Journal of Telematics and Informatics (JTI)
Vol.4, No.2, September 2016, pp. 39~44
ISSN: 2303-3703

Data Mining Sales Optimizations Using Sequential Minimal Optimization Algorithm

Dedy Kurniadi, Sam Farisa Chaerul Haviana
Department of Informatics Engineering, Faculty of Industrial Technology
Sultan Agung Islamic University
e-mail: ddy.kurniadi@unissula.ac.id

Abstract

Tightness of business nowadays requires businessman to be able to develop their business to compete with the other companies, this study was conducted to obtain data accurate on the type of clothing combinations that are favored by the consumers to optimize sales at convection companies, using data mining methods and technique of classification this data is classify into four classes namely, well-liked, liked, enough and dislike. To solve classification problems, this study used Sequential Minimal Optimization (SMO), SMO Algorithm can solve quadratic programming problems without requiring a large matrix and to solving the optimization SMO selected from the smallest optimization in every steps. Optimum accuracy obtained in this study were obtained from Correctly Classified Instance of 80.9% from 3072 record set of well-liked classes that is the class with type of combinations clothes polo and embroidery, then the level of measurement of consistency coefficient values using kappa statistic obtained for 0.73% where the data in the class showed a consistent value, from these data type are most well-liked combinations can optimize sales by 70.3%

Keywords*: data mining, SMO, Correctly Classified Instance, Kappa Statistic*

1. Introduction

Industry nowadays has a growing business very rapidly along with the era of globalization that demands the intense competition to make the industrial sector should be able to apply information technology, the rapid development of information technology is increasingly affecting the level of competition among companies to one another, the impact of the application of information technology course buildup analysis of data generated by the company, but the data analysis are often only used to the extent of the report only. Competition in business today requires companies leaders to find the right solution in order to further optimize the company's revenue in addition to the solutions that are taken must be based on data and facts on the ground in order to making the right decisions, the amount of data contained in a company requires create a model of the big data that can provide data quality is nice and with high accuracy is so important because it relates to the accuracy of decision-making is vital as they relate to the determination of the planning and strategy of the company in determining plans and strategies (strategic plan) the company's sales, decision making factors of course requires information that is large enough to be analyzed and generate the pattern for decision making [1]–[3].

However, the amount of data generated by the company can not be maximized up to the stage of data mining so that the data information is hidden and that there has not been formed or patterned with a maximum, by utilizing computer science, data was only used as a report can be processed using data mining methods with classification techniques, classification technique is a technique that brings together an object into one of the specific categories in which of these categories have been determined beforehand, classification techniques in data mining using the training set in which there are a set of records and the record consists of several kinds of attributes and the attributes key is the class attribute, extracting information using classification techniques utilizing big data is processed using Sequential Minimal Optimization (SMO) algorithm to identify information that can be used, this algorithms itself is often used to solve problems related to

optimization [4]–[6]. In the process of optimization SMO algorithm uses 2 Lagrange multipliers, its function is to find the most optimal Lagrange which are used to change the Support Vector Machine and obtained a new optimal value. The advantages obtained by using SMO is in the process to solving 2 Lagrange multiplier by means of analytical, SMO itself does not require a large matrix in solving the problem, most of the SVM in training requires considerable memory but not with SMO, which do SMO is to avoid the great matrix in its calculations, the accuracy resulting from SMO algorithm is quite high, SMO is able to solve optimization problems and provide high accuracy so that the results obtained are more accurate [7].

2. Research Method

2.1. Data Mining

Data mining is a process in which there is a series of processes to seek information in the data in the form of additional information that could be used as an added value in the form of information that is usually unknown, information extracted from a database by performing excavation techniques using particular patterns which aims to manipulate the data to produce information valuable [8].

There are many techniques in data mining include classification techniques used in this study, the technique of classification itself is a technique for systematically compiling data from a category or class of instance data is taken based on the attributes contained in the data, in the classification usually used a attributes category, the workings of the classification technique uses two stages, stages that are used first is learning, at the stage of learning classifier built based on several categories or classes that have been determined from the data, at this stage is also called the stage of training step, classifier built using classification algorithms are analyzing a training set [9].

$$X = \{X_1, X_{2,} \dots X_n\} \quad (1)$$

Where:

$X = Tuple$
$X_1 \dots X_n = Attribute\ Vector\ Dimension$

Each tuple is assumed to enter into the category or class that had previously been determined by the attributes, the first stage of mapping is called learning the steps or stages of a pattern of classification or classification for mapping to be predictable.

$$y = f(x) \quad (2)$$

Where:

$y = class\ label\ on\ tuple\ X$

$f(x) = functoin\ mapping$

Second stage of classification using a classifier that has been built in the previous stage and then the processed data to the process of classification is to perform the measurement accuracy of the data from predictive classifier [6]. The classification technique itself included on supervised learning because classes are already available, the analysis of data using knowledge discovery in databases (KDD) as for the techniques used to process the data are as follows:

a. Data Integration
   Data required in the data mining process is very much could be from several databases, data integration is to identify entities which are unique, data integration done carefully because all the processing begins with data integration so that the result is not distorted.

b. Data Selection
   Stage performed on data selection is collecting data to be processed, the data obtained from convection companies, data to be processed is a sales report data from the years 2009 to 2015 these data contain information items that are purchased by the customer from the start item disliked until the items are well-liked, this data will be used to process mining.
c. Pre-processing
   Pre-processing stages objective is to purge the database so that there are no redundancies but it is also for the sake of cleaning missing value.
d. Data Transforms
   Data transformation is advanced stages of pre-processing is data which has been cleaned then merge into a format appropriate to the process used in data mining
e. Pattern Evaluation
   The process of pattern Evaluation is to pull patterns that exist in knowledge based, the results of this phase is the form of patterns or predictive models and then the result evaluated whether the data are indeed achieved and in accordance with the pattern of data mining or not.

2.2. Sequential Minimal Optimization (SMO)

Sequential Minimal Optimization (SMO) is a highly efficient algorithm for solving problems related to optimization, SMO algorithm uses vector at each step of the solution, the optimization that occurs during the use of the support vector machine (SVM) can be solved by SMO without using a large matrix, excellence the SMO itself is way to resolve it using 2 Lagrange, SMO choose 2 Lagrange who then jointly optimized and produce or find the optimal values are used to update the value of SVM as the newest optimal values [7]. SMO solve the problem of quadratic programming related to optimization of the analytical, there are two parameters were optimized and the other parameters are allowed to remain in accordance with the needs of process data, profits from SMO by not using a large matrix from the training data that arise in the process of training using the SVM can be stored in computer memory, so the process can be done more quickly and minimize the computing time of SMO [6]. in the process to update 2 Lagrange multiplier is, first SMO calculate the constraints of the multipliers are to be solved into minimal constraints, SMO optimize 2 Lagrange multipliers in order to meet the linear equality constraint in every step taken which it is not possible to use one of the Lagrange multipliers [10].

$$\alpha_2^{new} = \alpha^2 + \frac{y_2(E_1 - E_2)}{\eta} \tag{3}$$

Where $$\eta = K(x_1^{\rightarrow}, x_1^{\rightarrow}) + K(x_2^{\rightarrow}, x_2^{\rightarrow}) - 2K(x_1^{\rightarrow}, x_2^{\rightarrow}) \tag{4}$$

And then the updated Lagrange is now clipping tho gain the value from the formula number 4, to fulfill every linear equality constraint.

$$\alpha_2^{new,clipped} = \begin{cases} H & if \quad \alpha_2^{new} \geq H; \\ \alpha_2^{new} & if \ L < \alpha_2^{new} < H; \\ L & if \quad \alpha_2^{new} \leq L \end{cases} \tag{5}$$

Where $L = \max(0, \alpha_2 + \alpha_1 - C) \quad H = \min(C, \alpha_2 + \alpha_1)$ when $y1 = y2$
$L = \max(0, \alpha_2 - \alpha_1) \quad H = \min(C, C + \alpha_2 - \alpha_1)$ when $y1 \neq y2$

$$\alpha_1^{new} = \alpha_1 + s(\alpha_2 - \alpha_2^{new,clipped}) \tag{6}$$

Where $s = y_1 y_2$

3. Results and Analysis

The first stage in data mining is preprocessing, this process aims to process raw data into good data quality, these stages are necessary as there are data incomplete, noisy and inconsistent then conducted several stages of data cleaning, Data integration, transformation and data reduction in table 1 shows the attributes contained in the data.

Table 1. Forming Attributes

No	Attributes Name
1	Price
2	Quality
3	Materials
4	Size
5	Kind
6	Type
7	Class

Then after a specified attribute next step is the formation of a class and counted how many distinct classes contained in the label and then used as a weighting of each of these classes, table 2 shows the phase weighting on the label.

Table 2. Weighting Class

No	Label	Count	Weight
1	Disliked	994	994.0
2	Enough	386	386.0
3	Liked	604	604.0
4	Well-Liked	1088	1088.0

After all the data has been through the stages of preprocessing the next step is performed stage normalization algorithm using Sequential Minimal Optimization (SMO) with kernel linear, at this stage data need to be normalization first, the result of the normalization with SMO algorithms uses kernel linear shown in Table 3, the weighting using the formula of linear kernel. Formula to normalize: $K(x, y) = < x, y >$

Table 3. Normalization

Class : Disliked, Enough, Liked, Well-liked	
Normalizers	Weight
Price=expensive	-1.5002
Price =enough	-1.5003
Price =cheap	0.5001
Price =discount	2.5004
Quality=good	0.0004
Quality=enough	2.0005
Quality=deficient	-0.20002
Quality=bad	-2.0007
Material=smooth	0
Material=rough	-0.0001
Material=dry fit	0.0002
Material=bad	-0.0001
Size=XL	-0.10001
Size=L	0.0001
Size=M	0

After normalization, then the next process is the validation and measurement accuracy of calculations produced, the validation process using the confusion matrix, this process is commonly used in the visualization of supervised learning models, table 5 shown confusion matrix.

Table 4. *Confusion Matrix*

a	b	c	d	Explanation
627	16	35	16	a = Disliked
67	276	29	14	b = Enough
75	28	438	63	c = Liked
111	35	97	945	d = Well-Liked

Confusion matrix contains the actual information on the classification system, then from the confusion matrix make a calculation of precision recall and f-measure to determine the level of accuracy and success of the amount of data that has been classified [11] [12], the result of the calculation shown in table 4.

$$P = \frac{TP}{TP+FP} \tag{7}$$

$$R = \frac{TP}{TP+FN} \tag{8}$$

$$F_1 = 2 \times \frac{precision \times recall}{precison+recall} \tag{9}$$

From the calculation precision, recall and f-measurement then obtained results that indicate the level of accuracy shown in table 5, table detailed accuracy class accuracy shown the value of all the classes and the average value.

Table 5. Detail Accuracy Class

Prec	Recall	F-Measure	Class
0.78	0.93	0.85	Disliked
0.77	0.71	0.74	Enough
0.73	0.72	0.72	Liked
0.90	0.77	0.83	Well-Liked
0.815	0.80	0.80	Avg

Cross validation is a method used to evaluate the regression models with predictive capability and to choose which model is the best, in a cross validation method to predict by dividing the data into two parts, the first part is the part that is used to create a regression model and the other is part of validating to know how well the resulting ability to predict the model.

Table 6. Cross Validation

Stratified	data	hasil
Correctly Classified Instance	2486	80.9%
Incorrectly Classified Instance	586	19.0%
Kappa Statistic	3072	0.73%

In Table 6 shows the kappa statistic stands at 0.73 significant value, of these results indicates that the value of the coefficient of the kappa statistic shows the correlation and the results of the processing of the data is consistent. Results of the correctly classified instance showed 80.9% and classified incorrectly instance by 19.0%.

Table 7. Sales Data

Year	Production	Sales
2012	3072	1.540
2013	3072	1.504
2014	3072	1.795
2015	3072	2.160

Table 5 shows that the class is the preferred class preferably has a coefficient value close to 1, the calculation results of tables 4 and 5 and then implemented with the sales data from the years 2009 to 2015 in which the range of the production is already inclined to a model which is preferred, the table 6 is a calculation that is performed with the use of the data that has been implemented and table 7 shows an increase in sales after using SMO algorithm, sales data showed an increase of 70.3%.

4. Conclusion

Sequential Minimal Optimization (SMO) Algorithms generate a high degree of accuracy Correctly Classified Instance 80.9% from the total of 3072 data. The coefficient of kappa statistic is 0.73% the value of kappa statistic indicates a consistent value of the best regression model. Sequential Minimal Optimization (SMO) Algorithms can be used to analyze sales data and is able to optimize sales by 70.3%.

References

[1] M. Jorge, C. Ismael, R. Bibiano, S. Manuel, and P. Mario, "A Data Quality in Use model for Big Data," *Futur. Gener. Comput. Syst.*, vol. 11, pp. 1–7, 2015.

[2] A. Hosseinzadeh, R. Smyth, A. Valadkhani, and V. Le, "Analyzing the Efficiency Performance of Major Australian Mining Companies Using Bootstrap Data Envelopment Analysis," *Econ. Model.*, vol. 57, pp. 26–35, 2016.

[3] J. Ghattas, P. Soffer, and M. Peleg, "Improving Business Process Decision Making Based On Past Experience," *Int. J. Comput. Sci.*, vol. 59, pp. 93–107, 2014.

[4] Y. Chen and C. Su, "Distance-based Margin Support Vector Machine For Classification," *Appl. Math. Comput.*, vol. 283, pp. 141–152, 2016.

[5] H. Jia, "Large-Scale Data Classification Method Based on Machine Learning Model," *Int. J. Database Theory Appl.*, vol. 8, no. 2, pp. 71–80, 2015.

[6] X. Wu and R. Srihari, "New v-Support Vector Machines and their Sequential Minimal Optimization," in *Proceedings of the Twentieth International Conference on Machine Learning*, 2003, no. Washington D.C.

[7] D. Vadicherla and S. Sonawane, "Decision Support System For Heart Disease Based On Sequential Minimal Optimization In Support," *Int. J. Eng. Sci. Emerg. Technology.*, vol. 4, no. 2, pp. 19–26, 2013.

[8] N. Kerdprasop, P. Kongchai, and K. Kerdprasop, "Constraint Mining in Business Intelligence : A Case Study of Customer Churn Prediction," *Int. J. Multimed. Ubiquitous Eng.*, vol. 8, no. 3, pp. 11–20, 2013.

[9] P. Kaur, M. Singh, and G. Singh, "Classification and Prediction Based Data Mining Algorithms To Predict Slow Learners In Education Sector," *3rd Int. Conf. Recent Trends Comput.*, vol. 57, pp. 500–508, 2015.

[10] S. Lakshmishree and J. Kumarnath, "Reconfigurable Sequential Minimal Optimization Algorithm for High- Throughput MIMO-OFDM Systems," *Int. J. Sci. Res. Eng. Technol.*, vol. 2, no. 12, pp. 876–880, 2014.

[11] J. Davis and M. Goadrich, "The Relationship Between Precision-Recall and ROC Curves," in *International Conference on Machine Learning*, 2006.

[12] D. M. W. Powers, "Evaluation : From Precision , Recall And F-Measure To Roc , Informedness , Markedness & Correlation − R," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011.