Recommender Systems Using Support Vector Machines

Sung-Hwan Min and Ingoo Han

Graduate School of Management, Korea Advanced Institute of Science and Technology,
207-43 Cheongrangri-dong, Dongdaemun-gu, Seoul 130-722, Korea
shmin@kgsm.kaist.ac.kr

Abstract. Due to the explosion of e-commerce, recommender systems are rapidly becoming a core tool to accelerate cross-selling and strengthen customer loyalty. There are two prevalent approaches for building recommender systems – content-based recommending and collaborative filtering (CF). This study focuses on improving the performance of recommender systems by using data mining techniques. This paper proposes an SVM based recommender system. Furthermore this paper presents the methods for improving the performance of the SVM based recommender system in two aspects: feature subset selection and parameter optimization. GA is used to optimize both the feature subset and parameters of SVM simultaneously for the recommendation problem. The results of the evaluation experiment show the proposed model's improvement in making recommendations.

1 Introduction

Recommender systems are the information filtering process to supply personalized information by predicting user's preferences to specific items. In a world where the number of choices can be overwhelming, recommender systems help users find and evaluate items of interest [11]. Due to the explosion of e-commerce, recommender systems are rapidly becoming a core tool to accelerate cross-selling and strengthen customer loyalty. To date, a variety of techniques for building recommender systems have been developed. These techniques can be classified into two main categories: content-based filtering and collaborative filtering (CF). CF is the most successful recommendation technique, which has been used in a number of different applications such as recommending movies, articles, products, Web pages [1, 10]. CF is built on the assumption that a good way to predict the preference of the active consumer for a target product is to find other consumers who have similar preferences, and then use those similar consumer's preferences for that product to make a prediction [6].

Support Vector Machines have attracted most interest in the last few years which is developed by [16]. SVM implements the principle of Structural Risk Minimization by constructing an optimal separating hyperplane in the hidden feature space, using quadratic programming to find a unique solution. Compared with most other learning techniques, SVM has shown remarkable results in pattern recognition [13], text categorization [9], speaker identification [12] and financial time series prediction [2, 4, 8]. But only a small number of studies used SVM in recommendation problem. Furthermore there has been few research which uses optimized SVM model in recommendation problems. In general, the performance of SVM model is sensitive not only to some parameters but also to feature subset. Therefore it is very critical to optimize SVM model for more accurate performance.

This study proposes an SVM based recommender system. Furthermore this study presents the methods for improving the performance of the SVM based recommender system in two aspects: feature subset selection and parameter optimization. GA is used to optimize both the feature subset and parameters of the SVM simultaneously for more accurate recommendations.

2 Background

2.1 Genetic Algorithm

Genetic Algorithm (GA) is an artificial intelligence procedure based on the theory of natural selection and evolution. GA uses the idea of survival of the fittest by progressively accepting better solutions to the problems. It is inspired by and named after biological processes of inheritance, mutation, natural selection, and the genetic crossover that occurs when parents mate to produce offspring [5]. GA differs from conventional non-linear optimization techniques in that it searches by maintaining a population (or data base) of solutions from which better solutions are created rather than making incremental changes to a single solution to the problem. GA simultaneously possesses a large amount of candidate solutions to a problem, called population. The key feature of a GA is the manipulation of a population whose individuals are characterized by possessing a chromosome.

Two important issues in GA are the genetic coding used to define the problem and the evaluation function, called the fitness function. Each individual solution in GA is represented by a string called the chromosome. Initial solution population could be generated randomly, which evolve to the next generation by genetic operators such as selection, crossover and mutation. The solutions coded by strings are evaluated by the fitness function. Selection operator allows strings with higher fitness to appear with higher probability in the next generation [7]. Crossover is performed between two selected individuals, called parents, by exchanging parts of their strings, starting from a randomly chosen crossover point. This operator tends to enable to the evolutionary process to move toward promising regions of the search space. Mutation is used to search further space of problem and to avoid local convergence of the GA [15].

2.2 Support Vector Machine(SVM)

Support Vector Machine (SVM) is a novel learning machine introduced by [16]. The SVM has emerged in recent years as powerful techniques both for regression and classification. The SVM learns a separating hyperplane to maximize the margin and to produce good generalization ability. Recently the SVM has been successfully applied in many areas and has shown remarkable results.

The SVM is based on the Structural Risk Minimization principle for which error-bound analysis has been theoretically motivated. The SVM performs pattern recognition for two-class problems by determining the separating hyperplane with maximum distance to the closest points of the training set. These points are called support vectors. In its simplest linear form, an SVM is a hyperplane that separates a set of positive examples from a set of negative examples with maximum margin. Suppose N observations has a pair $(\mathbf{x}_1, y_1)$, $(\mathbf{x}_2, y_2)$, $(\mathbf{x}_3, y_3)$, …, $(\mathbf{x}_n, y_n)$: a vector

$x_i \in R_n, y_i \in \{-1, 1\}$. The task of the SVM is to learn mapping the pair, $x_i \rightarrow y_i \in \{-1, 1\}$. The formula for the output of a linear SVM is $u = \vec{w} \cdot \vec{x} - b$ where $\vec{w}$ is the normal vector to the hyperplane, and $\vec{x}$ is the input vector. In the linear case, the margin is defined by the distance of the hyperplane to the nearest of the positive and negative examples. Maximizing the margin can be expressed as an optimization problem: minimize $\frac{1}{2} \| \vec{w} \|^2$ subject to $y_i(\vec{w} \cdot \vec{x} - b) \geq 1, \forall i$ where x_i is the ith training example and y_i is the correct output of the SVM for the ith training example.

The algorithms for solving linearly separable cases can be extended so that they can solve linearly non-separable cases as well by either introducing soft margin hyperplanes, or by mapping the original data vectors to a higher dimensional space where the new features contain interaction terms of the original features, and the data points in the new space become linearly separable [3]. In the non linear problem, we first map the data to some other Euclidean space H, using mapping, $\Phi : \mathbf{R}^d \mapsto \mathrm{H}$. Then instead of the form of dot products, "kernel function" K is used such that $K(\mathbf{x}_i, \mathbf{y}_i) = \Phi(\mathbf{x}_i) \bullet \Phi(\mathbf{x}_j)$.

3   Hybrid GA-SVM Model

This study presents the methods for improving the performance of an SVM based recommender system in two aspects: feature subset selection and parameter optimization. GA is used to optimize both the feature subset and parameters of SVM simultaneously for recommendation model.

3.1   Optimizing Feature Subset

Feature subset selection is essentially an optimization problem, which involves searching the space of possible features to find one that is optimum or near-optimal with respect to certain performance measures such as accuracy. In a classification problem, the selection of features is important for many reasons: good generalization performance, running time requirements and constraints imposed by the problem itself.

In the literature there are known two general approaches to solve the feature selection problem: The filter approach and the wrapper approach [14]. The distinction made depending on whether feature subset selection is done independently of the learning algorithm used to construct the classifier (i.e., filter) or not (i.e., wrapper). In the filter approach, feature selection is performed before applying the classifier to the selected feature subset. The filter approach is computationally more efficient than a wrapper approach. Wrapper approach train the classifier system with a given feature subset as an input and estimate the classification error using a validation set. Although this is a slower procedure, the features selected are usually more optimal for the classifier employed.

In a classification problem, feature subset selection plays an important role in the performance of prediction. Furthermore its importance increases when the number of features is large. This study seeks to improve SVM based recommender system. We propose the GA as the method of feature subset selection in the SVM system. This study uses the wrapper approach to select optimal feature subset of the SVM model using GA.

3.2   Optimizing the Parameters of SVM

One of the big problems in SVM is the selection of the value of parameters that will allow good performance. Selecting appropriate values for parameters of SVM plays an important role in the performance of SVM. But, it is not known beforehand which values are the best for one problem. Optimizing the parameters of SVM is crucial for the best prediction performance.

This study proposes GA as the method of optimizing parameters of SVM. In this study, the radial basis function (RBF) is used as the kernel function for SVM based recommender system. There are two parameters while using RBF kernels: C and δ^2. These two parameters play an important role in the performance of SVMs [2]. In this study, C and δ^2 are encoded as binary strings, and optimized by GA.

3.3   Simultaneous Optimization of SVM Using GA

In general, the choice of the feature subset has an influence on the appropriate kernel parameters and vice versa. Therefore feature subset and parameters of SVM need to be optimized simultaneously for the best prediction performance. The proposed model optimizes both feature subset and parameters of SVM simultaneously. The overall procedure of the proposed model starts with the randomly selected chromosomes which represent feature subset and parameters of SVM. Each new chromosome is evaluated by sending it to the SVM model. The SVM model uses the feature subset and parameters in order to obtain the performance measure (e.g. hit ratio). This performance measure is used as the fitness function and is evolved by GA.

The chromosomes for the feature subset are encoded as binary strings standing for some subset of the original feature set list. Each bit of the chromosome represents whether the corresponding feature is selected or not. 1 in each bit means the corresponding feature is selected, whereas 0 means it is not selected. The chromosomes for parameters of SVM are encoded as a 16-bit string which consists of 8 bits standing for C and 8 bits standing for δ^2. Fig. 1 shows examples of encoding for GA.

Each of the selected feature subsets and parameters is evaluated using SVM. This process is iterated until the best feature subset and values of parameters are found.

The data set is divided into a training set and a validation portion. The training set (T) consists of both T_1 and T_2.

GA evolves a number of populations. Each population consists of sets of features of a given size and the values of parameters. The fitness of an individual of the population is based on the performance of SVM. SVM is trained on T_1 using only the features of the individual and the values of parameters of the individual. The fitness is

the prediction accuracy of the SVM model over T_2. At each generation new individuals are created and inserted into the population by selecting fit parents which are mutated and recombined.

The fitness function is represented mathematically as follows:

$$Fitness = \frac{\sum_{i=1}^{n} H_i}{n} \tag{1}$$

where H_i is 1 if actual output equal to the predicted value of the SVM model, otherwise H_i is zero.

(a) Feature Subset

| 1 | 0 | 1 | 1 | 0 | ... | | 0 | 0 | 0 | 0 | 1 |

| f_1 | f_2 | f_3 | f_4 | f_5 | ... | f_n |

Feature subset $= \{f_1, f_3, f_4, f_n\}$

(b) Parameters $-\ \delta^2$ and C

| 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |

$C = 2^0*0 + 2^1*1 + 2^2*1 = 6$

| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |

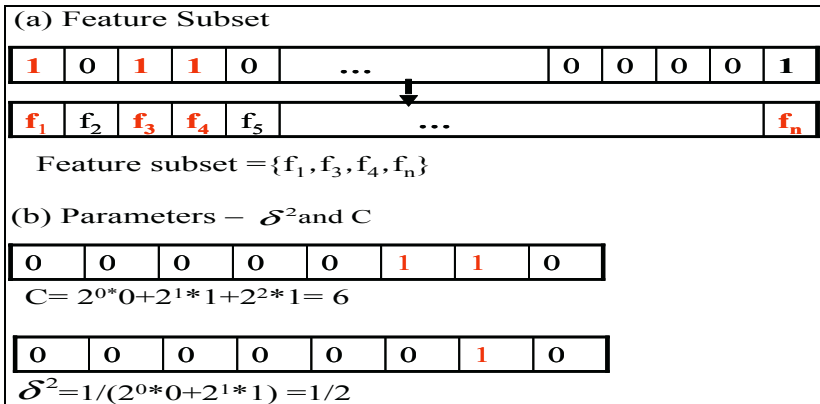$\delta^2 = 1/(2^0*0 + 2^1*1) = 1/2$

Fig. 1. Examples of Encoding for GA

4   Experimental Evaluation

We conducted experiments to evaluate the proposed model. For experiments we used the EachMovie database, provided by Comaq Systems Research Center (http://www.research.compaq.com/SRC/eachmovie). The dataset contains explicit rating data provided by each user for various movies. The EachMovie dataset has rating information on 1,628 movies by 72,916 users during an 18 month period from 1996. The ratings in the EachMovie are discretized into six levels, as 0, 0.2, 0.4, 0.6, 0.8 and 1. In the following, we defined a move "interesting" to an individual customer if his/her preference rating for this movie is greater than 0.5.

We used hit ratio as our choice of evaluation metric to report prediction experiments. First we selected 1000 users with more than 100 rated items. We divided the data set into a training set and a test portion.

We use the term, "GA-SVM" model as the proposed model which is simultaneous optimization of SVM using GA. To compare the performance of the proposed GA-SVM model we used the traditional CF algorithm as the benchmark model. The traditional CF recommendation employs the Pearson nearest neighbor algorithm.

Table 1 shows the classification accuracies of various parameters in SVM. The experimental results show that the prediction performance of SVM is sensitive to various parameters. Table 2 describes the average prediction accuracy of each model. In

SVM, we used the best result on the validation set out of results of Table 1. In Table 2, the proposed GA-SVM model shows better performance than SVM and TCF model.

The McNemar tests are used to examine whether the proposed model significantly outperforms the other models. This test is a nonparametric test for two related samples using the chi-square distribution. The McNemar test assesses the significance of the difference between two dependent samples when the variable of interest is a dichotomy. It is useful for detecting changes in responses due to experimental intervention in "before-and-after" designs.

Table 3 shows the results of McNemar test. As shown in Table 3, GA-SVM outperforms SVM and TCF with the 5% statistical significant level.

Table 1. Classification accuracy of SVM

C	δ^2						
	1	10	30	50	80	100	200
1	65.02	64.51	63.51	60.47	60.98	59.97	58.45
10	64.51	63	65.03	65.52	64.51	64.01	64.51
30	58.96	63	62	65.02	64.01	64.51	64.01
50	57.95	62.49	62.5	61.99	64.51	64.51	64.51
70	58.96	63.5	64.01	61.99	64.01	64.01	63.5
90	58.45	64.51	63	61.99	62.49	64.01	63
100	58.96	66.03	63	62.49	62.49	64.01	63
150	60.47	64.51	62.49	63.5	62.49	62.49	63
200	58.45	65.02	63	63.5	61.99	63	63.5
250	57.44	64.01	63	63	61.99	62.49	63

Table 2. Average prediction accuracy

Model		Hit Raio(%)
Proposed Model	GASVM	67.85
	SVM	66.03
Traditional Model	TCF	65.58

Table 3. p values of McNemar test

Model		p-value	
		SVM	TCF
Proposed Model	GASVM	0.032[a]	0.017**
	SVM	.	0.363
Traditional Model (TCF)			

[a] Significant at the 5% level

5　Conclusion

Due to the explosion of e-commerce, recommender systems are rapidly becoming a core tool to accelerate cross-selling and strengthen customer loyalty. This study focused on improving the performance of recommender system by using data mining techniques. This paper proposed an SVM based recommender system. Furthermore this study presented the methods for improving the performance of an SVM based

recommender system in two aspects: feature subset selection and parameter optimization. GA was used to optimize both the feature subset and parameters of the SVM simultaneously.

We conducted an experiment to evaluate the proposed model on the EachMovie data set and compared them with the traditional CF algorithm. The results show the proposed model's improvement in making recommendations.

In our future work, we intend to optimize kernel function, parameters and feature subset simultaneously. We would also like to expand this model to apply to the instance selection problems.

References

1. Breese, J.S., Heckerman, D., Kadie, C. (1998). Empirical Analysis of Predictive Algorighms for Collaborative Filtering. Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI-98), pp. 43-52.
2. Cao, L., Tay, F.E.H., (2001), Financial Forecasting Using Support Vector Machines, Neural Computing & Applications 10. pp. 184-192.
3. Cortes, C., and Vapnik, V., Support vector networks. Machine Learning, 20, 273-297, 1995.
4. Fan, A., Palaniswami, M., (2000). Selecting Bankruptcy Predictors Using A support Vector Machine Approach, Proceeding of the International Joint Conf. on Neural Network 6. 354-359.
5. Goldberg, D. E., Genetic Algorithms in Search, Optimization and Machine Learning (Addison-Wesley, New York 1989)
6. Herlocker, J.L., Konstan, J.A. and Riedl, J., (2000). Explaining collaborative filtering recommendations. Proceedings on the ACM 2000 Conference on Computer Supported Cooperative Work, (pp. 241–250). Philadelphia.
7. Holland, J. H., Adaptation in natural and artificial systems (The University of Michigan Press, Ann Arbor, 1975)
8. Huang, Zan., Che, Hsinchun., Hsu, Chia-Jung., Chen, Wun-Hwa., Soushan Wu, (2004). Credit rating analysis with support vector machines and neural networks: a Market comparative study, Decision Support Systems 37. 543-558.
9. Joachims, T. (1997). Text Categorization with Support Vector Machines, Technical report, LS VIII Number 23, University of Dormund
10. Sarwar,B.M., Konstan,J.A., Borchers,A., Herlocker,J.L., Miller,B.N., Ried1,J. (1998). Using filtering agents to improve prediction quality in the grouplens research collaborative filtering system. Proceedings of CSCW'98. Seattle, WA.
11. Schafer, J.B., Konstan, J.A. and Riedl, J. (2001). Electronic Commerce Recommender Applications. Data Mining and Knowledge Discovery 5(1/2), pp. 115-153.
12. Schmidt, M. S. (1996). Identifying Speaker with Support Vector Networks, In Interface '96 Proceedings, Sydney.
13. Sclkopf, B., Burges, C., Vapnik, V., (1995). Extracting support data for a given task. In U.M Fayyad and R. Uthurusamy, editors, Proceedings, First International Conference on Knowledge Discovery & Data Mining. AAAI Press, Menlo Park, CA.
14. Sun, Z., Bebis, G., Miller, R., Object Detection using Feature Subset Selection, Pattern Recognition 27 (2004) 2165-2176
15. Tang, K. S., Man, K. F., Kwong, S., He, Q., Genetic Algorithms and Their Applications, IEEE Signal Processing Magazine 13 (1996) 22-37.
16. Vapnik, VN., (1995). The Nature of Statistical Learning Theory (New York, Springer-Verlag.
17. http://www.research.compaq.com/SRC/eachmovie