Recommendation-based Business Intelligence
Architecture to Empower Self Service Business Users

Safwan Sulaiman⁴, Jorge Marx Gómez¹

¹Oldenburg University, Business Informatics/VLBA, Oldenburg, Germany
{safwan.sulaiman, jorge.marx.gomez}@uni-oldenburg.de

Abstract. Enterprises live today in a very competitive business environment, which is influenced by globalization, technical progress and volatile markets. Therefore, business intelligence systems are widely adopted to support enterprises to ensure a competitive edge against competitors by providing their decision makers with the required information at the right time. However, business users still face significant difficulties while performing both ad-hoc and sophisticated data analysis. In this paper, we propose a new approach for self-service business intelligence as a major support to business users by providing them with guided recommendations while they do their own data analysis. The main idea behind this approach is to transfer the knowledge of the power users, which is represented as analysis paths, to the business users in form of recommendations or suggestions to help them when struggling with business intelligence systems and their complex data analysis.

Keywords: Recommender System, Self-Service, Business Intelligence, Knowledge Transfer Model.

1 Introduction

In the current economic environment, enterprises face constant challenges to run their businesses in a very competitive business environment within a dynamic market [1]. Due to the rapid changes in the external and internal conditions of present day economic life, information has become a very important production factor for enterprises, who have become very dependent on reliable information [2]. Therefore, to ensure continued success in their businesses and to guarantee a sustained competitive edge, enterprises recognize the importance of providing their employees with relevant and timely information by using Business Intelligence (BI) tools effectively [3–5]. To respond to this situation enterprises have widely implemented and integrated BI systems [6].

In spite of the big market and the importance of BI and analytics for enterprises, various market studies and surveys have shown that the usage of BI systems is still very low. The Gartner Report showed that less than 30% of the potential users use BI systems [7]. Eckerson explain in his study about business driven BI that the BI penetration is minimal, only 26% of employees use BI tools [8]. According to the BI
survey 15 conducted by BARC Research, the median percentage of employees using BI in companies reached 13% [9].

In the literature, several user acceptance models for BI showed that the complexity of the tools is considered as biggest challenge for adoption of BI [10, 11]. In addition, due to their flexibility and powerfulness, BI systems are still too complex for business users, who face significant challenges in performing adequate ad-hoc analyses [12]. Alpar and Schulz pointed that when you offer business users more flexibility this requires more BI skills from them [13]. Consequentially, due to the high demand for reporting and data analysis, power users become bigger bottlenecks for an enterprise. The large amount of business users’ requests leads to the inability of power users to react sufficiently rapidly and accordingly business users will make decisions without relying on information derived from data analysis, possibly making decisions based on their good feeling or experience, which lead sometimes to wrong decisions [8, 13, 14].

To react to this situation, a new trend in the BI domain has emerged called self-service BI. The Data Warehouse Institute (TDWI) defined self-service BI as the facilities within the BI environment that enable BI users to become more self-reliant and less dependent on the IT organization [15]. In this paper, we propose a new approach towards the realization of self-service BI, where the business users will be empowered and supported while performing sophisticated data analysis to get the information they need at the time they need it, without relying on or asking the IT department or the power users. This approach mainly supports business users while they do their own data analysis by providing them with guided recommendations. These recommendations are derived from the extracted knowledge of the power users (analysis paths). This is not to be confused with other similar researches, its goal being to help and support business users while using BI systems to improve the adoption and usage rate of BI in enterprises and not the automation of the decision making process based on old decisions from power users.

The remainder of this paper is structured as follows. First, an overview of the theoretical foundation and related work are provided in Section 2. After that, the conceptual architecture of the recommendation-based self-service BI will be explained in detail in Sections 3. Next, a prototype as a proof of concept and the evaluation of this research will be presented in Section 4. Finally, in Section 5, the conclusion of this paper and its result are summarized and the outline for future work presented.

2 Theoretical Foundation and Related Work

In this section, self-service BI is first described in more detail, to indicate its importance. After that, the recommender systems will be also described as an important part of this research. Finally, relevant related works are introduced.

2.1 Self-Service BI

According to the result of BI trend monitor 2017, self-service BI is one of the most important trends for BI practitioners in their work [16]. To meet the time-to-insight
required by the current competitive business environment, many enterprises want to democratize analytics capabilities via self-service BI [6]. Self-service BI has been on organizations’ wish lists for a long time as IT departments struggle to satisfy the ever-growing demands and requirements from business users for faster changes and flexibility for ad-hoc reporting and analysis to meet their BI needs [14, 16]. Imhoff and White argued that BI providers have already achieved a high level of ease of use for reporting and simple analytics. However, more complicated analytics still need be enhanced and made easy to use and create. The realization of self-service BI should focus on one or more of the following four main objectives: providing easy access to source data for reporting and analysis, building easy-to-use BI tools and improved support for data analysis, providing fast-to-deploy and easy-to-manage data warehouse options such as cloud computing, and designing simpler and customizable user interfaces [15].

2.2 Recommender System

Due to the ever-increasing amounts of data and the heterogeneity of the users, a Recommender System (RS) is of great scientific importance and is described in numerous researches [17]. The main goal of the RS is to provide recommendations for items most likely to be of interest to a specific user for helping him within various decision making processes, such as what product to buy, what movie to see, etc. [17]. RS was already wide-spread in the E-commerce domain, however, it is also used in numerous application domains like web search, computational advertisements, social media recommendations [17]. In addition, RS is increasingly being used in the area of Technology Enhanced Learning (TED) to provide learners with appropriate and individualized knowledge that supports all their learning activities and helps them work independently and more flexibly to achieve a specific learning goal [18]. This area of research is close to the problem defined above. Therefore, in the following section we will introduce some related work that has some similarities with our approach.

2.3 Related Work

In this section, we present some interesting work, which plays a special role for our research. These are published in the area of TED.

Zaiane described a concept to help e-learning users through intelligent recommendations navigation and operation of the software system [19]. The recommendations should be generated by means of an access history, which is derived from the server log files. This work tried to transfer the RS functionality from E-commerce to the E-Learning domain by using web-mining techniques to extract and find learning activities to be suggested to the learner for improving the course material navigations within the E-learning system. In the method described, these patterns are analyzed dynamically, so that the user receives recommendations for further possible steps directly from the current situation. In addition, the log sequences are compared by the user with other log sequences from other users in order to be able to provide him with additional interesting content specific to his needs.
Khrib et al. describe a concept for recommendations, which provides automatic and personalized learning material in E-learning systems [20]. The learning material is referenced via a URL in the e-learning system and archived via a log file or a database so that it can be used for later recommendations. This work describes two approaches to modeling user models. On the one hand, there is the creation of a user model by explicit information about the properties of a user. On the other hand, an implicit approach is used for the automatic generation of a user model based on the navigation history and the activities within a user’s session. Users with the same interests are then grouped so that they can benefit from the behavior of others.

These approaches contain important characteristics for our work. However, analyzing log files of web servers to develop a RS for BI systems is not sufficient because many queries contain important information related to the usage of a BI system in the POST and GET parameters, which must also be considered for further recommendations. Furthermore, the focus of these approaches is to personalize a web site or E-learning system based on the user model of the learner or the user, while in our case the focus is to extract the analysis paths of the power users and then to suggest them to the business users independent of a specific business user.

3 Recommendation-based Architecture for Self-Service Business Intelligence

In this section, the conceptual architecture of the new self-service BI system will be explained. First, the main concept of the knowledge transfer model, which represents the main incentive of this architecture, will be described. After that, the first part of the knowledge transfer model, which is the capturing and extracting the knowledge of the power users, will be illustrated. Following this, the second part of the knowledge transfer model, which is the application of the extracted knowledge to the business user, will be explained in detail, this includes the description of the main components and algorithm of the RS.

3.1 Knowledge Transfer Model

The main idea behind this architecture emerged from the transferring of the knowledge from the power user (source of knowledge) to the business user (receiver of knowledge) [21]. The model of transferring knowledge consists of two phases. The first one is the externalization of the knowledge by conversion from tacit to explicit mode. The second phase is the internalization of the knowledge by conversion from explicit to tacit mode [21, 22]. The extraction and capturing of the knowledge of power users were already done: First, due to the shortcomings of the methods of analyzing the log files of web servers to extract the knowledge of the power user, a new tracing mechanism for the power users’ interactions while they use the BI system was conceptualized and developed [23]. The interactions and their metadata are stored in trace files. Then, based on applying sequential pattern mining algorithms on the stored trace files, the analysis paths of the power users were extracted [24]. An analysis path is defined as a sequence
of steps required to reach a specific goal (e.g. multidimensional view of business data) or to perform a specific data analysis within a BI system. In addition, analysis paths also include BI related metadata to enrich the business user with more information about what steps should be done within BI systems [24]. Analysis paths are stored in a knowledge repository. One of the main advantages of this approach is that the knowledge repository is remains up-to-date. As a consequence of power users use of the BI system, new analysis paths are always extracted and added to the knowledge repository. In the next sections, we will focus on the second phase of the knowledge transfer model, which is applying the extracted knowledge to business users through recommendations to help and empower them while using a BI system.

3.2 Conceptual Architecture

BI system have become more and more accessible via the web browser (web based BI system) and so are easily available for all employees an enterprise without the need of any installation on the client side. This influenced our conceptual architecture of the new self-service BI system. Accordingly, the new system is conceptualized and developed as a client-server architecture and typically consists of three layers, the presentation, the application (logic) and the data layer. Figure 1 depicts the architecture of the system and its components. On the left side of the figure, we can see the internet browser, which represents the access point of the BI system for power and business users. It was decided was to integrate the RS within the browser as a plug-in. In the following three sections, the three layers and the functionality of the architecture’s components are explained.

![Conceptual architecture of the recommendation-based self-service BI system](image)

**Figure 1.** Conceptual architecture of the recommendation-based self-service BI

**Presentation Layer.** This RS layer is represented by the browser plug-in. This latter dynamically inserts the resources of the RS into the BI system so that the RS can be directly initialized in the browser by the business user to enable her/him to work with
it interactively. Especially in the presentation layer, communication with the observer and the visualization of the recommendations play an important role. Furthermore, interaction options such as clicking on or evaluating a recommendation are also made in this layer. In this case, the interactions with the event listener are dynamically recorded and passed on to the application layer. The presentation layer essentially represents the user interface as well as the interaction possibilities of the RS. It includes the following main components:

- **Observer**: is responsible to trace the interactions with its BI metadata of the business user. After this, these are passed by a connector to the sequence matcher to be compared with the analysis paths from the knowledge repository. Observer aims to know what the business user is trying to do within the BI system.

- **Suggestion Viewer**: is responsible for the visualization of the resulting recommendations for the business user. It gets the recommendations list sorted by its priorities from the sequence matcher. After every interaction of the business user with the BI system, suggestion viewer first removes all previous recommendations and then adds the new recommendations to the current page dynamically sorted by their priorities. It also allows the business user to evaluate the last recommendation and to send the evaluation to the evaluator in the application layer.

**Application Layer.** This layer is represented by the web services and it is the link between the presentation layer and the data storage layer, in this case the knowledge repository. In this layer, the main logic and thus the core functions of the RS, which are the sequence matcher and the evaluator, are executed.

- **Sequence Matcher**: is responsible to match the interactions obtained from the observer with the analysis paths from the knowledge repository, and then to build a list of recommendations to be sent later to the business user. The function and the algorithm of the sequence matcher will be described in detail later in this section.

- **Evaluator**: is used to evaluate the recommendations so that future results can be generated better with the help of collected evaluations of the business users. The evaluator represents the evaluation module of the RS. At this point, the analysis paths suggested to the business user by the RS can be evaluated directly or indirectly. Indirect evaluation is done by clicking on a recommendation, while the direct evaluation takes place via a conscious evaluation by the business user who clicks on a recommendation to evaluate the last interaction performed with a thumb up or down. After the feedback for the recommendation made by the business user is prepared for the data layer, after that, the feedback is passed on to the knowledge repository.

**Data Layer.** The data storage layer of the recommender engine is represented by the knowledge repository. The data retention layer thus provides the RS with all the information necessary to generate new recommendations.

- **Knowledge Repository**: stores all analysis paths and their Metadata, which are already extracted from the power users. This component could be realized via a relational database.
**Sequence Matcher Algorithm.** Sequence matcher represent the heart of the RS, therefore, we will explain its functionality and the algorithm behind it in more detail. It is responsible for the search for matches between the analysis paths of the power users and the sequence of business user’s interactions that are passed from the observer.

When the sequence matcher is called after a business user interaction, the call contains information about the current business user session, such as the user role, the dimension name, the support level, and the interaction sequences of the business user from the Observer. The support level specifies how many interactions must be matched between the analysis paths and the business user interactions. This helps to reduce the amount of recommendations when lot of analysis paths come from the knowledge repository as a result of the matching. Figure 2 illustrates the functionality of the support level and sorting of the recommendations. Each interaction of a business user with the user interface of the BI system is represented with a unique ID “PXY” like “P05”.

![Figure 2. Recommendations matching and sorting algorithm](image)

Figure 2 shows that the click sequence of the business user requires a support level of three interactions. In this example P05, P11 and P32 must occur in the analysis paths from the knowledge repository. If this is not the case, the analysis paths from the knowledge repository are not considered further; all other analysis paths and their meta-information are temporarily stored in an array for the further generation of recommendations. If all analysis paths from the knowledge repository have been searched successfully, the array with the analysis paths that have fulfilled the support level is analyzed and the entries are prioritized. This analysis and prioritization process will be described with the help of Figure 2 in the following steps:

1. **Same recommendations are summarized.** In this step, all recommendations will be combined into a recommendation that would return the same value. In Figure 2, it can be seen that the click sequence of the business user occurs in three analysis paths from the knowledge repository, which are AP3, AP7 and AP8. Two analysis paths would return the same recommendation “P40” and the third analysis path would have the recommendation “P18” (marked blue). In order to avoid a double recommendation for the user, the two analysis paths (AP3 and AP7) are combined...
with the same recommendations in this step of the sequence matcher. Please note the
Meta information, such as the AP weight (click frequency and business user’s
evaluation), which is used to prioritize the recommendations. For this reason, the AP
weights are added when the same recommendations are combined, so that the
weighting for the analysis paths remains summarized during the summarization
process. The result of this step is that P40 will be recommended with an AP weight
5+8 =13. Then the step P18 will be recommended as second recommendation with
the AP weight 10.

2. **The recommendations are sorted according to the call / click frequency.** In this
step, the entries with the highest click frequency are sorted in descending order, since
it can be assumed that entries with a high click frequency also provide particularly
good results because they have already been called up several times by other
business users. The click frequency is, on the one hand, a good indicator as to
whether a click sequence has often been clicked on by business users, and on the
other hand, it represents the individual evaluation of a business user as this user has
the possibility to evaluate it directly after clicking on an analysis path. In the case of
a positive evaluation, the analysis path is increased by one counter, while a negative
value is reduced by two. The reduction by two counters is because the clicking
frequency is already increased by clicking on the analysis path and this step is first
undone after a negative evaluation and an additional counter has to be subtracted.

3. **The recommendations are sorted by Metadata.** If the recommendations according
to step 2 still have the same prioritization because of the same click frequencies, it
is checked whether there are any recommendations that have the metadata with the
same cube or dimension names, with which the business user is currently interacting.
If so, they are preferred because the probability is greater that the recommendations
from the same cube are more appropriate for the business user. Moreover, it is also
checked whether there are any recommendations made by the same user group (user
role) under the same results. If this is the case, these are preferred since the
probability is that recommendations from the same user group will be more adequate
for the business user.

### 4 Prototypical Implementation and Evaluation

In this Section, the prototype and its technologies will be first explained, followed by
the description of process flow of the RS. After that, a business scenario is defined to
demonstrate the artifact based on the prototype in workshops to validate the utility and
applicability of the artifact.

#### 4.1 Prototype of the Recommendation based Self Service BI

Based on the conceptual architecture explained above, a prototype for recommendation
based self-service BI was implemented as a proof of concept. Pentaho Business
Analytics Platform was chosen because it is an open source BI system under the GPL
license. Moreover, it includes different BI tools like reporting, dashboard and OLAP
analytics, which can be accessed via a Web browser. For the implementation of the server-side functionalities, PHP web services were used for sequence matcher and the evaluator. In addition, knowledge repository was realized using MySQL database. The client application has been implemented as an extension for the Google Chrome browser. Moreover, the plugin itself was implemented using JavaScript, CSS3, HTML5 and JQuery. In addition, JSON has been used as a lightweight format for exchanging data between the Web server and Web pages.

**Figure 3. Recommender System for Pentaho Business Analytics**

**Process Flow:** After the business user has logged-in to the BI system, the plug-in is called and all resources such as the frameworks used, the program code or the libraries are loaded and then integrated into the frontend of the BI system. As soon as the RS has been successfully loaded, the business user in the BI system receives a fixed area on the right side of the browser as we can see in Figure 3, which will be used to display the recommendations in the future. This part could be also hidden if the business user does not want to see it. As we can see at the top on the right side in Figure 3 a business user can change the support level that leads changing the number of steps are sent to the sequence matcher. Then, below this, we can see the recommendations sorted by the priority, as explained above. We see in each recommendation first the number of the recommendation followed by an element that is the same symbol as in the Pentaho Business Analytics and can be clicked, because it was implemented to allow the business user to call the function within the RS. Finally, we can see the description of the element. In the lower right hand of the figure, the business user can evaluate the previous recommendation with a thumb up or down.

**4.2 Evaluation and Discussion**

This research follows the design science research framework [25]. Therefore, the evaluation of this work and its utility and applicability was done based on the selection of the appropriate evaluation methods suggested by Hevner et al.. To achieve that,
functional (black box) and structural (white box) testing of the artifact were done by performing various tests of the whole functionalities and interfaces of the prototype to ensure the correctness and consistency of the artifact. After that, an illustrative business scenario was defined to demonstrate the utility, applicability and the organizational impact of the artifact using the prototype. These were presented in form of workshops in two companies. The first was done with one of the directors in an IT company. The second workshop was addressed to two BI power users from a production company [25, 26].

Based on a fictive sales dataset that includes the following (product, category, city, region, country and year as dimensions and price, units, turnover and profit as facts), a multidimensional data model was created and loaded into Pentaho Business Analytics. The business scenario for the expert interviews was to perform data analysis to let the business users display the turnover of all products on a graph every year. As we already mentioned in Section 3, the first phase of our approach is to extract the analysis paths of the power users. This is done by training the system to extract the analysis paths of its power users that are stored in the knowledge repository. After that, the workshops were organized in three parts. Firstly, the new concept was presented in front of the experts. Secondly, based on the business scenario, the prototype was demonstrated to show how the business user is guided by providing him with the appropriate recommendations to accomplish his data analysis. Finally, using a questionnaire-based expert interview, the architecture and its prototype was evaluated.

In terms of the utility of the new approach, the feedback that we got in the first workshop was that the recommendations were user-friendly and represented to support the business user while performing data analysis. In addition, this will improve the analytical skills of business users every time they use the system. In the second workshop, the experts said that business users always forget how to get the information they need, even if they already got the information before from the power users. Therefore, the experts concluded that our approach could help business users to do their ad-hoc analysis by themselves without asking the power users every time.

In term of the applicability of the approach, the director confirmed the applicability of the artifact and said this can offer several potentials. In the second workshop, the application of the artifact in their company was intensively discussed. The conclusion was that this approach could be applicable, but with extra operational costs, because our prototype is implemented using Pentaho Business Analytics and it cannot be transferred one to one to their company. This requires analyzing the interfaces of their BI system to adjust our prototype to suit their BI system. We were aware of this point from the beginning, but in the research, the prototype was implemented with the goal of demonstrating the new approach in many companies regardless what specific BI system they use. The conceptual architecture can be instantiated for every Web-based BI system.

Regarding the improvement of the recommendations: In the first workshop, the expert suggested that the RS could be extended by enabling business users to write comments for other business users on specific recommendations to enhance the functionality of the system. In the second workshop, based on their experience with the business users, the experts suggested improving the system with a search functionality
to allow business users to find a specific result. These recommendations and suggestions from the two workshops will be discussed and considered in our future work.

5 Conclusion

In this paper, we have presented a novel approach to realize the concept of self-service BI. It is designed to transfer the knowledge of power users to business users via recommendations. Such recommendations guide the business users while they perform their own ad-hoc data analysis to empower them to be more self-reliant and less depending on power users or IT department. This will help enterprises to improve the usage rate and the adoption of their BI systems. The conceptual architecture was described in details including the algorithm of generating the recommendations. Furthermore, the prototype and a business scenario were used in demonstration and evaluation phases to firstly illustrate the functionalities of the new approach and secondly to evaluate its results. The evaluation was done using a questionnaire-based expert interview. The target group of this interview were experts in two companies in IT and production domains. The feedback gained from the answers of the experts in the interviews will be incorporated in future work. Finally, we plan to evaluate this work in further companies and to obtain more feedback to enable the improvement of our artifact and to prove its applicability to a wide variety of BI companies.

References

7. James Richardson: Business Intelligence Adoption Trends (2011)