Report on the outcomes of a Virtual Mobility[1]

Action number: CA19130

Reference: E-COST-GRANT-CA19130-453dc334

Grantee name: Miroslav Hudec

Virtual Mobility Details

Title: The synergy of statistical approaches and fuzzy logic approaches in mining patterns from p2p loan data

Start and end date: 01/03/2024 to 30/08/2024

Description of the work carried out during the VM

Description of the virtual collaboration and activities carried out during the VM, with focus on the work carried out by the grantee. Any deviations from the initial working plan shall also be described in this section.

(max. 500 words)

The collaboration started with the exploration of Bondora P2P data. We found a publicly available data set of 266,483 records. The activities were carried out according to the VM application: First, we conducted an analysis of a data set followed by a statistical analysis. Next, the AI models of fuzzy functional dependencies were applied on the selected attributes. It was followed by the work on another AI model: linguistic summaries. The findings were examined considering the benefits for diverse categories of data users and building a business intelligence dashboard to provide an explanation by short story to augment statistical figures.

As in the case of research projects, our schedule for parts was not strict, but methodological. The collaborative perspective was reflected in the fact that the project team members are from institutions included in this COST initiative from seven countries. The project team members cooperated in the steps mentioned above. The tasks within this small project were not managed by the so-called waterfall method, but by the so-called spiral method, as the findings in each step influenced the other steps.

The guarantee has contributed to the coordination activities and with the significant proportion of text in the article currently in SSRN, available upon request. The main problem was the coordination in integrating different views, terminologies and knowledge from fuzzy logic, statistics, and financial fields in short period. A slight deviation is in the article sent to the SSRN instead of to journal. The limited time has not allowed for submitting to a top-tiered journal, and get acceptance in due time. Journals with a

[1] This report is submitted by the grantee to the Action MC for approval and for claiming payment of the awarded grant. The Grant Awarding Coordinator coordinates the evaluation of this report on behalf of the Action MC and instructs the GH for payment of the Grant.

rapid review exist, but the quality is questionable and moreover for several team members such journals are not accepted by their universities.

In our opinion, this VM has achieved declared goals from the scientific perspective, but more from the networking perspective, as we have created a multidisciplinary research team for possible future cooperation in science and practical applications. The main problem was the coordination in integrating different views, terminologies and knowledge from fuzzy logic, statistics, and financial fields. This work has prepared background for applying to future scientific activities.

The link to SSRN is: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4941256

All required files (data, calculations and the whole article) available upon request.

Description of the VM main achievements and planned follow-up activities

Description and assessment of whether the VM achieved its planned goals and expected outcomes, including specific contribution to Action objective and deliverables, or publications resulting from the VM. Agreed plans for future follow-up collaborations shall also be described in this section.

(max. 500 words)

The main objective was to explore the benefit and weak points of the synergy of statistical and fuzzy logic approaches (AI methods) in the mining and interpretation of valuable information from financial data. Activities were partially related to deliverable 4: exploring P2P database for mining valuable knowledge and integrate it for business intelligence reporting; deliverable 7: explaining influence of attributes on P2P lending as well as on explaining patterns linguistically, and deliverable 11 by merging statistical and AI (namely fuzzy logic) to explain revealed information. First, we explored on adjustable toy data, whether correlation coefficient, fuzzy functional dependencies and logistic summaries are able to provide consistent information from different angles. Next, we adopted these approaches on the P2P lending data (266, 483 records). For instance, between two pairs of the selected attributes, we recorded low correlations with a smaller difference. Consequently, linguistic summaries explained this difference in subdomains and also the direction of dependence. However, the existing algorithms for fuzzy functional dependencies were not able to finish task due to the extensive number of pair-to-pair comparisons.

Fuzzy logic approaches belong to the XAI, so the results are explained linguistically. But they are more computationally demanded. Thus, the input from statistics is very relevant to focus on particular attributes for further explanation. The correlation reveals whether it is positive or negative, but not the direction. Fuzzy functional dependencies do not reveal whether correlation or dependency it is positive or negative, but the direction of dependency (formula is asymmetric). In addition, this approach is less sensitive to imprecision and vagueness in data. Linguistic summaries are able to explain the direction of dependency on particular subdomains of attributes, especially without a sharp distinction between subdomains, i.e. when it says that about half of young people have a high amount of the lending, but most of elderly have a small amount. It is expected that if the number of elderly clients increases, the correlation will increase, but if the number of young increases, the correlation might decrease.

The future work will focus on explaining the other statistical figures like skewness and outlier tails linguistically, developing more powerful algorithms for fuzzy functional dependencies (due to the combinatorial explosion problem of the existing ones), and practical realisation of business intelligence dashboards.

Due to the limited time for this research and our intent to submit to the high-quality journal, the preliminary article is available in SSRN. Thus, the other scientists can benefit and might contact us if they are interested. Anyway, we will send it to the journal.