



Case Study

Customer satisfaction prediction with Michigan-style learning classifier system

Keivan Borna¹  · Shokoofeh Hoseini²  · Mohammad Ali Mehdi Aghaei³

© The Author(s) 2019 [OPEN](#)

Abstract

Many different classification algorithms can be used in order to analyze, classify and predict data. Learning classifier system (LCS) which is known as a genetic base machine learning system, combines the machine learning with evolutionary computing and other heuristics to produce an adaptive system that learns to solve a particular problem. This paper uses the Michigan style LCS, in the context of bank customer satisfaction to classify customers into two different groups: unsatisfied/satisfied customers. Three different Rule Compaction strategies are used to compare the rule population's accuracy and micro/macro population size. The result specifies features that mostly influence prediction.

Keywords Learning classifier system · Evolutionary algorithm · Michigan style LCS · Machine learning · Prediction

1 Introduction

[Learning] Classifier Systems (LCSs) [1, 2] are a kind of Rule-Based system (RBS) [3, 4] with general mechanism for parallel rule processing, adaptive generation of new rules, and testing the effectiveness of existing rules. These mechanisms approach to more reliable without “brittleness” learning systems in AI. For a further understanding of what is the LCS see [1, 5, 6]. This paper indicates the reason of using LCS as a Genetic Base Machine Learning (GBML) [7, 8] system for prediction. A preprocessing step is required to prepare dataset. Experimental results are conducted by applying three Rule Compaction algorithms [9, 10] on a dataset which consists of customer's satisfaction information in Santander Bank [11]. Section 2 indicates the eagerness of using LCS. The proposed method is presented in Sect. 3 and the concept of Rule Compaction and their algorithm is presented in Sect. 4, experimental results and evaluation are discussed in Sect. 5, and finally Sect. 6 is devoted to the conclusions.

2 Why using LCS?

LCS algorithms in general, constitute a unique alternative to other well-known machine learning strategies that follow the classic paradigm of seeking to identify a ‘best’ model that can individually be applied to the entire dataset. There are a lot of LCS implementation [12] that causes prediction/classification. Here are the advantages that encourage us to use LCS [13, 17].

- Model free: They make limited assumptions about the environment, or the patterns of association within the data [17].
- Ensemble Learner: is to build a predictive learning systems by integrating multiple learner to improve the performance and accuracy. Majority Voting and averaging are two of the applicable ensemble methods [17].
- Stochastic Learner are Non-deterministic learning with advantage in large-scale or high complexity in compare with deterministic.

✉ Keivan Borna, borna@khu.ac.ir; Shokoofeh Hoseini, hoseini.shokoofeh@gmail.com; Mohammad Ali Mehdi Aghaei, shabar.aghaei@gmail.com | ¹Department of Computer Science, Faculty of Mathematics and Computer Science, Kharazmi University, Tehran, Iran. ²Department of IT Management, Faculty of Management, Kharazmi University, Tehran, Iran. ³Department of IT Management, Research and Science Branch Islamic Azad University, Tehran, Iran.



- Implicitly Multi-objective: is a characteristics of obtaining general and accurate rules with implicit and explicit pressures, encouraging maximal generality/simplicity [17].
- Interpretable: LCS rules are logical IF:THEN statements, interpretable to human [14].

3 Proposed method

Figure 1 shows the proposed method phases. Starting from preprocessing the raw dataset, then applying three rule compaction strategies separately on the processed dataset. After obtaining the predicted results, a comprehensive evaluation is investigated and presented in Sect. 5, while the subsection 3.1 discusses the dataset used, subsection 3.2 presents the preprocessing steps required to prepare the dataset, and the subsection 3.3 illustrates the reasonable configuration parameter for applying LCS.

3.1 The dataset

The dataset consists of 369 anonymized features, excluding the ID/target column. So a challenge with this dataset is what each feature means—thus little domain knowledge or intuition is used.

3.2 The preprocessing steps

Figure 2 shows five sub-steps which applied in the preprocessing steps. The first step is to remove duplicate

columns. There are several columns which have a single constant value which are removed in second step. Then strongly-correlated columns are identified and only ones in the training dataset are remained. The value (0.85) is chosen as the threshold for high correlation in the third step. There is a massive mismatch between the numbers of satisfied customers (96%) versus unsatisfied ones (4%). In fourth step we balanced the two classes. Synthetic Minority Over-sampling Technique (SMOTE) [15] is used for balancing the classes. SMOTE implementation is available in the R package DMwR. The number of satisfied customers outnumber the unsatisfied ones by roughly a factor of 24.27. After preprocessing steps the balanced dataset’s records yield to 147,392, and the number of features yield to 143, excluding ID and Target.

The last step is to convert all attribute values into binary format, because the LCS implementation acts as rule-base system (like other GBML systems) and has been coded to handle binary values.

3.3 LCS configuration

The arbitrary configurations and their values are discussed.

- Learning Iteration: is one of the most critical run parameters. In this case, LCS iterates over instances as twice as the folded dataset size (23,826) which occurs two epochs and generates more reliable rules [9].
- Maximum Population Size: must be specified by initial trial and error, in this case maximum population size of 7000 is applied [9].

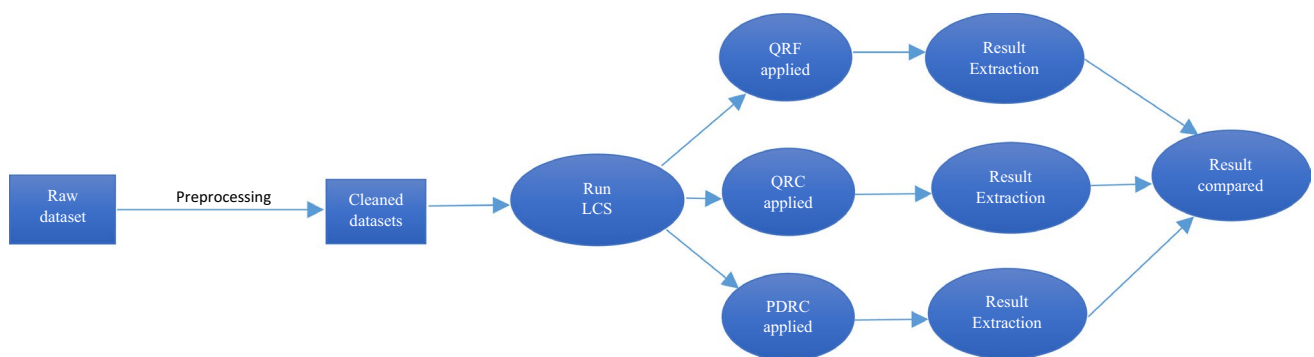


Fig. 1 Proposed method steps



Fig. 2 Preprocessing steps

- **Cross Validation:** The fivefold cross validation (CV) is determined and serially per-formed a complete run, then evaluated on each training and testing dataset to have a better predictions.
- **Attribute Tracking/Feedback:** Attribute tracking (AT) and Attribute feedback (AF) are used to guide the algorithm to more intelligently explore reliable attribute patterns [16].

4 Rule compaction strategies

Three rule compaction strategies (QRC, QRF, and PDRC) [9] are applied and the rule population, macro/micro population size and accuracy are compared.

- **Quick Rule Compaction (QRC):** It modifies two miner Rule Compaction strategies (Fu1, Fu2) which sorts the rules decreasingly by fitness (or accuracy) then for all in-instances in a dataset calculate MatchCount and considers any rule that has this parameter greater than zero.
- **Quick Rule Filter (QRF):** QRF is simply a filter which scans the rule population and deletes any rule with an accuracy ≤ 0.5 . Additionally, a rule is also deleted if it covers (i.e. matches) less than two instances in the dataset.
- **Parameter Driven Rule Compaction (PDRC):** there are three different rule parameter (accuracy, numerosity and generality). These parameter updated during LCS iteration. In PDRC algorithm these parameters are considered in rule compaction strategy as follows: Find the best rules which have the highest value of the product of accuracy and numerosity and generality.

5 Comparisons and experimental results

LCS algorithm is applied in conjunction with three rule compaction strategies and at-tribute tracking/feedback to a dataset containing more than 147,392 records and 143 features. Fivefold cross validation (CV) is employed to measure average testing accuracy and account for over-fitting. With fivefold CV, twice the each fold training dataset

size (235,826), the LCS-based algorithm are completed followed by the same number of runs for each of the three rule compaction strategies. Experiments are run with The ExSTraCS [17].

Statistical analysis For each experiment, the value of training accuracy, test accuracy, macro population size, micro population size, rule generality, and the rule compaction time are reported. Results over fivefold CV are averaged.

Table 1 shows QRF method is the fastest and QRC gives the better accuracy. The difference between micro and macro population size is a good reference to understand the characteristic of the rule population. The higher difference between micro and macro population size shows the stronger and more reliable rules exist in the population [17].

Attribute Tracking and Attribute Feedback are also applied. With these mechanism three summary statistics are introduced in [18] which can be use in knowledge discovery to identify attributes that are of particular importance in making class prediction.

These statistics include the specificity sum, the accuracy sum, and the attribute tracking global sum. Attributes that consistently have the highest sums for these three metrics are likely to be most important for making accurate predictions [17].

In this experiment, 20 top attributes which have the highest value of above metrics are selected then the common attributes are chosen as an important attributes. Table 2 shows the common attribute set of the chosen metrics within all three rule compaction and none rule compaction. In this experiment these are the most important attributes.

Table 1 Comparison of different experiments

RC Strategy	Training Accuracy	Testing Accuracy	Macro PopSize	Micro PopSize	Generality	Rule compaction time
None	0.7821268	0.782142	2943	7000	0.98983	0
PDRC	0.8819462	0.881681	142	1717	0.99295	6.769883
QRC	0.8840805	0.884339	507	2257	0.99079	1.590283
QRF	0.84784910	0.84555980	2422	5933	0.98968	3.34E-05

Table 2 Attributes with highest sums in three metrics

The most influenceable attributes
num_var43_recib_ult1
Var15
Var36
num_meses_var39_vig_ult3

6 Conclusion

This paper analyzed and compared three rule compaction strategies and applying them on a dataset containing more than 147,392 records. The data represent customer satisfaction information of the Santander Bank. A comprehensive comparison is conducted after obtaining the results. The results showed that QRC makes better accuracy whereas QRF is running faster. Then we indicate the most important attributes by applying attribute tracking and attribute feedback mechanisms and extract four most important attributes for prediction.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Holland JH (1975) *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor
- Lanzi P-L, Riolo R (2000) A roadmap to the last decade of learning classifier system research. In: Lanzi P-L, Stolzmann, Wilson SW (eds) *Learning classifier systems: from foundations to applications*. Springer, New York, pp 33–62
- Bassel GW, Glaab E, Marquez J, Holdsworth MJ, Bacardit J (2011) Functional network construction in Arabidopsis using rule-based machine learning on large-scale data sets. *Plant Cell* 23(9):3101–3116
- Holland JH (1986) The possibilities of general-purpose learning algorithms applied to parallel rule-based systems. In: *Machine learning, an artificial intelligence approach*, vol 2. pp 593–623
- Urbanowicz RJ, Brown WN (2018) *Introduction to learning classifier system*, 1st edn. Springer, Berlin
- Janikow CZ (1993) *A knowledge-intensive genetic algorithm for supervised learning*. Kluwer Academic Publishers, Boston
- Bonelli P, Parodi A, Sen S, Wilson S (1990) NEWBOOLE: a fast GBML system. In: *Machine learning: proceedings of the seventh international conference*, Austin, Texas. Morgan Kaufmann, pp 153–159
- Goldberg DE (1989) *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley Publishing Company, Inc, Boston
- Tan J, Moore JH, Urbanowicz RJ (2013) Rapid rule compaction strategies for global knowledge discovery in a supervised learning classifier system. In: *ECAL*. MIT Press, Cambridge
- Dixon PW, Corne DW, Oates MJ (2002) A ruleset reduction algorithm for the XCS learning classifier system. In: *International workshop on learning classifier systems*. Part of the lecture notes in computer science book series (LNCS, vol 2661). Springer, Berlin, Heidelberg, pp 20–29
- Kaggle (2015) Santander customer satisfaction. <https://www.kaggle.com/c/santander-customer-satisfaction>
- Urbanowicz RJ, Moore JH (2009) Learning classifier systems: a complete introduction, review, and roadmap. *J Artif Evol Appl* 2009:1
- Urbanowicz R, Browne W (2015) Introducing rule-based machine learning: a practical guide. In: *Proceedings of the companion publication of the annual conference on genetic and evolutionary computation*, Madrid, Spain — July 11–15, 2015. ACM, pp 263–292
- Urbanowicz RJ, Granizo-Mackenzie A, Moore JH (2012) An analysis pipeline with statistical and visualization-guided knowledge discovery for michigan-style learning classifier systems. *IEEE Comput Intell Mag* 7(4):35–45
- Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 16:321–357
- Urbanowicz R, Granizo-Mackenzie A, Moore J (2012) Instance-linked attribute tracking and feedback for Michigan-style supervised learning classifier systems. In: *Proceedings of the 14th international conference on genetic and evolutionary computation conference*, pp 927–934. ACM
- Urbanowicz RJ, Moore JH (2015) ExSTraCS 2.0: description and evaluation of a scalable learning classifier system. *Evol Intell* 8(2–3):89–116
- Urbanowicz RJ, Granizo-Mackenzie A, Moore JH (2012) An analysis pipeline with statistical and visualization-guided knowledge discovery for michigan-style learning classifier systems. *Comput Intell Mag* 7(4):35–45

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.