

Company Recommendation Model:

Empowering the Accounting System and Publicly Available Data

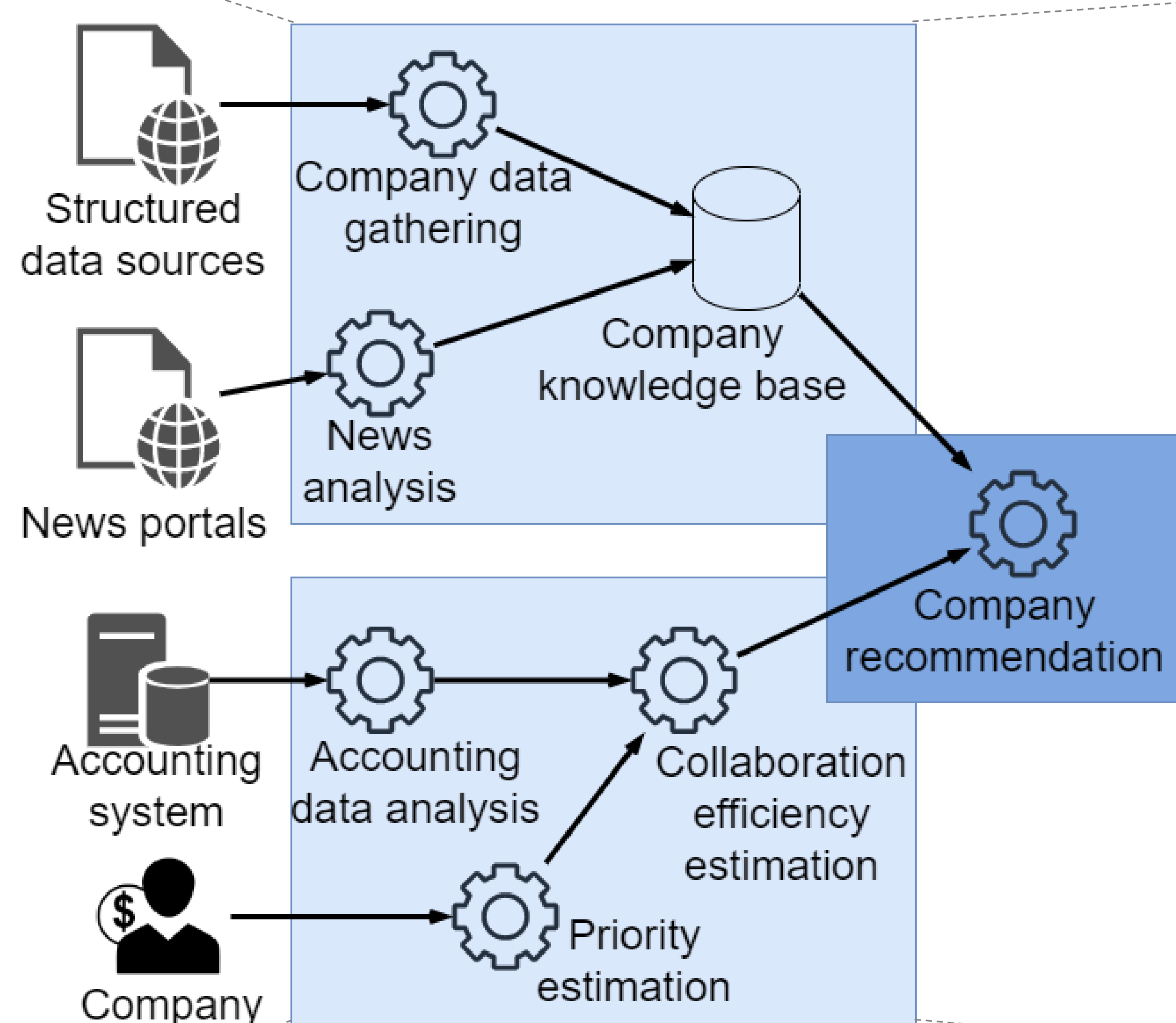
Rokas Štrimaitis, Pavel Stefanovič, Simona Ramanauskaitė, Asta Slotkienė

Abstract. Business management requires constant decision-making. Usually, the selection of suitable partners for collaboration is very intuitive, experience-based skill or requires the usage of data analytics. The most important aspect to get an explainable recommendation for possible collaboration between companies is data. In this research, a company recommendation model is created to incorporate both the accounting system and publicly available data. The accounting system data is used to estimate the collaboration effectiveness of existing collaboration cases. The publicly available data include company registration data and news portal article texts. The news article data include the detection of the company mentioning, estimation of its sentiment and context category. Separate models were developed for different data extraction and analysis. A combined company recommendation system was designed. Model validation with gathered test cases demonstrated 70% recommendation accuracy.

Data from 4 company databases were extracted (via APIs or parsing of web page structured data).

Article texts from 5 Lithuanian news portals constantly parsed and analyzed:

- Company mentioning identified by using SpaCy Named Entity Recognition model for organization and person. Stemmed mentioning and company keyword database versions compared. Company recognition f-score - 79%, mentioning linking to database company - 83%.
- Company mentioning sentences analyzed by learning AI model with >10000 labelled news sentences dataset:
 - Sentence 3-class sentiment estimated by using LSTM model and text preprocessing. The f-score reached 71%.
 - Sentence 10-class context categories assigned by using a combined similarity matrix and feature transformation model. The model reached 20% full match and 80% partial match accuracy.



For publicly available data the collaboration efficiency (average of last 6 months) was assigned. This dataset is used as training data.

Feature engineering showed the news portal data currently is not perspective in the analyzed situation (a small portion of companies are mentioned in the media). 8 companies defining publicly available features were used for KNN model (k=7).

To achieve 70% of the 10 recommended companies for collaboration will have the top10 efficiency between the validation records, at least 16 training data records were needed.

The AHP method is used to define more than 20 accounting system data fields (structured into a 3-level hierarchy) for each company. No significant difference between the 8 analyzed Lithuanian companies defined.

Collaboration between companies was evaluated using the 3-class scale (effective, neutral, not effective) by representatives of analyzed companies (223 evaluations). AHP-based collaboration efficiency score aligns with the 3-class scores and reached 74% for monthly aggregated accounting data.

Conclusions. Company collaboration efficiency by using the AHP method for monthly aggregated company accounting data and discretizing it into 3 classes (using efficiency quartiles of all company efficiency scores from the last 6 months) allows collaboration evaluation to be similar to the company owner or presenter opinion in 74% tested cases. Automated efficiency evaluation then is used for recommendations of companies for collaboration. While news portal data analysis showed sufficient results for an automated summary of company mentions, mentioning sentiment and content, its usage for company recommendation does not increase the recommendation accuracy. Integrating the most general publicly available company data such as region, the number of employees, activity class and other features with company collaboration score, automated, not adapted to specific case model is able to achieve 70% recommendation accuracy.

Partly funded by the project "Extension of Accounting and Business Management Systems using the artificial intelligence methods" (No. 01.2.1-LVPAK-856-01-0083).