*Sensors & Transducers, Vol. 170, Issue 5, May 2014, pp. 262-266*

**Sensors & Transducers**



*http://www.sensorsportal.com*

# Improvement of Proximal Support Vector Machine and Its Application in a New Method of Making a Mixed Refrigerant in the Ground Source Heat Pump System

**[*] Yan Manfu, Wang Jiuhai**
Tangshan Normal University, Hebei province Tangshan City, Hebei Province,
156 North Road Construction, 063000, China
[*] Tel.: 18633115758, fax: 8603153863012
[*] E-mail: 3005@tstc.edu.cn

*Received: 27 February 2014 /Accepted: 30 April 2014 /Published: 31 May 2014*

**Abstract:** To solve the problem of the refrigerant performance in a ground source heat pump system, the existing proximal support vector machine [1] has been improved and updated into a Weighted Proximal Support Vector Machine (PSVM) model. Meanwhile, through analyzing and researching into the commonly-used refrigerants in the ground source heat pump system, a new method of making a mixed refrigerant for the system is put forward by applying PSVM. *Copyright © 2014 IFSA Publishing, S. L.*

**Keywords:** Ground Source Heat Pump, Refrigerant, Support Vector Machines, Proximal support vector machine, Multi-component mixed refrigerant.

## 1. Introduction

Ground source heat pump [2] is a highly efficient air-conditioning system which takes advantage of shallow-underground geothermal resource (also known as Ground Energy, including groundwater, soil water or surface water) to provide both heat and cooling energy. It transfers heat energy from low temperature to high temperature by inputting a little amount of high-grade energy (e.g. electricity). In the ground source heat pump system, geothermal energy is used as the heating source of the heat pump in winter and the cooling source of air conditioning in summer. Ground source heat pump emit no exhaust gas, water and waste residues and thus is considered as an ideal green technology with renewable energy as well as a sustainable development technology. Originally, it was first proposed and named as "Ground Source Heat Pump" by the Swiss Zoelly in 1912. As the name suggests, ground source heat pump is one type of the heat pump, similar to air source heat pump. "Ground source" refers to the low-level heat of the heat pump from the ground. The ground source heat pump can be classified into ground-coupled heat pump (GCHP) and water-source Heat Pump (WSHP) based on different ways of using the low-level heat.

Ground-source heat pump system is a mechanical vapor compression/refrigeration cycle system, through which heat can be discharged into or absorbed from the surface layer. Therefore, the refrigerant performance is the main factors influencing the performance of GSHP system, and its influence degree is relevant to the heat pump unit, operating conditions and the type of refrigerant. Taking constant temperature of soil as cold and heat source, the operation of ground source heat pump is relatively stable; the best performance of refrigerant is decided by the refrigerating mode working condition. At present, the refrigerant used in Ground source heat pump systems is classified according to the refrigerant component classification [3]: single refrigerant and mixed refrigerant [4]. The refrigerant mixture is generally divided into two categories, namely the non-azeotropic mixture refrigerant and the azeotropic mixture refrigerant. The properties of non-azeotropic mixed refrigerant are close to the ideal solution. The formation condition of ideal solution is that the two kinds of component molecules have similar structure. Azeotropic mixture is a non-ideal solution with large deviation. Common Azeotropic mixture refrigerant contains many kinds, $R_{500}$, $R_{501}$, $R_{502}$, $R_{503}$, $R_{504}$, $R_{505}$, $R_{506}$, $R_{507}$, etc. non-azeotropic mixed refrigerant contains $R_{12}$, $R_{22}$, $R_{11}$, $R_{13}$, $R_{14}$, $R_{21}$, $R_{30}$, $R_{40}$, etc. At present, the majority of mixed refrigerants contain two components. The development of multi-component refrigerant mixture with traditional method is very complicated and has certain difficulties due to the refrigerating capacity, compression ratio, and energy consumption and refrigeration coefficient factors related to refrigerant performance. Development of multi-component mixed refrigerant must guarantee the refrigeration performance first, and then consider the economic costs, reducing pollution and corrosion of equipment. To solve these problems needs to combine calculated method and experimental method. In fact, this is an optimization problem. Here we try to use a new method of data mining - optimized by experimental research and support vector machine, in order to get the ideal results.

## 2. The Feasibility Analysis of Support Vector Machine and Its Application in Making a Mixed Refrigerant in the Ground Source Heat Pump System

### 2.1. Support Vector Machine

Support Vector Machine (SVM), which is a new data mining method, was brought up by Cortes and Vapnik in 1995 and has already been one significant achievement of Machine Learning Research in recent years [5]. The theoretical foundation of SVM is Statistical Theory [6] and Optimization Theory [7]. It has been successfully applied to the military, economy and other fields in developed countries such as the United States to solve problems including pattern classification, regression analysis, estimation function, etc. Common principles are concluded from existing observation samples and utilized to predict future data or data that cannot be observed or collected. In other words, Support Vector Machine is a process where, given a training set, an optimization model is built to obtain optimized solutions for creating a decision function which can be applied to practical problems to make optimal decisions.

### 2.2. The Feasibility Analysis of Support Vector Machine and Its Application in Making a Mixed Refrigerant in the Ground Source Heat Pump System

Ground source heat pump system employs a great variety of refrigerants, including $R_{500}$, $R_{501}$, $R_{502}$, $R_{503}$, $R_{504}$, $R_{505}$, $R_{506}$, $R_{507}$, $R_{12}$, $R_{22}$, $R_{11}$, $R_{13}$, $R_{14}$, $R_{21}$, $R_{30}$, $R_{40}$, etc. Each refrigerant has its own characteristics, such as heat transfer performance, corrosion, price, toxicity on human body, leakage and potential risks. Under some actual condition, a single refrigerant or two mixed refrigerants are used. Users mix different proportion of refrigerant solutions. No related research studied how to determine the concentration proportion of refrigerant solutions to achieve the best effect before since a variety of factors are required to be involved into the research and difficulty exists in quantitative analysis with traditional techniques and approaches.

Here, a new data mining method--support vector machine is created to solve this problem. N kinds of refrigerants are denoted as $a_1$, $a_2$ ... $a_n$. $a_i$ refers to the i-th type of refrigerant and also represents the amount of the i-th refrigerant with certain concentration. Thus, a n-dimensional vector ($a_1$, $a_2$, ..., $a_n$) is constructed. For each component, when some certain concentration and amount is given, there will be one n-dimensional vector. A standard denoted by A for refrigerant performance is set according to design requirements of a ground source heat pump system and geographic conditions (such as surface temperature, ground corrosion resistance, etc.) If a result is higher than or equal to A, the concentration proportion is considered to be eligible; if lower than A, it is not eligible. In the experiment of L number of n-dimensional vectors, the number of results more than and equal to A is recorded as $l_1$ and each result is labeled as +1, while results less than A are counted as $l_2$ and each result is labeled as -1. In this way, we get L training points including $l_1$ positive points and $l_2$ negative points, which constitute a training set $T=\{(x_1,y_1), ..., (x_l,y_l)\}$, where $x_i=(a_{i1},a_{i2},...,a_{in})$, $y_i=\pm1$, $i=1,2,...,l$. To find the optimal proportion, a decision function f(x) is derived from support vector machines and the optimal solution can be achieved from this function f(x). That is, SVM-based decision function f(x) is the key to figuring out the optimal proportion of refrigerant mixture with better performance, low cost and weak corrosion.

## 3. Weighted Proximal Support Vector Machine

For practical considerations, we always meet the cases of unbalance between positive and negative point numbers of training set. To solve the problem hereby, we follow the train of thought on constituting weighted Support Vector Machine to constitute weighted Support Vector Machine Model. By introducing parameter $C_+$ to positive point, and introducing parameter $C_-$ to negative point, the original optimization problem therefore is:

$$\min_{w,\eta,b} \quad \frac{1}{2}(\|w\|^2+b^2)+\frac{C_+}{2}\sum_{y_i=1}\eta_i^2+\frac{C_-}{2}\sum_{y_i=-1}\eta_i^2, \tag{1}$$

$$s.t. \quad y_i((w\cdot x_i)+b)=1-\eta_i, i=1,2,\cdots,l, \tag{2}$$

Or more commonly, if the importance of each point is known before, it is workable to introduce different punishing parameter $C_i$ to each training point, then the original optimization problem therefore is

$$\min_{w,\eta,b} \quad \frac{1}{2}(\|w\|^2+b^2)+\frac{1}{2}\sum_{i=1} C_i\eta_i^2, \tag{3}$$

$$s.t. \quad y_i((w\cdot x_i)+b)=1-\eta_i, i=1,2,\cdots,l, \tag{4}$$

Theorem 3.1 Topic (3) - (4)'s dual problem is

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j((x_i\cdot x_j)+1)+\frac{1}{2}\sum_{i=1}^{l}\frac{\alpha_i^2}{C_i}-\sum_{i=1}^{l}\alpha_i \tag{5}$$

Topic (3) - (4)'s Lagrange function is

$$L(w,b,\eta,\alpha)=\frac{1}{2}(\|w\|^2+b^2)+\frac{1}{2}\sum_{i=1}^{l}C_i\eta_i^2-\sum_{i=1}^{l}\alpha_i(y_i((w\cdot x_i)+b)+\eta_i-1), \tag{6}$$

where $\alpha\in R^l$ is Lagrange to multiply subvector, draw Lagrange function on the minimum of $w,b,\eta$, then the following conditions are coming:

$$w=\sum_{i=1}^{l}\alpha_i y_i x_i, \tag{7}$$

$$b=\sum_{i=1}^{l}y_i\alpha_i, \tag{8}$$

$$\eta_i=\frac{\alpha_i}{C_i}, i=1,\cdots,l., \tag{9}$$

$$y_i((w\cdot x_i)+b)+\eta_i-1=0, i=1,\cdots,l., \tag{10}$$

Substitute (7)-(9) in Lagrange function and draw the maximum of α, unconstrained dual problem comes out hereby (5),

Topic (5) is strict convex quadratic programming, whose optimum relation is

$$\alpha=(I_C+Y(XX^T+ee^T)Y)^{-1}e=(I_C+HH^T)^{-1}e, \tag{11}$$

where $X\in R^{l\times n}$ is matrix constituted by inputting, $x_i\in R^n, i=1,\cdots,l$ $Y\in R^{l\times l}$ is diagonal element which is $y_i,\cdots,y_l$ respectively, other elements are matrixes of zero, $e=(1,\cdots,1)^l$, diagonal

$$I_C=diag\{\frac{1}{C_1},\cdots,\frac{1}{C_l}\}, \tag{12}$$

and

$$H=Y[X,e]. \tag{13}$$

By introducing kernel function $K(x_i,x_j)$ to replace the inner product $(x_i\cdot x_j)$ of topic (5), Weighted Centroid Support Vector Machine will come. Specific algorithm is as follows:

1) Set known training set

$$T=\{(x_1,y_1),\cdots,(x_l,y_l)\}\in(\mathrm{x}\times\mathrm{y})^l,$$

where $x_i\in \mathrm{x}=R^n, y_i\in \mathrm{y}=\{-1,1\}, i=1,\cdots,l$;

2) Choose appropriate parameter $C_i, i=1,\cdots,l$; choose appropriate kernel function $K(x,x')$;

3) Constitute and solve optimization problem optimum relation $\alpha^*$ comes out;

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j(K(x_i,x_j)+1)+\frac{1}{2}\sum_{i=1}^{l}\frac{\alpha_i^2}{C_i}-\sum_{i=1}^{l}\alpha_i, \tag{14}$$

4) b is confirmed by formula (8), constituting classification decision function:

$$f(x)=\mathrm{sgn}(\sum_{i=1}^{l}\alpha_i y_i K(x,x_i)+b).$$

## 4. The Application of Weighted PSVM in Making a Mixed Refrigerant in the Ground Source Heat Pump System

In the running ground source heat pump system, eight refrigerants are mixed, that is to take n=8; 1000 trials of the mixed refrigerants with different proportions are tested, that is to take l=1000. Each trial is labeled as qualified or unqualified by ±1 so that the training set T={($x_1$,$y_1$),...,($x_{1000}$,$y_{1000}$)} is collected, where $x_i$= ($a_{i1}$,$a_{i2}$,...,$a_{i8}$), $y_i$=±1, i=1, 2,...,1000. Thus the actual training decision-making is done; the decision-making information systems based on weighted PSVM is derived.

### 4.1. Data Preprocessing

By observing these data, it can be seen that the value of some indicator data is small, but some indicator data is very large, therefore, the data shall be standardized, and the standardization method here is minimum-maximum standardization method and the formula of it is:

$$[x_j]_i'=([x_j]_i-\min_{j=1,\cdots,100}([x_j]_i))/(\max_{j=1,\cdots,100}([x_j]_i-\min_{j=1,\cdots,100}([x_j]_i)) \tag{15}$$

The dataset can be standardized into D' by this method.

Then the dataset D' shall be divided into 2 parts according to the proportion of 7:3 randomly, in which one part shall be Training Set T and the amount of training point of it shall be recorded as l (l=700 here); the other part shall be Test Set S and the amount of training point of it shall be recorded as m (m=300 here). Let the amount of positive points of training set be $T_+$, while the negative points $T_-$, and the amount of positive points of Test Set is $S_+$, while negative points is $S_-$. By observing the data, it can be seen that the amount of negative points, namely the points that do not satisfy the profit requirement, is 260, while the amount of positive points, namely the points that satisfy the profit requirement, is 440, which reflects the disproportion of these 2 types of points. Therefore, we shall provide different penalty parameters of $C_+$ and $C_-$ to two types of points, but $C_+$ and $C_-$ shall be confirmed according to the following formula:

$$C_+ = C \times \frac{T_-}{l}, \quad C_- = C \times \frac{T_+}{l}, \tag{16}$$

In this, C>0 shall be the given parameter beforehand.

## 4.2. Model Option

Aiming at the classification problems above, first, we shall chose the proper arithmetic model, and select 3 kinds of Support Vector Machine models respectively, and the first one is the Weighted Proximal Support Vector Machine presented in front, and the problem need to be solved is

$$\min_{\omega,\eta,b} \quad \frac{1}{2}(\|\omega\|^2 + b^2) + \frac{C_+}{2}\sum_{y_i=1}\eta_i^2 + \frac{C_-}{2}\sum_{y_i=-1}\eta_i^2, \tag{17}$$

$$s.t. \quad y_i((\omega \cdot x_i) + b) = 1 - \eta_i, i = 1, 2, \cdots, l, \tag{18}$$

And its dual problem is

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j (K(x_i, x_j) + 1) + \frac{1}{2C_+}\sum_{y_i=1}\alpha_i^2 + \frac{1}{2C_-}\sum_{y_i=-1}\alpha_i^2 - \sum_{i=1}^{l}\alpha_i \tag{19}$$

The second one is the weighted reasoning Support Vector Machine model, the third one is the weighted standard Support Vector Machine model [8], the corresponding parameters shall be selected after the 3 models are confirmed above, which includes the kernel functions $k(x, x')$ and C, C*, and the parameters in kernel function. We chose the kernel function as the radial base kernel function.

$$K(x, x') = \exp(-\frac{\|x - x'\|^2}{\sigma^2}), \tag{20}$$

The parameters need to be selected shall be C, C* and $\sigma$.

Chose the best optimum parameters for each model according to method of lattice, which means the numeric area of C and C* {0.1,1,10,100,1000,10000} and the numeric area of $\sigma$ {0.1,0.2,0.5,1,2,5} to constitute the parameter group $(C, C^*, \sigma)$, and for the value of each group of parameters, calculate LOO error [9], and then chose the correspondingly LOO error parameter group $(\overline{C}, \overline{C}^*, \overline{\sigma})$ as the optimum parameter, and the result is that the optimum parameter group $(\overline{C} = 10, \overline{\sigma} = 1.5)$ corresponds with the weighted Proximal Support Vector Machine model, and the weighted reasoning Support Vector Machine corresponds with the parameter group $(\overline{C} = 100, \overline{C}^* = 100, \overline{\sigma} = 2)$, and the weighted standard Support Vector Machine corresponds with the optimum parameter $(\overline{C} = 10, \overline{\sigma} = 5)$.

## 4.3. Result

Substitute the 3 groups of optimum parameters gotten above into the corresponding models, and get the final decision function and judge the points of Detected Set S, the result is shown in the Table 1:

**Table 1.** Result Comparison.

| Detected Set | C-SVC | TSVC | PSVC |
|---|---|---|---|
| Detected Precision | 83 % | 87 % | 89 % |
| Rate of Misdeclaration | 3 % | 0 % | 1 % |
| Rate of Detection | 67 % | 68 % | 79 % |

C-SVC in this diagram is the weighted standard Support Vector Machine, TSVC is the weighted reasoning Support Vector Machine and PSVC is the weighted Proximal Support Vector Machine.

In these, the detected precision is the ratio of the amount of the correct detected samples of Test Set and the total amount of samples of Test Set; the rate of misdeclaration is the ratio of the amount of normal samples which are mistaken as unusual samples and the amount of the total normal samples; the rate of detection is the ratio of the amount of detected unusual data samples and the total amount of abnormal samples.

From this result we can see that among the three support vector machine classifiers, the most accurate way to solve the problem of decision-making for the refrigerant mixing ratio is the weighted PSVM. Therefore, in ground source heat pump system, given a refrigerant mixing ratio program, to test the program eligibility, we can simply enter data into the decision-making system. If the output is 1, it is executable; if the output is -1, the refrigerant concentration or dosage needs to be adjusted (the low price, small corrosion refrigerant should be adjusted first), the best ratio program will be obtained until the decision-making system outputs 1.

## 5. Conclusions

In summary, it is feasible to apply support vector machines to the decision-making for the refrigerant mixing ratio. A variety of support vector machines will improve the decision-making process on different levels, while specific models should be designed based on actual issues for high precision. As shown above, support vector machines can provide optimal decision-making schemes to classification-related problems. The results of recent years of research and experimentation present that the optimized decision-making system based on a support vector machine model brings better decision-making effects in management and economic benefits to enterprises.

## Acknowledgements

This work was supported in part by a grant from Natural Science Fund Project of Hebei Province, China (NO. A2012105021), Tangshan Science & Technology Research and Development Project (No. 11140202A and No. 10130202A-4).

## References

[1]. Fung G, Mangasariam O.L, Proximal Support Vector Machine classifiers, in *Proceedings of the KDD*, San Francisco, CA, USA, 2001.

[2]. S. Yunxia Qu, Linhua Zhang, Zhaohong Fang, Anjia Li, The Choice of Ground coupled Heat Pump System Antifreeze, *Energy-Saving and Environmental Protection*, Issue 8, 2002, pp. 20-23.

[3]. A. Desheng Xu, Houde Han, Refrigeration and Air Conditioner, *Shanghai Jiao Tong University Press*, May 1997.

[4]. S. Xinglong Guo, Xinnan Song, Optimization on oil vapor condensing recovery system by mixed refrigerant cycle, *Cryogenics & Superconductivity*, Issue 3, 2012, pp. 45-49.

[5]. S. Cortes C, V. Vapnik, Support Vector Networkers, *Machine Learning*, Issue 20, 1995, pp. 273-295.

[6]. S. Vapnik V, Statistical learning theory, *Wiley*, New York, 1998.

[7]. S. Yuquan Hu, Cause of Operation Research, Beijing, *Tsinghua University Press*, 2003.

[8]. Manfu Yan Support Vector Machines for Classification and Its Application, *Beijing China Agricultural University*, 2005 (Doctor's Degree Paper).

[9]. S. Naiyang Deng, Yingjie Tian: New Method in Data Mining -- Support Vector Machine, *Science Press*, Beijing, 2004.

[10]. A. Deng Naiyang, Tian Yingjie, Support Vector Machine -- Theory, Algorithm and Expansion, *Beijing, Science Press*, 2009, pp. 63-64.