Monitoring Recommender Systems: A Business Intelligence Approach

Catarina Félix[1,2], Carlos Soares[1,3], Alípio Jorge[2,4], and João Vinagre[2,4]

[1] INESC TEC, Portugal
[2] Faculdade de Ciências da Universidade do Porto, Portugal
[3] Faculdade de Engenharia da Universidade do Porto, Portugal
[4] LIAAD-INESC TEC, Portugal
{cfo,joao.m.silva}@inescporto.pt, csoares@fe.up.pt, amjorge@fc.up.pt

Abstract. Recommender systems (RS) are increasingly adopted by e-business, social networks and many other user-centric websites. Based on the user's previous choices or interests, a RS suggests new items in which the user might be interested. With constant changes in user behavior, the quality of a RS may decrease over time. Therefore, we need to monitor the performance of the RS, giving timely information to management, who can than manage the RS to maximize results. Our work consists in creating a monitoring platform - based on Business Intelligence (BI) and On-line Analytical Processing (OLAP) tools - that provides information about the recommender system, in order to assess its quality, the impact it has on users and their adherence to the recommendations. We present a case study with Palco Principal[1], a social network for music.

1 Introduction

Websites like Amazon[2] or eBay[3] incorporate Recommender Systems (RS) that can use, for example, history from users purchases to train a recommendation model able to predict new products that they may be interested in.

Because the behavior of users may change [1, 2], it is possible that a RS degrades its quality over time. Also, as it is possible to have many different recommendation models (by using different algorithms or parameterization) for a given problem. Therefore, even if the performance of the current RS does not degrade, it is possible that a different model will provide better recommendations. All these factors generate the need for methods and systems to monitor or support the management of RS.

The goal of this work is to develop a Business Intelligence-based system to monitor RS. Our approach consists of implementing BI tools such as dashboards, reports and OLAP that provide information on the behavior of the RS used on

[1] http://www.palcoprincipal.com
[2] http://www.amazon.com/
[3] http://www.ebay.com/

B. Murgante et al. (Eds.): ICCSA 2014, Part VI, LNCS 8584, pp. 277–288, 2014.

a site. These tools are based on a data warehouse (DW), which is the central component of our proposal. We also illustrate the use of the proposed approach with the system that was developed for Palco Principal, a social network for music.

2 Business Intelligence

Business Intelligence is based on tools and techniques to store, manage and analyze large amounts of business data to support the decision making process. The data is stored in data warehouses (DW) and some of the tools that are commonly used for analysis include OnLine Analytical Processing (OLAP) and data mining.

A DW is a database used for storing the data needed for decision making processes in an organization. It is usually separated from the organization's operational database because the two databases are organized in a different way, as they serve different purposes.

The DW is used for information processing, allowing the analysis of consolidated and historical data. The data is stored in the DW by ETL (Extract, Transform and Load) processes: they extract data from the operational system, transform it in order to suit the needs of the decision support system and load the data into the DW. The data warehouse consists of dimension and fact tables. A fact table is the primary table in the dimensional model and also stores numerical measurements (measures) that represent the indicators that are relevant to support decisions, and the foreign keys that connect to the dimension table's primary keys [3]. The dimension attributes serve as the source for query constraints, grouping and report labels, making the dimension tables the entry point to the fact table. A generic example is: to answer the question "What was the store's revenue, by user age?", the fact table would contain the measure "revenue" and one of the dimension tables would have a "user" attribute.

2.1 Pentaho

Pentaho [4] is a Business Intelligence suite that includes a wide range of analytics tools, including data integration and On-line Analytical Processing (OLAP) analysis. It is based on open source projects and there are two versions available: Pentaho Enterprise and Pentaho Community.

Pentaho's Business Analytics Platform is Web-based, which enables it to be accessible anywhere and using any platform. Its interface allows the user to create, edit or view analysis: charts that illustrate the DW's information. Here, the administrator can also manage users, database connections and scheduled tasks.

The data integration tool of Pentaho is Spoon [5], from the Kettle [5] project. It is used to create the DW and also the ETL processes used to populate it with data from the operational databases. The ETL processes transform the data according to the requirements of the monitoring system. We can use the

Schema Workbench [5] to design the cube, with its dimensions and measures, to be used by the OLAP and other analytics tools. For data analysis, Pentaho uses the Saiku plugin [6], which is a web based analytics solution that allows users to perform OLAP analysis. Pentaho's workflow can be viewed in Fig. 1.

3 Recommender Systems

Recommender Systems [7] allow websites to dynamically display up-to-date content that suit the preferences of the user, also satisfying the users' complex and diverse needs and behaviors. This is usually accomplished by collecting site information and storing it into a database. The data typically concerns the characteristics of the content in the site (content data: e.g. the products sold in the web site), user (user data: e.g. age, gender, location) and the interactions between the user and the site (usage data: e.g. products browsed and bought, alongside with the users that performed those actions). The RS analyses information from this database and uses it to generate recommendations. Examples of services using Recommender Systems are Amazon.com and eBay [8], which recommend new products to the user, based on his browsing and purchasing history.

To generate the recommendation, as we can see in Fig. 2, the system can use different algorithms and parameter configurations to originate the models. The next phase consists in selecting the model to be used, and it can be performed manually or automatically. The selected model will generate the recommendations that will be sent to de web site.

An example of a recommendation technique is Collaborative Filtering. It suggests new items to a user based on the items the user likes and the opinion of users with similar tastes. This technique can predict the likelihood of a user being interested in one item and, then, recommend him the items he will probably like most, provided that the items haven't been yet rated (or bought, accessed, etc.) by the user [9].

Monitoring recommender systems has been recognized as an important part of the system itself. This can done obtrusively by conducting surveys or focused studies [10, 11] or by seamlessly collecting and analyzing user behavior data [12, 13]. Our work follows the later approach. We propose a business intelligence system to provide the manager of the recommender system to explore user behavior data in response to recommendations made by the system. This approach has the advantage of not having to explicitely ask questions to the users and also not making strong assumptions concerning the questions that the site manager might ask.

4 BI Architecture for Monitoring RS

Monitoring Recommender Systems enables the assessment of the behavior of recommendation models. Although, the most important aspect is the quality (e.g. whether a recommendation is followed by the user), other aspects may be

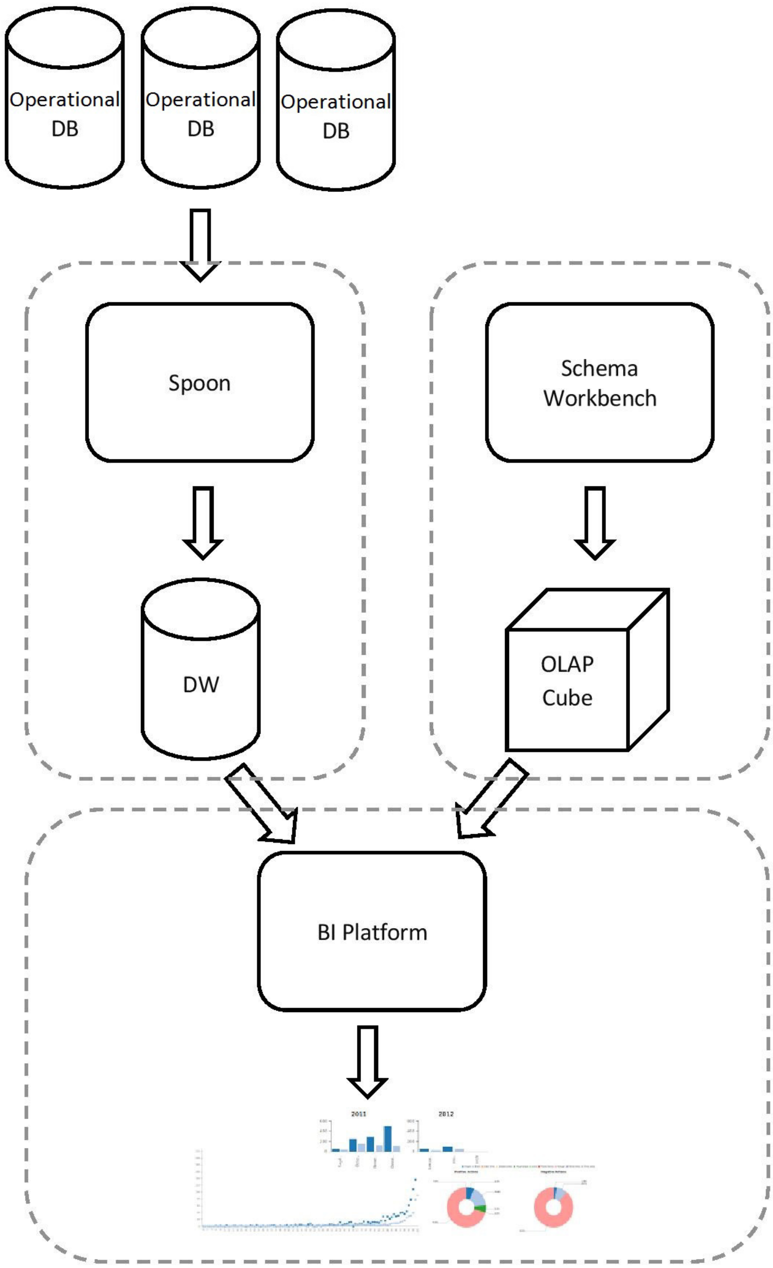

Fig. 1. Pentaho's Workflow

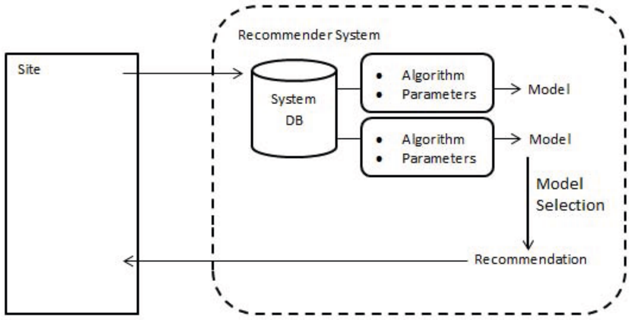

Fig. 2. Recommendation Generation Workflow

interesting as well (e.g. variety of the recommendations). Additionally, monitoring may focus on the evaluation of a single model over time or it may compare different models. This way, the website administrator can switch between recommenders in order to achieve the business goals. He/she may also take measures to improve the ones that are not behaving the way they are intended to. To do this a tool is required to support the analyses of the recommendations and help to find the circumstances under which they are performing better or worse.

Monitoring a RS enables:

- Evaluating recommendation performance: we can measure the acceptance rate of the recommendations made by this system and compare it to other values from previous points in time;
- Evaluating the recommender operation mode: this leads to being able to adapt the recommenders to the users;
- Managing recommenders: if the monitor shows that recommender A is performing worse than usual, the site administrator can switch to recommender B and see how it behaves;
- If integrated with the the BI system for the whole business, the RS monitoring system may help understand how it is contributing (or not) the the business goals.

For example, Fig. 3 shows a comparison of the number of positive actions (accepted recommendations) and negative actions (rejected recommendations) for each of the recommendation models used. The models here are represented by the date in which the system has started to use them. As we can see in the chart, some models originate more positive actions, while the negative ones are nearly the same for every model. Using this information, the site administrator can select the model to be used to generate the recommendations.

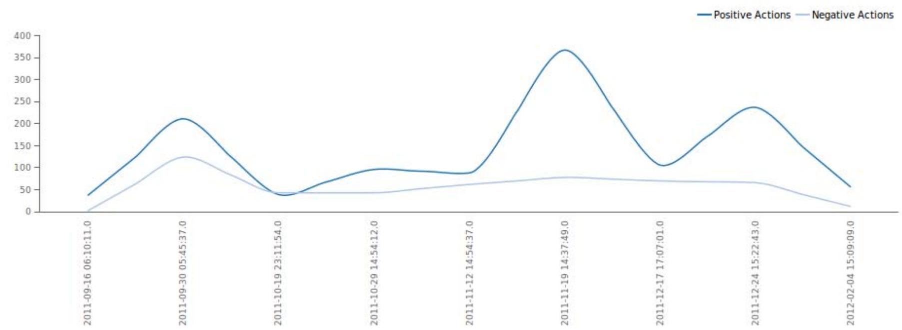

Fig. 3. Comparison of positive and negative actions for recommendation model

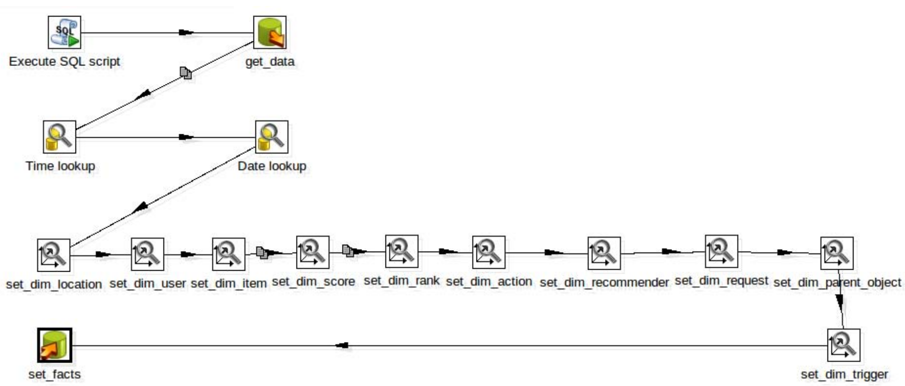

Fig. 4. Spoon Transformation

5 Case Study: Music Portal

Palco Principal (PP) is a social network based on music. Its members are Internet users with interests on that kind of content: artists and listeners. The artists upload their music to the website, along with metadata characterizing the uploaded tracks, such as genre and pace.

The listeners, upon registration, are invited to choose their preferred music genre and, then, the Recommender System can start recommending content based on the user's preferences. After listening to the recommended items, the user can either add them to a playlist or to a blacklist. These factors will hereafter be considered by the RS to generate new recommendations.

To make the most effective use of recommendations, the manager of the web site needs to know how the recommender system is behaving and how users are responding to it. Therefore, every recommendation that is shown is stored in a database. This data is processed in Spoon, using the process shown in Fig. 4, and loaded to the data warehouse represented in Fig. 5.

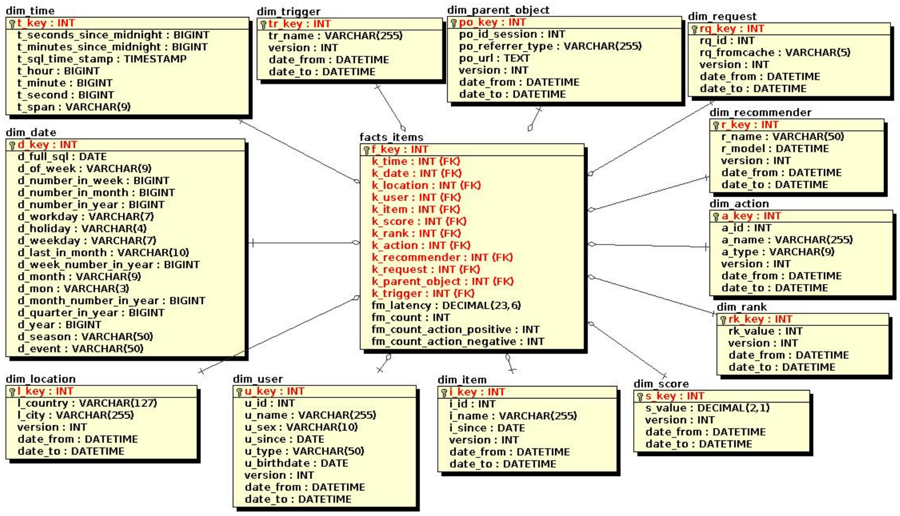

Fig. 5. PP's data warehouse schema

The data warehouse used for this job (Fig. 5) can be represented by a star schema: it is composed by a fact table surrounded by several dimension tables. This design style was chosen because it provides better query performance. This is accomplished due to the faster aggregation capabilities, when compared to the snowflake schema (that is composed by various stars that are connected to each other). This performance increase is achieved because queries can be written with simple inner joins between the fact table and the dimension tables.

Here we have the fact table surrounded by the dimension tables: *dim_time*, *dim_date*, *dim_location*, *dim_user*, *dim_item*, *dim_score*, *dim_rank*, *dim_action*, *dim_recommender*, *dim_request*, *dim_parent_object* and *dim_trigger*. The dimension tables are then linked by the foreign keys in the fact table, which also stores some measurements: *latency* (time interval between the recommendation and the user reaction) and indicators to the type of the action (positive or negative). These attributes are the ones we want to measure for the analysis. For each item recommended, a row in the fact table is created, and it is either linked to the corresponding rows on the dimension tables (if they already exist) or the dimension tables rows are created and then the fact row is linked to them.

For the proper functioning of the monitoring system we expect certain data to be imported from the database of the operational system of the website: the content of the site, the users and the offered recommendations. This way we can relate the recommendations with the characteristics of the items and of the users. We can relate any of the attributes from the recommendations with any of the attributes of the users and/or the items, as shown in Fig. 6, which shows a fragment of the schema in Figure 5.

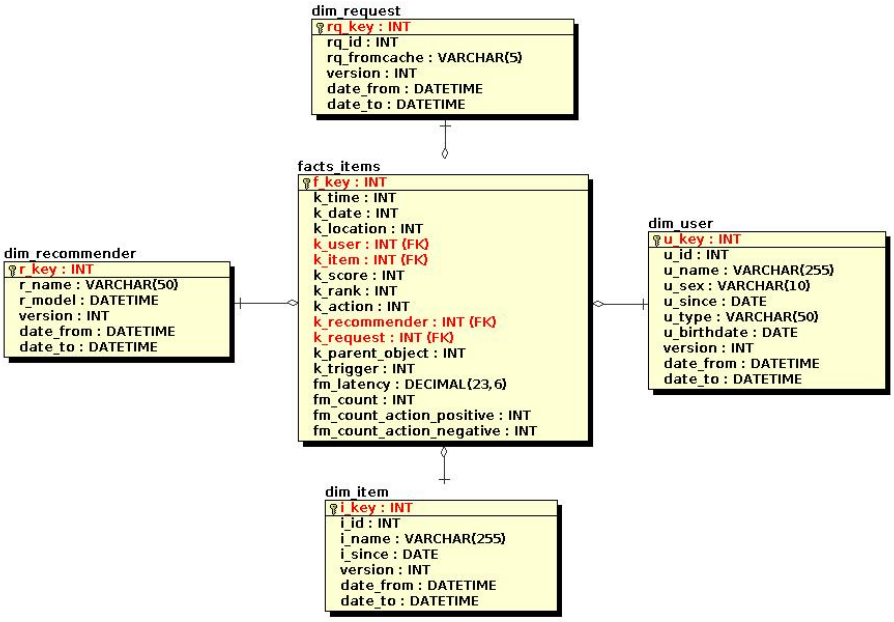

Fig. 6. Relation between Users, Items and Recommendations

5.1 Recommendation Monitoring Functionalities

With this DW we can, for example, analyze the difference between positive (adding an item to a playlist) and negative (adding the item to the blacklist) actions, taking into consideration the date when the recommendations were made, as shown in Fig. 7. In this figure, we can see that the number of positive actions is always bigger than the negative ones, but also that the system registered more actions (positive or negative) during 2011 than during 2012. With these results the administrator could decide to switch back to the recommender used in 2011, if it was a different one, in order to obtain better results.

Another possible analysis can be in terms of the time of the day when the recommendations are made and the actions on those recommendations are performed, as depicted in Fig. 8. In the chart, we can see a peak of actions in the afternoon. This can be a consequence of the global usage of the site. However, if during this time of the day the system was using a different recommender, the administrator can decide to use that recommender for the rest of the day, expecting to increase user interaction with the website. On the other hand, if the recommendation model is the same, this could mean that different models should be used at different times of the day. He could also take precautions (in terms of system resources) so that during that period there will not occur problems in the website due to an overload of requests.

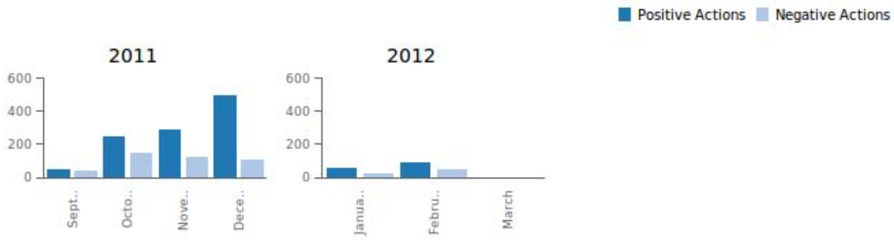

Fig. 7. Positive and Negative Actions by Date (Year and Month)

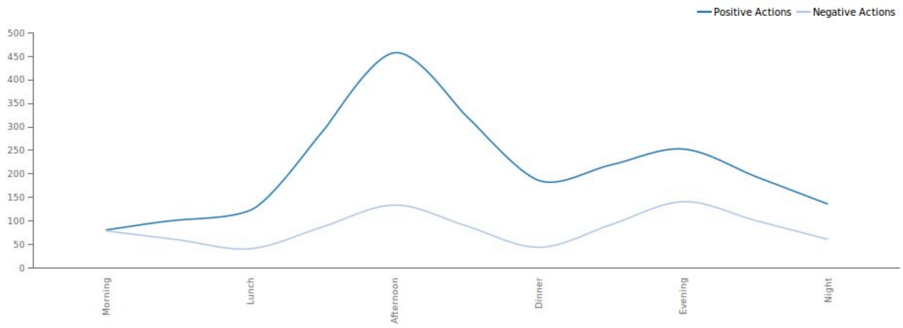

Fig. 8. Positive and Negative Actions by time of day

The site administrator can also analyze reactions to recommendations according to the gender of the user Fig. 9 shows that, while the gender has nearny no impact in negative actions, male users tend to perform more positive actions than females.

The administrator can also view the distribution of actions in terms of the user countries. Fig. 10 shows that most actions are made by users from Portugal (70.4% of positive and 88.8% of negative actions), followed by Brazil, Angola and Mozambique. This is expected because the site was first launched in Portugal and the other portuguese-speaking countries.

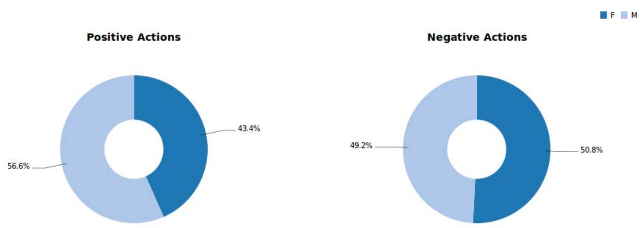

Fig. 9. Positive and Negative Actions by user gender

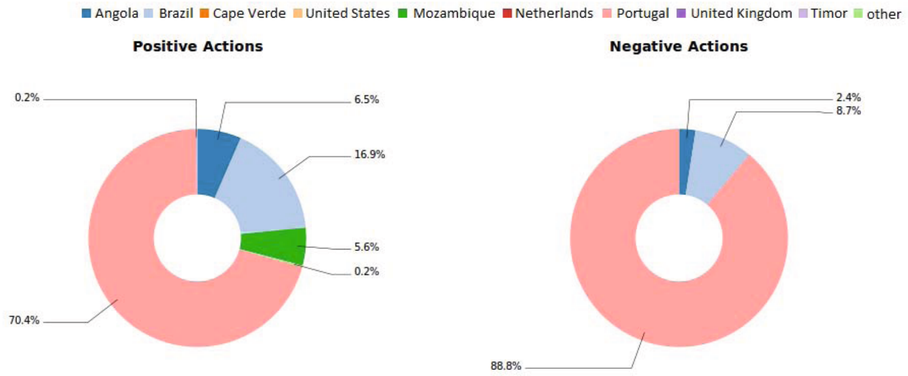

Fig. 10. Positive and negative actions by User's Country

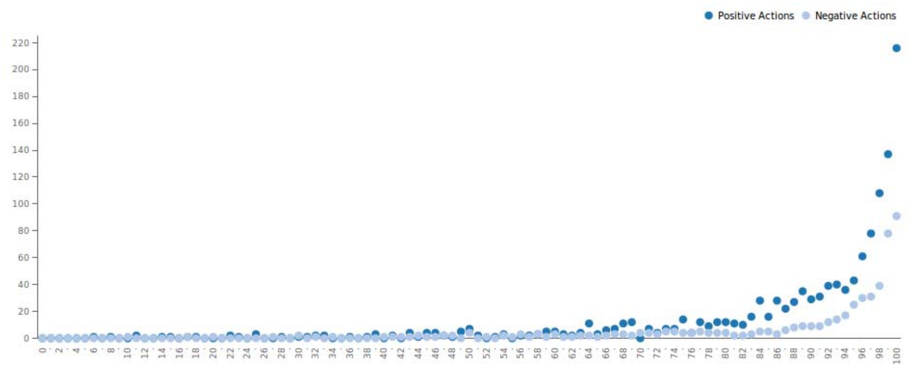

Fig. 11. Positive and negative actions by Item's Rank in Previous Recommendations

We may also want to know if the rank of the item in the recommendation affects the likelihood of it being accepted by users. Fig. 11 shows that when the rank is high (i.e., the recommendation is stronger), there are more actions on the item. The administrator can then tune the recommendation system based on this fact.

In this system, a recommender model is identified by the date and time when it started being used and not only by the name of the algorithm that generated it. This happens because new models may be generated using the same algorithm (and even the same parameter values), using different sets of data. This is necessary for several reasons, including the fact that there are constantly new users and items, and the model information needs to be updated. Also, the recommendations can be made from cache memory, when similar requests have been made. In Fig. 12 we can see the difference between the newly

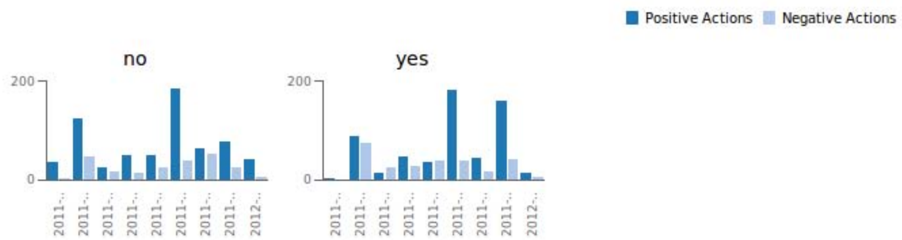

Fig. 12. Positive and negative actions by Recommender Model and whether the Recommendation was (chart with title "yes") or not (chart with title "no") made from the cache

calculated recommendations (left chart) and the ones made from cache memory (right chart).

5.2 Summary

Using Business Intelligence tools, the site administrator can monitor and assess the performance of the recommendation systems, simply by analyzing charts containing data from the acceptance or rejection of the recommendations. These analyses can help the administrator in tasks such as switching on and off the recommenders, or tuning their parameters differently, in order to improve the system's performance.

The system can also be used to analyze other aspects that were not illustrated here. Information such as, for instance, what are the most recommended items, what type of music is most often recommended, may also be useful for the site administrator.

6 Conclusions and Future Work

With the introduction of recommender systems in e-business and social websites, and since the performance of the recommendation models may vary over time (due to, for example, changes in user behavior), there is a need to develop tools for monitoring them.

In this paper we have proposed the use of Business Intelligence Tools in order to complete that task. Using these tools, system administrators may assess the online performance of the system and use this information to improve the overall system's quality, either by switching between algorithms, triggering model updates or fine-tuning parameters.

Despite the difficulty in evaluating our proposal, it would be worthwhile considering a process that allows the site manager to score the utility of each feature at each moment.

It would also be useful to have a tool or process that would grant the administrator the capacity of measuring the impact of the monitoring architecture and tools on a production system.

Additionally, it would be useful to automate the management process. This means the ability to dynamically select different recommendation algorithms, calibrate parameters or, at least, support the site administrator in that task by providing recommendations. One approach is metalearning, which consists of modeling the relationship between data and the performance of the algorithms [14]. Such a tool could use the data store in the monitoring database described in this paper.

Acknowledgments. This work is partially funded by FCT/MEC through PIDDAC and ERDF/ON2 within project NORTE-07-0124-FEDER-000059.

References

1. Billsus, D., Pazzani, M.: User modeling for adaptive news access. User Modeling and User-Adapted Interaction 10(2-3), 147–180 (2000)
2. Koychev, I., Schwab, I.: Adaptation to drifting user's interests. In: Proceedings of ECML2000 Workshop: Machine Learning in New Information Age, pp. 39–46 (2000)
3. Kimball, R.: The data warehouse toolkit: The Complete Guide to Dimensional Modeling. Wiley (2006)
4. Pentaho: Pentaho website, `http://www.pentaho.com/` (accessed: January 17, 2014)
5. Pentaho: Pentaho community website, `http://community.pentaho.com/` (accessed: January 17, 2014)
6. Saiku: Saiku website, `http://meteorite.bi/saiku` (accessed: January 17, 2014)
7. Resnick, P., Varian, H.R.: Recommender systems. Commun. ACM 40(3), 56–58 (1997)
8. Schafer, J.B., Konstan, J., Riedl, J.: Recommender systems in e-commerce. In: Proceedings of the 1st ACM Conference on Electronic Commerce. EC 1999, pp. 158–166. ACM, New York (1999)
9. Sarwar, B., Karypis, G., Konstan, J., Riedl, J.: Item-based collaborative filtering recommendation algorithms. In: Proceedings of the 10th International Conference on World Wide Web. WWW 2001, pp. 285–295. ACM, New York (2001)
10. Nowak, M., Nass, C.: Effects of behavior monitoring and perceived system benefit in online recommender systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI 2012, pp. 2243–2246. ACM, New York (2012)
11. Pu, P., Chen, L., Hu, R.: A user-centric evaluation framework for recommender systems. In: Proceedings of the Fifth ACM Conference on Recommender Systems. RecSys 2011, pp. 157–164. ACM, New York (2011)
12. Domingues, M.A., Gouyon, F., Jorge, A.M., Leal, J.P., Vinagre, J., Lemos, L., Sordo, M.: Combining usage and content in an online recommendation system for music in the long tail. IJMIR 2(1), 3–13 (2013)
13. Kohavi, R., Longbotham, R., Walker, T.: Online experiments: Practical lessons. IEEE Computer 43(9), 82–85 (2010)
14. Brazdil, P., Giraud-carrier, C., Soares, C., Vilalta, R.: Metalearning: Applications to Data Mining. In: Cognitive Technologies. Springer (2009)