Towards Collaborative Business Intelligent Framework: Crowdsourcing Approach

Olga Cherednichenko,[1] Maryna Vovk,[2]
Oleksandr Sutiahin[2]

[1] Université de Lyon, Lyon 2, UR ERIC, France
olga.cherednichenko@univ-lyon2.fr
[2] National Technical University «Kharkiv Polytecnic Institut», Ukraine
maryna.vovk@khpi.edu.ua, sutiahin.oleksandr@cs.khpi.edu.ua

Abstract Collaborative Business Intelligence goes beyond the organization frame and collecting people in one virtual space in order to give their opinion concerning certain issue results in an opportunity to boost the decision-making process and make it more reasonable. Strangers solve one problem without direct interaction and use the developed results. Tourists often search for information from different sources and compare found facts. It takes a lot of time and effort. So, having an application, where users could find relevant information and use it for decision-making is essential. The aim of the paper is to develop an application for tourists using which they could make a decision about where and why to go. We propose an approach to collect and clear data related to tourism, preprocess and structure tourism data convenient interface and tools for collaborative analysis.

Keywords: business intelligent system, collaborative BI, framework, crowdsourcing

University of Maribor Press

DOI https://doi.org/10.18690/um.fov.3.2023.16
ISBN 978-961-286-722-5

1 Introduction

The decision-making process is complex and, as a rule, depends significantly on the information that the person who makes the decision owns. Today's world is characterized by huge volumes of accumulated data in various domains. These data can be really helpful in terms of preparing and making the decisions. However, this data are collected and stored by different unrelated software systems, stored in different formats, providing different levels of access and security. In addition, these data may be incomplete, contradictory, unreliable. To solve problems associated with the processing of large volumes of data, they turn to data analysts. Business intelligence (BI) helps you gain valuable insights and make strategic decisions. Business intelligence tools analyze historical and current data and present the results in intuitive visual formats. A significant obstacle to achieving the effect of using the accumulated data is the lack of direct communication between technical specialists and decision makers and business process analysts. The solution to this problem is to apply the approach of collaborative business intelligence (CBI).

The numerous free BI sites exist now (Business Intelligence Software, n.d.). However, the majority of such kind tools mostly focus on dashboards and visualization and have limited functionality. Moreover, they are still out of the reach of small companies, non-governmental organizations, researchers, independents such as journalists or active citizens. Also, a demand for collaborative tools for data analysis grows.

As the analysis shows, the need for data research arises not only from business in order to increase its profits, or the government to solve national problems, but also from society and individual citizens to understand and justify socially significant or private decision-making. In this case, it is quite difficult to organize the interaction of potential users, decision makers, and technical specialists in data analysis. The project BI4people (Business Intelligence for the people, n.d.) is created to help solve these problems. The aim of BI4people is to bring the power of Online analytical processing (OLAP) interactive analysis to the largest possible audience, by implementing the data warehousing process in software-as-a-service mode, from multisource, heterogeneous data integration to intuitive analysis and data visualization (Fahad, 2022).

Social networks, quizzes, brainstorms are tools for collaborative BI. Even chatting in Teams with two-three coworkers to solve some bug is CBI. So main idea is to collect people in one virtual or real space and encourage them leave their comment or opinions for general purpose. Moreover, reusing another collaborators' results or comments makes general BI - CBI.

2 Related works

Collaborative BI combines the traditional BI and collaborative technology to satisfy the specific needs of business enterprises' collaborative decision-making and business process (Liu, 2017). So, using collaborative BI improves quality of business analysis and engage different colleagues participate in it, improving their knowledge of different business departments (Liu, 2016).

There are some researches that are oriented on creating application which allow generating data from heterogenous sources. Business Intelligence stands up for an agile solution using metadata across different systems. Paper (Varga, 2018) presents an RDF-based (Resource Description Framework) metadata metamodel. A flexible solution was designed which can be smoothly spread among heterogeneous systems. It works with metadata processing semi-automation. The developed model can be used for both new metadata models and existing ones.

Artificial intelligence plays the leading role in archiving metadata. Developed intelligent framework (Gosaibi, 2020; Latreche, 2020) for an organization that contains Chatbot, Sentiment Analysis, and Key Performance Indicators and can improve the performance of a company. Using a framework, organizations can build a mechanism for their workforce to retrieve meaningful information.

Business Intelligence Network (Rizzi, 2012) provides an opportunity to expose querying functionalities focused on sharing business information for the decision-making process. Decentralization, scalability, and full autonomy of peers are its main features. It was also an attempt to create a business intelligence platform for tourism analysis (Bustamante, 2020). BI platform combined four collaborative data sources (Twitter, Openstreetmap, Tripadvisor and Airbnb). The development was quite successful, but the main drawback is data quality.

We can divide collaborative BI for three types: internal analysis, partnership in data, partnership in analysis (Kaufmann, 2014). Internal - collaborative BI that is implemented within single business, coworkers from different departments are gathered to solve some problems, for example lead of some IT team can have meeting with someone from accountant department to share opinions about their new financial product. Partnership in data is when business tries to gather external data, the source can be sharable for different companies, or it can be bench of sources connected to each other. Partnership in analysis is when one company can retrieve information about made decisions of another company.

3 Methods and Materials

Planning traveling tourists usually search for information from different sources, such as Google maps, feedback at booking sites, and compare found facts. It takes a lot of time and effort. Thus, having a certain application, where users could find relevant information and use it for decision-making is crucial. There are a lot of open sources where such kind of information can be found but it is given in various formats. Tourists would like to make the right decision based on other people's experiences. The obstacle is that "right" and "proper" decisions would be different for different people. Some pursue the goal to have the cheapest trip, for others convenience has the highest priority, for the third group is substantial to as many attractions as possible in the same area. Someone has a business trip and is constrained by the time frame. Each of mentioned cases needs extra information to solve a task of right decision-making. Most of the tourists are not experts in data analysis and reasonable decision-making. So, they need an easily understandable tool with data visualization, which is able to propose possible solutions and ideas.

The usage of open data on the Internet and social media helps people to find necessary information for decision-making. Searching for data and giving any feedback leads to collaboration and creating value together. Such an approach is called crowdsourcing. The term crowdsourcing was first used by Jeff Huff in 2006 in Wired magazine (Howe, 2006). Crowdsourcing means engaging the general public, through the Internet, in research and solving public problems. Crowdsourcing presumes a process of combining individuals, paid or unpaid who are joined together with a shared interest and are able to increase results thanks to

their aggregated actions. This approach can offer researchers access to new opportunities for co-creation, task optimization, and cost reduction.

Our goal is to create a framework that will allow us to implement a virtual space for collaborative data exploration and decision making. However, how to involve various people who are not familiar with each other and not related to business interests in the general process of brainstorming and discussing ideas? It's a challenge. In this study, we propose to use the idea of crowdsourcing to obtain enough data to evaluate a collaborative decision-making strategy.

We propose to consider the following main stages of the research project. First, a domain must be defined in which collaborative analysis and BI can be modeled. In this study, we justify the tourism as a domain that can attract a large number of users with different research questions about tourism data. Considering different goals, preferences, experience and conditions, different users will access the same data with different requests forming the content of collaborative session. Second, data sources, data collection, and data access methods need to be identified. Combining data from different sources requires solving the problems of data consolidation, cleaning, and standardization. Structuring and preparing for a possible subsequent OLAP analysis requires a thorough study of the domain and the development of an information model. Tourist information can be found in reviews on booking sites, travel agencies, in the form of labels and descriptions on maps. Data for analysis can be supplemented, for example, with information about the weather or exchange rates. To successfully implement a crowdsourced collaborative BI prototype, it is necessary to prepare and constantly update data that is relevant to user requests. Thirdly, it is necessary to develop a convenient interface for visualizing data and organizing interaction in the virtual space. At the beginning, it is proposed to consider a structured system of comments and quality assurance from users that are created in the process of data analysis as such a collaborative space. In the future, such a system can be supplemented with a chatbot that will provide information support for a user who is not technically prepared to solve data analysis tasks. Fourth, one of the main stages is the formation of a knowledge base of collaborative decision-making cases. At this stage, you need to develop a model for collecting data about each session, including user behavior and the results of his research, as well as interaction with other users. In addition, it is necessary to develop a CBI case model and a similarity assessment model for such cases. Fifth, it is necessary to develop

and deploy a prototype for collaborative BI. This stage also includes the choice of how to promote such a service, since an important task is to attract as many users as possible. The final stage is associated with the processing, analysis, and summarizing of the collected data about user behavior. We believe that as a result we will be able to create a CBI framework and prove models and technologies for supporting virtual space, which will expand the functionality of the BI4people project platform.

4 Results

The main idea of the given research is to develop a CBI framework, but it's really complicated task. We propose to perform a sub-task and observe users' behavior while they are doing data analysis and to collect information. Then we will summarize the data collected and create a framework. As a case-study, we choose the tourism, as far as a lot of people are travelling and are interested in information about different touristic places. The goal of tourist data analysis is to decide about where and why to go. Different people rely on different criteria when choosing a trip destination. Coordinated decisions take into consideration the most important criteria and sometimes new criteria based on other tourists' experiences. For example, some people choose hotels according to cleanness marks and another user can try to do the same, even if he didn't account for that before.

So, we develop the software that implements some virtual space as a partnership in analysis where the users can share their problems, comments and the software should gather and analyze their data then visualize data in order to help clarify dependencies between different data and in the end the user must comment how he used this analysis.

User workflow can consist of next acts (fig. 1): the user (e.g. his name is John) can choose topic that he is interested in, then he writes his question, such as "I want to see dependency between tourists' reviews rating and weather condition in Paris during the summer", and software must firstly check whether this question has been asked before and provide to John results of previous users. If it's completely new request John should see chart with weather and rating axes. Then he should add comment like "Okay, I want to visit it in second part of August". His comment is saved for this question and creates use case.

So inexperienced person can check what was answered to his question even before he posts it and just follow the recommendations, like using Stack overflow forum (https://stackoverflow.com/) to solve programming problems.

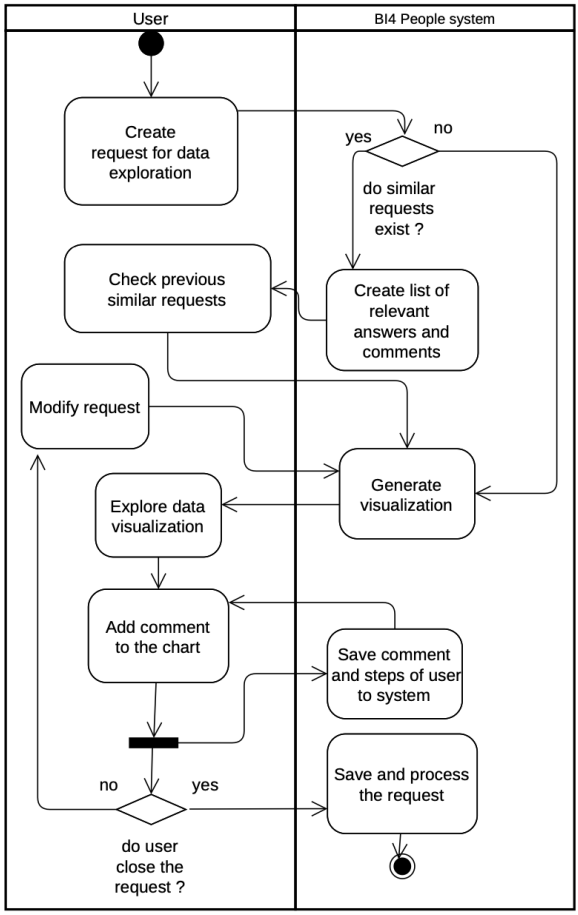

Figure 1: The main workflow

Using external data, we can add more information that leads to better decision. For tourism app we can gather data from Twitter and weather forecast to understand which time is the best to visit some places.

Usually for BI OLAP cubes are used. OLAP cubes are computer-based technique of analyzing data to look for insights. The term cube here refers to a multi-dimensional dataset, which is also sometimes called a hypercube if the number of dimensions is greater than three. The app (fig.2) consists of three large parts: frontend application, backend application, Cube JS instance. Frontend part will be implemented with React JS framework because of simple connecting to Cube JS instance. In this app user can request some chart that help him perform BI analysis in toursim area, moreover the system is able to investigate whether the request has some similar questions from previous users. Cube JS has API to create appropriate result set of data to create charts in React web app. Source of the data for chart is MongoDB. MongoDB will be filled by open source data gathered by open data preprocessor (Google maps, trip adviser, weather forecast websites). Also the user has to leave comment after investigation of the chart, the comment should include information how the chart helps user in his study. So the comment, the request and chart data set will be sent to Java backend REST API and saved in Postgres DB. This saved use case will be proposed for next user with similar request. All parts will be dockerized due to easy deployment and CI/CD implementing.

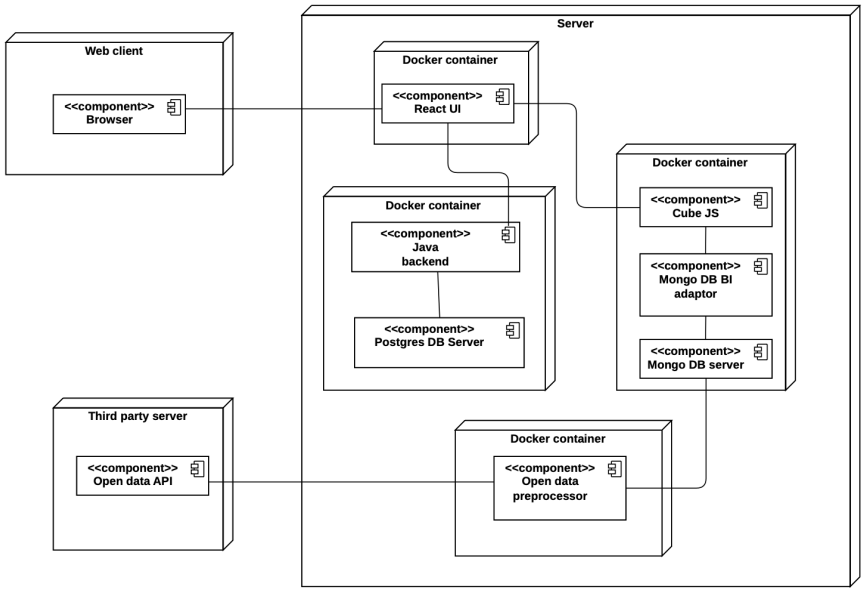

Figure 2: The deployment diagram

The fragment of data model is depicted on the figure 3. Main entity is place that charactarized by next fields: name, city name, price (it scales from 1 to 5 point), categories (park, museum), address, total rank in Google Maps (from 0 to 5 stars) and attendance or occupation – this field show how busy this place during specific hour of certain day of week. Next part is review, it's comment with rating that some user created for the place. The review consists of review text, count of stars (rating of place), published date time which is set in separate table. Data for those tables is crawlered from Google Maps. Last table is weather forecast, that contains information about weather in certain city – temperature, date and time, conditions (cloudy, sunny), comfort index. The weather forecast data are grabbed from open-source weather APIs.

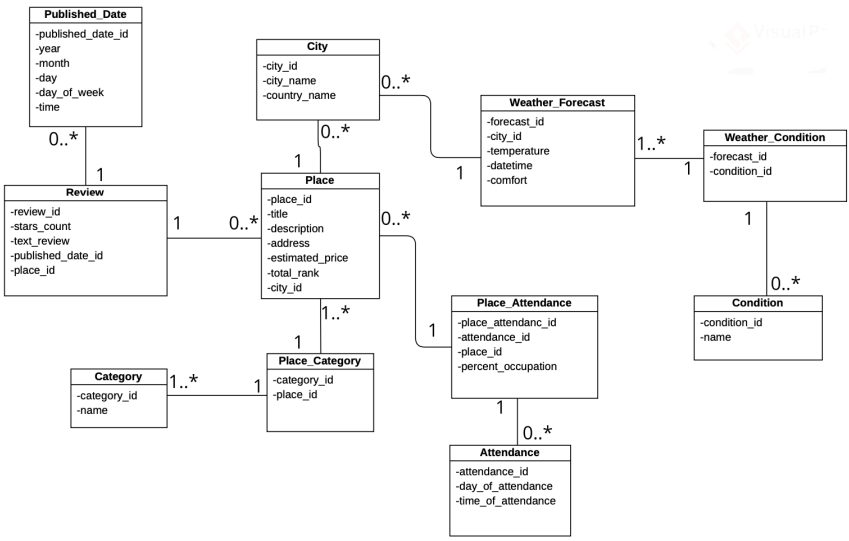

Figure 3: Data model (fragment)

As a result, we are creating a prototype application for data exploration in the tourism domain. We identified data sources and data collection methods, built a domain information model, and identified tools for implementing the software.

5 Discussion and Conclusion

In this paper we have presented an approach how to create, approbate and estimate collaborative decision-making models. BI systems are vastly used as a tool to support decision-making in different kind of organizations. CBI give even more opportunities for reasonable decision-making as they allow using external information from various sources. We propose to create application, which should be used in tourist sphere, because significant number of people travel all over the world. It's presupposed that they would be interested in using of such application, that integrates data from collaborative sources, the social networks, open map platforms and booking sites. We are collecting and processing data, developing convenient interface and tools for collaborative analysis. The next step is to implement a prototype.

Acknowledgements

The research study depicted in this paper is funded by the French National Research Agency (ANR), project ANR-19-CE23-0005 BI4people.

References

Business Intelligence for the people – BI4people https://anr.fr/Project-ANR-19-CE23-0005

Business Intelligence Software https://www.predictiveanalyticstoday.com/open-source-free-business-intelligence-solutions/

Bustamante, A., Sebastia, L., Onaindia, E. (2020) BITOUR: A Business Intelligence Platform for Tourism Analysis. ISPRS Int. J. Geo-Inf., 9, 67Rizzi, S. (2012) Collaborative Business Intelligence: eBISS 2011, LNBIP 96, 186–205.

Fahad, M., Darmont, J., Favre C. (2022) The Collaborative Business Intelligence Ontology (CBIOnt), 18e journées Business Intelligence et Big Data (EDA -22), Clermont-Ferrand, Octobre 2022; RNTI, Vol. B-18.

Gosaibi, A., Sait, A. et al. (2020) Developing an Intelligent Framework for Improving the Quality of Service in the Government Organizations in the Kingdom of Saudi Arabia (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 12

Howe, J. (2006) The rise of crowdsourcing. Wired magazine 14 (6), 1-4, 6101

Kaufmann, J., Chamoni, P. (2014) Structuring Collaborative Business Intelligence: A Literature Review. 47th Hawaii International Conference on System Science

Latreche, O., Boukraa, D. (2020) Self-Service, On-Demand Creation of OLAP Cubes over Big Data: Metadata-Driven Approach 2020 IEEE International Conference on Big Data

Liu, T., Shen, A., Hu, X., Tong, G., Gu, W. (2017) The Application of Collaborative Business Intelligence Technology in the Hospital SPD Logistics Management Model. Iran J Public Health, Vol. 46, 6, 744-754.

Liu, T., Shen, A., Hu, X., Tong, G., Gu, W. (2016) SPD-based Logistics Management Model of Medical Consumables in Hospitals. Iran J Public Health, Vol. 45, No.10, 1288-1299

Varga, J., Romero, O., Pedersen, T. B., & Thomsen, C. (2018). Analytical metadata modeling for next generation BI systems. Journal of Systems and Software, 144, 240–254