A Comprehensive Reference Model for Personalized Recommender Systems

Matthias Breyer, Kawa Nazemi, Christian Stab, Dirk Burkhardt, and Arjan Kuijper

Fraunhofer Institute for Computer Graphics Research,
Fraunhofer Str. 5, 64283 Darmstadt, Germany
{matthias.breyer,kawa.nazemi,christian.stab,dirk.burkhardt,
arjan.kuijper}@igd.fraunhofer.de

Abstract. Existing reference models for recommender systems are on an abstract level of detail or do not point out the processes and transitions of recommendation systems. However, this information is relevant for developers to design or improve recommendation systems. Even so, users need some background information of the calculation process to understand the process and accept or configure these systems proper. In this paper we present a comprehensive reference model for recommender systems which conjuncts the recommendation processes on an adequate level of detail. To achieve this, the processes of content-based and collaboration-based systems are merged and extended by the transitions and phases of hybrid systems. Furthermore, the algorithms which can be applied in the phases of the model are examined to identify the data flow between these phases. With our model those information of the recommendation calculation process can be identified, which encourages the traceability and thus the acceptance of recommendations.

Keywords: recommender system, recommender reference model, recommendation calculation, calculation information, data flow.

1 Introduction

In internet-based information systems the user is confronted with a growing amount of data, which makes it difficult to identify the relevant information. This is not only a problem for information seekers but also for information providers. In particular for commercial enterprises selling products on the internet, the demand for offering adequate information to a user is crucial. The financial success of their online selling platform directly depends on the user-centered identification and presentation of relevant information [10].

To assist the user in finding relevant information, three complimentary methods can be applied: classification and structuring, search-engines, or recommendation systems. The classification and structuring approach is suitable for an explorative system in which the user can browse through and easily gather an overview of the information. This may be time consuming and ineffective for non-explorative usage scenarios. Search engines are an established approach commonly users are familiar with. The main drawback of search engines is the need for appropriate search queries

M.J. Smith, G. Salvendy (Eds.): Human Interface, Part I, HCII 2011, LNCS 6771, pp. 528–537, 2011.

the user has to declare. Especially if the user has no clue about available information this problem deteriorates. To overcome these issues recommender systems (RS) can be integrated to identify and present information artifacts to the users which match their individual needs and preferences. Using RS the user can benefit from these recommendations with a minimum of inputs specifically stated to the RS. Furthermore, recommendation systems as merchandising tools individually advertise products and thus can enable increasing sales.

In this paper we present a comprehensive reference model for personalized recommendation systems. The model unifies relevant aspects of existing and established PRS reference models. The objective of this reference model is to show the overall interplay of the components and coevally accenting the impact of input parameters and transferred data in the recommendation calculation process. As far as we know there exists no such reference model on this level of detail.

The resulting PSR model is a connected and comprehensive reference model for existing personalized recommendation approaches comprising relevant detail information. The comprehensive reference model is especially suitable for further developments and optimizations of PRS.

In the next section we introduce basic information for recommender systems. The third section will give an overview about related works on which the presented reference model for recommender systems is built on. The comprehensive reference model itself will be described in section four. The last section summarizes the paper.

2 Basics for Recommender Systems

For a common understanding of the construction of a comprehensive reference model for recommender systems some basic information is essential. In this section these basics will be described.

2.1 User Feedback as Input Data

To obtain user and usage data as input data for recommender systems basically three approaches exists which can be applied in combination [8]:

Explicit feedback/direct learning technique. The user explicitly states preference or socio-demographic information in input fields of the 'conversational system'. The main drawback of this approach is users have to be motivated to actively perform these inputs [10]. Furthermore, users have to interpret and understand the feedback mechanisms correct in order to achieve proper recommendations [2].

System-driven explicit feedback/partially direct learning technique. The user explicitly states preferences of socio-demographic information in input fields, but the system determines when and which information is necessary. The objective is to minimize the number of explicitly stated information thus minimizing user interactions which are not related to his originally task.

Implicit feedback/indirect or transparent learning technique. User preferences are deduced by analyzing the interactions ('behavioral usage data', like mouse movements, selections, orders, etc.) of the user [18].

2.2 Classification Criteria: Personalization and Functional Input Sources

Recommender systems are mainly classified in personalized, non-personalized, and supporting systems [1]. In between personalized and non-personalized systems can be settled a fourth category named group-based recommendation systems [7]. In this paper we focus on personalized recommender systems (PRS) which predict individual preference values for information artifacts [13].

For PRS various classification criteria were identified in previous works, of which the criterion by functional input seems to be used most often. To support comparability this criterion will be used in this paper equally. This criterion partitions PRS in content-based, collaboration-based, and hybrid approaches [1,11]. Content-based PRS (item-based/feature-based) extract significant features out of the content (domain-space), based on which similar information artifacts are identified to these artifacts the user rated positive (implicit or explicit, see 2.1). In contrast, collaboration-based RPS compare characteristics from the user models itself to recommend information artifacts ranked positive by similar user models (targeted customer inputs and/or community inputs). Hybrid PRS combine both of these approaches to overcome drawbacks of each separately.

With the evolving Web2.0 and social communities the class of group-based PRS (social-based, social-tagging-based) is stated as a fourth category [10]. Due to the fact these systems use similar functional input sources these systems will here be classified as collaboration based systems. Furthermore, there exist the categorizations of rule-based and knowledge-based systems. However, these require predefined rules and knowledge, so they are not fully automatic and will not be considered further on [5].

3 Related Work

After this short introduction to the basics for recommendation systems and the classification criteria applied in this paper, in this section related works will be discussed. These related works are also the basics for the designed comprehensive reference model in the subsequent section.

3.1 Process Models for Recommender Systems

In the field of recommendation systems some models exist which describe the process of recommendation calculation in different abstraction levels. Often, these models are influenced by the point of view the model is developed for. To obtain an overview of these models the most important are presented here, which lay the groundwork for the comprehensive reference model.

Abstract model of recommendation systems. [5] described this model which consists of four components: (i) background knowledge: knowledge domain and previously gathered user/usage data (see section 2.1), (ii) user/usage data of the current interaction session, (iii) an recommendation algorithm which calculated user preference approximations based on the background knowledge and the current user/usage data, and (iv) presentation of the recommendations.

Model including activities and information flows. [18] present an extended recommendation model which includes activities between the involved responsibility parties and the resulting information flows between these parties. This model is defined on a very abstract level of detail to cover as many recommendation activities as possible. For a concrete usage scenario these activities significantly vary. Furthermore, not each of the activities of information flows has to be concretized or realized.

Conceptional classification model. A particularized description model is presented in [15]. In this model the characteristics of recommendation systems are differentiated in functional in- and outputs (see section 2.2), recommendation method (see section 3.3), and additional design decisions. The model differentiates the functional inputs in those that describe the current user and those representing the whole community. Additional design decisions are determined by the recommendation system provider (before runtime) but determine the recommendation quality and the kind of presentation. Thus, these decisions have a major impact for the users' perception of the recommendations which furthermore constitute the success of recommendation systems.

3.2 Recommendation Calculation Processes for Recommender Systems

In this section the processes for recommendation calculation will be described which are used to predict preference values for the current user. The presented calculation processes are oriented to the differentiation of functional input sources (see section 2.2). The subsequent section 3.3 lists calculation functions which can be utilized within the calculation processes presented here.

Calculation process of content-based systems. Content-based systems calculate preference values based on objective properties, the features. Using these features an item-to-item-correlation is constructed based on the preferences of the current user [5]. The objective is to identify unrated information artifacts which is alike rated information (information the system has preference values for).

The calculation process of content-based systems is composed of two phases, as shown in Fig. 1 (left). In the first phase 'feature extraction' objective properties are extracted for the information artifacts, with the aim of identifying the best describing characteristics of this artifact. In the second phase 'filtering and preference value prediction' a heuristic or model-based mathematical function is used to calculate the similarity of informational artifacts. The applied functions can be distinguished in exact-match and best-match approaches.

Calculation process of collaboration-based systems. Collaboration-based recommendation systems recommend information artifacts to the current user based on community data, the user/usage data of other users. The process is composed of three phases, as shown in Fig. 1 (right). In the first phase a user-to-user-correlation is constructed based on a comparison of acquired preferences of these users (or user account). The aim of the second phase is to identify a homogeneous neighborhood group of mostly similar users. In the last phase the preference value predictions are calculated. The resulting recommendations for the current user are these information artifacts which are preferred by other users which have mostly similar preferences to the current one [5,13,18].

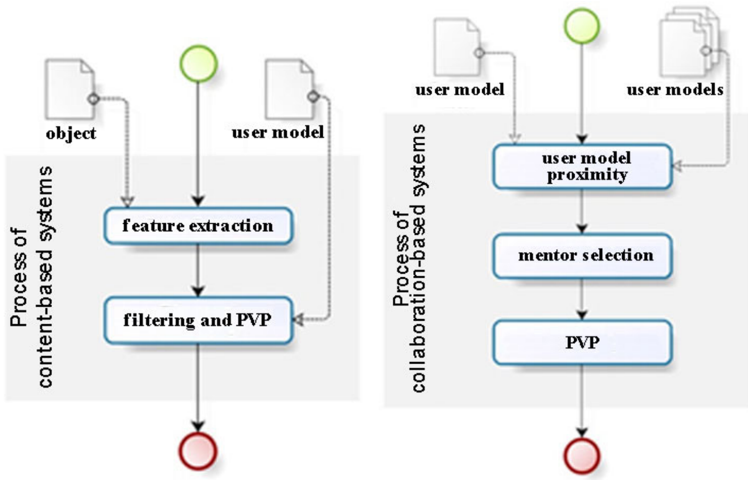

Fig. 1. Calculation processes for content-based (left) and collaboration based (right) systems

Hybrid systems. Hybrid recommendation systems combine the approaches of content-based and collaboration-based processes. The objective of the combination is to reduce weaknesses and problems when applying the above processes separately [3]. [1] identified four categories of combinations for hybrid systems:

(i) Combination of results: both processes are integrated in independently, the resulting preference predictions are combined afterwards.

(ii) Content-based aspects in collaboration-based methods: content-based user models are utilized to calculate homogeneous neighborhood groups.

(iii) Collaboration-based aspects in content-based methods: the predominant approach is a dimension reduction of content-based user models based on collaboratively constructed neighborhood groups.

(iv) Unified model: both processes are modeled in one joint model which is used to calculate the preference value predictions. This model can be probabilistic, rule-based, or knowledge-based.

3.3 Computation Functions in the Recommendation Process

In the phases of the recommendation calculation processes mathematical functions can be applied. In this section these some applicable functions will be introduced, which is necessary to identify the information flows in the comprehensive reference model for recommender systems.

Feature extraction. For the phase of feature extraction the text-based approach TF-IDF is mentioned which identifies relevant terms of documents and their significances. Using latent semantic analysis or respectively latent semantic indexing the quality of this feature extraction can be increased [1].

Proximity calculation. To calculate the proximity between user models five strategies can be distinguished [1,9,12,13,14,16,17]:

1. 1. distance-based: Lq-distance, Mahalanobis-distance,
2. 2. correlation-based: (constrained/weighted) pearson correlation,
3. 3. vector-space-based : cosine-based approach, adjusted cosine similariy, TF-IDF-schema, rocchio algorithm,
4. 4. rule-based: boolean matching, decision trees, association rule discovery, and
5. 5. probabilistic: Hidden Markov Model, dynamic Bayesian networks, Bayesian belief networks, naïve Bayesian classifier, maximum-entropy-method, probabilistic relational model, probabilistic latent semantic analysis.

Classification. A classification leads to groups of 'similar' objects based on a metric function, which is here based on features. Here three approaches can be distinguished [4,6,19,20]:

1. 1. Nearest Neighbor: kNN,
2. 2. linear classifier: (modified) least squares, Support Vector Mashines, modified naïve Bayes, and
3. 3. Cluster-algorithms: Tree-Clustering, Expectation-Maximization-Algorithm, K-Means, Gibbs Sampling, Monte-Carlo-Methods, Horting-Algorithm

Aggregation. Aggregation functions unite multiple values to one. Approaches are the simple average (means), weighed sum, adjusted weighed sum or the regression model. A similar approach are providing the (bi-polar) slope one-algorithms [1,13,14].

4 Comprehensive Reference Model for Recommender Systems

In this section the comprehensive reference model for recommender systems will be presented. The model is based on the Recommendation Calculation Processes for Recommender Systems (see section 3.2) and summarizes the main aspects of related work reference models (see section 3.1). Furthermore the information flows in the reference model will be identified.

4.1 Aggregation of Existing Process Models

The existing process models are described in section 3.2. To achieve the comprehensive reference model, the processes of content- and collaborative-based recommendation systems had been combined and extended by the transitions and phases of hybrid systems.

The resulting comprehensive reference model for recommendation systems is presented in Fig. 2. The entry point for this model is the green node at the upper side, the endpoint with the resulting preference value predictions is the red node at the bottom side. The yellow squares are decision points which path has to be followed. The decision itself is mainly determined by design conditions. Input data for the reference model are the objects itself, or to be precise, features of these objects for which the preference value predictions are to be calculated, as well as user models, containing processed user information (like interactions or selection/buy decisions). These user models are differentiated in the user model for the current user for whom the preference values are to predict, and the user models of other users.

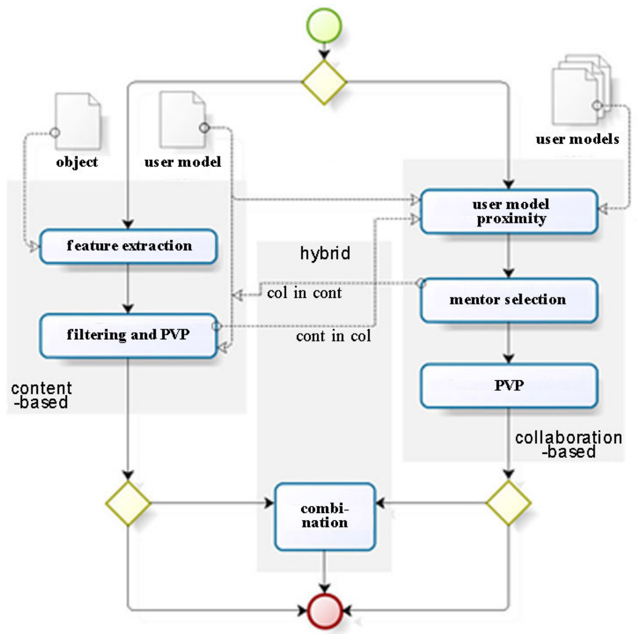

Fig. 2. Comprehensive reference model for recommendation systems

In Fig. 2 the process of content-based recommendation systems is presented in the left grey box, whereas the process of collaboration-based recommendation systems is visible in the right grey box. In between these two boxes transitions for possible hybrid combinations are displayed to illustrate the intersections between these two classes of recommendation systems. The smaller grey box at the bottom side of the model represents the class of hybrid recommendation systems which combines the results the two processes calculated independently. The residual category of unified models (see section 3.2) is not distanced in form of transitions or phases, but is a form auf instantiation of the whole process model.

4.2 Data Flow in the Transitions of the Reference Model

In Fig. 2 some transitions are displayed between the phases of the recommendation processes. These been examined for their data flow. Therefore, the computation functions described in section 3.3, which may be executed in the phases, had been taken into account. These computation functions expect specific input data and process them to a specific output data. Therefore these data is passed between the connected phases when recommendations are calculated. In Fig. 3 these annotated data flow in the transition is presented.

The features of the objects may be extracted using feature extraction algorithms. The resulting feature vector is passed on to the phase ‘filtering and preference value prediction’. In conjunction with the feature vector of the current users’ user model distance values are calculated using proximity calculation algorithms. The results of the content-based recommendation process are the preference value predictions.

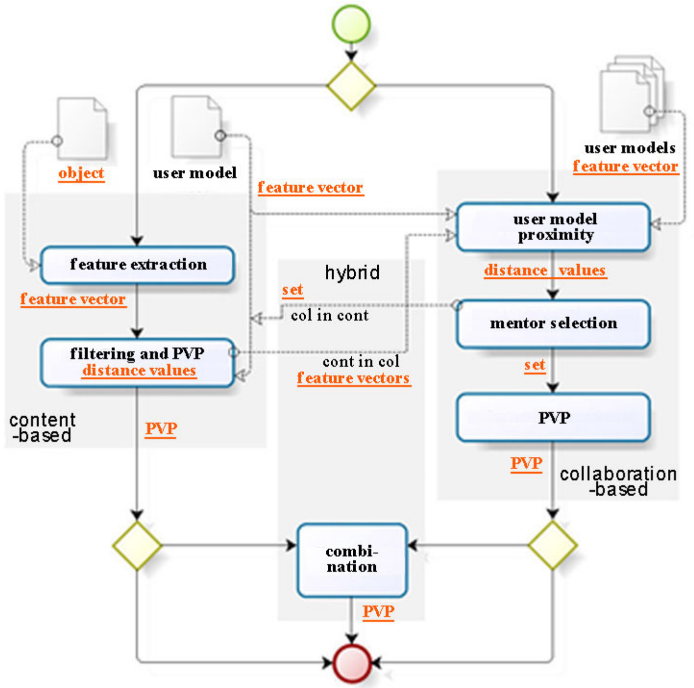

Fig. 3. Comprehensive reference model for recommendation systems extended by annotations of the data flow at the transitions

For the collaboration-based recommendation process both feature vectors of the current users' user model and the user models of other users are the input for the phase 'user model proximity' calculation. Here again proximity calculation algorithms are applied which generate distance values. These values flow into the phase 'mentor selection' where classification algorithms are applied. Using the resulting user model set describing users likewise similar preferences, in the phase 'preference value prediction' these predictions are calculated.

For hybrid systems combining the processes the one approach is to hand over the set of mentors to the content-specific phase 'filtering and preference value prediction'. The other approach of bringing aspect of one process into the other process is to pass the resulting feature vector of the current users' user model (containing the process result of the objects feature vector) to the collaboration-specific phase 'user model proximity calculation'. The category of combining the results uses aggregation functions to merge the calculated preference value predictions.

5 Conclusion

For the success of recommendation systems in applications it is crucial the current user accepts recommended objects. The acceptance depends on the traceability, which in turn depends on the information the user may perceive. Using the model presented in this paper those information of the recommendation calculation process can be identified which should be depicted to the user to enhance traceability. So the user can better understand why objects had been recommended. This may increase the success and returns for the vendors utilizing the system.

In this paper we presented a comprehensive reference model for recommender systems. This reference model conjuncts the processes of content-based and collaboration-based recommendation systems and extends these processes with the transitions and phases of the category of hybrid systems. Furthermore, the algorithms which can be applied in the phases of the model had been examined to identify the data flow between these phases.

The resulting reference model presents the state of the art approaches of recommender systems to calculate preference value predictions. Using this reference model, developers can systematically construct new recommendation systems or develop further improvements and extensions of existing recommendation systems.

Acknowledgments. The here described work was developed as a part of the Core-Technology-Cluster for Innovative User Interfaces and Visualizations of the THESEUS Program, a 60-month program partially funded by the German Federal Ministry of Economics and Technology.

References

1. Adomavicius, G., Alexander, T.: Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE Trans. on Knowledge and Data Engineering 17(6), 734–749 (2005)
2. Anick, P.: Using terminological feedback for web search refinement: a log-based study. In: SIGIR 2003: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 88–95. ACM Press, New York (2003)
3. Balabanovic, M., Yoav, S.: Fab: content-based, collaborative recommendation. Communications of the ACM 40(3), 66–72 (1997)
4. Breese, J.S., Heckerman, D., Kadie, C.: Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In: Proceedings of the 14th Conference of Uncertainty in Artificial Intelligence, pp. 43–52. Morgan Kaufmann, San Francisco (1998)
5. Burke, R.: Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction 12(4), 331–370 (2002)
6. Iyengar, V.S., Zhang, T.: Empirical Study of Recommender Systems Using Linear Classifiers. In: Cheung, D., Williams, G.J., Li, Q. (eds.) PAKDD 2001. LNCS (LNAI), vol. 2035, pp. 16–27. Springer, Heidelberg (2001)
7. McCarthy, K., Salamo, M., Coyle, L., Mc-Ginty, L., Smyth, B., Nixon, P.: Group recommender systems: a critiquing based approach. In: Paris, C., Sidner, C.L. (eds.) IUI, pp. 267–269. ACM Press, New York (2006)
8. Micarelli, A., Gasparetti, F., Sciarrone, F., Gauch, S.: Personalized Search on the World Wide Web. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 195–230. Springer, Heidelberg (2007)
9. Min, Sunghwan und Ingoo Han: Optimizing Collaborative Filtering Recommender Systems. In: Szczepaniak, Piotr S., Janusz Kacprzyk, Adam Niewiadomski (Ed.): AWIC, LNCS Vol. 3528, pp. 313-319, Springer, (2005).

10. Neumann, A.W.: Recommender Systems for Scientific and Technical Information Providers. Doctor thesis, University of Karlsruhe, Germany (2008)
11. Resnick, P., Varian, H.R.: Recommender Systems – Introduction to the Special Section. Communications of the ACM 40(3), 56–58 (1997)
12. Rocchio, J.J.: Relevance Feedback in Information Retrieval. In: Salton, G. (ed.) The SMART Retrieval System – Experiments in Automatic Document Processing. Prentice Hall, Englewood (1971)
13. Runte, M.: Personalisierung im Internet - Individualisierte Angebote mit Collaborative Filtering. Doctor thesis, University of Kiel, Germany (2000)
14. Sarwar, B., Karypis, G., Konstan, J., Riedl, J.: Item-based Collaborative Filtering Recommendation Algorithms. In: Proceedings of the 10th International World Wide Web Conference, pp. 285–295 (2001)
15. Schafer, J.B., Konstan, J.A., Riedl, J.: E-Commerce Recommendation Applications. Data Mining and Knowledge Discovery 5(1-2), 115–153 (2001)
16. Shardanand, U., Maes, P.: Social Information Filtering: Algorithms for Automating 'Word of Mouth'. In: Proceedings of the Conference on Human Factors in Computing Systems (1995)
17. Su, X., Khoshgoftaar, T.M.: A Survey of Collaborative Filtering Techniques. Advances in Artificial Intelligence, 1–20 (2009)
18. Terveen, L., Hill, W.: Beyond Recommender Systems: Helping People Help Each Other. In: Carroll, J. (ed.) HCI in the New Millenium. Addison-Wesley, Reading (2001)
19. Ungar, L., Foster, D.: Clustering Methods for Collaborative Filtering. In: Workshop on Recommendation Systems. AAAI Press, Menlo Park (1998)
20. Zhang, T., Iyengar, V.S.: Recommender systems using linear classifiers. J. Mach. Learn. Res. 2, 313–334 (2002)