HCI for Recommender Systems: the Past, the Present and the Future

André Calero Valdez
RWTH Aachen University
Campus Boulevard 57
52074 Aachen, Germany
calero-valdez@comm.rwth-aachen.de

Martina Ziefle
RWTH Aachen University
Campus Boulevard 57
52074 Aachen, Germany
ziefle@comm.rwth-aachen.de

Katrien Verbert
KU Leuven
Celestijnenlaan 200A
B-3001 Leuven, Belgium
katrien.verbert@cs.kuleuven.be

ABSTRACT

How can you discover something new, that matches your interest? Recommender Systems have been studied since the 90ies. Their benefit comes from guiding a user through the density of the information jungle to useful knowledge clearings. Early research on recommender systems focuses on algorithms and their evaluation to improve recommendation accuracy using F-measures and other methodologies from signal-detection theory. Present research includes other aspects such as human factors that affect the user experience and interactive visualization techniques to support transparency of results and user control. In this paper, we analyze all publications on recommender systems from the scopus database, and particularly also papers with such an HCI focus. Based on an analysis of these papers, future topics for recommender systems research are identified, which include more advanced support for user control, adaptive interfaces, affective computing and applications in high risk domains.

Keywords

Recommender Systems, Human-Computer Interaction, Uncertainty, Risk

1. INTRODUCTION

You have most certainly seen the output of a recommender system. The online retailer Amazon suggests your next purchase by letting you know what others bought together with the product you're currently viewing ("Frequently bought together."). And maybe you have also experienced the limitations of such systems by getting suggestions that made no sense to you at all.

Several recommendation techniques, such as content-based, knowledge-based, collaborative filtering and their hybridizations, are discussed in state-of-the-art surveys [1, 5], including their merits and limitations. Typical fields of application are recommending movies, music or related products in e-commerce. These application domains are similar: each recommendation is important in creating additional revenue and of relatively low risk. Recommender systems are not typically used to propose medical procedure, emergency plans in nuclear plants or stock trading, where uncertainty and risk are high. Why is this the case?

This paper reflects on the development of recommender systems by reviewing publications from scopus. We look back at the past to reflect on present research and try to predict future research topics for recommender systems based on current trends in ICT and a short bibliometric analysis of the field of recommender systems.

RecSys '16 September 15-19, 2016, Boston , MA, USA

ACM ISBN 978-1-4503-4035-9/16/09.

DOI: http://dx.doi.org/10.1145/2959100.2959158

2. THE PAST

Tapestry [8] was the first system that employed *collaborative filtering* to filter the large amount of emails that reached people in Xerox PARC. It included a system based on annotating emails and later filtering emails according to these annotations. The novelty was the collaborative aspect of this annotation process.

The term *recommender system* appears for the first time in a 1997 article by Paul Vesnick and Hal R. Varian [25] and describes recommender systems as a tool for decision making and not just for information retrieval. They already argue in terms of evaluating costs (of false positives/negatives) beyond mere measures from signal detection theory.

Since then research on recommender systems has drastically increased. Over 9,400 articles and 76 reviews can be found on scopus[1]. The trend shows an ever increasing amount published on recommender systems (see Fig. 1). Additionally, we see two peaks in review papers about recommender systems (see Fig. 1). The first peak in 2005 and the second in 2010.

The first review paper on recommender systems by Herlocker et al. [11] focused on evaluation of recommender systems. Another review by Adomavicius and Thuzhilin [1] discusses the use of content-based, collaborative and hybrid approaches and proposes possible extension for research. The authors also mention the evaluation related problems with measures of signal detection theory such as the famous F1-Measure. Such metrics focus on coverage and accuracy [1], while at the same time criteria like usefulness and quality as well as, explainability, trustworthiness, scalability and privacy issues may play bigger roles in applications.

[1]http://www.scopus.com (as of April 10^{th} 2016)

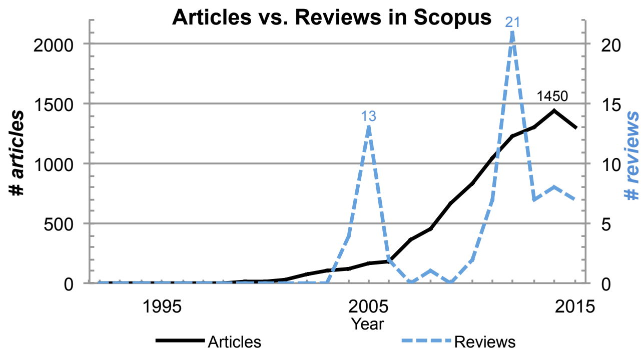

Figure 1: The amount of articles published on recommender systems has been constantly increasing, while reviews have had two spikes.

3. THE PRESENT

In recent years, researchers have become more aware of the fact that effectiveness of recommender systems goes beyond recommendation accuracy [16]. Thus, research on these human factors has gained increased interest, for instance by combining interactive visualization techniques with recommendation techniques to support transparency and controllability of the recommendation process. Visualization leverages visual representations to facilitate human perception, while interaction stresses user involvement through dialogue with the system. In a recent study, 24 interactive recommender systems were analysed that use an interactive visualization technique to support user interaction [10]. A large share of these systems focus on transparency of the recommendation process to address the *black box* issue. Here, the overall objective is to explain the inner logic of a recommender system to the user in order to increase acceptance of recommendations. A good example of this approach is PeerChooser [19], a visual interactive recommender that uses a graph-based representation to show relationships between users and recommended items from a collaborative filtering recommender system. Similarly, SmallWorlds [9] allows the exploration of relationships between recommended items and friends with a similar profile using multiple aspects. This way, users can explore different relationships to gain transparency and ultimately increase the chance of finding items.

In addition, TasteWeights [3] allows users to control the impact of the profiles and behaviors of friends and peers on the recommendation results. TasteWeights provides an interface for such hybrid recommendations. The system elicits preference data and relevance feedback from users at runtime in order to adapt recommendations. SetFusion [22] is a recent example that allows users to fine-tune weights of a hybrid recommender system. SetFusion uses a Venn diagram to visualize relationships between recommendations.

Results of this survey indicate that many interesting prototypes have been developed and evaluated that demonstrate the importance of this research for the community. Nevertheless, many open research challenges exist that provide interesting opportunities for research in this field.

4. THE FUTURE

Trying to predict the future is only helpful when the prediction increases the knowledge about the future beyond what is already known. *Safe bets* are uninteresting predictions, while *wild guesses* may seem ridiculous from hindsight. Therefore using *safe bets* from other fields that may relate to recommender research may prove to be an interesting middle-ground.

4.1 Bibliometric analysis

From the Scopus database we searched for all documents from the field of computer science that contained the search term “recommender system” (n=9,432). From this data-set we extracted the author keywords. We removed all stop words, punctuation and mapped common noun phrases to single-word phrases (e.g. user model to user_model). We then counted frequencies for all years and plotted the relative frequency of selected terms to analyze possible trends regarding our selection (see Fig. 2). The terms were chosen to contrast HCI related keywords (in the left column) with more algorithm related keywords (in the right column).

4.2 Where should we direct future research?

When taking a closer look at publication counts (see Fig. 1) we can see two spikes in review publishing curve. The most cited works in the first spike are the previously mentioned reviews [1, 11]. The same year (2005) also marks a change in interest in topics (see Fig. 2), focusing on adaptive recommender systems and the user model. The second change in topics can be seen in 2009, where trust becomes a heavily published keyword. The second spike in reviews three years later marks another change in interest, where the most frequently cited reviews [7, 15] focus on machine learning and user focus.

Surely developments in improving algorithms are important and some interesting directions were proposed by Park et al. [21]. On the other hand, some researchers in recent reviews [15, 10] have argued for focusing on the user of the recommendation system. Following the latter argument, we think that HCI related aspects are underrepresented in literature (see Fig. 2) and yet critically important for future applications of recommender systems.

User control.

Current research on interactive recommender systems is focused on controlling the importance of different recommendation parameters or revising the user profile [10]. Such control is insufficient to establish a good level of trust and to address privacy concerns [2]. Further research is needed to support more advanced levels of interaction, such as controls that define which data can be tracked and taken into account for which purposes. Examples of such more advanced levels of control have been researched extensively by the visual analytics (VA) community. Whereas these examples are promising, research is needed to apply these VA concepts, methods and techniques to the field of recommender systems and casual users.

Adaptive.

The aim of adequate user control is to increase recommendation accuracy [24] by incorporating user input and feedback. Previous research shows however that user satisfaction does not always correlate with high recommendation

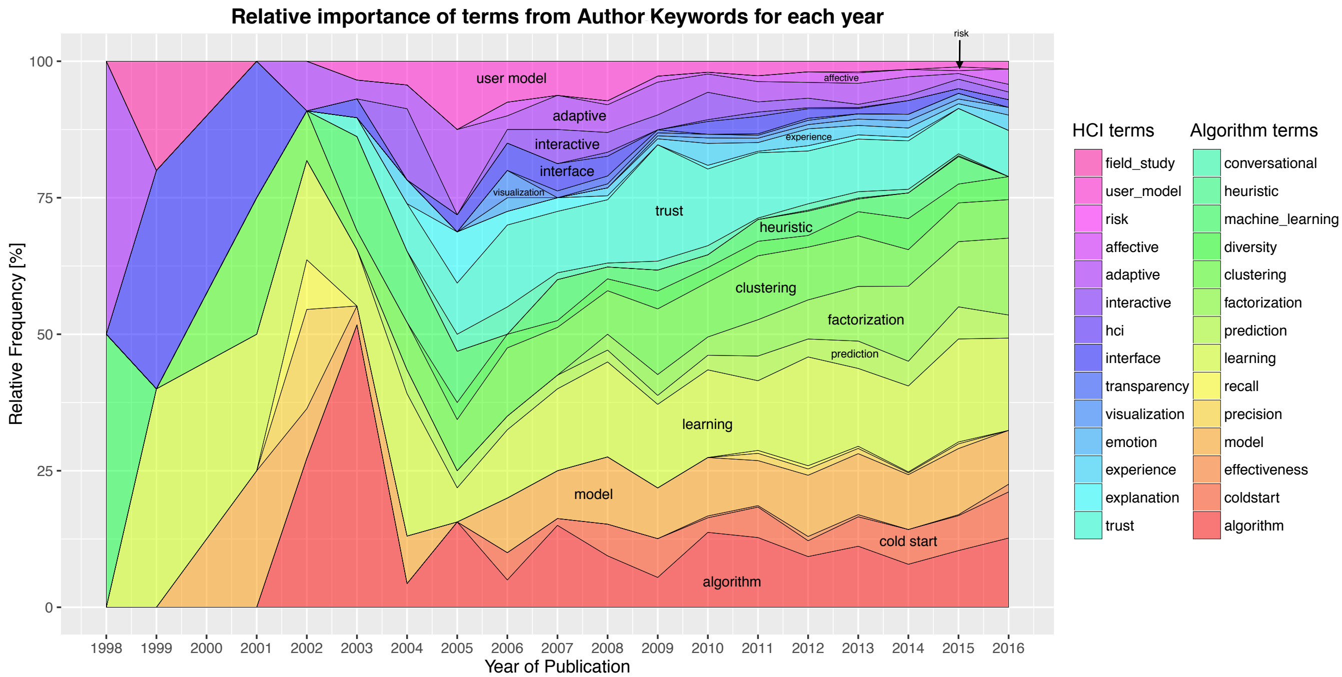

Figure 2: This picture show the relative importance of our search terms for each year. Importance is derived from occurrence in author keywords from 9,432 publications.

accuracy and that other factors, such as the knowledge level of the user [14] and her interests [12], need to be considered. Current recommender systems interfaces are *static*, i.e. they do not tailor the interface to these user characteristics. There is a need to adapt recommender systems and their user interfaces to these different personal and situational characteristics. Research that has been conducted by the adaptive hypermedia, adaptive visualization and personalized search communities provides a promising starting point to address this challenge.

Affective.

Emotions play a crucial role in decision making [23]. An interesting future line of research is to experiment with novel sensing technologies to capture behavioral data (physiological data, facial expressions, speech, ...) in order to detect emotions and to adapt recommendations based on emotional responses. Although measurement of emotions in a controlled laboratory environment is well studied for years by a large number of research groups [20], multimodal emotion recognition in real world environments is still a challenging task [17]. A good review of existing methods has been reported by Hrabal [13]. As none of the methods have yet led to successful subject- and situation-independent emotion recognition, interactive methods that enable users to revise detected variables seem promising. The challenge is to research the development of a next generation of recommender systems that can incorporate both automatically acquired data and revisions by end-users as a basis to tailor recommendations based on current contextual needs of the user, including emotions that are key in decision making.

High-risk domains.

The biggest risk a user of e-commerce faces is spending money for an undesired product. Thus product recommendations are of very specific risk. Other domains have a higher level of uncertainty and risk. Giving recommendations in domains were choices must be made under uncertainty and risk is intricately more challenging. Risk-aware algorithms [4] or predictions of risks [6] have been investigated only recently and not extensively. How uncertainty and risk of a recommendation should be visualized or communicated has not been investigated yet, but may be crucial for the application of these in high-risk domains such as medicine [18].

5. CONCLUSIONS

Looking at the big trends in ICT we see that Big Data and thus more advanced techniques from AI (e.g. deep learning) will become available and then applied to recommender systems. These do not only play a role in improving algorithms but also new interactions paradigms. In the future we could analyze not just the transactions of users, but also patterns of interaction (e.g. mouse movement, keystrokes, facial expressions etc.). These ultimately lead to new research questions for new adaptive interfaces and how the user controls these recommender systems. When recommending in domains of risk or uncertainty new visualizations will be necessary to improve the trust and understanding of recommendations. Using more intimate data such as facial expression will also bring new problems of privacy and user acceptance.

6. LIMITATIONS

The bibliometric analysis was conducted using a pre-defined set of keywords. We have put efforts into mapping similar terms to our terms, nevertheless some keywords might have been overlooked. Looking at the relative frequencies skews the data in favor of this set of keywords. Thus our findings represent upper thresholds for relative importance.

7. ACKNOWLEDGMENTS

The authors thank the German Research Council DFG for the friendly support of the research in the excellence cluster “Integrative Production Technology in High Wage Countries”. The work of Katrien Verbert is supported by the KU Leuven research council as part of the starting grant “Flexible Interaction with Intelligent Systems” (grant agreement STG/14/019)

8. REFERENCES

[1] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6):734–749, 2005.

[2] F. Bakalov, M.-J. Meurs, B. König-Ries, B. Sateli, R. Witte, G. Butler, and A. Tsang. An approach to controlling user models and personalization effects in recommender systems. In *Proceedings of the 2013 int. conf. on Intelligent user interfaces*, pages 49–56. ACM, 2013.

[3] S. Bostandjiev, J. O’Donovan, and T. Höllerer. Tasteweights: a visual interactive hybrid recommender system. In *Proceedings of the sixth ACM conference on Recommender systems*, pages 35–42. ACM, 2012.

[4] D. Bouneffouf, A. Bouzeghoub, and A. L. Ganarski. Risk-aware recommender systems. In *Neural Information Processing*, pages 57–65. Springer, 2013.

[5] R. Burke. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331–370, 2002.

[6] R. Conforti, M. de Leoni, M. La Rosa, W. M. van der Aalst, and A. H. ter Hofstede. A recommendation system for predicting risks across multiple business process instances. *Decision Support Systems*, 69:1–19, 2015.

[7] P. Domingos. A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78–87, 2012.

[8] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61–70, 1992.

[9] B. Gretarsson, J. O’Donovan, S. Bostandjiev, C. Hall, and T. Höllerer. Smallworlds: visualizing social recommendations. In *Computer Graphics Forum*, volume 29, pages 833–842. Wiley Online Library, 2010.

[10] C. He, D. Parra, and K. Verbert. Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications*, 56:9–27, 2016.

[11] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1):5–53, 2004.

[12] Y. Hijikata, Y. Kai, and S. Nishida. The relation between user intervention and user satisfaction for information recommendation. In *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, SAC ’12, pages 2002–2007, New York, NY, USA, 2012. ACM.

[13] D. Hrabal. *Emotion Classification in Human-computer Interaction on the Basis of Physiological Data*. PhD thesis, Universität Ulm. Medizinische Fakultät, 2013.

[14] B. P. Knijnenburg, N. J. Reijmer, and M. C. Willemsen. Each to his own: How different users call for different interaction methods in recommender systems. In *Proc. of the Fifth ACM Conf. on Recommender Systems*, RecSys ’11, pages 141–148, New York, NY, USA, 2011. ACM.

[15] J. A. Konstan and J. Riedl. Recommender systems: from algorithms to user experience. *User Modeling and User-Adapted Interaction*, 22(1-2):101–123, 2012.

[16] S. M. McNee, J. Riedl, and J. A. Konstan. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI’06 extended abstracts on Human factors in computing systems*, pages 1097–1101. ACM, 2006.

[17] S. Meudt and F. Schwenker. Enhanced autocorrelation in real world emotion recognition. In *Proceedings of the 16th International Conference on Multimodal Interaction*, ICMI ’14, pages 502–507, New York, NY, USA, 2014. ACM.

[18] G. Naik, H. Ahmed, and A. G. Edwards. Communicating risk to patients and the public. *Br J Gen Pract*, 62(597):213–216, 2012.

[19] J. O’Donovan, B. Smyth, B. Gretarsson, S. Bostandjiev, and T. Höllerer. Peerchooser: visual interactive recommendation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1085–1088. ACM, 2008.

[20] M. Ouwerkerk. Unobtrusive emotions sensing in daily life. In *Sensing Emotions*, pages 21–39. Springer, 2011.

[21] D. H. Park, H. K. Kim, I. Y. Choi, and J. K. Kim. A literature review and classification of recommender systems research. *Expert Systems with Applications*, 39(11):10059–10072, 2012.

[22] D. Parra and P. Brusilovsky. User-controllable personalization: A case study with setfusion. *International Journal of Human-Computer Studies*, 78:43–67, 2015.

[23] R. W. Picard, S. Papert, W. Bender, B. Blumberg, C. Breazeal, D. Cavallo, T. Machover, M. Resnick, D. Roy, and C. Strohecker. Affective learning - a manifesto. *BT Technology Journal*, 22(4):253–269, 2004.

[24] P. Pu, L. Chen, and R. Hu. Evaluating recommender systems from the user?s perspective: survey of the state of the art. *User Modeling and User-Adapted Interaction*, 22(4-5):317–355, 2012.

[25] P. Resnick and H. R. Varian. Recommender systems. *Communications of the ACM*, 40(3):56–58, 1997.