Credit risk evaluation in peer-to-peer lending with linguistic data transformation and supervised learning

József Mezei[1,2]
[1]School of Business and Management
Lappeenranta University of Technology
Lappeenranta, Finland
[2]F-Secure Corporation, Helsinki, Finland
Jozsef.Mezei@lut.fi

Ajay Byanjankar[3], Markku Heikkilä[3],
[3]Faculty of Business, Social Sciences and Economics
Åbo Akademi University
Turku, Finland
abyanjan@abo.fi, maheikki@abo.fi

Abstract

The widespread availability of various peer-to-peer lending solutions is rapidly changing the landscape of financial services. Beside the natural advantages over traditional services, a relevant problem in the domain is to correctly assess the risk associated with borrowers. In contrast to traditional financial services industries, in peer-to-peer lending the unsecured nature of loans as well as the relative novelty of the platforms make the assessment of risk a difficult problem. In this article we propose to use traditional machine learning methods enhanced with fuzzy set theory based transformation of data to improve the quality of identifying loans with high likelihood of default. We assess the proposed approach on a real-life dataset from one of the largest peer-to-peer platforms in Europe. The results demonstrate that (i) traditional classification algorithms show good performance in classifying borrowers, and (ii) their performance can be improved using linguistic data transformation.

1 Introduction

Peer-to-peer lending (P2P) is a form of online microfinancing that allows individuals to lend or borrow virtually without financial intermediaries and collateral. P2P lending has seen a rapid growth due to the ease in receiving credit without having to deal with any financial intermediaries [1]. As the main advantage, there is a low cost of borrowing money over P2P lending as the operating cost for P2P lending is low as the platform operates online, where most of the processes are automatized [2]. In addition, it facilitates quick loan decisions as the operations take place on the Internet that connects borrowers and lenders instantly [3]. P2P lending allows borrowers with short credit history an easy access to credit with a lower interest rate than traditional banks. Furthermore, the cost is reduced from the unbundling of unnecessary services that are coupled with traditional intermediaries [4]. Small scale borrowers, such as individuals and small firms and borrowers that are placed at the long tail of credit, are mostly attracted to P2P lending due to nonrequirement of collateral for the loans and lack of financial intermediaries [2, 5].

Considering all the highlighted issues, it is a highly relevant problem to develop novel methods to assess borrowers in P2P platforms offering decision support for investors. Motivated by the developments in the general finance domain, in this paper we focus on utilizing various machine learning techniques to classify borrowers as default or non-default based on the data available at the time of the loan application. Methods and approaches based on machine learning and linguistic modeling have reached attention in finance, as one of the most important classes of problems in financial risk analysis is the identification and prediction of risk incidents, such as fraud, bankruptcy or default. In P2P lending decisions, intelligent classification methods (that find the patterns of realized risky incidents from the data and produce a set of rules to predict further incidents) are based on predetermined binary risk classes of non-default and default. As it is in the interest of P2P investors to minimize losses caused by incidents of default, these classes supervise the training of the algorithm to find those rules that show the best performance in classifying data as the basis of future lending decisions.

In this article, we propose to use the combination of traditional machine learning algorithms and linguistic data transformation in order to improve credit risk evaluation in P2P lending. In this vein, our main objective is to *improve the utilization of classification algorithms in assessing credit risk in peer-to-peer lending with fuzzy sets-based linguistic data transformation.*

The rest of the paper is structured as follows. In Section 2, related literature is summarized from the domains

URI: http://hdl.handle.net/10125/50056
ISBN: 978-0-9981331-1-9

of P2P lending and the role of machine learning and fuzzy sets based linguistic modeling in finance. In Section 3, we describe an improved version of the approach presented in [6] as the basis of assessing credit risk in P2P lending. The performance of the proposal is assessed on a real life dataset in Section 4. Finally, conclusions and future research directions are discussed in Section 5.

2 Literature review

In this section relevant literature is presented. First, we look at the literature on P2P lending from the perspective of the methodologies applied to assess the risk associated to loan contracts. Secondly, we discuss the potential of utilizing machine learning and particularly motivate using fuzzy sets and linguistic modeling in the finance domain.

2.1 Peer-to-peer lending and credit risk

P2P lending, although being an attractive alternative to conventional banking, has some major problems concerning credit risk for lenders. According to Li et al. [7], P2P lending faces various risks, such as default risk, operational risk and policy risk. P2P platforms acting as simple intermediaries with no credit services have low risk, whereas high risk is placed on the lenders. Wang and Greiner [8] identify the fundamental problem of P2P lenders in the risk of loss on investment, since loans are not secured with a collateral. Credit risk in P2P lending can be high, which leads to lenders' investments being in high risk of default. In order to reduce credit risk, P2P platforms deploy various protection mechanisms, such as capital protection and recovery of arrears [7]. In addition, information asymmetry is a significant issue existing in P2P lending, meaning that some borrowers tend to show fraudulent behavior that leads to the misinterpretation of credit risks by lenders [9]. It is a difficult task to follow the credit risk management measures applied by conventional banking to strengthen the trust in borrowers in P2P lending due to the high associated cost and the fact that they operate online and borrowers and lenders do not have any physical contact with each other [10, 11]. According to Lee and Lee [12], lenders in P2P lending are not professional investors and they lack appropriate skills to evaluate credit risk; additionally, the lack of collateral exposes them to a high risk of loss. Due to the lack of skills in evaluating credit risk and fear of selecting risky loans, most lenders tend to follow herding behavior and consequently finance loans with high number of bidders.

In most of the cases, investors make the decision to lend to borrowers based on the demographic and financial information provided by borrowers. Klafft [1] suggests to lend to loans that have no delinquencies, debt to income rate below a certain value and no credit inquires. In addition, the decision to lend is highly affected by the trust between lenders and borrowers, where the trust is developed through exchange of messages [13]. Furthermore, the trust in borrowers is highly affected by the quality of information they provide and the quality of services and security protection provided by the P2P platform [14]. Lin et al.[15] identify powerful social networking to be a significant factor in making an informed decision for reducing default risk.

Credit risk management has been an extensively researched topic in the field of finance and there has been a need for better means of evaluating the credit risk with the increase in credit applications. Credit risk is a challenging and complex task to manage and evaluate and is significantly important in financial risk management [16]. A wide range of statistical methods are applied to model credit risk for classifying borrowers by means of credit scoring [17]. The popular methods applied in credit risk modeling include logistic regression [18], neural networks [16], support vector machines [19], decision trees [20], discriminant analysis [21], and *k*-nearest neighbor [22]. In addition, survival analysis [23][24] is also a popular method applied in credit risk management to predict the time to default.

The growth of P2P lending has been fascinating with the growth being at a high rate especially in the last couple of years. However, credit risk existing in the business is also high even though there are different approaches applied by the platforms in controlling the risk. In recent years, one can identify several contributions in the literature with the focus on understanding the credit risk and its measures in reducing the risk in P2P lending. Emekter et al. [11] evaluated the credit risk of borrowers from a leading P2P platform in the United states, Lending club, to help lenders to make a more accurate decision. They applied a non-parametric test to identify the significance of borrowers' characteristics on a loan being default and modeled the default risk of borrowers with a binary logit regression. Furthermore, they utilized the Cox Proportional Hazard model to empirically examine the relationship between the loan duration and probability of default. Following a similar approach, Lin et al. [10] performed risk evaluation of borrowers from a P2P lending platform in China. They applied logistic regression model to evaluate credit risk based on demographic features and the corresponding loan information. Their results suggest that borrowers with low default risk are young female adults, with long working time, stable marital status, low loan amount, low debt to income, high education and no default history.

Cinca et al. [2] performed an empirical study on data from Lending Club to analyze credit risk in P2P lending. They first studied the factors predicting default risk with univariate tests and survival analysis. Their results show that the most important factors explaining default risk are loan grade, interest rate, loan purpose, income, credit history and borrowers' indebtedness. The study further applied logistic regression to model the default risk identifying the grade assigned to loans as the most significant determinant of default. Byanjankar et al. [25] evaluated credit risk in P2P lending with a credit scoring model using artificial neural networks. The model is used to classify the loan applications into default and non-default groups based on their characteristics. The model is proven to be effective in screening the loan applications and outperformed logistic regression in identifying default loans. Similarly, Jiang and Li [26] also built a credit risk evaluation model for P2P lending with back propagation neural networks.

Malekipirbazari and Aksakalli [3] proposed a random forest based classification model for improved identification of good borrowers. Their study compares the random forest model with other machine learning models, such as logistic regression, support vector machine, and k-nearest neighbor classifier. The comparison shows that random forest outperforms other methods in predicting borrowers' risk status. Vedala and Kumar [27] proposed a classification technique to classify good and bad borrowers applying both hard and soft information. They applied Naive Bayes classifier to classify data with only hard information and later compared it with multi-relational Naive Bayes classifier using both hard and soft information.

Jin and Zhu [28] applied a data-driven approach to build credit risk models. They applied random forest for feature selection and then built and compared five different models, including two neural network models, two decision tree models and one support vector machine model. Their result revealed that support vector machine achieved the best performance but with only a slight improvement. Furthermore, they found that loan term, loan amount, annual income, loan grade, debt to income ratio and revolving line utilization play an important role in predicting default. Cinca and Nieto [29] developed a profit scoring model for P2P lending using multivariate linear regression and decision trees. They compared the results of applying profit scoring to credit scoring with a logistic regression. The results revealed that profit scoring models performed better in identifying profitable investments.

2.2 Linguistic modeling and machine learning in finance

The main objectives of mathematical linguistic models based on the general theory of fuzzy sets are to examine linguistic data statistically, to classify the data based on patterns found and to extract knowledge that is operational from the point of view of some other objective. P2P investment decisions prediction of default is an interesting application of linguistic models. Many times data on financial events, such as loan application, repayment or default, consist of sets of thousands of records with both numerical and textual data. Combining both numerical and textual data into a meaningful data set is essential for finding meaningful content for learning patterns and training of the supervised models for P2P default risk analysis.

Lin et al. [30] survey over 100 machine learning articles, 18 baseline classifiers and a number of method combinations, both ensemble and hybrid, and test a number of approaches in the setting of bankruptcy prediction and credit scoring. In the article, numerical and categorical features of 3 financial datasets are used for testing. The interesting result is that no single winner is found but single baseline, ensemble and hybrid methods perform best over the German (credit data), Australian (credit approvals) and Japanese (credit screening) datasets, respectively.

In combination with traditional methodologies, the tools of fuzzy set theory have been applied to machine learning in various ways. Cheng and Hoang [31] propose a hybrid fuzzy instance based classifier for contractor default prediction ($FICDP$) to minimize risk of default related to selecting contractors. $FICDP$ is used in the process of balancing the data set containing only a small number of defaults by oversampling this minority class. Ghandar and Michalewicz [32] perform an experimental evaluation of how the predictive capability relates to interpretability of fuzzy rule based systems obtained using Multi-Objective Evolutionary Algorithms (MOEA) by predicting whether the Bombay Stock Index will rise or fall based on momentum indicators.

In their research, Hajek and Henriques [33] list 27 studies of how various machine learning methods perform in binary classification of financial statements. In their experiment, they consider 32 financial variables in financial reports and 8 linguistic variables in managements discussion and analysis parts of financial reports with 14 machine learning methods. Linguistic variables are calculated as relative frequency counts and feature analysis is performed resulting in 30 subsets of features implying that six financial variables and one linguistic variable should be used in classification. The results show considerable improvements in classification with respect to both accuracy and miscalculation cost for all the methods when linguistic variables are used.

When a machine learning algorithm, such as an artificial neural network [25], is used, patterns found in P2P loan events data define what knowledge can be retrieved from the dataset. For classification, distances between individual instances in features play a crucial role because they define measures of similarity (or dissimilarity) (cf. [34]). For linguistic data, definition and modeling of linguistic expressions and distances between expressions makes a big difference in the quality of trained classification models, because it is often difficult to transform the semantic content of an expression into numeric value(s) usable for finding distances for classification without losing knowledge contained in the expression. Linguistic expressions are often characterized by semantic imprecision, and should be correctly expressed as containing such imprecision in learning models as well. In this sense, imprecise linguistic expressions should be modeled in the context of analyzed data. This can be done with fuzzy linguistic models and specifically by transforming both categorical, numerical and other data sources into fuzzy linguistic values.

Hüllermeier [35] identifies four problems areas in the interpretation of distances from data to decision support with fuzzy tools:

1. Perception that logical and analytical structure of the model as well as the meaning of the parameters are clear - which they often are not;

2. High dimensionality of the model makes understanding of the model more difficult;

3. Association of the fuzzy rules generated in the data-driven process with linguistic labels used in classification - as well as semantic labels used for interpretation - is not clear;

4. Presentation of the model to the user is not straightforward.

Furthermore, Hüllermeier [35] states that classical *knowledge-driven* fuzzy modeling finds the rules with manual analysis by deductive inference. In such a case, when expert and knowledge-based support systems are developed, the number of data-points is relatively small and consists of collections of expert knowledge. On the other hand, for fuzzy machine learning, *data-driven* fuzzy modeling generates the output inductively by model fitting. For inductive generation of decision rules, there are a number of problems that call for attention.

In fuzzy modeling, global model accuracy and interpretability are two conflicting modeling objectives. Zhou and Gan [36] discuss this phenomenon in the context of data-driven fuzzy models and present a unified perspective to tackle this problematics with the concepts low-level interpretability and high-level interpretability. In the approach described in the following section, while the interpretability of the results and underlying data is more straightforward using a linguistic transformation (discretization) of data, classification accuracy remains high.

3 Methodology

In this section, we will describe the methodology utilized in the classification task of predicting credit risk in the domain of P2P lending. We start by describing the basis of the proposal described in a recent paper [6] focusing on linguistic data transformation as a preprocessing step of a classification procedure. Subsequently, we extend the methodology by introducing interval-valued fuzzy sets in determining the optimal number of linguistic labels in the data transformation step.

3.1 Linguistic data transformation for classification tasks

As we reasoned in the previous section, fuzzy linguistic modeling methods can potentially enhance various machine learning algorithms applicable in various domains of finance. Specifically, the use of a linguistic transformation as a type of discretization step can improve the classification performance of supervised learning methodologies. In particular, an optimal granularity value as the basis of transforming the available data before utilizing classification algorithms can potentially improve the final results.

In [6], the steps of determining optimal linguistic label sets for transforming data are summarized as follows (as depicted in Figure 1):

1. Define the linguistic label set used for each variable in the dataset

2. Transform data into labels from the selected set using a multigranular fuzzy linguistic transformation function. One can employ either range transformation (numeric variables) or multigranular transformation

3. Carry out the classification learning process using the training set

4. Evaluate the obtained machine learning model using the test data. If the classification performance is above a predefined threshold (in terms of accuracy,

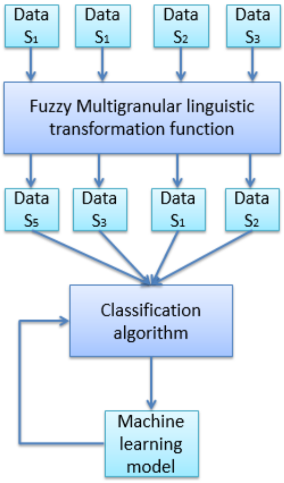

Figure 1: Classification process using multi-granular fuzzy linguistic modeling ([6])

AUC, etc.), the used label set is stored as the optimal one

5. If the classification performance is not sufficiently good, redefine linguistic label sets used for the data and repeat the process from step 3

3.2 Choosing the optimal number of granules with fuzzy entropy

An important step of the presented algorithm is the choice of the number of granules in the linguistic label sets. A typical approach is to use balanced linguistic label sets as they reflect various important properties with the added simplicity in terms of interpretation and ease of transformation. As a straightforward approach, the same number of labels can be used for each variables, optimizing the overall performance through changing a single parameter. This basic method can be extended by specifying optimal label sets for individual variables. For this purpose, (fuzzy) entropy-based discretization, a widely used approach in machine learning literature, can be applied. After partitioning the support interval of a variable into an optimal number of (not necessarily equal) sub-intervals, one can specify (triangular) fuzzy numbers corresponding to each interval reflecting the strength of the observation belonging to the sub-group. The main advantage of this approach is to correctly identify observations on the border of two sub-intervals: while traditional discretization approaches assign these points to only one group, with fuzzy membership, they can (partially) belong to two different subsets of the support.

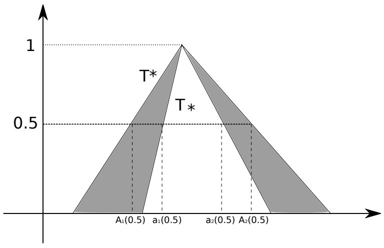

Figure 2: Triangular interval-valued fuzzy number

This basic approach is developed in more details for example in [37], and particularly using fuzzy entropy in combination with machine learning techniques in [6]. In this article, we propose to use interval-valued fuzzy sets for representing the membership of belonging to a (fuzzy) subset of the support of a variable in the data transformation step. Interval-valued fuzzy sets extend traditional fuzzy sets by introducing a second level of imprecision. In the traditional case, while the exact value of an object is not precisely identified, the membership function estimates the possibility of a specific value being the real value of the underlying object. In the interval-valued case, we assume even the membership values to be ill-known and hence represented by intervals from $[0,1]$. A typically used triangular interval-valued fuzzy number is depicted in Figure 2.

In the following, we describe how the original proposal presented in [6] can be extended to incorporate interval-valued fuzzy entropy based on the developments in [38, 37]. The main idea of the approach is to calculate the fuzzy entropy for a variable with respect to the output classes. In the algorithm, the number of labels in the label set is increased in each iteration. The optimal value of the number of individual labels is identified when the improvement in the entropy remains below a given threshold value.

In the following, we assume that there are n data points $\{x_1, x_2, ..., x_n\}$. In general, one can assume that there are m classes $C_1, C_2, ..., C_m$ to which the observations belong; in our application domain, two classes will be defined simplifying the presented general procedure: default and non-default loans. In the traditional fuzzy approach, the match degree D_j of the fuzzy set A with respect to class C_j is defined as

$$D_j = \frac{\sum_{x_i \in C_j} \mu_A(x_i)}{\sum_{x_i \in X} \mu_A(x_i)}.$$

The fuzzy entropy of the elements of the class C_j on A is the following:

$$H_j(A) = -D_j \log D_j.$$

Finally, the fuzzy entropy for the entire set is defined as

$$H(A) = \sum_{j=1}^{m} H_j(A).$$

When incorporating interval-valued fuzzy sets, one can proceed by defining the upper and lower match degrees as follows:

$$D_j^U = \frac{\sum_{x_i \in C_j} \mu_{A^U}(x_i)}{\sum_{x_i \in X} \mu_{A^U}(x_i)}, D_j^L = \frac{\sum_{x_i \in C_j} \mu_{A^L}(x_i)}{\sum_{x_i \in X} \mu_{A^L}(x_i)},$$

where A^U and A^L stand for the upper and lower membership function of the interval-valued fuzzy set, respectively. The final match degree is determined as the weighted average of the upper and lower match degrees, and the entropy is specified

$$H_j(A) = -D_j \log D_j.$$

A different approach, utilized in this paper, is to define match degree with an entropy measure directly used for interval-valued fuzzy sets, not by combining values obtained from the individual membership functions. In this paper, we use the entropy specified for interval-valued fuzzy numbers in [39]:

$$H(A) = \sum_{i=1}^{n} \left(\mu_A^U(x_i) - \mu_A^L(x_i) \right).$$

A pseudo code summarizing the steps of the proposed approach for one feature variable, F, is specified in the following:

```
procedure LINGUISTIC CLASSIFICATION WITH INTERVAL-VALUED FUZZY DATA TRANSFORMATION($x_i, C_i, F, n, m$)
    Specify the maximum number of allowed labels in the label set as $t$
    Record the initial entropy value and the optimal number of labels as $L = 1$
    for $k \in [1, t]$ do
        Divide the $x_i$ values into $t$ disjoint intervals such that the proportion of data-points is equal in all subsets and specify symmetrical interval-valued fuzzy numbers on each subinterval
        Calculate the match degrees and the overall entropy for the specified interval-valued set with respect to class information
        If the entropy is lower then the actual value of $E$, update $E$ and $L$
    end for
    return $E$, $L$ and the interval-valued fuzzy sets created in step $L$ of the loop
end procedure
```

4 Data collection and analysis

In this section we present the data used in the credit scoring classification task through a brief descriptive analysis. Afterwards, the most important concepts in evaluating classification results are introduced, followed by the results of the analysis.

4.1 Data description

In order to validate the usefulness of the linguistic data transformation approach, we utilized data from one of the most popular European P2P lending platforms, Bondora. The data is publicly available and updated on a daily basis [1].

The retrieved data consists of data points as loan applications recorded in the Bondora system in the time period between February 2, 2009 and June 13, 2017. The data comprises of demographic and financial information of borrowers, and the associated loan transactions. The data comprises of 37,008 observations. The data sample included 35.68% defaulted loans.

In the analysis, the following variables are used as predictors of the event of the loan defaulting:

- NewCreditCustomer: binary variable indicating whether the customer has prior credit history in Bondora (67.03% new customers)
- Age: age of the borrower (mean: 37.91 years)
- Gender: gender of the borrower (53.24% male, 46.6% female)
- Country: residency of the borrower(61.37% from Estonia, 20.3% from Spain, 17.53% from Finland)
- AppliedAmount: the amount of the loan applied for (mean: 2893.25 euro)
- Interest: maximum interest rate accepted in the loan application (mean: 18.65%)
- LoanDuration: the length of the loan term (mean: 41.3 months)
- UseOfLoan: purpose of the loan applied (categorical variable with numerous options, with business purpose being the most frequent)
- MaritalStatus: current marital status (the most frequent marital status is single with 33.05%)
- EmploymentStatus: current employment status (81.23% full-time employee)

[1] The dataset can be downloaded from https://www.bondora.com/en/public-reports

- IncomeTotal: total monthly income of the borrower (mean: 1881 euro)

4.2 Classification algorithms and performance measures

In this paper, the linguistic data transformation procedure is combined with the following widely used classification procedures:

- **Neural networks** [40]: replicating the information processing processes taking place in a brain by capturing patterns relevant in a phenomenon using a non-linear representation. While they offer a black box solution as one cannot obtain an explanation for a given output value, neural networks are one of the most widely used machine learning tools in supervised learning.

- **Classification trees** [41]: a sequence of splitting the support of selected variables in each step results in a decision tree in which every leaf is assigned one of the output classes based on the number of observations from each class ending up in the specific leaf.

- ***k*-nearest neighbor** (*kNN*) [42]: a class is assigned to a new observation based on the proportion of the output classes among the k data points closest to the new observation. The parameter k, the number of neighbors to be accounted for in the class assignment is specified in advance; the model is in general very sensitive to the choice of k.

The traditional way of evaluating binary classification problems is through the confusion matrix which is specified through the following four values: (i) true positive (TP), i.e. positive cases classified correctly as positive; (ii) true negative (TN), i.e. negative cases classified correctly as negative;(i) false positive (FP), i.e. negative cases classified incorrectly as positive;(iv) false negatives (FN), i.e. positive cases classified incorrectly negative.

A basic measure of prediction performance is accuracy $(TP + TN/(TP + TN + FP + FN))$: the proportion of correctly classified cases. A popular performance measure is the area under the curve (AUC), which is based on the Receiver Operating Characteristic (ROC). The ROC curve [43] depicts the true positive and false positive rates based on the threshold chosen in case of a probabilistic classifier output to determine the output class, and AUC measures the area under the ROC curve. The maximum value of AUC is 1, and the closer the value is to 1, the higher the probability is that the classifier assigns the right class to the data point.

In order to account for the slightly imbalanced class distribution, two additional evaluation approaches are considered in the analysis.

- Cohen's kappa (κ) [44]: relying on the concepts of expected and observed accuracy, this measure offers an evaluation metric less sensitive to class distribution differences as the result of taking into consideration the chance of random class assignments.

- Relative usefulness (U_{rel}) [45]: taking into consideration type I errors ($T_1 = FN/(FN + TP)$) and type II errors ($T_2 = FP/(TN + FP)$) and specifying the preference between making these errors as μ, the loss function can be defined as $L(\mu) = \mu T_1 P_1 + (1 - \mu) T_2 P_2$, with P_1 and P_2 denoting the probabilities of positive and negative cases, respectively. The absolute usefulness is calculated by comparing the loss function to the simple model of assigning every observation to the more frequent class, while Relative usefulness compares absolute usefulness with a model of loss function value 0.

4.3 Results

The data transformation and its combination with the three classification algorithms have been performed using the R software environment [46]. In particular we utilized the caret package [47] for training the models and identifying the optimal parameters in the classification procedures. We used 10-fold cross-validation procedure to ensure the validity of the results. The data transformation process was implemented by the authors and combined with the available tools in the caret package. The results are evaluated based on classification accuracy and the AUC measure. The results are reported in Tables 1, 2, and 3. In the three tables, we compared the classification performance of the algorithms with: (i) the original variable values, (ii) all the variables are transformed using the same, fixed amount of labels (9, 5 or 3), (iii) each variable is transformed using a unique number of labels based on traditional fuzzy entropy, and (iv) each variable is transformed using a unique number of labels based on interval-valued fuzzy entropy. Additionally, in the tables AUC values which significantly improve (at the 0.1 level using a t-test) over the traditional model without linguistic data transformation are marked with *.

On a general level, we can claim that the performance of the classification after the data is transformed using fuzzy sets is superior to the original results. This illustrates the benefits that the combination of machine learning and fuzzy sets theory can offer to the domain of risk assessment in P2P lending. The improvement proposed in this paper can been seen to offer the best performance

Measure	Acc	AUC	κ	U_{rel}
Original	78.5%	0.822	0.79	0.54
9 granules	76.4%	0.810	0.73	0.48
5 granules	79.8%	0.845*	0.84	0.57
3 granules	77.2%	0.820	0.78	0.51
Entropy	78.7%	0.824	0.80	0.54
New approach	**80.02%**	**0.855***	0.69	0.54

Table 1: Results based on neural network classification

Measure	Acc	AUC	κ	U_{rel}
Original	75.5%	0.782	0.75	0.52
9 granules	**77.4%**	**0.803** *	0.79	0.55
5 granules	75.8%	0.789	0.76	0.53
3 granules	72.2%	0.756	0.71	0.49
Entropy	76.7%	0.799*	0.77	0.53
New approach	77.2%	0.801*	0.79	0.54

Table 2: Results based on classification trees

Measure	Acc	AUC	κ	U_{rel}
Original	74.5%	0.749	0.69	0.47
9 granules	75.4%	0.766*	0.72	0.50
5 granules	73.8%	0.722	0.65	0.45
3 granules	72.2%	0.715	0.63	0.44
Entropy	75.7%	0.770*	0.73	0.51
New approach	**75.8%**	**0.770***	0.74	0.51

Table 3: Results based on *k*-nearest neighbors

in combination with neural networks and *k*-nearest neighbors. Regarding particular methods, as it can be expected, data transformation contributes to the smallest extent in case of *k*-nearest neighbors algorithm, as it relies on the distance between data points which is significantly affected by a transformation. The best performance is obtained by combining neural networks and interval-valued fuzzy entropy-based linguistic transformation, resulting in accuracy higher than 80% and *AUC* of 0.855.

5 Conclusions

The task of classifying loan applications in P2P lending relying on the likelihood of default is a highly relevant and important task in finance as P2P platforms are increasing in use continuously. In this article, we have introduced linguistic data transformation combined with machine learning into the domain of P2P lending. We successfully illustrated on a real life data set how the use of interval-valued linguistic labels and entropy-based discretization can improve the classification performance of traditional supervised learning methods in machine learning. The results show that particularly the combination of neural networks and linguistic data transformation based on entropy of interval-valued fuzzy sets is effective in identifying risky loan applications.

As only one data set has been utilized to assess the validity of the presented approach, the most important future research direction is to test the impact of linguistic data transformation on classification algorithms with different datasets. The presented observations cannot be directly generalized to other P2P lending platforms in various countries, but the use of cross-validation and robust performance measures ensures the potential of the methodology. Additionally, as there are numerous other existing definitions of entropy for interval-valued fuzzy sets, a numerical comparison of different measures is an important task to be done in future research.

References

[1] M. Klafft, “Online peer-to-peer lending: a lenders’ perspective,” pp. 371–375, 2008.

[2] C. Serrano-Cinca, B. Gutierrez-Nieto, and L. López-Palacios, “Determinants of default in p2p lending,” *PloS one*, vol. 10, no. 10, p. e0139427, 2015.

[3] M. Malekipirbazari and V. Aksakalli, “Risk assessment in social lending via random forests,” *Expert Systems with Applications*, vol. 42, no. 10, pp. 4621–4631, 2015.

[4] Y. Demyanyk and D. Kolliner, “Peer-to-peer lending is poised to grow,” *Federal Reserve Bank of Cleveland*, 2014.

[5] C. R. Everett, “Group membership, relationship banking and loan default risk: the case of online social lending,” 2010.

[6] J. A. Morente-Molinera, J. Mezei, C. Carlsson, and E. Herrera-Viedma, “Improving supervised learning classification methods using multi-granular linguistic modelling and fuzzy entropy,” *IEEE Transactions on Fuzzy Systems*, vol. PP, no. 99, pp. 1–1, 2016.

[7] J. Li, S. Hsu, Z. Chen, and Y. Chen, “Risks of p2p lending platforms in china: Modeling failure using a cox hazard model,” *The Chinese Economy*, vol. 49, no. 3, pp. 161–172, 2016.

[8] H. Wang and M. E. Greiner, “Prosper: the ebay for money in lending 2.0,” *Communications of the Association for Information Systems*, vol. 29, no. 1, p. 13, 2011.

[9] H. Yum, B. Lee, and M. Chae, “From the wisdom of crowds to my own judgment in microfinance through online peer-to-peer lending platforms,” *Electronic Commerce Research and Applications*, vol. 11, no. 5, pp. 469–483, 2012.

[10] X. Lin, X. Li, and Z. Zheng, “Evaluating borrowers default risk in peer-to-peer lending: evidence from a lending platform in china,” *Applied Economics*, pp. 1–8, 2016.

[11] R. Emekter, Y. Tu, B. Jirasakuldech, and M. Lu, “Evaluating credit risk and loan performance in online peer-to-peer (p2p) lending,” *Applied Economics*, vol. 47, no. 1, pp. 54–70, 2015.

[12] E. Lee and B. Lee, “Herding behavior in online p2p lending: An empirical investigation,” *Electronic Commerce Research and Applications*, vol. 11, no. 5, pp. 495–503, 2012.

[13] L. Gonzalez and Y. K. Loureiro, “When can a photo increase credit? the impact of lender and borrower profiles on online peer-to-peer loans,” *Journal of Behavioral and Experimental Finance*, vol. 2, pp. 44–58, 2014.

[14] D. Chen, F. Lai, and Z. Lin, “A trust model for online peer-to-peer lending: a lenders perspective,” *Information Technology and Management*, vol. 15, no. 4, pp. 239–254, 2014.

[15] M. Lin, N. R. Prabhala, and S. Viswanathan, “Judging borrowers by the company they keep: friendship networks and information asymmetry in online peer-to-peer lending,” *Management Science*, vol. 59, no. 1, pp. 17–35, 2013.

[16] A. Khashman, “Credit risk evaluation using neural networks: Emotional versus conventional models,” *Applied Soft Computing*, vol. 11, no. 8, pp. 5477–5484, 2011.

[17] D. J. Hand and W. E. Henley, “Statistical classification methods in consumer credit scoring: a review,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 160, no. 3, pp. 523–541, 1997.

[18] T.-S. Lee and I.-F. Chen, “A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines,” *Expert Systems with Applications*, vol. 28, no. 4, pp. 743–752, 2005.

[19] T. Bellotti and J. Crook, “Support vector machines for credit scoring and discovery of significant features,” *Expert Systems with Applications*, vol. 36, no. 2, pp. 3302–3308, 2009.

[20] H. Ince, B. Aktan, *et al.*, “A comparison of data mining techniques for credit scoring in banking: A managerial perspective,” *Journal of Business Economics and Management*, no. 3, pp. 233–240, 2009.

[21] D. West, “Neural network credit scoring models,” *Computers & Operations Research*, vol. 27, no. 11, pp. 1131–1152, 2000.

[22] B. Twala, “Multiple classifier application to credit risk assessment,” *Expert Systems with Applications*, vol. 37, no. 4, pp. 3326–3336, 2010.

[23] B. Baesens, T. Van Gestel, M. Stepanova, D. Van den Poel, and J. Vanthienen, “Neural network survival analysis for personal loan data,” *Journal of the Operational Research Society*, vol. 56, no. 9, pp. 1089–1098, 2005.

[24] J. Banasik, J. N. Crook, and L. C. Thomas, “Not if but when will borrowers default,” *Journal of the Operational Research Society*, pp. 1185–1190, 1999.

[25] A. Byanjankar, M. Heikkilä, and J. Mezei, “Predicting credit risk in peer-to-peer lending: A neural network approach,” in *IEEE Symposium Series on Computational Intelligence, SSCI 2015, Cape Town, South Africa, December 7-10, 2015*, pp. 719–725, 2015.

[26] D. Jiang and X. Li, "The study on the credit risk assessment of borrower in p2p network of china," in *Proceedings of the Tenth International Conference on Management Science and Engineering Management*, pp. 1619–1630, Springer, 2017.

[27] R. Vedala and B. R. Kumar, "An application of naive bayes classification for credit scoring in e-lending platform," in *Data Science & Engineering (ICDSE), 2012 International Conference on*, pp. 81–84, IEEE, 2012.

[28] Y. Jin and Y. Zhu, "A data-driven approach to predict default risk of loan for online peer-to-peer (p2p) lending," in *Communication Systems and Network Technologies (CSNT), 2015 Fifth International Conference on*, pp. 609–613, IEEE, 2015.

[29] C. Serrano-Cinca and B. Gutiérrez-Nieto, "The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (p2p) lending," *Decision Support Systems*, vol. 89, pp. 113–122, 2016.

[30] W.-Y. Lin, Y.-H. Hu, and C.-F. Tsai, "Machine learning in financial crisis prediction: a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 421–436, 2012.

[31] M.-Y. Cheng and N.-D. Hoang, "Evaluating contractor financial status using a hybrid fuzzy instance based classifier: Case study in the construction industry," *IEEE Transactions on Engineering Management*, vol. 62, no. 2, pp. 184–192, 2015.

[32] A. Ghandar and Z. Michalewicz, "An experimental study of multi-objective evolutionary algorithms for balancing interpretability and accuracy in fuzzy rulebase classifiers for financial prediction," in *Computational Intelligence for Financial Engineering and Economics (CIFEr), 2011 IEEE Symposium on*, pp. 1–6, IEEE, 2011.

[33] P. Hajek and R. Henriques, "Mining corporate annual reports for intelligent detection of financial statement fraud – a comparative study of machine learning methods," *Knowledge-Based Systems*, vol. 128, pp. 139–152, 2017.

[34] D. J. Hand, H. Mannila, and P. Smyth, *Principles of data mining*. MIT press, 2001.

[35] E. Hüllermeier, "Does machine learning need fuzzy logic?," *Fuzzy Sets and Systems*, vol. 281, pp. 292–299, 2015.

[36] S.-M. Zhou and J. Q. Gan, "Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modelling," *Fuzzy Sets and Systems*, vol. 159, no. 23, pp. 3091–3131, 2008.

[37] H.-M. Lee, C.-M. Chen, J.-M. Chen, and Y.-L. Jou, "An efficient fuzzy classifier with feature selection based on fuzzy entropy," *IEEE Trans. Syst. Man Cyber., Part B: Cybernetics*, vol. 31, no. 3, pp. 426–432, 2001.

[38] J. Mezei, J. A. Morente-Molinera, and C. Carlsson, "Feature selection with fuzzy entropy to find similar cases," in *Advance Trends in Soft Computing*, pp. 383–390, Springer, 2014.

[39] P. Burillo and H. Bustince, "Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets," *Fuzzy sets and systems*, vol. 78, no. 3, pp. 305–316, 1996.

[40] A. Bahrammirzaee, "A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems," *Neural Computing and Applications*, vol. 19, no. 8, pp. 1165–1195, 2010.

[41] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and regression trees*. CRC press, 1984.

[42] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[43] A. P. Bradley, "The use of the area under the roc curve in the evaluation of machine learning algorithms," *Pattern recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.

[44] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[45] P. Sarlin, "On policymakers loss functions and the evaluation of early warning systems," *Economics Letters*, vol. 119, no. 1, pp. 1–7, 2013.

[46] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.

[47] M. Kuhn, "Caret package," *Journal of Statistical Software*, vol. 28, no. 5, pp. 1–26, 2008.