Customer Relationship Management by Semi-Supervised Learning

Siavash Emtiyaz, Shilan Rahmani Azar, Behzad Boukani and Mohammad Trick

Young Research Club, Sardasht Branch, Sardasht Islamic Azad University, Sardasht, Iran

Abstract: With the increase of customer information and the rapid change of customer requirements, the need for automated intelligent systems is becoming more vital. An automated system reduces human intervention, improves the quality of information extracted and provides fast feedback for decision making purposes. This study investigates the use of a technique, semi-supervised learning, for the management and analysis of customer-related data warehouse and information. The idea of semi-supervised learning is to learn not only from the labeled training data, but to exploit also the structural information in additionally available unlabeled data. The proposed semi-supervised method is a model by means of a feed-forward neural network trained by a back propagation algorithm (multi-layer perceptron) in order to predict the category of an unknown customer (potential customers). In addition, this technique can be used with Rapid Miner tools for both labeled and unlabeled data.

Key words: Customer Requirement · Neural network · Data minig · Customer Relationship Management · Semi-Supervised Learning

INTRODUCTION

For a long time, the focus of modern companies has been shifted from being product oriented to customer-centric organizations. In the industry it is common to maintain their existing customers which is more cost-effective than attracting new ones [1]. Feng *et al*. [2] point out that about 5% increase in customer retention leads to a 25–95% increase in company profit. Therefore, companies are focusing attention on building relationships with their customers in order to improve satisfaction and retention. This implies that companies must learn much about their customers' needs and demands, their tastes and buying propensities, etc., which is the focus of Customer Relationship Management (CRM) [3]. For CRM purposes, data mining techniques can potentially be used to extract hidden information from customer databases or data warehouses. CRM includes methodologies and products (e.g. software) that maintain, manage and optimize the customer's relationship with companies. In brief, the goal of CRM is to provide an optimal balance between a company's investment and the customers' satisfaction [2]. This study introduces a semi-supervised learning technique to enhance CRM processes and their efficiency. Machine learning algorithms have been shown to be practical methods for real-world recognition problems. They also have proven to be efficient in domains that are highly dynamic with respect to many values and conditions. Various machine learning algorithms have been proposed in the literature [3, 4]. This paper explores the possibility of using semi-supervised Learning for analytical CRM. The reality is that the situation for operational CRM application is no different from other real-world problems [5]. In this application, unlabeled data can be collected by automated means from various databases, while labeled data requires human experts or other limited or expensive categorization resources. The fact that unlabelled data is readily available, or inexpensive to collect, can be appealing and one may want to use them. Despite the natural appeal of using unlabeled data, it is not obvious how records without labels; that can help to develop a system for predicting the labels. This paper introduces a feed forward neural network model, which combines a large set of unlabeled data with a small set of labeled records, to boost the performance of classification. The algorithm is discussed in detail in the following sections.

Corresponding Author: Behzad Boukani, Young Research Club, Sardasht Branch,
Sardasht Islamic Azad University, Sardasht, Iran.

Related Work: Recently, analytical CRM is gaining momentum and some approaches have been proposed in the literature.Al- Mudimig *et al*. [6] categorized CRM components into three groups: adaptive learning, forward looking and optimization. Using these components, they have formulated CRM in such a way that state-dependent factors were taken into consideration. Using the proposed framework, they were able to learn the evolution of customer demand, as well as the dynamic effect of its marketing. Xu and Walton [7] have discussed the latest findings of CRM systems and their application. In addition, they have introduced an analytical CRM system for customer knowledge acquisition. Eelectronic CRM has also been investigated in [8]. In other study [2], hidden information from the logs to automatically tailor ranking functions for a particular user group or collection was extracted. Baesens *et al*. [9] have introduced a method with which they were able to predict the customers' future spending patterns. Predicting customer behavior has also been investigated using a transactional database [10].

Methodology: Machine learning algorithms have been successfully applied in various domains including analytical CRM. Semi-supervised learning (SSL) is halfway between supervised and unsupervised learning. A set of labeled data is used for designing an initial classifier, which is then used for labeling the remaining unlabelled data. Once this is done a classifier is constructed on the basis of both the original and newly labeled data.In addition, it is a well-known fact that when the problem domain is too complex, a large number of training records is required for classification. However, in CRM, similar to other real-world applications, unlabeled data is abundant and readily available, while labeled data requires human experts or other expensive categorization tools. In this paper, we introduced a semi-supervised learning technique that requires only a small size of labeled data for training. Additional labeled records automatically will be collected.

The Data: We examined the proposed algorithm on two datasets. These datasets representing customer relationship management problems have been used in experiment. First dataset is related to bank dataset. It contained 1000 records and 32 variables. Dataset was divided into a training subset of 700 records and a test subset of 300 records. Then, the training subset was divided into labeled and unlabelled subsets of 216 and 440, respectively. The training data consisted in two classes (good and bad).

Secondary dataset is related to insurance dataset. It contained 8802 records and 22 variables. Dataset was divided into a training subset of 6000 records and a test subset of 2802 records. Then, the training subset was divided into labeled and unlabelled subsets of 2250 and 3750, respectively. The training data consisted in two classes (yes and no).

Proposed Semi-Supervised Algorithm: In this paper, we introduced a semi-supervised learning method that can be employed in analytical CRM systems. Various semi-supervised learning methods have been proposed. Self-training is a commonly used technique for semi-supervised learning. In self training a classifier is first trained with the small amount of labeled data. The classifier is then used to classify the unlabeled data. Typically the most confident unlabeled points, together with their predicted labels, are added to the training set. The classifier is re-trained and the procedure repeated. Note the classifieruses its own predictions to teach itself. The procedure is also called self-teaching or bootstrapping (not to be confused with the statistical procedure with the same name) [11]. Another popular approach is co-training. Co-training assumes that:

- Features can be split into two sets
- Each sub-feature set is sufficient to train a good classifier
- The two sets are conditionally independent given the class.

Initially two separate classifiers are trained with the labeled data, on the two sub-feature sets, respectively. Each classifier then classifies the unlabeled data and 'teaches' theother classifier with the few unlabeled examples (and the predicted labels); they feel most confident. Each classifier is retrained with the additional training examples given by the other classifier and the process repeats. In co-training, unlabeled data helps by reducing the version space size. In other words, the two classifiers (or hypotheses) must agree on a much large unlabeled data as well as the labeled data. In this paper, we used self-training algorithm for customer behavior modeling (Figure 1) [11].
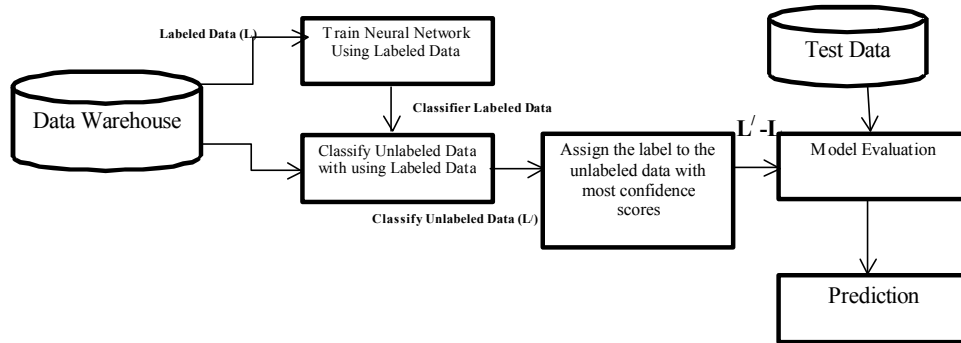
Fig. 1: Architecture of the proposed Algorithm

The proposed method is multi-layer perceptron neural network algorithm in which a back propagation (BP) algorithm [8] is used to classifier. Multi-Layer Perceptronnetworks consists of multiple layers of computational units, usually interconnected in a feed-forward way. Each neuron in one layer has direct connections to the neurons of the subsequent layer. In many applications the units of these networks apply a sigmoid function as an activation function.Each layer consists of units which receive their input from units from a layer directly below and send their output to units in a layer directly above the unit. There are no connections within a layer [12, 13].

Multi-layer networks use a variety of learning techniques, the most popular being back-propagation. Here, the output values are compared with the correct answer to compute the value of some predefined error-function. By various techniques, the error is then fed back through the network. Using this information, the algorithm adjusts the weights of each connection in order to reduce the value of the error function by some small amount. After repeating this process for a sufficiently large number of training cycles, the network will usually converges to some state where the error of the calculations is small. In this case, one would say that the network has learned a certain target function. To adjust weights properly, one applies a general method for non-linear optimization that is called gradient descent [13]. Consider a network with a single real input X and network function F. The derivative F'(X) is computed in two phases:

- Feed Forward: the input X is fed into the network. The primitive functions at the nodes and their derivatives are evaluated at each node. The derivatives are stored.

- Back propagation: the constant 1 is fed into the output unit and the network is run backwards. Incoming information to a node is added and the result is transmitted to the left of the unit. The result collected at the input unit is the derivative of the network function with respect to X [12, 13].

$$E(\vec{w}) \equiv \frac{1}{2} \sum_{d \in D} \sum_{k \in outputs} (t_{kd} - O_{kd})^2 \tag{1}$$

In This Paper, We define the structure of the neural network with the parameter list hidden layers. Each list entry describes a new hidden layer. The key of each entry must correspond to the layer name [8]. The value of each entry must be a number defining the size of the hidden layer. A size value of -1 indicates that the layer size should be calculated from the number of attributes of the input example set. In this case, the layer size will be set to (number of attributes + number of classes)/4. If do not specify any hidden layers, a default hidden layer with sigmoid type and size (number of attributes + number of classes)/4 will be created and added to the net. If only a single layer without nodes is specified, the input nodes are directly connected to the output nodes and no hidden layer will be used. The used activation function is the usual sigmoid function. Therefore, the values ranges of the attributes should be scaled to -1 and +1. The type of the output node is sigmoid if the learning data describes a classification task and linear for numerical regression tasks [12].

Experiment and Result: The proposed semi-supervised algorithm was carried out using Rapid Miner 5.1 Rapid Miner is unquestionable the world-leading open-source system for data mining. It is available as a stand-alone application for data analysis and as a data mining engine

Table 1: An overview labeled data to unlabeled data with most confidence

Raw No	1	2	3	4	5	6	7	8	9	10	11	12
Confidence (no response)	0.172	0.016	0.899	0.994	1.000	0.132	0.952	0.995	0.864	0.012	0.936	0.000
Confidence (response)	0.817	0.945	0.061	0.001	0.000	0.743	0.031	0.005	0.142	0.995	0.121	1.000
Assign Labeled Data	Response	Response	no response	no response	no response	Response	no response	no response	no response	Response	no response	Response

Table 2: Confusion matrix for two class classifier [14]

		Predicted	
		Negative	Positive
Actual	Negative	a	b
	Positive	c	d

for the integration into own products. It provides data mining and machine learning procedures including: data loading and transformation (ETL), data preprocessing and visualization, modeling, evaluation and deployment. Design environment contain three primary operators:

- Neural network: This operator learns a model by means of a feed-forward neural network trained by a back propagation algorithm.
- Apply Model: This operator applies a model to unlabeled data. Models usually contain information about the data they have been trained on. This information can be used for predicting the value of a possibly unknown label (Table 1)
- Performance: This operator delivers list of performance values automatically determined in order to fit the learning task type.

In addition, design environment contains two secondary operators:

- Nominal to Numerical: This operator maps all non numeric attributes to real valued attributes. Nothing is done for numeric attributes, binary attributes are mapped to 0 and 1. In this case, we maps name attribute to real attribute.
- Set Role: This operator can be used to change the role of an attribute of the input *Example Set*. In this case, we change Label attribute for prediction. In the test presented here, we split data as random 30% for test and 70% for training. Furthermore, we use confusion matrix for compare the result. A confusion matrix contains information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using the data in the matrix. Table 2 shows the confusion matrix for a two class classifier. The entries in the confusion matrix have the following meaning in the context of our study:

- Is the number of correct predictions that an instance is negative,
- Is the number of incorrect predictions that an instance is positive,
- Is the number of incorrect of predictions that an instance negative and
- Is the number of correct predictions that an instance is positive

- The accuracy (AC) is the proportion of the total number of predictions that were correct. It is determined using the following equation:

$$c = \frac{a+b}{a+b+c+d}$$

- The recall or true positive rate (TP) is the proportion of positive cases that were correctly identified, as calculated using the following equation:

$$TP = \frac{d}{c+d}$$

- The false positive rate (FP) is the proportion of negatives cases that were incorrectly classified as positive, as calculated using the following equation:

$$FP = \frac{b}{a+b}$$

- The true negative rate (TN) is defined as the proportion of negatives cases that were classified correctly, as calculated using the following equation:

$$TN = \frac{a}{a+b}$$

- The false negative rate (FN) is the proportion of positives cases that were incorrectly classified as negative, as calculated using the following equation:
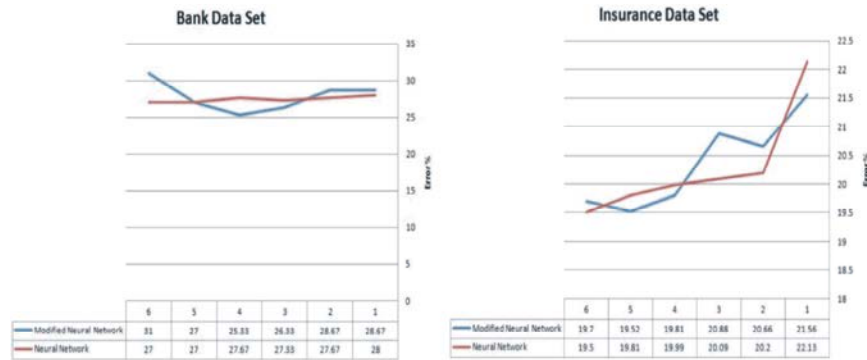
$$FN = \frac{c}{c+d}$$

Fig. 2: Aplot comparison to show the percent error in bank data set and insurance data set
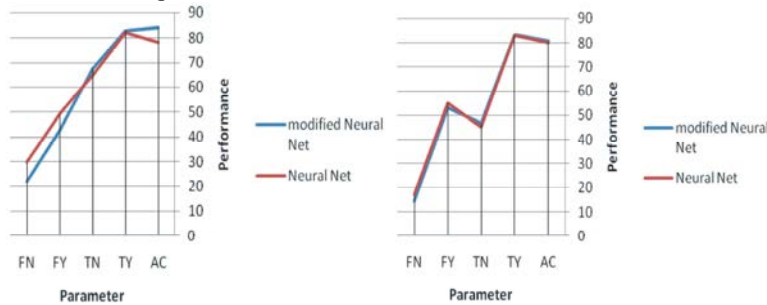

Fig. 3: Test performance on the bank data set(Left) and Test performance on the insurance data set(Right)

Table 3: Test Performance (%) for proposed semi-supervised algorithm on the bank data set

Operator	Accuracy(%)	True Good (%)	True Bad (%)	False Good (%)	False Bad (%)
Neural Net	78.24	82.26	64.88	49.12	29.75
Proposed algorithm	84.27	83.10	67.34	42.66	21.87

Table 4: Test Performance(%) for proposed semi-supervised algorithm on the Insurance data set

Operator	Accuracy (%)	True Yes (%)	True No (%)	False Yes (%)	False No (%)
Neural Net	79.90	82.98	45.00	55.00	17.02
Proposed algorithm	80.79	83.44	46.80	53.20	14.56

We trained one-hidden layer of neural net operator in rapid miner with sigmoid of hidden units according following formula: ((number of attributes + number of classes)/2) + 1). Furthermore, using the error-back propagation algorithm, training cycle 500, training rate 0.3 and error epsilon 1.0E-5. Rapid miner is open source system for data mining and we modified this operator codes and we could get a better result than the previous. We modified trained with one-hidden layer of neural net operator in rapid miner with sigmoid of hidden units according following formula:(Number of attributes + Number of classes)/4.

In Figure 2, we observe a plot comparison to show the percent error in the performance of neural network algorithm for both modified neural network and neural network. Therefore, we have used above the formula for this research. Blue line is modified neural network, which is defined by the following formula:

$$(\text{Number of attributes} + \text{Number of classes})/ X \quad X=1,2,3,4,5,6$$

Also, the red line is the former neural network which is defined by the following formula:

$$((\text{Number of attributes} + \text{Number of classes})/X)+ 1 \quad X=1,2,3,4,5,6$$

As can be seen in Figure 2, in parts three and four our algorithm has the lowest percentage of error.

Results are presented in Table 3. Table 3 shows the test result on the bank data set. These results have been 500 training cycle. Figure 3 shows the test performance (%) for proposed semi-supervised algorithm on the bank data set. Also, in Figure 3 we observe a significant increase in the performance of proposed semi-supervised. Table 4 summarized the obtained test results on insurance data set.
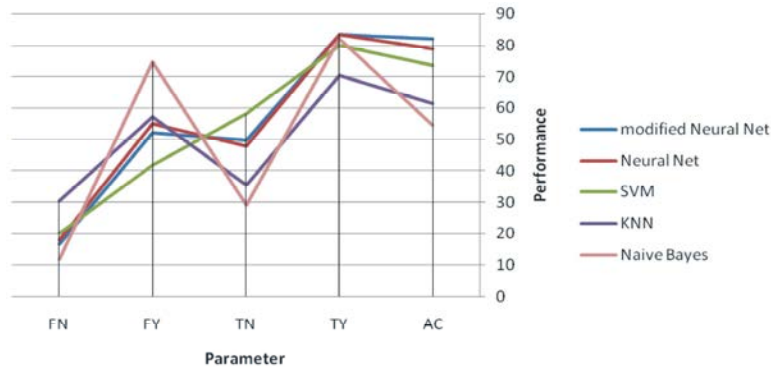
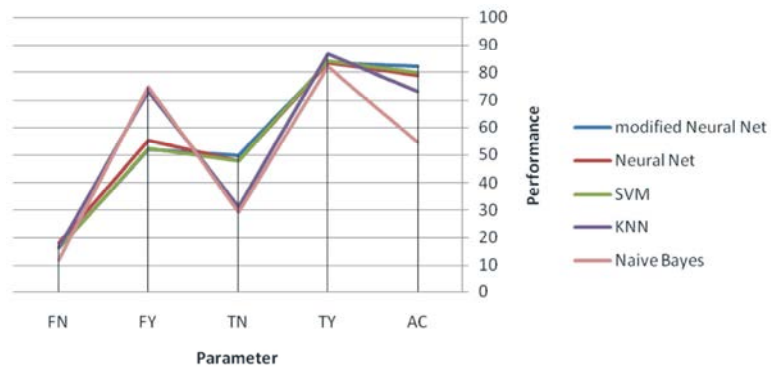Fig. 4: Testperformance on the Bank data setwith other algorithms

Fig. 5: Testperformance on Insurance datasetwith other algorithms

Table 5: Test Performance (%) for proposed semi-supervised algorithm and other algorithms on the bank data set

Operator	Accuracy(%)	True Good (%)	True Bad (%)	False Good (%)	False Bad (%)
Proposed algorithm	75.67	82.28	62.22	40.76	19.12
Neural Net	71.33	82.21	58.65	43.32	22.77
SVM	73.67	79.91	58.14	41.86	20.09
KNN	61.60	70.48	35.41	57.22	30.50
Naive Bayes	54.67	78.84	37.64	63.39	22.16

Table 6: Test Performance (%) for proposed semi-supervised algorithm and other algorithms on the insurance data set

Operator	Accuracy (%)	True Yes (%)	True No (%)	False Yes (%)	False No (%)
Proposed algorithm	82.19	83.44	49.80	52.20	16.56
Neural Net	78.80	83.33	47.90	55.00	18.02
SVM	79.87	84.00	47.70	52.14	15.98
KNN	73.19	86.48	31.19	72.95	16.57
Naive Bayes	54.60	82.15	29.18	74.73	11.95

We trained the neural network classifier with the training data and tested on records. The lowest error we achieved 5.33%. The proposed algorithm illustrated similar generalization capability to the classifier trained with the original training records, with the difference of approximately 4%. Now, we want compare our proposed semi-supervised algorithm with other algorithms in rapid miner. Results are presented in Tables 5 and 6 on the bank and insurance data sets, respectively.

CONCLUSION

The most important results obtained from the above techniques are reorganization of data and information for easier access, yielding more effective Efficiency and better data classification in order to obtain best results from them. Data mining and CRM integration has advantages and lots of benefits for companies that need to discover

profitability of some customers than other customers. Data mining can identify important customers in data warehouse.

Our approach has follow advantages: firstly, semi-supervised machine learning techniques are used to automatically construct customer behavior modeling to improve accuracy. Secondary, a neural network can be used for data visualization purposes. Data visualization is an important step in data analysis and can provide valuable insights into all steps of classification.

Future work will also examine the use other classification methods, such as fuzzy neural network, to classifier labeled data of varyingconfidences. We are also interested in benchmarking our method on other operational CRM databases.

REFERENCES

1. Huang, S.M., I. Kwan and S.H. Li, 2003. Web Mining for CRM an Empirical Study of computer Game Service Company. International Journal of Electronic Business Management, 1(1): 36-45.

2. Feng, D., Z. Zhang, F. Zhou and J. Ji, 2008. Application study of data mining on customer relationship management in E-commerce. in Computer-Aided Industrial Design and Conceptual Design, 2008. CAID/CD 2008. 9th International Conference on. 2008: IEEE.

3. Dara, R.A., T. Khan, J. Azim, O. Cicchello and G. Cort, A Semi-Supervised Approach to Customer Relationship Management. Intelligence and Soft Computing.

4. Ding, X., Y. Li and Y. Zhao, 2008. A framework of user model based on Semi-supervised techniques. in e-Business Engineering, 2008. ICEBE'08. IEEE International Conference on. 2008: IEEE.

5. Al-Mudimigh, A.S., Z. Ullah and F. Saleem, 2009. Data mining strategies and techniques for CRM systems. in System of Systems Engineering, 2009. SoSE 2009. IEEE International Conference on. 2009: IEEE.

6. Al-Mudimigh, S.Z., F. Ullah and N. Al-Aboud, 2009. Efficient Implement of Data Mining: Improve Customer's Behavior. in IEEE Intel ligent Systems, College of Computer and Information Sciences, King Saud University, Riyadh Kingdom of Saudi Arabia.

7. Xu, M. and J. Walton, 2005. Gaining customer knowledge through analytical CRM. Industrial management & data systems, 105(7): 955-971.

8. Mithas, S., M.S. Krishnan and C. Fornell, 2005. Why do customer relationship management applications affect customer satisfaction? Journal of Marketing, pp: 201-209.

9. Baesens, B., G. Verstraeten, D. Van den Poel, M. Egmont-Petersen, P. Van Kenhove and J. Vanthienen, 2004. Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers. European Journal of Operational Research, 156(2): 508-523.

10. Buckinx, W., G. Verstraeten and D. Van den Poel, 2007. Predicting customer loyalty using the internal transactional database. Expert Systems with Applications, 32(1): 125-134.

11. Zablah, A.R., D.N. Bellenger and W.J. Johnston, 2004. An evaluation of divergent perspectives on customer relationship management: Towards a common understanding of an emerging phenomenon. Industrial Marketing Management, 33(6): 475-489.

12. Galushkin, A.I., 2007. Neural networks theory. Vol. 420. 2007: Springer Berlin.

13. Zainuddin, Z., N. Mahat and Y.A. Hassan, 2005. Improving the convergence of the backpropagation algorithm using local adaptive techniques. World Academy of Science, Engineering and Technology, 1: 79-82.

14. Provost, F. and R. Kohavi, 1998. Guest editors' introduction: On applied research in machine learning. Machine learning, 30(2): 127-132.