ENGINE Centre, Wrocław University of Technology
Wyb. Wyspiańskiego 27, building A1, room 203k
50-370 Wrocław, Poland
tel.: +48 71 320 3453, email: engine@pwr.wroc.pl

WT2.18b: University of Granada

Report

Activities performed during the visit

in University of Granada, Granada, Spain

period: 29.09.2014 -20.11.2014

author: Bartosz Krawczyk

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 316097 [ENGINE].

ENGINE Centre, Wrocław University of Technology
Wyb. Wyspiańskiego 27, building A1, room 203k
50-370 Wrocław, Poland
tel.: +48 71 320 3453, email: engine@pwr.wroc.pl

Personal Information

Mr./Ms. Bartosz Krawczyk, faculty member of Department of Systems and Computer Networks, Faculty of Electronics, Wroclaw University of Technology, Poland visited University of Granada, Granada, Spain in the period from 29.09.2014 to 20.11.2014 in

order to carry out research and training activities in the field of machine learning, one-class classification and imbalanced data.

Information about Seminars

The seminar presentation was organized on 14.10.2014

the date

It was entitled:

Clustering-Based Ensemble of One-Class Classifiers

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 316097 [ENGINE].

ENGINE Centre, Wrocław University of Technology
Wyb. Wyspiańskiego 27, building A1, room 203k
50-370 Wrocław, Poland
tel.: +48 71 320 3453, email: engine@pwr.wroc.pl

Description of scientific activities

(Please describe value added to the ENGINE project i.e. new knowledge, new skills with respect to the objectives of the project, the assigned common area of future cooperation with the partner, plans for common research, projects, publications and how it could be used in the scope of ENGINE)

During this stay new research topics for collaboration were discussed. They can be summarized under the following points:

- Usage of dynamic classifier selection with novel threshold-based pruning for decomposing multi-class problems with one-class classifier ensembles. A multi-class problem can be tackled with a divide and conquer approach, by reducing it to a number of simpler sub-problems. Transforming a multi-class dataset into a set of binary problems is a popular technique to deal with a high number of classes. Together with our partner at University of Granada, we are developing novel methods for decomposing multi-class problems with one-class classifiers. In this scenario, each class has assigned a given one-class classifier that learns its properties. We lose the information about other classes, but gain high robustness to noise, class imbalance and complex distributions. During this stay, a new method was developed based on a dynamic selection of classifiers. This is based on an assumption, that not in every case we need all of the classifiers from the pool. For each incoming object, we select only a subset of classifiers corresponding to the classes with highest a priori probabilities. By this, we are able to reduce the complexity of the classification task and improve the robustness of our system in comparison to using all of possible class combinations.
- Novel approach for handling big data with one-class classifiers, based on evolutionary prototype reduction in order to discard objects with low importance for formed classifier. Processing datasets with a high number of instances can be time-consuming and complex for standard classifiers. Additionally, not all of training data can contribute new, useful information to the constructed classifier. Based on this observation, a new method was proposed for prototype reduction in one-class classification problems. We use an efficient differential evolution algorithm in order to select valuable samples for training single-class classifiers. By using a fast metaheuristic, we can process very large datasets in an reasonable time. To evaluate the subset of examples, we use a dedicated classifier consistency measure. It is fully unsupervised and does not require an access to any counterexamples (which is impossible in one-class scenarios). Experimental analysis had shown, that with this approach we are able to build efficient one-class classifiers on highly reduced datasets (achieving 60-90% of data reduction without the loss of final accuracy).
- Application of non-parametric statistical tests for pruning ensembles, in order to select the smallest pool of classifiers that are statistically significantly better than other candidates. Currently, the ensemble selection is done according to some selected measure (accuracy / AUC / diversity etc.) using maximum rule i.e., the combination of classifiers that return the best performance is selected as the final

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 316097 [ENGINE].

ENGINE Centre, Wrocław University of Technology
Wyb. Wyspiańskiego 27, building A1, room 203k
50-370 Wrocław, Poland
tel.: +48 71 320 3453, email: engine@pwr.wroc.pl

combined classifier. However, differences between different combinations can be very small. During the collaboration with University of Granada, we worked on using non-parametric statistical analysis in order to improve the interpretability of ensemble pruning procedure. If two or more combinations return similar results according to the statistical analysis, we propose to select the setting with a lower number of classifiers. Experimental results have proven that our approach allows us to maintain the accuracy of previously used methods, while constructing significantly smaller ensembles.

- Tackling real-life medical problems (early breast cancer detection) with an ensemble of classifiers dedicated to the imbalanced classification domain. We have addressed a real-life problem in cooperation with University of Granada and Hospitals from Wroclaw, aiming at creating a fully automatic decision support system for pathologists. It requires only an raw image from fine needle biopsy to work. Fully automatic algorithms conduct segmentation, nuclei detection, feature extraction and finally, the classification step. However the problem is difficult due to the high imbalance ratio between the groups of patients – we have much less examples coming from the highest malignancy case. That is why, an extra attention must be paid to this class. We used a novel EUSBoost ensemble, that combines boosting with evolutionary undersampling in each iteration. By this, each classifier is trained on a balanced set of objects, and most difficult objects get an extra attention due to the properties of boosting scheme. Experiments carried out on a large dataset collected by authors show, that the proposed approach delivers excellent performance according to several measures (sensitivity / AUC / F-measure) and outperforms 15 state-of-the-art algorithms dedicated to the imbalanced classification.

During the stay one JCR journal publication was prepared and submitted:

- **Krawczyk B.**, Galar M., Jeleń Ł. & Herrera F., Evolutionary Undersampling Boosting for Imbalanced Classification of Breast Cancer Malignancy. *Artificial Intelligence in Medicine*. **(Impact Factor 1.356)**

while three other are currently being finished and will be submitted along December.

Information referring to the intellectual property

(the generally binding low in this area in the visited country and procedures of patenting);

In Spain, a patent is obtained by filling an application as referred to in Article 21 of Law No. 11 of 20 March 1986, and set to one of three institutions:

- (a) Directly at the Industrial Property Registry.
- (b) At the Provincial Delegations of the Ministry of Industry.
- (c) At those Autonomous Regional Communities recognized as competent.

The application requires the following documents:

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 316097 [ENGINE].

ENGINE Centre, Wrocław University of Technology
Wyb. Wyspiańskiego 27, building A1, room 203k
50-370 Wrocław, Poland
tel.: +48 71 320 3453, email: engine@pwr.wroc.pl

- (a) A request for a patent of invention.
- (b) The name and surname(s) or company name, nationality, and address of the applicant. Legal persons shall be identified by their corporate name or a name in accordance with the legal regulations governing them. Natural persons shall state their national identity card number.
- (c) The title of the invention, free of fanciful language, setting out the technical designation for the invention as clearly and succinctly as possible, and in accordance with the claims.
- (d) A designation of the inventor or inventors; if the applicant is not the inventor or sole inventor, a statement attesting to how rights on the invention have been acquired shall be submitted.
- (e) A list of accompanying documents submitted with the application.

The description included with the patent application must contain:

- (a) The title of the invention as it appears on the application form.
- (b) Mention of the technical field to which the invention relates.
- (c) An indication of the prior art before the priority date, known to the applicant, which may be useful in understanding the invention and drawing up the report on the state of the art citing, to the extent possible, the documents reflecting the prior art.
- (d) A disclosure of the invention as characterized in the claims in such terms that the technical problem and its solution can be understood, stating, if applicable, the advantages of the invention in reference to the prior art.
- (e) A brief description of the drawings, if any.
- (f) A detailed description of at least one way of carrying out the invention, which may, as appropriate, be illustrated with examples and references to the drawings, if any.
- (g) An indication of the way in which the invention is capable of being applied in industry, when this is not obvious from the description or nature of the invention.

Description of the cooperation between universities and industry

(how it is organized in partner's organization, the sources of funding, the opinions about drawbacks and strengths of existing solution).

Currently at University of Granada, Spain there is established an internal organ for cooperation between the University and Industry (similar to Technology Transfer Center in Wroclaw University of Technology), which manages most of the possible collaborations, all the necessary official documentation and support for scientist. However, collaboration between the researchers and industry are also possible outside of this unit, in form of applied grants or contracts for specific tasks. Such a collaboration appears very often in University of Granada and is treated as an additional source of income and a way of practical testing of the developed methods. Commercialization of research results is highly encouraged.

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 316097 [ENGINE].

ENGINE Centre, Wrocław University of Technology
Wyb. Wyspiańskiego 27, building A1, room 203k
50-370 Wrocław, Poland
tel.: +48 71 320 3453, email: engine@pwr.wroc.pl

However, the number of patents is quite low and the collaboration is more oriented on delivering a product / algorithm / analysis / solution to a problem.

Other activities

.....

REMARK: Apart from this information also a program of the visit and the presentation in electronic version should be given to the project office (please send all of them to Urszula.Markowska-Kaczmar@pwr.wroc.pl). Please respond to the points 1-5 for outgoing visit and points 1-3 for incoming visit. Point 6 is for extra activities that are not put in points 1-5.

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 316097 [ENGINE].