

REZUMAT

În cadrul cercetării de după anul 2007, ne-am propus elaborarea modelelor de evaluare a performanțelor instituțiilor financiare din România pe baza tehnicilor de inteligență computațională sau tehnicilor de Data Mining.

Am împărțit contribuțiile la cercetarea privind utilizarea metodelor de Inteligență Computațională (IC) pentru evaluarea performanțelor instituțiilor financiare nebancare (IFN-urilor) în două părți în funcție de sarcina de Data Mining (DM) îndeplinită.

Contribuții la metodele de IC utilizate pentru sarcina de DM de clasterizare:

- investigarea și compararea diferitelor metode de IC (Self Organising Map - SOM, C-Means, Fuzzy C-Means - FCM, Weighting FCM) pentru sarcina de clasterizare;
- propunerea metodologiei în două faze de obținere a clasterelor de performanță a IFN-urilor pe baza algoritmului SOM: în prima fază, aplicarea algoritmului SOM și obținerea unui număr mare de clastere "brute", iar în faza a doua gruparea clasterelor "brute" în supra-clastere sau clastere "reale" pe baza metodei U-matrix propusă de Ultsch (1993);
- propunerea unor elemente de noutate referitoare la aplicarea tehnicilor fuzzy de clasterizare: caracterizarea automată a fiecărui claster pe baza variabilelor calitative construite prin transformarea variabilelor cantitative (efectul de paraghie, credite/active, rata de rentabilitate) cu ajutorul termenilor calitativi: foarte scăzut, scăzut, mediu, ridicat, foarte ridicat și identificarea observațiilor "nesigure" pe baza gradelor de apartenență a observațiilor la diferitele clastere;
- introducerea algoritmului Weighting FCM care alocă mai bine observațiile "nesigure" în clastere în comparație cu algoritmi SOM și FCM clasic;
- propunerea unei metodologii de comparare a tehnicilor de clasterizare;
- utilizarea rezultatelor clustering-ului pentru dezvoltarea în continuare a unor modele de clasificare (idee sugerată și de alți autori: Witten și Franck, 2000; Costa, 2000; De Andres, 2001) poate fi considerată ca o implementare a propunerilor din literatură de specialitate. Cu toate acestea, deși această combinație (clasterizare+clasificare) a fost sugerată și în altă parte, doar câteva studii au implementat-o eficient.

Contribuții la metodele de IC utilizate pentru sarcina de DM de clasificare:

- investigarea și compararea diferitelor modele de clasificare hibride. Modelele hibride se bazează pe o metodologie în două faze: prima fază constă în aplicarea unei tehnici de clasterizare și construirea variabilei "clasa de performanță", iar în a doua fază se modelează legătura dintre variabila "clasa de performanță" și variabilele de intrare prin construirea clasificatorilor hibridi;
- introducerea unei proceduri pentru determinarea celei mai adecvate arhitecturi a unei rețele neuronale pentru un set de date de antrenare dat;
- noi metode de reantrenare a rețelelor neuronale perceptron multistrat (algoritmi genetici);
- introducerea unei metode standard de comparare a diferitelor abordări ale sarcinii DM de clasificare. Această metodă standard de comparare constă în următorii pași metodologici:
 1. verificarea cerințelor referitoare la dimensiunea setului de date și tratamentul datelor lipsă;
 2. elaborarea modelului de clasificare (bazat pe o anumită tehnică);
 3. testarea modelului (pe baza ratei de acuratețe de antrenare);
 4. interpretarea rezultatelor;
 5. validarea modelului (pe baza ratelor de acuratețe de testare)

În cadrul articolelor care abordează problema elaborării modelelor de clasificare am prezentat unele modalități practice de setare a parametrilor pentru fiecare abordare a sarcinii de clasificare. După cum sugerează și De Andres (2001), am menționat că alegerea modelului hibrid de clasificare depinde de context și de problema ce urmează a fi abordată, în cazul nostru, evaluarea comparativă a performanțelor IFN-urilor.

Implicații manageriale:

- caracterizarea sectorului IFN-urilor pe baza tehnicilor de DM descrise mai sus și identificarea traiectoriei fiecărei IFN, precum și previzionarea evoluției viitoare a acesteia;
- propunerea de modele alternative la cele existente (CAAMPL) referitoare la evaluarea performanțelor instituțiilor financiare;
- propunerea setului de 12 indicatori pentru caracterizarea performanțelor IFN-urilor din punct de vedere al adecvării capitalului, calității activelor și profitabilității;
- colectarea datelor necesare calculării acestor indicatori. Astfel, setul de date îmbunătățit cuprinde 22 de indicatori în valoare absolută colectați trimestrial între 2007 și 2012 (21 trimestre) pentru IFN-urile din Registrul Special (aprox. 68), ajungându-se, în total, la un număr de aproximativ 990 de observații. Pe baza unei părți a acestor indicatori în valoare absolută au fost calculați 12 indicatori de performanță.

SUMMARY

In our research after the year 2007, we developed models for assessing comparatively the performance of financial institutions in Romania based on some computational intelligence or Data Mining techniques.

I have divided the contribution to the research of using the Computational Intelligence (CI) methods for assessing the non-banking financial institutions' (NFIs') performance in two categories depending on which Data Mining (DM) task is addressed.

The contributions to the CI methods used to address the DM clustering task:

- the investigation and comparison of different CI methods (Self Organising Map - SOM, C-Means, Fuzzy C-Means - FCM, Weighting FCM) for the clustering task;
- the proposal of the two-phase methodology for obtaining the NFIs' performance clusters based on the SOM algorithm: in the first phase the application of SOM algorithm obtaining a certain number of "raw" clusters, and, in the second phase, the grouping of the "raw" clusters in supra-clusters or "real" clusters based on the U-matrix method proposed by Ultsch (1993);
- the proposal of new approaches with respect to the application of fuzzy clustering techniques: the automate characterization of each cluster based on the qualitative variables generated by transforming the quantitative ones (leverage, loans/assets, return on assets) using qualitative terms: very low, low, average, high, very high and the identification of „uncertain” observations based on the membership degrees of the observations in each cluster;
- the introduction of Weighting FCM which better allocates the „uncertain” observations into the clusters compared to the SOM and classic FCM algorithms;
- the proposal of a new methodology to assessing comparatively the clustering techniques;
- the use of clustering results to further developing some classification models (the idea was suggested elsewhere, e.g.: Witten și Franck, 2000; Costa, 2000; De Andres, 2001) could be considered as an implementation of the approaches followed in the literature. poate fi considerată ca o implementare a propunerilor din literature de specialitate. However,

eventhough this combination (clustering+classification) has been suggested elsewhere, few studies have efficiently implemented it.

The contributions to the CI methods used to address the DM classification task:

- the investigation and comparison of different hybrid classification models. The hybrid models are based upon a two-stage methodology: the first stage consists of the application of a clustering technique in order to build the “performance class” variable, and in the second stage we model the relationship between the “performance class” variable and the input variables by constructing hybrid classifiers;
- the introduction of a new procedure for determining the most adequate neural network architecture for a given training dataset;
- new methods for re-training the multi-layer perceptron neural networks (the genetic algorithms);
- the introduction of a standard method for assessing comparatively the different approaches for the DM classification task. This standard method consists of the following methodological steps:
 1. the verification of dataset dimensionality and missing values requirements;
 2. building the classification model (based on some technique);
 3. model testing (based on the training accuracy rates);
 4. the interpretation of results;
 5. model validity (based on the testing accuracy rates)

Within the papers that tackle the problem of building classification models I have presented some practical ways to set the parameters in the case of each approach for performing the classification task. As it is suggested in De Andres (2001), I mention that the choice of the hybrid classification model is context and problem-dependent, in our case, the problem of assessing comparatively the NFIs’ performance.

Managerial implications:

- the characterization of the NFIs’ sector based on the DM techniques described above and the identification of each NFI’s trajectory, as well as the prediction of its future evolution;
- the proposal of some alternative models to the existing ones (CAAMPL) regarding the assessment of financial institutions’ performance;
- the proposal of a set of 12 financial ratios for characterizing the NFIs’ performance from the perspective of capital adequacy, assets’ quality and profitability;
- the collection of necessary data for calculating these financial ratios. Thus, the improved dataset consists of 22 absolute-valued indicators collected quarterly between 2007 and 2012 (21 quarters) for the NFI from the Special Register (approx. 68), totalling a number of approximately 990 observations. Based on a part of these absolute-valued indicators, I have calculated 12 performance financial ratios.