

# Recommender Systems in Computer Science and Information Systems - a Landscape of Research

Dietmar Jannach<sup>1</sup>, Markus Zanker<sup>2</sup>, Mouzhi Ge<sup>3</sup>, and Marian Gröning<sup>1</sup>

<sup>1</sup> TU Dortmund, 44227 Dortmund, Germany

{Dietmar.Jannach, Marian.Groening}@tu-dortmund.de

<sup>2</sup> Alpen-Adria-Universität Klagenfurt, 9020 Klagenfurt, Austria

Markus.Zanker@aau.at

<sup>3</sup> Bundeswehr University Munich, 85577 Neubiberg, Germany

Mouzhi.Ge@unibw.de

**Abstract.** The paper reviews and classifies recent research in recommender systems both in the field of Computer Science and Information Systems. The goal of this work is to identify existing trends, open issues and possible directions for future research. Our analysis is based on a review of 330 papers on recommender systems, which were published in high-impact conferences and journals during the past five years (2006-2011). We provide a state-of-the-art review on recommender systems, propose future research opportunities for recommender systems in both computer science and information system community, and indicate how the research avenues of both communities might partly converge.

## 1 Introduction

Rooted in the fields of information retrieval (IR), machine learning (ML) and decision support systems (DSS), recommender systems (RS) have emerged as a research field of their own during the last twenty years. Recommender systems propose ranked lists of items (that are subsets of a larger collection) according to their presumed relevance to individual users. Relevance is determined from explicit and implicit user feedback such as ratings on items, commercial transactions or explicitly stated requirements [1]. With the rapid growth of electronic commerce, the ubiquity of mobile information access and the advent of the Social Web, the interest in RS research has grown enormously during the past years. This is for example documented by the rapidly growing ACM Recommender Systems conference series as well as by the publication of various focused journal special issues and books. The reasons for this high attractiveness of the field are manifold and include highly visible competitions such as the Netflix prize, increased industrial interest or the new application opportunities for recommendation techniques in the Mobile and Social Web. Based on what is the dominant goal of a recommender system, research can be considered from a variety of different perspectives:

- The IR perspective addresses the problem of information overload. The purpose of a RS is therefore to identify the items relevant to a user’s information need.
- The ML perspective on recommender systems is to learn a model that predicts the user feedback on a specific item as accurately as possible. The goal is therefore to reduce the error between the predicted and the actual feedback (i.e. typically a rating value) of a given user.
- From the point of view of DSS, recommender systems can be understood as tools supporting consumers in their decision making process. Therefore, the ultimate goal of a RS is to increase the quality of decisions made. Measuring the achievement of this goal is less straightforward compared to the goals of the two previous perspectives. A variety of factors actually influence the users’ decision making such as their appreciation of the system, trust in the information and service provider, experience and domain expertise, word-of-mouth as well as their real preferences. Researchers in marketing, e-commerce and information systems research might be interested in proxies such as the impact of such systems on customer behavior, loyalty and sales figures.

Given this diversity of research perspectives, the goal of our work is to review and classify recent research in recommender systems in order to quantify the research interests and identify opportunities for future research. In addition, this analysis should serve as a basis to understand limitations of current research practice in this field. As RS are IT applications we naturally limit our analysis to publications in the neighboring fields of Computer Science (CS) and Information Systems (IS). In the next section, we will describe the methodology of our literature analysis. In Section 3 we present detailed findings and conclude with a discussion on under-researched areas.

## 2 Methodology

We systematically evaluated all publications of a pre-defined set of high-impact journals and conferences in the fields of Computer Science (CS) and Information Systems (IS) during the period from January 2006 to July 2011. We included both journal articles as well as full papers appearing in conference proceedings. In particular, we considered those journals, where special issues on recommender systems have appeared. Table 1 lists the publication outlets, their type, i.e. journal (jrnl) or conference proceedings (proc), their presumed belonging to either CS or IS and the respective number of publications considered for further analysis. In total 330 publications have been identified, out of which 73 appeared in journals and 257 in conference proceedings. 65 publications (~20% from total) appeared in outlets that belong to the IS community, if such an attribution is permitted. Not astoundingly, the newly established ACM conference series on RS is the single most important publication venue in this field. We classify publications according to the following scheme presented in Table 2. The classification task has been performed by the authors. We will discuss the possible evaluations for each class attribute in conjunction with the results in the next section.

| Name  | Type  | Field | Nbr. |
|---|-------|-------|------|
| ACM Conf. on Human Factors in Comp. Syst. (CHI) | proc  | CS    | 13   |
| ACM Conf. on Recommender Syst. (RecSys)         | proc  | CS    | 86   |
| Int. Conf. on Int. User Interfaces (IUI)        | proc  | CS    | 17   |
| Int. Conf. on Knowl. Disc. and DM (SIGKDD)      | proc  | CS    | 22   |
| Int. Conf. on Res. and Dev. in IR (SIGIR)       | proc  | CS    | 33   |
| Int. Conf. on World Wide Web (WWW)              | proc  | CS    | 21   |
| Int. Joint Conf. on AI (IJCAI)                  | proc  | CS    | 13   |
| AAAI Conf. on AI (AAAI)                         | proc  | CS    | 10   |
| Int. Conf. on Data Mining (ICDM)                | proc  | CS    | 5    |
| Americas Conf. on Information Systems (AMCIS)   | proc  | IS    | 8    |
| European Conf. on Information Systems (ECIS)    | proc  | IS    | 6    |
| Int. Conf. on Information Systems (ICIS)        | proc  | IS    | 7    |
| Med. Conf. on Information Systems (MCIS)        | proc  | IS    | 5    |
| Pac. Asia Conf. on Information Systems (PACIS)  | proc  | IS    | 11   |
| ACM Trans. on Intell. Syst. and Techn. (TOIST)  | jrnal | CS    | 6    |
| ACM Trans. on the Web (TWeb)                    | jrnal | CS    | 5    |
| AI Comm.  | jrnal | CS    | 12   |
| IEEE Intelligent Systems                        | jrnal | CS    | 14   |
| Int. Jrnal. of Human Computer Studies (IJHCS)   | jrnal | CS    | 5    |
| World Wide Web (WWW)                            | jrnal | CS    | 3    |
| Dec. Supp. Syst. Jrnal. (DSS)                   | jrnal | IS    | 9    |
| Inf. Syst. Res. (ISR)                           | jrnal | IS    | 3    |
| Int. Jrnal. of Electronic Comm. (IJEC)          | jrnal | IS    | 7    |
| Jrnal. of Mgt. Information Systems (JMIS)       | jrnal | IS    | 7    |
| Mgt. Information Systems Quarterly (MISQ)       | jrnal | IS    | 2    |

**Table 1.** Considered publication outlets

|                         |  |
|-------------------------|--|
| Research contribution   | What is the main contribution of the paper? For instance proposing a novel algorithm.                          |
| Recommended items       | For instance, media and entertainment resources, people or diverse e-commerce products                         |
| Recommendation paradigm | Collaborative filtering, content-based filtering or knowledge-based recommendation techniques.                 |
| Research method         | E.g. experimental research on datasets, studies involving real users in lab or field conditions, formal proof. |
| Data sets               | Which data sets are used in the paper?   |
| Evaluation measures     | Employed metrics and choice of a baseline.   |

**Table 2.** Classification scheme

## 3 Results

### 3.1 Research contribution

We limit our taxonomy of research contributions first to *constructive* contributions that developed a novel technical artifact, notably an algorithm or recommendation technique, and *empirical* research that advanced the body of theory

by, for instance, hypotheses driven user-involved studies in the lab or the real-world (experimental as well as observational research approaches). Not surprisingly empirical research plays compared to the other categories a relatively more important role in the IS field while in contrast CS publications focus heavily on algorithmic improvements.

| Type of contribution | IS outlets | CS outlets  |
|----------------------|------------|-------------|
| Technical artifacts  | 24 (36.9%) | 189 (71.3%) |
| Empirical research   | 21 (32.3%) | 18 (6.8%)   |
| Both                 | 9 (13.8%)  | 43 (16.2%)  |
| Other                | 11 (16.9%) | 15 (5.7%)   |
| Total                | 65 (100%)  | 265 (100%)  |

**Table 3.** Research contributions

Furthermore, the following observations can be made based on an additional content analysis.

- More than 25% of CS papers were related to recommendation in the context of the Social and Semantic Web, while only 6% of IS papers considered this context.
- The CS community also addressed questions on non-functional requirements such as scalability or privacy (15% of all CS papers). These areas have been mostly ignored in IS research.
- Cold-start recommendations, i.e. proposing items to users that newly entered the system or recommending novel items, is an issue in both fields. About 10% of the CS papers and about 5% of the IS papers referred to this issue.
- Questions of user interface design (CS 5.8%, IS 12.3%) and transparency (CS 6.8%, IS 10.8%) are relatively more relevant to IS research.
- Topics such as group recommendations, context-awareness, diversity of recommendations, multi-criteria and knowledge-based recommendations, as well as methodological questions still play a very small role in both communities.

Summarizing, according to our analysis the IS community focuses more on the user perspective and the interplay of computerized systems and users whereas research in CS more often takes an algorithmic perspective. Even though questions regarding human computer interaction in RS have been addressed in the early CS literature, see for example Swearingen and Sinha [2], there is still more work required in this area. Topics such as user-centric evaluation, human decision-making and user interaction, have only recently gained more attention also in the CS field. This trend can be observed from recent publications such as Pu et al. [3] and Knijnenburg et al. [4], or from recent journal special issues, for example, a special issue on measuring the impact of recommendation of personalization on user behavior [5]. This can be considered as an emerging trend of CS and IS in RS research.

### 3.2 Recommended items

The next aspect we consider in our analysis is related to the application domain and the recommended items. In the CS field, the main application areas are media and entertainment (>45% from CS total), social networks (>25%), as well as general e-commerce and browsing and search (each 10%). In the IS field, recommendations in e-commerce play a dominant role (>50% from IS total). Beside media and entertainment (12%) also digital libraries (11%) and the use of RS within the organization (12%), e.g. for team collaboration or staffing, were relevant. We subsumed the most popular item category of movie recommendations under media and entertainment even though this problem could be also subsumed under the e-commerce umbrella. Figure 1 depicts this distribution of IS and CS publications over these item categories.

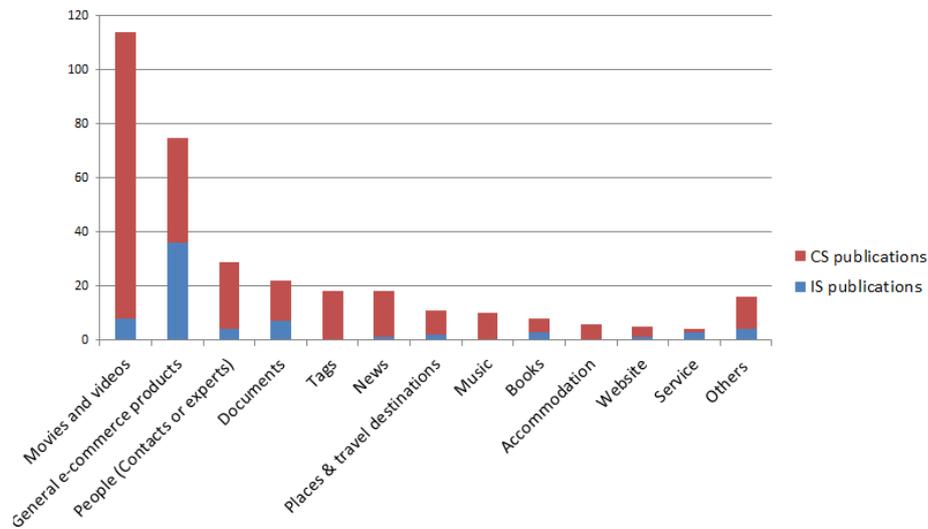


Fig. 1. Recommended items

The availability of public datasets for evaluation such as MovieLens, Netflix or data from Social Web platforms seems to strongly bias the choice of researchers on which application domains to work on. This is particularly the case for the CS field, while IS researchers focus on the recommendation of shopping goods and documents. With respect to avenues for future research, it would be of interest to see if and how the algorithm models of movie recommenders can be transferred to other commercial and business domains. The fact that over 40% of CS papers are focusing on the movie domain shows that other benchmark datasets are badly required. Such datasets would help to stimulate new research directions in particular for CS research and prevent the community from further optimizing the predictive accuracy for the media & entertainment domain. While

general e-commerce and intra-organizational document retrieval is at the core of IS researchers' interest, the advent of the Social Web is not yet reflected in their work. However, we can assume that once the intra-organizational use of social networks becomes more popular, RS in social networks will also move into the focus of IS researchers.

### 3.3 Recommendation paradigms

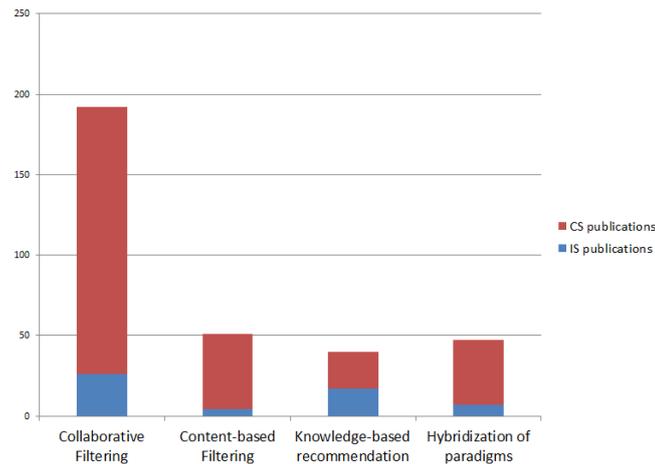
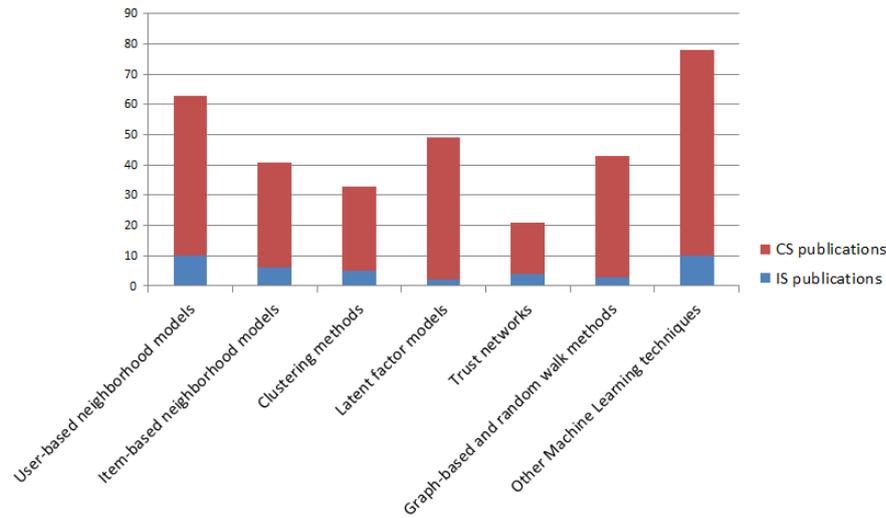


Fig. 2. Recommendation paradigms

The recommendation paradigm determines the principal underlying mechanism how a RS computes recommendations [6, 7]. Collaborative filtering, for instance, reasons that users who had similar opinions in the past will also more likely agree in the future, while content-based filtering determines recommendable items based on their similarity with items the user has liked in the past. While these two mechanisms are most popular, the analysis of our sample reveals significant differences in the research practice of CS and IS communities (see Figure 2). In particular knowledge-based recommendation systems that exploit explicitly codified domain expertise are comparably popular in IS research but play a minor role in CS research. In IS research, RS are often referred to as Recommendation Agent [8] or Digital Advisor [9]. IS papers on decision support or expert systems are usually based on constraints and interactive preference elicitation because business decisions are more often guided by strict rules rather than user experience. In our review, we found that for both IS and CS research, two prominent types of knowledge-based systems are constraint-based and critiquing approaches [10, 11]. Hybridizations of different paradigms play only a minor role in both fields. Considering the importance of collaborative filtering technique in RS, we looked into more details of the used algorithms. Figure 3



**Fig. 3.** Collaborative filtering techniques

shows a more detailed picture of the algorithms used in collaborative filtering in both IS and CS fields. Some papers fall into several categories at the same time.

The analysis shows that classical RS techniques such as nearest-neighbor and clustering are still relevant in today's RS research. During the past ten years and particularly boosted by the Netflix prize, various types of latent factor models including Singular Value Decomposition (SVD) and Latent Dirichlet Allocation (LDA) have become popular and are nowadays often used as a baseline algorithm for comparative evaluations in CS [7]. In IS research, however, these models only play a minor role. In addition, we found that various machine learning methods such as probabilistic or regression models are relevant in both fields. As for the future research, a better understanding of different aspects of latent factor approaches that go beyond predictive accuracy is required. It is for example unclear how recommendations based on these models can be explained to the users in order to increase the user's trust. Also, we know little about how these methods perform with respect to the diversity, serendipity or novelty of the generated recommendations. Furthermore, as early work such as Balabanovic & Shoham [12] has showed that combining different techniques or sources of knowledge can advance the system performance, more research is required in the selection of recommender algorithms to construct high performance RS. Also, recommender strategies may perform differently in different situations [13]. For example, the study of Jannach & Hegelich [14] revealed that the choice of the most effective recommender depends on the specific situation and goal of the user. In the future it is interesting to further understand how to select different recommender strategies in line with various situations.

### 3.4 Research method

In this section, we will discuss the research design of the selected papers, which includes research methods, underlying theories, data sets, evaluation metrics and data analysis methods. Figure 4 classifies the sampled publications according to the employed research methods. However, note that we did not use a comprehensive taxonomy of research methods, but denoted the most commonly used terminology in RS research.

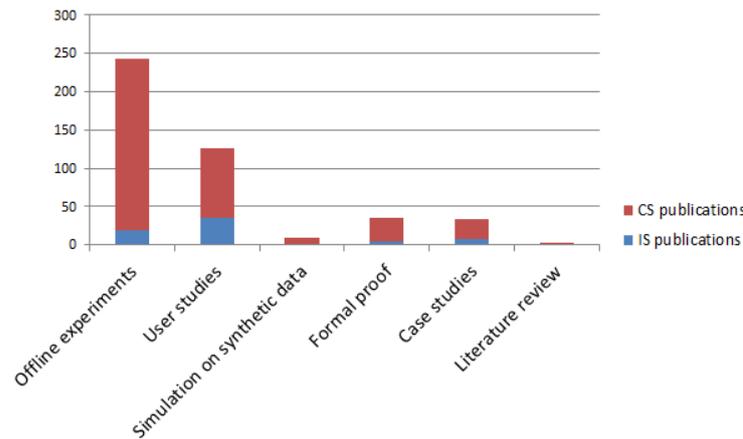


Fig. 4. Research methods

Over the last five years, two third of the selected papers in CS are based on offline experiments using historical user data. In IS research, in contrast, user studies and user-involved experiments in conjunction with questionnaires or interviews are more popular. Simulation and formal proofs play a minor role in both fields. While most CS papers typically improve the RS performance, IS papers often try to explain a phenomenon or theories related to RS. With respect to hypothesis development and social theories, IS research is more based on theoretical considerations than CS. In about half of the IS papers, theories such as the Technology Acceptance Model, Social Agency, Social Presence or the Theory of Reasoned Action are mentioned as underlying theories.

As offline experimentation on data sets is the most important method, we provide an overview on the most popular types of data (Figure 5). MovieLens, Netflix, Eachmovie, IMDB as well as a major share of Yahoo! data denote data sets from the movie domain. This documents the dominating role of benchmark data from media and entertainment that was employed by at least 50% of all papers that are based on offline evaluations. Besides movies, recommendations in the context of Web 2.0 (included in *Others*) also played an important role in the past five years. For instance, the corresponding datasets can be derived from Epinions, Delicious, YouTube, CiteULike, Bibsonomy, Flickr, Orkut, Twitter, Digg, or Wikipedia. In total, about 25% of CS papers were evaluated based

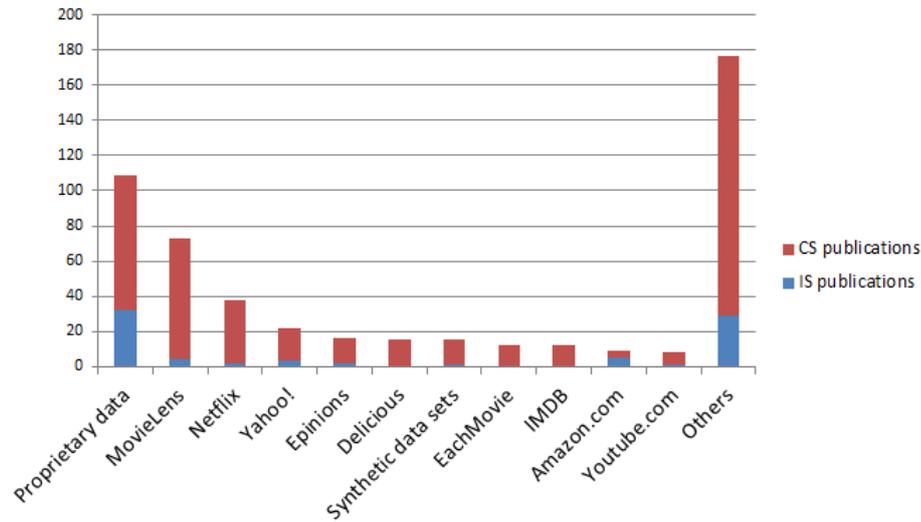


Fig. 5. Datasets

on Web 2.0 data. To some smaller extent, news recommendation is also a relevant topic in CS research. In addition, a small number of IS experiments were done in the e-commerce domain, based for example on the data obtained from Amazon.com.

|  | IS outlets | CS outlets |
|--|------------|------------|
| <b>IR measures</b>                     |            |            |
| Precision and Recall                   | 12         | 115        |
| F1                                     | 2          | 20         |
| Rank measures (e.g. NDCG)              | 9          | 27         |
| ROC curve                              | 1          | 11         |
| Area under ROC                         | 0          | 8          |
| <b>ML measures</b>                     |            |            |
| Mean absolute Error                    | 6          | 57         |
| Root Mean Squared Error                | 0          | 49         |
| <b>Application quality</b>             |            |            |
| Computation time                       | 2          | 28         |
| Coverage metrics                       | 2          | 28         |
| <b>Decision support quality</b>        |            |            |
| Perceived utility or user satisfaction | 11         | 7          |
| Online conversion                      | 3          | 12         |
| Diversity metrics                      | 0          | 10         |

Table 4. Evaluation measures

Table 4 quantifies which metrics are used in the literature to assess the quality of recommendations such as the whole quality of RS or factors used to test research hypotheses. It is structured according to the perspectives denoted in the introduction.

- Classic Information Retrieval (IR) metrics measuring classification accuracy such as Precision, Recall and their harmonic mean F1 as well as ROC curves and the area below the curve are particularly popular in the CS community. Rank measures that consider the position of items in recommendation lists such as the normalized discounted cumulated gain (NDCG) or half-life utility also fall into this category, but they are only used to a minor extent in RS research.
- In Machine Learning (ML), traditionally aggregate deviations between actual and predicted rating values are measured. The Mean Absolute Error (MAE) metric averages the deviations between predicted and actual rating values, while the Root Mean Squared Error (RMSE) puts more weight to larger deviations. Although several authors have critiqued these measures that, for instance, combine prediction errors for items the user hates as well as for highly recommended items, a considerable share of work still argues its contribution by improvements in terms of these error measures.
- The third section in Table 4 groups measures describing aspects of the technical *application quality*. Coverage measures determine the share of the user population that can actually receive recommendations, while computation time is an important aspect for the system’s responsiveness.
- Finally, we identified measures that focus on quality aspects of a system’s decision support capabilities. For instance, perceived utility is a well known concept from technology acceptance research. Online conversion rates do measure short-term commercial success and the persuasive traits of a recommendation system, but do not measure medium and long-term satisfaction with the bought item. Furthermore, diversity metrics measure if a system provides a broad view on the offered choices.

Most CS papers are more homogenous with respect to the applied metrics and focus only on the accuracy of recommendations. This can again be explained by the existence of standardized benchmark problems. For the metrics that go beyond accuracy, only diversity measures are applied in recent literature. Even though the problems of using only accuracy metrics in RS have been discussed in the last few years, for example in [15], more evaluation metrics such as novelty, popularity and serendipity are not largely applied so far. Future RS research can therefore focus on evaluating and improving RS by considering a variety of metrics to achieve an overall high quality. In IS research, most papers are focused on measuring the perceived quality such as utility or user satisfaction and explaining the latent relationships and effects based on user-centric evaluation. As typically practiced in social science research, papers are often organized by proposing specific research hypotheses and validated via a corresponding experimental design. The appropriate evaluation method therefore depends on the design of the proposed models or hypotheses. This may lead to the effect that

a multitude of different measures are proposed but barely reused. We observed that individual metrics which were used in about 30% of the papers were barely used in other papers. Therefore future research can focus on proposing validated and widely accepted measuring instruments for more rigorous research.

## 4 Discussion

The large amount of papers which aim at optimizing the accuracy of predictions based on historical data in the movie domain underlines the strong need for the CS research community to focus both on other domains as well as on other types of experimental designs to understand the real impact of recommender systems on users. Even for the supposedly well-understood movie domain, it is not definitely clear in which situations and to which extent lower RMSE values finally lead to higher user satisfaction or sales. Consider for example a RS that recommends the fifth sequel of a movie to a user who has liked all other movies in this series. Such a prediction might be accurate but not valuable. The context of movie consumption (Am I watching a movie alone or with friends? Am I looking for entertainment or for intellectual challenge?) is largely not taken into account in the majority of papers. In these areas, only the time of the day or week was considered as a context factor to some extent in recent works. The same holds for the development of the user's taste over time and long-term user models. In future work, more focus should therefore be put on context-aware RS, see for example [16]. The rapid development of the Social and Semantic Web has the potential to further boost the field of RS. On the one hand, new application areas arise, for example the recommendation of tags, pictures, friends or links. At the same time, more and more data is available for exploitation by recommendation algorithms, as users are increasingly willing to contribute, participate, or interact, for example, on social networks, review platforms or blogging sites. One final observation in the context of RS evaluation is that theoretical considerations about the computational complexity or questions of scalability of algorithms are covered only by a small fraction of research papers in both fields. Most of the recent algorithmic approaches are model-based, rely on an offline learning phase and support the efficient generation of predictions. However, in particular in the context of the Social Web and the massive amounts and different types of data that have to be processed, appropriate strategies to efficiently compute models and predictions might become more relevant in the future.

## 5 Conclusions

Our literature review indicated the importance of recommender systems in the fields of Information Systems and Computer Science. Given the different roots of the fields, CS researchers focus more on algorithms whereas IS researchers are more interested in the systems-perspective and the effects of RS on the

users. Correspondingly, different research designs and methods dominate in the two communities as has been documented by this work. As an outlook, we see evidence that increased mutual exchange of results from the two communities can help further advance the research of recommender systems. In the survey of Xiao and Benbasat [8] the authors discuss the role of recommender systems for example in the context of consumer research and marketing, human decision making, electronic commerce or human computer interaction. Only few works in CS focus on these topics today. Thus we see the development of techniques that exploit the insights from these different areas as a field of future RS research for the CS community. In parallel, the IS community can benefit from incorporating recent algorithmic results from the CS community.

## References

1. Zanker, M., Jessenitschnig, M.: Case-studies on exploiting explicit customer requirements in recommender systems. *UMUAI* **19**(1-2) (2009) 133–166
2. Swearingen, K., Sinha, R.: Beyond algorithms: An HCI perspective on recommender systems. In: *ACM SIGIR 2001 Workshop on Recommender Systems*. (2001)
3. Pu, P., Chen, L., Hu, R.: A user-centric evaluation framework for recommender systems. In: *Proc. ACM RecSys'11*. (2011) 157–164
4. Knijnenburg, B., Willemsen, M., Gantner, Z., Soncu, H., Newell, C.: Explaining the user experience of recommender systems. *UMUAI* **22**(4) (2012) 441–504
5. Zanker, M., Ricci, F., Jannach, D., Terveen, L.G.: Measuring the impact of personalization and recommendation on user behaviour. *IJHCS* **68**(8) (2010) 469–471
6. Jannach, D., Zanker, M., Felfernig, A., Friedrich, G.: *Recommender Systems - An Introduction*. Cambridge University Press (2011)
7. Ricci, F., Rokach, L., Shapira, B., Kantor, P.B., eds.: *Recommender Systems Handbook*. Springer (2011)
8. Xiao, B., Benbasat, I.: E-commerce product recommendation agents: Use, characteristics, and impact. *MIS Quarterly* **31**(1) (2007) 137–209
9. Felfernig, A., Friedrich, G., Jannach, D., Zanker, M.: An Integrated Environment for the Development of Knowledge-Based Recommender Applications. *International Journal of Electronic Commerce* **11**(2) (06-7) 11–34
10. Zanker, M., Jessenitschnig, M., Schmid, W.: Preference reasoning with soft constraints in constraint-based recommender systems. *Constraints* **15**(4) (2010) 574–595
11. Felfernig, A., Friedrich, G., Jannach, D., Zanker, M.: Developing constraint-based recommenders. [7] 187–215
12. Balabanović, M., Shoham, Y.: Fab: content-based, collaborative recommendation. *Communications of the ACM* **40**(3) (1997) 66–72
13. Pilászy, I., Tikk, D.: Recommending new movies: even a few ratings are more valuable than metadata. In: *Proc. ACM RecSys'09*. (2009) 93–100
14. Jannach, D., Hegelich, K.: A case study on the effectiveness of recommendations in the mobile internet. In: *Proc. ACM RecSys'09, New York*, (2009) 41–50
15. McNee, S., Konstan, J.R.J.: Being accurate is not enough: How accuracy metrics have hurt recommender systems. In: *EA ACM CHI 2006*. (2006) 997–1001
16. Adomavicius, G., Tuzhilin, A.: Context-aware recommender systems. [7] 217–253