sigma^2)$. The mean of this distribution for a given x is modeled by $y(x)$ which is defined in Eq. (1) for SVM. The likelihood of the dataset can be written as

$$p(t|w, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp \left\{ -\frac{1}{2\sigma^2} \|t - \Phi w\|^2 \right\}, \quad (2)$$

where $t = (t_1 \dots t_N)$, $w = (w_0, \dots, w_N)$ and Φ is the $N \times (N + 1)$ ‘design’ matrix with $\Phi_{nm} = K(x_n, x_{m-1})$ and $\Phi_{n1} = 1$. In order to overcome the overfitting problem caused by the maximum-likelihood estimation of w and σ^2 from Eq. (2), w is assigned with the following prior conditional probability distribution:

$$p(w|\alpha) = \prod_{i=0}^N N(w_i|0, \alpha_i^{-1}), \quad (3)$$

with α as a vector of $N + 1$ hyperparameters.

The posterior probability over the weights can be obtained

from Bayes' rule:

$$p(w|t, \alpha, \sigma^2) = (2\pi)^{-\frac{(N+1)}{2}} |\Sigma|^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{2}(w - \mu)^T \Sigma^{-1} (w - \mu) \right\}, \quad (4)$$

with $\Sigma = (\Phi^T B \Phi + A)^{-1}$, $A = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$, and $\mu = \Sigma \Phi^T B t$, $B = \sigma^{-2} I_N$.

By integrating out the weights, the marginal likelihood can be obtained for the hyperparameters:

$$p(t|\alpha, \sigma^2) = (2\pi)^{-\frac{N}{2}} |B^{-1} + \Phi A^{-1} \Phi^T|^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{2} t^T (B^{-1} + \Phi A^{-1} \Phi^T)^{-1} t \right\}. \quad (5)$$

Values of α and σ^2 that maximize (5) cannot be obtained in a closed form with two alternative formulae for iterative re-estimation of α . First, by considering the weights as "hidden" variables, an EM approach gives:

$$\alpha_i^{new} = \frac{1}{\langle w_i^2 \rangle_{p(w|t, \alpha, \sigma^2)}} = \frac{1}{\sum_{ii} + \mu_i^2}. \quad (6)$$

Second, direct differentiation of (5) and rearranging gives:

$$\alpha_i^{new} = \frac{\gamma_i}{\mu_i^2}, \quad (7)$$

where $\gamma_i = 1 - \alpha_i \sum_{ii}$.

For the noise variance, both methods lead to the same re-estimate:

$$(\sigma^2)^{new} = \|t - \Phi \mu\|^2 / (N - \sum_i \gamma_i). \quad (8)$$

4. Granular Analysis of Meteorological Data

A quotient space granular model can combine itself with rough set theory [9] and fuzzy set theory [10, 11] organically [5]. It emphasizes the conversion and dependency of a universe with different grain sizes and can be helpful for data mining from different granularities.

Basic issues of granular computing may be studied from two related aspects: the construction of granules and computation with granules. The former deals with the formation, representation, and interpretation of granules, whereas the latter deals with the utilization of granules in problem solving [12].

4.1 Construction of Information Granules

Quotient space theory is used for considering a problem from different granules, and choosing a proper grain-size of a universe for problem solving.

Four general principles are discussed in quotient space theory for partitioning: 1) attribute-based method, 2) projection-based method, 3) the function-based method, 4) the constraint-based method [4]. This paper adopts (1) for partitioning the problem domain, and the domain X is constructed of meteorological time series data sets. The attributes of $f(x)$ are constructed by key meteorological factors: $f_1(x)$, $f_2(x)$, $f_3(x)$, where $f_1(x)$: average temperature; $f_2(x)$: rainfall; $f_3(x)$: sunlight, and the meteorological time attribute function $f_4(x)$ which is used for constructing the set of equivalence relation R . Against the characteristics of meteorological time series data sets, domain X is partitioned by $f_4(x)$. The original grain-size for X stored in a database is a day. Suppose that grain size L equals one day, 10 days, or one month, by which X can be fuzzy partitioned as $X(L)$. Additionally, these partitions can be constructed as hierarchical structures with three levels, which are $([X], [f], [T])_{day}$, $([X], [f], [T])_{10days}$, $([X], [f], [T])_{month}$.

If the grain size is too fine, the model cannot obtain good results. Conversely, if the grain size is too coarse, the model cannot reflect the characteristics of the meteorological time series sensitively. Thus, a proper partition may not be a single grain size, but some combination of different granularities should be required. Combination technology in quotient space theory will not only emphasize the key factors but will also not neglect the subordinate factors.

The combination in quotient space includes a combination of domains, a combination of structure, and a combination of attribute functions. This paper discusses the combination of domains.

Definition 4.1. Assume that R_1 and R_2 are any two equivalence relations on X . If R_3 is the lowest upper bound of R_1 and R_2 in the semi-order lattice, then R_3 is the combination of R_1 and R_2 . If $X_1 = \{a_i\}$ and $X_2 = \{b_j\}$ are two partitions with respect to R_1 and R_2 , respectively, then the combination of X_1 and X_2 can be represented by $X_3 = \{a_i \cap b_j | a_i \in X_1, b_j \in X_2\}$.

By the "no-solution preserving property" in quotient space model which states that if there does not exist a solution in the coarse-grained domain, there is no solution in the fine-grained domain, the computational complexity of problem solving can be reduced.

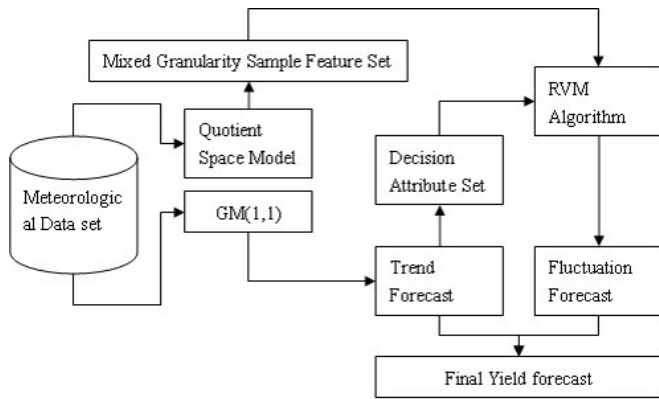


Figure 1. The structure of our prediction model.

4.2 Yield Forecast Model Based on Analysis of Meteorological Data

The structure of our model is depicted in Figure 1, and the detailed steps are presented as follows:

Step 1: Domain X , in which every sample x is part of an annual meteorological character time series for crop yield, that is every day’s sunlight, average temperature and rainfall in special period (s_1 to s_2), is partitioned by time attribute $f_4(x)$, which equals one day, 10 days, and one month. Thus, a quotient space with three levels is constructed

Step 2: A combination of the three levels can be conducted by a combination method. One possible granular analysis model is described below.

First, every month in the chosen period (s_1 to s_2) is analyzed with the different grain sizes and combined with a combination method. The key meteorological factors (g) are given, where $g \in \{\text{sunlight, averagetemperature, rainfall}\}$, and g_m represents the subordinate factors. Then, a combination is conducted for the key factors (g) with the finer grain size. By Definition 4.1, the result is the intersection grain size of 10 days, which is \bar{g}_a, \bar{g}_b and \bar{g}_c (the average value of the first 10 days, middle 10 days, and last 10 days), the g_m holds the value \bar{g}_m (average value of month). Finally, the time series in this month is: $\bar{g}_a, \bar{g}_b, \bar{g}_c,$ and \bar{g}_m . In this way every month in this period (s_1 to s_2) is discussed, and new feature data sets can be formed.

Step 3: $\hat{Y} = \hat{Y}_t + \hat{Y}_w$, where \hat{Y} is the crop yield, and \hat{Y}_t is the trend yield, \hat{Y}_w is the meteorological yield.

1) Annual \bar{y}_t can be obtained by grey model GM(1, 1), and regarded as the trend yield.

2) The fluctuation of the forecast: $z = (y_t - \bar{y}_t)/\bar{y}_t, t = 1, 2, \dots, n$, where y_t is the actual yield of year t , \bar{y}_t is the trend

yield of year t , and $y_t - \bar{y}_t$ is regarded as the meteorological yield. Thus, the meteorological yield can be separated in this way and the decision label will be determined depending on this fluctuation forecast.

Step 4: First, the meteorological feature data sets constructed from Step 2 are combined and normalized. Second, the decision labels computed in Step 3 are added to the last dimension of the training sample. After that, the new training sample data sets will be regarded as the input instance sets to construct the RVM model.

Step 5: The meteorological feature data sets that need to be forecasted can be input to the RVM model constructed by Step 4, which then outputs the fluctuation of forecast z .

Step 6: By a combination of the trend forecast and fluctuation forecast, the final result can be obtained.

5. Experiment and Results

This experiment concerns the yield forecast for winter wheat in Anhui province in China. Three districts are sampled to test the GM-RVM model based on quotient space theory presented in this paper, and the predicted years range from 1996 to 2000. The meteorological time series ranges from 1980 to 2000, after the combination with feature data sets obtained by the space quotient granular analysis and the decision label produced by the grey model. The total number of instances is 200, and the number of dimensions in each sample is 90.

In this experiment, a Gaussian kernel is chosen as the kernel function for RVM:

$$K(x, x_i) = \exp \left\{ -\frac{\|x - x_i\|^2}{2\sigma^2} \right\}.$$

As shown in Figures 2-4, the relative error of the forecast is from three different models, including the grey model (GM(1,1)), the GM-Markov model, which is the combination of GM(1,1) and the Markov model, and our GM-RVM based on quotient space theory. The GM-Markov model is similar to our model and also separates the yield prediction into two parts, including a trend prediction part, for which a grey model can be used, and a fluctuation prediction part, for which a Markov model can work. The relative error can be defined as the absolute error divided by the actual value.

As we can see from Figures 2-4, the forecast relative error for our model is relatively smaller than that of other models in most cases. Especially in 2000, the average relative error of GM(1,1) is above 15%, the error of the GM-Markov model is

Table 1. Average relative error for prediction in three districts with three models

	1996	1997	1998	1999	2000
GM(1,1)	-7.6%	9.13%	8.13%	-17.2%	17.5%
GM-Markov	-6.17%	7.6%	10.1%	-9.93%	10.4%
GM-RVM	-4.13%	6.67%	6.33%	-7.73%	6.3%

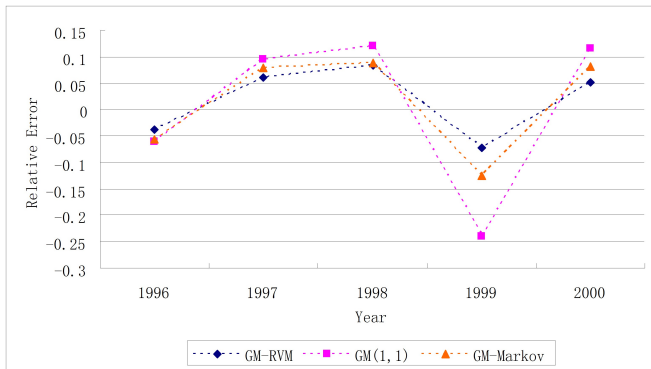


Figure 2. Relative error for GM-RVM, GM(1,1) and GM-Markov models in district 1.

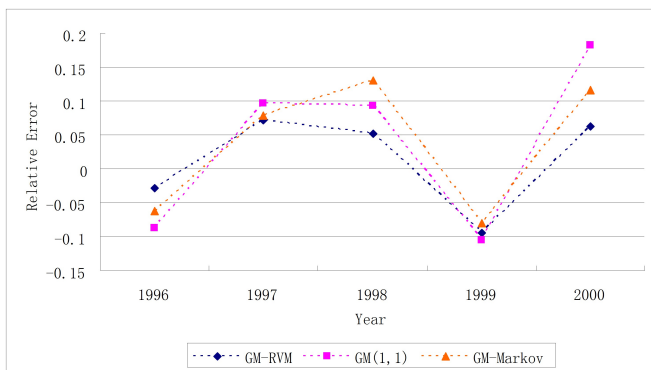


Figure 3. Relative error for GM-RVM, GM(1,1) and GM-Markov models in district 2.

above 10%, but the error of the GM-RVM is only 6.3%. Thus, the GM-RVM can improve the prediction performance. The average forecast error is shown in Table 1.

The results in Table 1 show that by using the model introduced in this paper, the forecast precision can be improved by a combination of the grey model and an RVM algorithm, compared with the other two models.

In order to discover if the granularity factor is important for the improvement of precision or whether a suitable grain size is chosen is important, the prediction of our model with mixed

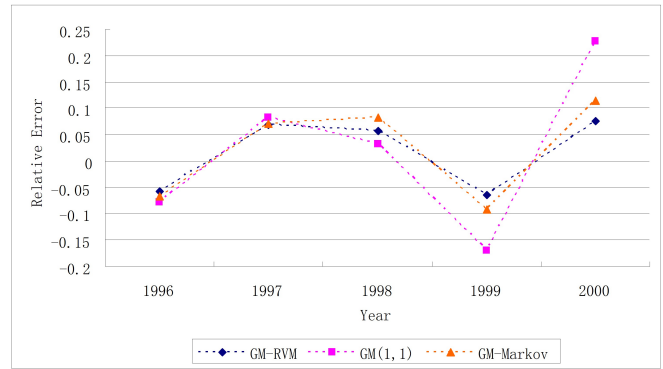


Figure 4. Relative error for GM-RVM, GM(1,1) and GM-Markov models in district 3.

Table 2. Yield forecast comparison in 2000 between mixed granularity and month granularity

Grain-size	Training number	Testing number	Relative error
Mixed	190	10	6.3%
Month	190	10	9.1%

granularity by quotient space analysis is compared with this model only using the coarse grain size of one month, which is shown in Table 2.

Table 2 shows the average relative forecast error in three districts using the model introduced in this paper with mixed and single-month grain sizes. The results show that the prediction performance can be improved by a combination of different granularities with quotient space theory, compared with only using a single-month granularity.

6. Conclusion

For the meteorological data sets, under the framework of the quotient space granular computing model, complicated information of meteorological time series is analyzed at different grain sizes. A model that combines a grey model (GM(1, 1)) and an RVM algorithm is introduced for meteorological data mining (yield forecast). Experiments on real-world data sets using this model show that this is an efficient way to obtain structural knowledge with granular analysis.

On the other hand, in a general situation, data sets associated with crop yield are limited and incomplete, and this is the bottleneck that restricts research for prediction and pattern reorganization in this area, which is also a cause for prediction error.

Consequently, how to deal with data mining in incomplete data sets efficiently is a direction that should be discussed in the future [13].

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgements

This paper is supported by the National Research Foundation (NRF) in Korea Government (Ministry of Science, ICT & Future Planning) (No. 2015R1A2A2A03004088).

References

- [1] J. L. Deng, "Introduction to grey system theory," *Journal of Grey System*, vol. 1, no. 1, pp. 1-24, 1989.
- [2] M. E. Tipping, "The relevance vector machine," in *Proceedings of 1999 Advances in Neural Information Processing Systems*, Denver, CO, 1999, pp. 652-658.
- [3] B. Zhang and L. Zhang, *Theory and Applications of Problem Solving*. Amsterdam: North Holland, 1992.
- [4] L. Zhang and B. Zhang, "The quotient space theory of problem solving," *Fundamenta Informaticae*, vol. 59, no. 2-3, pp. 287-298, 2004.
- [5] Z. Ling and Z. Bo, "Theory of fuzzy quotient space (methods of fuzzy granular computing)," *Journal of Software*, vol. 14, no. 4, pp.770-776, 2003.
- [6] S. J. Kim and I. Y. Seo, "A clustering approach to wind power prediction based on support vector regression," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 12, no. 2, pp. 108-112, 2012. <http://dx.doi.org/10.5391/IJFIS.2012.12.2.108>
- [7] F. A. Mianji and Y. Zhang, "Robust hyperspectral classification using relevance vector machine," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp.2100-2112, 2011. <http://dx.doi.org/10.1109/TGRS.2010.2103381>
- [8] Q. Duan, J. G. Zhao, Y. Ma, and K. Luo, "RVM and SVM for classification in transient stability assessment," in *Proceedings of 2010 Asia-Pacific Power and Energy Engineering Conference*, Chengdu, China, 2010, pp. 1-4. <http://dx.doi.org/10.1109/APPEEC.2010.5448612>
- [9] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*, Dordrecht, Netherlands: Kluwer Academic Publishers, 1991.
- [10] L. A. Zadeh, "Fuzzy logic = computing with words," *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 2, pp. 103-111, 1996. <http://dx.doi.org/10.1109/91.493904>
- [11] L. A. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets and Systems*, vol. 90, no. 2, pp. 111-127, 1997. [http://dx.doi.org/10.1016/S0165-0114\(97\)00077-8](http://dx.doi.org/10.1016/S0165-0114(97)00077-8)
- [12] Y. Y. Yao, "Granular computing: basic issues and possible solutions," in *Proceedings of the 5th Joint Conference on Information Sciences*, Atlantic City, NJ, 2000, pp. 186-189.
- [13] X. Wu, X. Zhu, G. Q. Wu, and W. Ding, "Data mining with big data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 1, pp. 97-107, 2014. <http://dx.doi.org/10.1109/TKDE.2013.109>



Young Im Cho received her B.S., M.Sc., and Ph.D. from the Department of Computer Science, Korea University, Seoul, Korea, in 1988, 1990 and 1994, respectively. She is a professor at Gachon University. Her research interest includes AI, Big data, information

retrieval, smart city etc.

Tel: +82-31-750-5800

Fax: +82-31-750-5662

E-mail: yicho@gachon.ac.kr