CHEMICAL ENGINEERING TRANSACTIONS
VOL. 66, 2018
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet
Guest Editors: Songying Zhao, Yougang Sun, Ye Zhou

ISBN 978-88-95608-63-1; **ISSN** 2283-9216
DOI: 10.3303/CET1866001

Study on the Importance of Multi-components of Slag-based Composite Cementitious Material by LSMI Feature Selection

Hongmei Liu[a], Chengbin Liu[*b], Rong Li[a]

[a]Department of Information Technology, Beijing Vocational College of Agriculture, Beijing 102442, China
[b]Department of Hydraulic and Architectural Engineering, Beijing Vocational College of Agriculture, Beijing 102442, China
70653@bvca.edu.cn

The slag-based compound cementitious material (SBCCM) consists of five components, according to the traditional test methods for trial, the workload is very heavy, the efficiency is low. A least-squares mutual information (LSMI) method was created for feature selection of SBCCM. The paper used the LSMI feature selection to sort the importance of the material components, selected the three components that play an important role in the material, and then used the orthogonal test method to quickly test out the optimal ratio of the material. The LSMI is a regression-oriented method capable of solving difficult probability density problems. It sorts the importance of input features by their dependence on output values, thereby screening out important features. The results showed that the LSMI was an effective feature selection method, and it can greatly reduce the workload in the preparation of the SBCCM.

1. Introduction

Slag-based composite cementitious material (SBCCM) is prepared with as many as five components, including superfine slag powder, building plaster, lump lime, magnesia and sodium chloride. The traditional trial preparation method often results in a huge workload. To enhance the efficiency, this paper identified the key components of the SBCCM through feature selection, and then adjusts the amount of these components to obtain the best mixture ratio.

Feature selection is an important data pre-processing method. By this method, a feature subset is selected to give full play to the related models and algorithms. One of the main application fields for feature selection lies in machine learning (Comon, 1994; Guyon and Elisseeff, 2003), In this field, the feature selection approaches include the linear regression model, the random forest, the support vector machine (SVM), etc.

In the linear regression model, the features are selected by the regression coefficient. The feature importance is positively correlated with the regression coefficient in the model. Thus, the coefficient equals zero if the corresponding feature is irrelevant to the output variables. To prevent overfitting and enhance generalization, the linear regression model is often added with the penalty factors like L1/Lasso and L2/Ridge regression. However, this model only applies to the case that the features are linearly correlated with the output variables but independent from each other. It is difficult to make a good prediction with the model if the features are associated with each other.

In the random forest, decision trees serve as the basic sorters, and the features are selected based on information gain or mutual information. Considering information gain, information gain rate and Gini coefficient, the random forest method is, in most cases, suitable for feature selection in classification problems. Nevertheless, the prediction accuracy may decline sharply in the evaluation of unimportant features.

Our research tackles a regression-oriented problem, in which the features are associated with each other but rather limited in number. The abovementioned methods either underperform in the selection of interconnected features, or loss stability in the evaluation of unimportant features. Suffice it to say none of them are applicable to our problem.

In light of the above, a least-squares mutual information (LSMI) method was created for feature selection of SBCCM. The LSMI is a regression-oriented method capable of solving difficult probability density problems (Kanamori et al., 2009; Van Hulle, 2005). It sorts the importance of input features by their dependence on

Please cite this article as: Liu H., Liu C., Li R., 2018, Study on the importance of multi-components of slag-based composite cementitious material by lsmi feature selection, Chemical Engineering Transactions, 66, 1-6 DOI:10.3303/CET1866001

output values, thereby screening out important features. Here, this method is adopted to analyse the components of SBCCM and sort the importance of the components.

2. LSMI feature selection

2.1 Overview

Mutual information (MI) refers to the relevance between two random variables, and the decrease in uncertainty of one random variable when the other random variable is given. If two random variables have no correlation, the MI will be 0; if one random variable eliminates the uncertainty of the other random variable, then the MI between them will be maximized. The maximum MI is called the mutual information entropy. Therefore, the MI not only determines the existence/absence of the correlation between two random variables, but also reflects the closeness of the correlation (Suzuki et al., 2009; Cover and Thomas, 1991; Kraskov et al., 2004; Torkkola, 2003).
For two random variables X and Y, their MI is defined as follows:

$$I(X;Y)=\sum_{y\in Y}\sum_{x\in X}p(x,y)\log(\frac{p(x,y)}{p(x)p(y)}) \tag{1}$$

where p(x,y) is the joint probability distribution function of X and Y; p(x) and p(y) are the marginal probability density functions of X and Y, respectively. The MI between parameters help to disclose the effect of each parameter on the overall strength of SBCCM.
In actual computation, it is difficult to estimate probability density. Thus, the LSMI method was employed to calculate the density ratio directly, avoiding the estimation of probability density. The specific procedure is as follows:

$$w(x,y)=\frac{p(x,y)}{p(x)\,p(y)} \tag{2}$$

In analytical forms, the LSMI can compute the MI in an effective manner. To estimate the density ratio, the mean square error (MSE) of the MI is obtained as:

$$\overline{I_S}(X,Y)=\frac{1}{n^2}\sum_{i,j=1}^{n}(\overline{w}(x_i,y_i)-1)^2 \tag{3}$$

In this research, the components of SBCCM related to and interacted with each other. The LSMI method was adopted to ascertain to degree of association between these components, and the effect of one component on the uncertainty of another.

2.2 Component sorting

As mentioned above, SBCCM is prepared from superfine slag powder, building plaster, lump lime, magnesia and sodium chloride. During the sampling, the mass (g) of each component in 100g geopolymer was taken as the test basis. The 3d and 38d folding strength and compressive strength (MPa) were obtained for each of the five main components. According to the international standard, the 28d compressive strength was taken as the output value, as shown in Table 1 (9 out of the 36 samples are listed).

Table 1: Measured data

SN	Nacl	Magnesia	Building plaster	Lump lime	Slag	3d folding strength	3d compression strength	28d folding strength	28d compression strength
1	2	8	5	5	80	3.56	8.93	4.69	21.34
2	2	8	5	10	75	1.55	7.22	4.5	21.25
3	2	8	5	15	70	0	8.63	6.05	27.36
4	2	8	10	5	75	3.15	10.84	6.82	34.58
5	2	8	10	10	70	0	5.53	3.09	20.4
6	2	8	10	15	75	0	5.33	5.02	20.05
7	2	8	15	5	70	3.47	7.69	3.94	19.7
8	2	8	15	10	75	0	5.74	3.52	17.15
9	2	8	15	15	60	0	4.67	2.98	16.78

Four groups of measured data were chosen as test samples, converted to .txt files, and inputted into the LSMI model for evaluation. The importance ranking of the components (in descending order) is as follows:

Magnesia > building plaster > sodium chloride > superfine slag powder> lump lime

3. Experiments and analysis

To verify the performance of the LSMI feature selection, a prediction model was created to simulate all the features against the output. The five features were divided into important ones (magnesia, building plaster, sodium chloride) and unimportant ones (superfine slag powder, lump lime). The prediction model combines the ridge regression in machine learning with cross validation.

3.1 Experiments

First, all components of SBCCM were involved in the prediction, with the predicted performance score of 0.828157483663, MSE of 0.593163876378 and RMSE of 0.770171329237. The predicted results are shown in Figure 1.

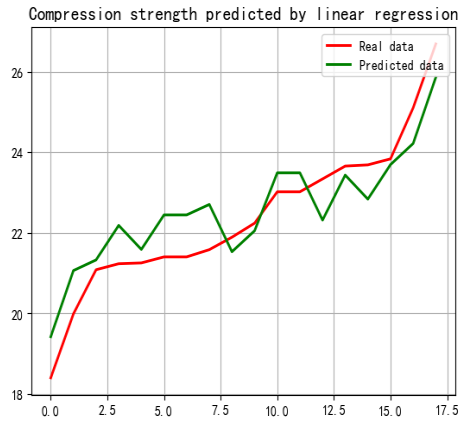

Figure 1: Predicted results of all components

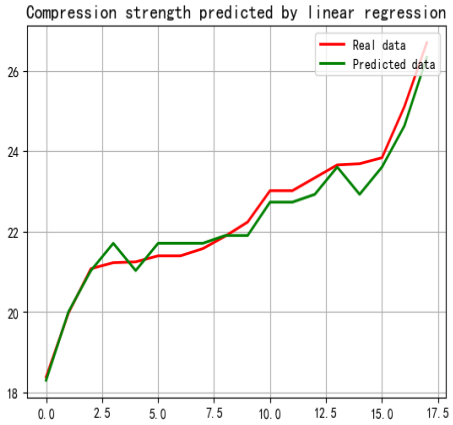

Figure 2: Predicted results of important components

Then, three important components, namely, magnesia, building plaster and sodium chloride were involved in the prediction, with the predicted performance score of 0.850993506536, MSE of 0.514338774553 and RMSE of 0.71717415915. The predicted results are shown in Figure 2.

Finally, the two unimportant components, including superfine slag powder and lump lime, were involved in the prediction, with the predicted performance score of -0.017261978477, MSE of 3.51137233852 and RMSE of 1.87386561378. The predicted results are shown in Figure 3.

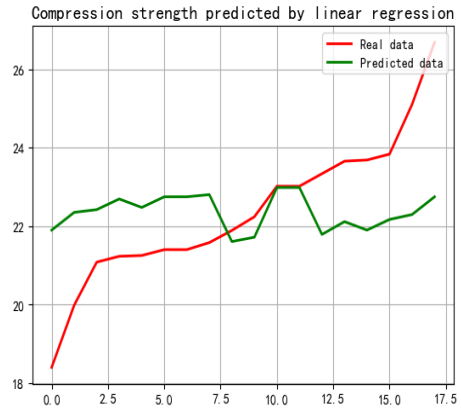

Figure 3: Predicted results of unimportant components

3.2 Results analysis

Table 2: Comparison of predicted results

Feature combination	Score	Mse	Rmse
Magnesia, building plaster, NaCl, slag, lump lime	0.828	0.593	0.770
Magnesia, building plaster, NaCl	0.851	0.514	0.717
Slag, lump lime	-0.017	3.511	1.874

According to the results of the 3 experiments in Table 2, the LSMI selection of important features is valid and feasible for regression problem with correlated features. The prediction model produced with the three important features manged to improve prediction performance, reduce variance and MSE, and increase the predictive ability. On the contrary, the prediction model produced with the two unimportant features underwent a sharp decline in prediction performance, as the variance reached 3.5.
From Figures 1, 2 and 3, it is observed that the prediction made according to important features is consistent with that made according to all features, while the prediction made according to unimportant features deviates greatly from that made according to all features. Therefore, it is meaningful to analyse the key components of SBCCM based on LSMI feature selection.

4. Importance mechanism of SBCCM components

In light of the foregoing analysis, some adjustments were made on magnesia, building plaster and sodium chloride. Based on Table 1, orthogonal tests were performed on magnesia with its content being 6, 8 and 10, building plaster with its content being 5, 10 and 15, and sodium chloride with its content being 0, 2 and 4. Finally, the optimal mixture ratio of SBCCM was obtained as: magnesia: building plaster: sodium chloride: lump lime: superfine slag powder = 8:10:2:5:75.

4.1 XRD results and analysis

In author's papers (Liu et al., 2014; 2015), the XRD diffraction image showed that the hydration products of SBCCM included: $3CaO•Al_2O_3•(0.5CaCl_2•0.5CaSO_4)•12H_2O$ (Calcium Aluminium Chloride Sulphate Hydrate) and $Ca_{1.5}SiO_{3.5}•xH_2O$ (Calcium Silicate Hydrate). In addition, the SBCCM also produced $0.8CaO•0.2Na_2O•Al_2O_3•3SiO_2•6H_2O$ (Unnamed zeolite), $Mg_9(SiO_4)_4(OH)_2$ (Magnesium Silicate Hydroxide) and $Ca_6Al_2(SO_4)_3(OH)_{12}•26H_2O$ (Ettringite) after hydration. The silicon dioxide in superfine slag powder were too many to be consumed up in the reaction. Thus, the test block still contained substantive silicon dioxide. The F diffraction peak stands for the unreacted silicon dioxide in the superfine slag powder.

4.2 Influence on strength from building plaster and sodium chloride

The principal component of magnesia is magnesia, with hydrated products including M-S-H gel, $Mg(OH)_2$, 518 phase and 318 phase. It has two reaction paths: one is that the active magnesia directly produces $Mg(OH)_2$ under the alkaline environment and then reacts with reactive silica in the slag; the other is that active magnesia in the slag compound curing agent may have magnesium oxychloride cement reaction with Cl iron in NaCl to produce $Mg(OH)_2$, 518 phase and 318 phase.

1. Hydration reaction with the first path

Calcium oxide in the slag compound curing agent produces calcium hydroxide through the following reaction:

$$CaO+H_2O=Ca(OH)_2 \quad (4)$$

Active magnesia will have the following reaction:

$$MgO+H_2O=Mg(OH)_2 \quad (5)$$

Mg^{2+} will first bond with OH^- to produce $Mg(OH)_2$ since $Mg(OH)_2$ has a lower solubility than $Ca(OH)_2$. It will react with active SiO_2 dissolved in replacement of $Ca(OH)_2$:

$$Mg(OH)_2+SiO_2+H_2O\rightarrow M\text{-}S\text{-}H \quad (6)$$

2. Hydration reaction with the second path

Magnesia and magnesium chloride solution will produce magnesium oxychloride compound, which has been studied by many scholars both at home and abroad.
Different amount of magnesium chloride will produce different magnesium oxychloride compounds. $MgO/MgCl_2$ mol ratio was used to represent the amount of magnesium chloride. When the amount of $MgCl_2$ is small, the mol ratio is higher and vice versa.

Matkovio studied steady hydration products of magnesium oxychloride cement under different mol ratios. When $MgO/MgCl_2$ was less than 4, the five-phase $Mg_3(OH)_5Cl•4H_2O$ was first produced, with instable property, and gradually developed into the triphase $Mg_2(OH)_3Cl•4H_2O$ with progress of duration. The lower the $MgO/MgCl_2$ is, the faster the transformation speed will be. This is to say with increase in the amount of magnesium chloride, after the five-phase product was produced, substantive magnesium chloride would be left and reacted with $Mg(OH)_2$ to produce $Mg_2(OH)_3Cl•4H_2O$. Thus it can be seen that the great the amount of magnesium chloride is, the faster the triphase production speed will be. Therefore, to enable production of stable five-phase products produced by magnesium oxychloride cement, it is of vital importance to control the amount of magnesium chloride.

When $MgO/MgCl_2$ was greater than 6, which means the amount of magnesium chloride was small, the amount of magnesia was very high, the triphase product of $Mg(OH)_2$ could also be produced under normal temperature. In the solution, however, substantive $Mg(OH)_2$ would be left for reacting with the triphase to produce the five-phase products. Hence the five-phase products can only be stable when $MgO/MgCl_2$ fell into the range of 4-6.

Cole et al. also studied stability of hydration products of magnesium oxychloride cement in the air. They found out that the five-phase ($Mg_3(OH)_5Cl•4H_2O$) or triphase ($Mg_2(OH)_3Cl•4H_2O$) was instable in the air and could be carbonized by CO_2 to produce $Mg(OH)_2•2MgCO_3•MgCl_2•6H_2O$. Walter-Levy pointed out since $Mg(OH)_2•2MgCO_3•MgCl_2•6H_2O$ had smaller solubility than hydration product of magnesium oxychloride, so that hydration product of magnesium oxychloride was easy to be carbonized in the air, to finally produce the stable carbon-magnesium chlorate compound $Mg(OH)_2•2MgCO_3•MgCl_2•6H_2O$.

Matkovic also studied stability of hydration products of magnesium oxychloride cement in the air and draw similar conclusions as Cole. When $MgO/MgCl_2$ was less than 4, however, the five-phase or triphase was carbonized into $Mg(OH)_2•2MgCO_3•MgCl_2•6H_2O$ slowly, with only carbonized surface and basically unaffected inside, so little change happened to strength of the test block. When the test samples were placed in the natural environment for a long term, especially affected by rain, the surface of the original test block would be carbonized into $Mg(OH)_2•2MgCO_3•MgCl_2•6H_2O$, which would resolve to $4MgCO_3•Mg(OH)_2•4H_2O$. Hydration products inside the test block were mainly the five-phase and also included some magnesium hydroxide and carbon magnesium chlorate.

Substantive research shows that in the magnesia, magnesium chloride and water systems, hydration products were influenced by many factors and were instable, making a distinctive feature of magnesium oxychloride cement. Yet above analysis concluded that when $MgO/MgCl_2$ ranked 4-6, the five-phase of hydration products could exist stably. In other conditions, however, hydration products would have transformation or carbonization, to change the internal crystallization structure of the test block, thereby to reduce strength.

Under room temperature 518 phase would change into $Mg(OH)_2$ and 318 phase, and the latter would change into $Mg(OH)_2$:

$$5Mg(OH)_2·MgCl_2·8H_2O(s)=3Mg(OH)_2·MgCl_2·8H_2O(s)+2Mg(OH)_2(s) \quad (7)$$

$$5Mg(OH)_2·MgCl_2·8H_2O(s)=5Mg(OH)_2(s)+Mg2+(aq)+2Cl-(aq)+8H_2O(l) \quad (8)$$

$$3Mg(OH)_2·MgCl_2·8H_2O(s)=3Mg(OH)_2(s)+Mg2+(aq)+2Cl-(aq)+8H_2O(l) \quad (9)$$

$$MgO+H_2O=Mg(OH)_2 \quad (10)$$

The magnesium hydroxide produced would continue to react with reactive silica to produce M-S-H gel.

4.3 Influence on strength from building plaster and NaCl

The concentration of Ca^{2+} and SO_4^{2-} in the building plaster affects the hydration degree of superfine slag powder. The SO_4^{2-} could react with calcium aluminate hydrate to produce needle-rod-like crystal woodfordite. It is obvious that the concentration is positively correlated with the amount of SO_4^{2-}, the production speed and strength of the product. Of course, there should not be an excessive amount of building plaster in the SBCCM. The reason is similar to that for the control of gypsum amount in cement. Too many building plasters may slow down the hydration speed and delay the coagulation. This is because the woodfordite included on the particle surface could prevent external moisture from entering the inside of the granule.

The hydration reactions were complicated by superfine slag powder, lump lime and gypsum:

$$CaO+H_2O=Ca(OH)_2 \quad (11)$$

$$xCa(OH)_2+SiO_2+mH_2O=xCaO•SiO_2•(m+x)\ H_2O \quad (12)$$
(active SiO_2) (C-S-H)

$$yCa(OH)_2+Al_2O_3+nH_2O=yCaO•Al_2O_3•(y+n)\ H_2O \quad (13)$$

(active Al_2O_3)

$$yCaO•Al_2O_3•(y+n)\ H_2O+CaSO_4+H_2O→Ca_6Al_2(SO_4)3(OH)_{12}•26H_2O \quad (14)$$

During the $yCaO•Al_2O_3•(y+n)\ H_2O$ reaction, a part of calcium aluminate hydrate reacted with gypsum to produce calcium sulfoaluminate hydrate (AFt), namely woodfordite. The other part of calcium aluminate hydrate reacted with sodium chloride to produce $3CaO•Al_2O_3•CaCl_2•10H_2O$ (Fridel salt):

$$yCaO•Al_2O_3•nH_2O+NaCl→3CaO•Al_2O_3•CaCl_2•10H_2O+NaOH \quad (15)$$

After Fridel salt was produced, Fridel salt continued to react with gypsum under the growing concentration of SO_4^{2-}, with SO_4^{2-} replacing some part of Cl^- to produce Kuzel salt:

$$3CaO•Al_2O_3•CaCl_2•10H_2O+CaSO_4•2H_2O→3CaO•Al_2O_3•(0.5CaCl_2•0.5CaSO_4)•12H_2O+CaCl_2 \quad (16)$$

By products (NaOH and $CaCl_2$) from the foregoing reaction continued to work as stimulator of the superfine slag powder, which guaranteed the sustained growth of strength of the net paste test block.

5. Conclusions

Focusing on the key components of SBCCM, the LSMI feature selection reveals that the important features of SBCCM are magnesia, building plaster and sodium chloride. The three key components were used for prediction and achieved good predictive results, proving LSMI's feature selection ability in regression problems. Based on the analytical results, the optimal mixture ratio of SBCCM was adjusted to magnesia: building plaster: sodium chloride: lump lime: superfine slag powder = 8:10:2:5:75. The net-paste test block produced at this mixture ratio had an increase in its strength. Therefore, it is concluded that the LSMI is an effective feature selection method, and it can greatly reduce the workload in the preparation of the SBCCM.

References

Comon P., 1994, Independent component analysis, a new concept, Signal Processing, 36(3), 287-314, DOI: 10.1016/0165-1684(94)90029-9

Cover T.M., Thomas J.A., 1991, Elements of Information Theory, John Wiley &Sons, Inc., N. Y., 429, DOI: 10.1198/jasa.2008.s218

Guyon I., Elisseeff A., 2003, An introduction to variable feature selection, Journal of Machine Learning Research, 3, 1157-1182, DOI: 10.1162/153244303322753616

Kanamori T., Hido S., Sugiyama, M., 2009, A least-squares approach to direct importance estimation, Journal of Machine Learning Research, 10, 1391-1445, DOI: 10.1145/1577069.1755831

Kraskov A., Stogbauer H., Grassberger P., 2004, Estimating mutual information, Physical Review E(3), 69(6), 066138, 16.

Liu C.B, Ji H.G., Liu J.H., He W., Gao C., 2015, Experimental study on slag composite cementitious material for solidifying coastal saline soil, Journal of Building Materials, 18(1), 82-87, DOI: 10.3969/j.issn.1007-9629.2015.01.015

Liu C.B., Fan J.L., Gao X.Q., Zhou Y.C, Chen X.D., 2014, Compressive-flexing resistance and microstructure of NaCl-mixed slag cementitious agent, Journal of Lanzhou University of Technology, 40(3), 130-134.

Liu C.B., Gao X.Q., Zhou Y.C., 2014, Research on hydration mechanism of slag composite cementitious material mixed with NaCl, Journal of Chemical and Pharmaceutical Research, 6(4), 845-849.

Liu C.B., Ji H.G., Liu J.H., 2014, Characteristics of slag fine-powder composite cementitious material-cured coastal saline soil, Emerging Materials Research, 36(6), 282-291, DOI: 10.1680/emr.14.00021

Suzuki T., Sugiyama M., Kanamori T., Sese, J., 2009, Mutual information estimation reveals global associations between stimuli and biological processes, BMC Bioinformatics, 10(1), S52, DOI: 10.1186/1471-2105-10-S1-S52

Torkkola K., 2003, Feature extraction by non-parametric mutual information maximization, Journal of Machine Learning Research, 3, 1415-1438, DOI: 10.1162/153244303322753742

Van Hulle M.M., 2005, Edgeworth approximation of multivariate differential entropy, Neural Computation, 17(9), 1903-1910, DOI: 10.1162/0899766054323026