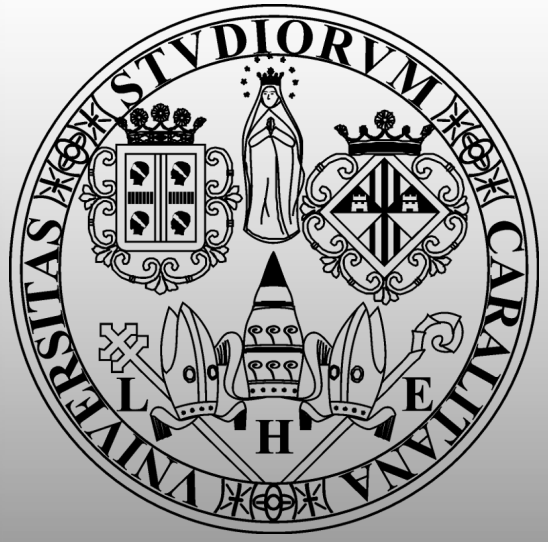


An Entropy Based Algorithm for Credit Scoring

Roberto Saia and Salvatore Carta
Department of Mathematics and Computer Science
University of Cagliari, Italy

International Conference on Research and Practical Issues of Enterprise Information Systems
13÷14 December 2016 - Vienna, Austria

Outline

- **Overview**
  Introduction ▪ State of the Art ▪ Limits
- **Intuition**
  Idea ▪ Usage ▪ Advantages
- **Implementation**
  Local Entropy ▪ Global Entropy ▪ Classification
- **Results**
  Experiments ▪ Conclusions ▪ Future Work

OVERVIEW

Introduction

- The main aim of **credit scoring models** is the classification of the loan applications as *reliable* or *unreliable*

Introduction

- They are usually **designed** on the basis of the past loan applications
- They represent a **crucial** instrument for the financial operators

State of the Art

- Many **techniques** have been designed to address the *Credit Scoring* problem, by exploiting several models, such as:
  - *Statistical* and *Data Mining*[1]
  - *Linear Discriminant*[2]
  - *Logistic Regression*[3]
  - *Neural Networks*[4]
  - *Genetic Program*[5]
  - *Decision-tree*[6]
  - … and many more

State of the Art

- Regardless of the adopted technique, the definition of effective *Credit Scoring* models represents a **hard challenge** due to several factors

- The most important of which is the **Imbalanced class distribution**[7]

Limits

- The **Imbalanced class distribution** exists because the number of *default* cases (*non-refunded loans*) is **much smaller** than that of the *non-default* ones (*refunded loans*)
- This reduces the **effectiveness** of the most widely used approaches[8]

NON-DEFAULT CASES **VS** DEFAULT CASES

INTUITION

Idea

- Since the **Shannon Entropy** (from now on referred simply as *entropy*) increases as the data becomes equally probable and decreases otherwise
- We exploit this concept in order to evaluate a new loan application (from now on referred simply as *instance*)

Idea

- Since the **Shannon Entropy** (from now on referred simply as *entropy*) increases as the data becomes equally probable and decreases otherwise
- We exploit this concept in order to evaluate a new loan application (from now on referred simply as *instance*)

The idea is to measure the *entropy* in the set of the past non-default instances, **before** and **after** we added to it a new instance to evaluate

Usage

- We calculate the **Local Maximum Entropy** (LME) which give us information about the entropy achieved by each **instance*** feature in the set of non-default cases

  (*Differences between instances in terms of single features*)

- We also calculate the **Global Maximum Entropy** (GME) which represents a meta-feature based on the *integral* of the area under the curve of the LME values

  (*Differences between instances in terms of all features*)

* It contains several information about the applicant, e.g., **gender**, **age**, **marital status**, **income**, **loan amount**, **other loans**, etc.

Usage

- Finally we define the **Difference in Maximum Entropy** (DME) approach that allows us to classify a new instance as *accepted* o *rejected*
- It was performing by exploiting the previous *Local* and *Global Maximum Entropy* (**LME** and **GME**) information
- It is based on the observation of **how changes** the **LME** and **GME** values when we add a *new* (*unevaluated*) instance to the set of the *past* non-default instances

Usage

A **larger** entropy indicates that the added new instance contains similar data in the features, increasing their level of equiprobability

Otherwise, a **lower** entropy indicates different data in the instance, and we consider it as a potential *default case*

Advantages

- The adoption of the proposed entropy-based approach presents several advantages:
  - It **overcomes** the imbalance class distribution issue by using only a class of data (*non-default instances*)
  - The use of only *non-default* cases, allows us to operate in a **proactive** way
  - It reduces the problems related to the **cold-start** issue[9]

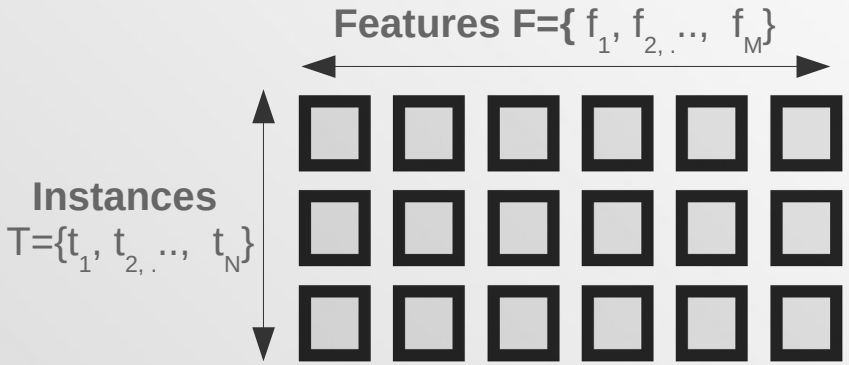
IMPLEMENTATION

Implementation

Step 1 of 3 – Local Maximum Entropy

We define a set **Λ** that contains the maximum value of entropy **H** achieved by each feature **f** in the non-default instances set **T**

$$\Lambda = \{\lambda_1 = max(H(f_1)), \lambda_2 = max(H(f_2)), \ldots, \lambda_M = max(H(f_M))\}$$

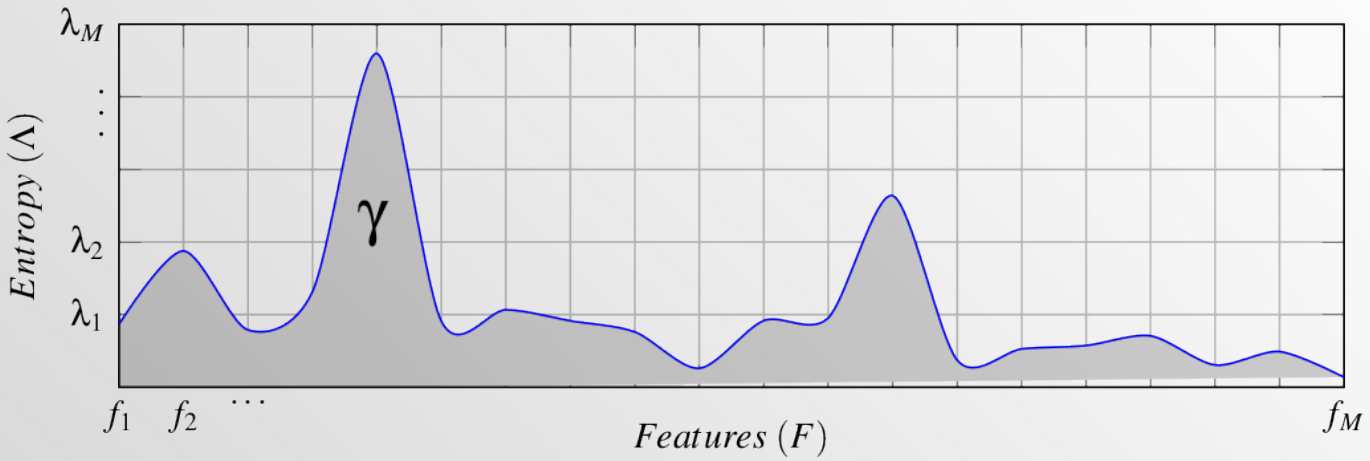


Implementation

Step 2 of 3 – Global Maximum Entropy

Subsequently, we calculate the integral γ of the area under the curve given by the previous set Λ

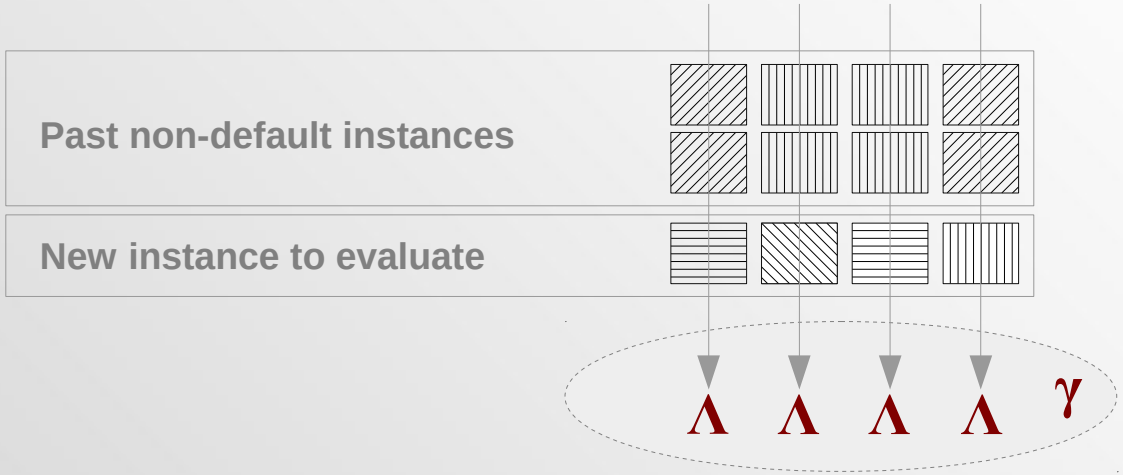
$$\gamma = \int_{\lambda_1}^{\lambda_M} f(x)\,dx \approx \frac{\Delta x}{2} \sum_{k=1}^{|\Lambda|} \left(f(x_{k+1}) + f(x_k)\right) \quad \textbf{with} \quad \Delta x = \frac{(\lambda_M - \lambda_1)}{|\Lambda|}$$



Implementation

Step 3 of 3 – Difference in Maximum Entropy

A new instance is classified as *accepted* or *rejected* on the basis of the Λ and γ values, compared them before and after that we added to the set of non-default past instances **T** the instance to evaluate



Lower values of Λ and γ indicate that the new instance contains different data in its features

RESULTS

Experiments

The **experiments** were performed by following the *k-fold cross validation* criterion, with *k=10*

- We used two **real-world datasets**[10] characterized by a *strong imbalanced* data distribution
- As competitor we use **Random Forests**, one of the most performing state-of-the-art approaches in this field[11]

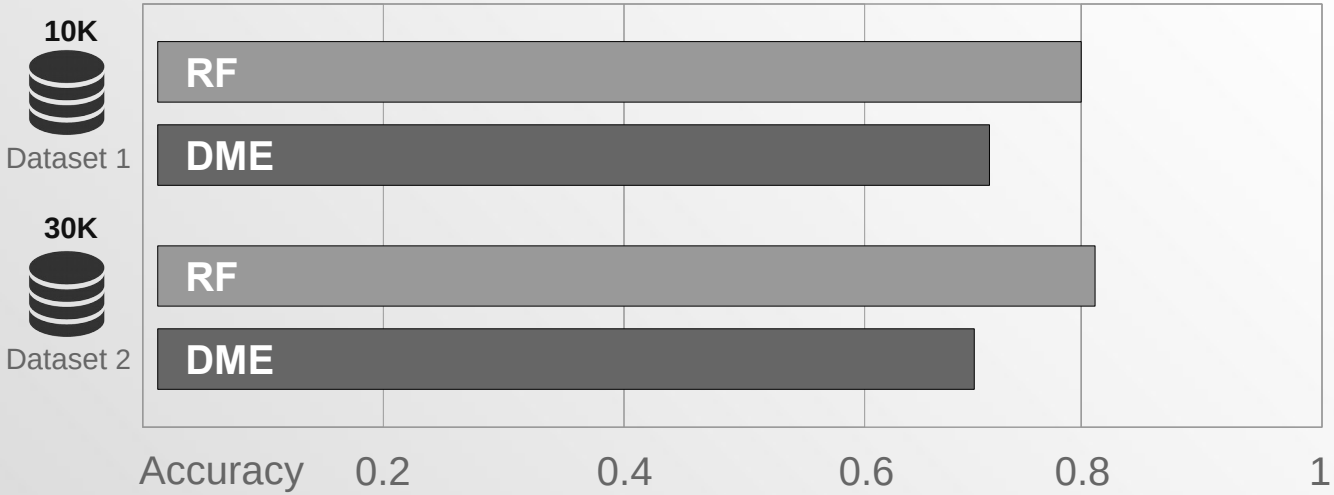
Experiments

- The *first two* experiments show the *general performance* achieved by our **DME** approach and Random Forests, in terms of ***Accuracy*** and ***F-score***
- The *third* experiment is instead aimed to evaluate our predictive model in terms of **AUC***

* Area Under the (Receiver Operating Characteristic) Curve

Experiments

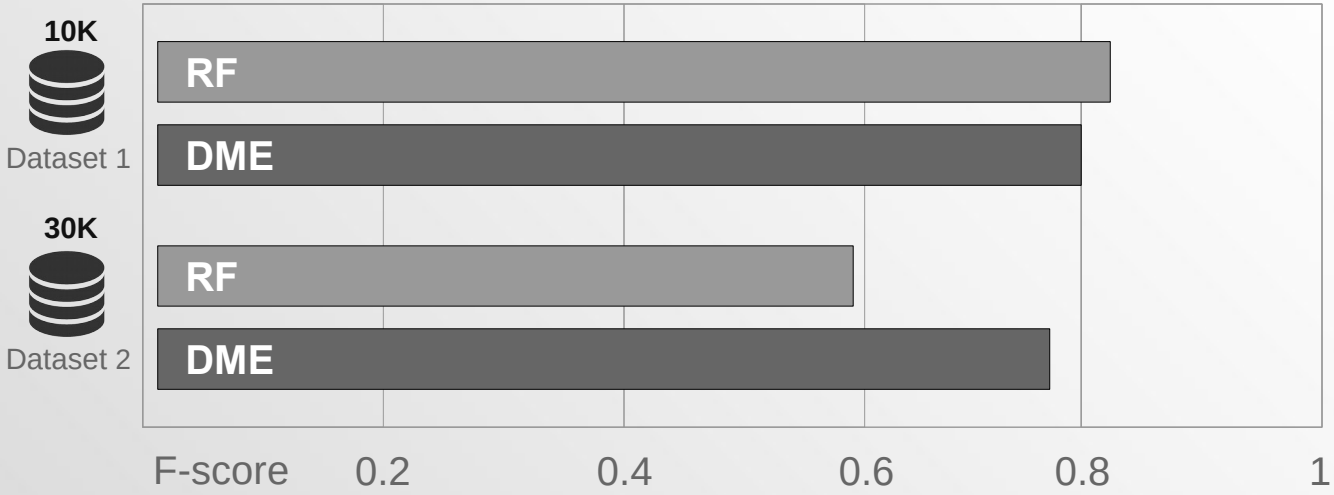
- The general performance of our **DME** approach in terms of *accuracy* is very close to those of Random Forests (**RF**)



- In spite of the fact that we did **not use** any past default instance to train our model

Experiments

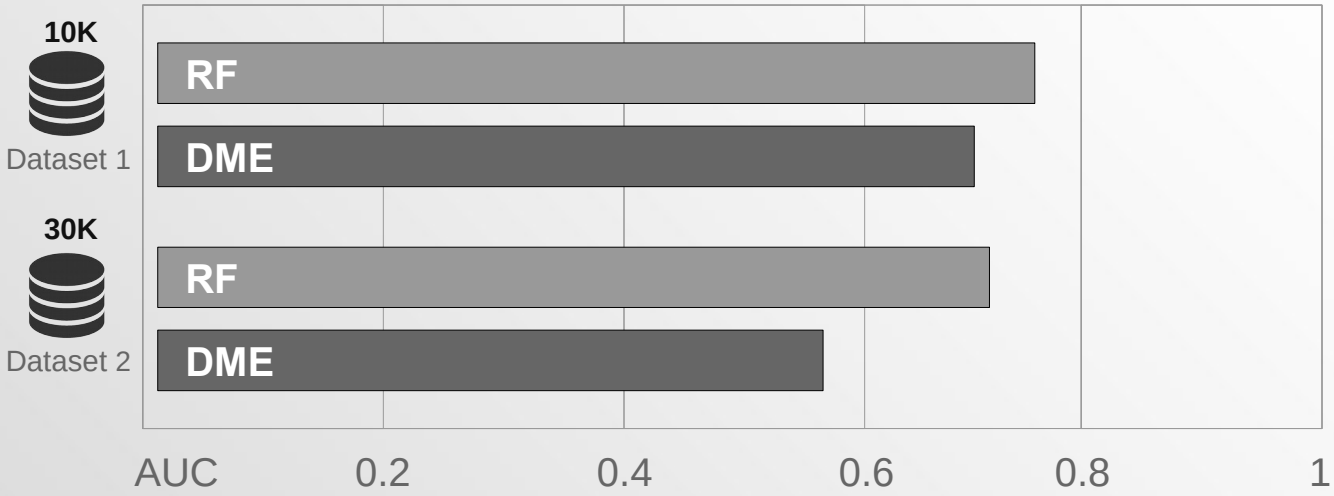
- Also in terms of *F-score*, the general performance of our **DME** approach is very close to those of Random Forests (**RF**)



- Differently from **RF**, the effectiveness of **DME** increases with the number of instances

Experiments

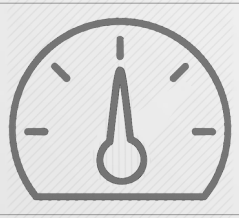
- The last set of experiments measures the performance in terms of **AUC** (*Area Under ROC Curve*)



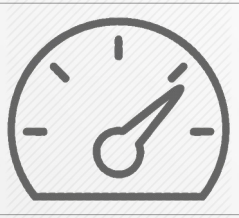
- Also in this case, the performance of our *predictive model* is very close to those of **RF**

Conclusions

- We proposed a novel approach of ***Credit Scoring*** that exploits an *entropy-based* criterion
- The experiments show **two** main results:



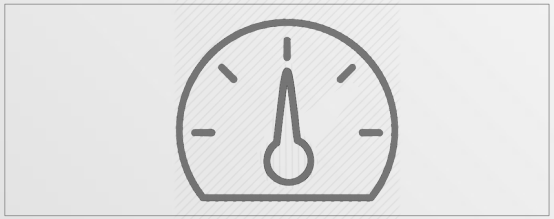
By using a small dataset, our performance is **similar** to one of the most performing state-of-the-art approach (*RF*), in spite of the fact that we do not use the past default instances



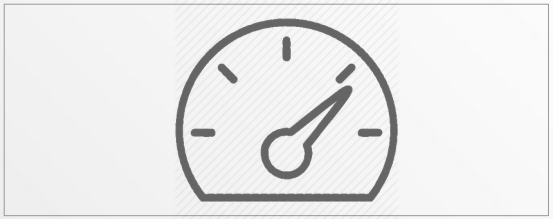
By using a larger dataset, our approach **outperforms** its competitor (*RF*), while continuing to not use past default instances during the training

Conclusions

- We proposed a novel approach of ***Credit Scoring*** that exploits an *entropy-based* criterion
- The experiments show **two** main results:



Proactivity



Improvement

Future Work

- A possible follow up of this work could be a new series of experiments performed by using **both** the *past default* and *non-default* cases
- This will allow us to investigate about the advantages and disvantages related to a **non-proactive** approach
- We also consider to test our approach in **heterogeneous** scenarios that involve different types of financial data

References

1) Chen, S. Y. and Liu, X. (2004). *The contribution of data mining to information science*. Journal of Information Science, 30(6):550–558.

2) Reichert, A. K., Cho, C.-C., and Wagner, G. M. (1983). *An examination of the conceptual issues involved in developing credit-scoring models*. Journal of Business & Economic Statistics, 1(2):101–114.

3) Henley, W. E. (1994). *Statistical aspects of credit scoring*. PhD thesis, Open University.

4) Desai, V. S., Crook, J. N., and Overstreet, G. A. (1996). *A comparison of neural networks and linear scoring models in the credit union environment*. European Journal of Operational Research, 95(1):24–37.

5) Ong, C.-S., Huang, J.-J., and Tzeng, G.-H. (2005). *Building credit scoring models using genetic programming*. Expert Systems with Applications, 29(1):41–47.

6) Davis, R., Edelman, D., and Gammerman, A. (1992). *Machine-learning algorithms for credit-card applications*. IMA Journal of Management Mathematics, 4(1):43–51.

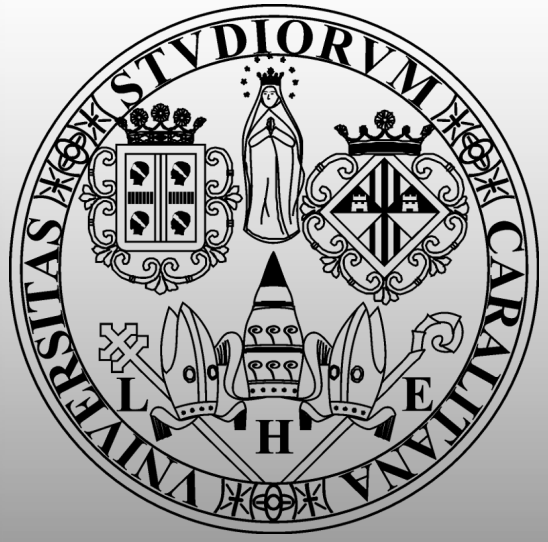
7) He, H. and Garcia, E. A. (2009). *Learning from imbalanced data*. IEEE Trans. Knowl. Data Eng., 21(9):1263–1284.

8) Japkowicz, N. and Stephen, S. (2002). *The class imbalance problem: A systematic study*. Intell. Data Anal., 6(5):429–449.

9) Attenberg, J. and Provost, F. J. (2010). *Inactive learning?: difficulties employing active learning in practice*. SIGKDD Explorations, 12(2):36–41.

10) *Default of Credit Card Client* and *German Credit*. UCI Repository of Machine Learning Databases, ftp://ftp.ics.uci.edu/pub/machine-learning-databases/.

11) Lessmann, S., Baesens, B., Seow, H., and Thomas, L. C. (2015). *Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research*. European Journal of Operational Research, 247(1):124–136.



An Entropy Based Algorithm for Credit Scoring

THANK YOU FOR YOUR ATTENTION

International Conference on Research and Practical Issues of Enterprise Information Systems
13÷14 December 2016 - Vienna, Austria