A Content Independent Model for Context Adaptation and Individualization in Information Retrieval

Thomas Mandl, Christa Womser-Hacker

University of Hildesheim
Information Science
Marienburger Platz 22
31141 Hildesheim, Germany
{mandl, womser}@uni-hildesheim.de

Abstract. The consideration of context in information retrieval is expected to improve retrieval effectiveness. Most current approaches rely on the similarity of content which can only partly capture the complexity of the task. In this paper, we present a model for context adaptation which can be realized independently from content. The model is built in the framework of the MIMOR model (Multiple Indexing and Method-Object Relations) for the long-term learning of user preferences in information retrieval. MIMOR integrates a fusion method and a relevance feedback processor into a learning model.

1 Introduction

The analysis of context in information retrieval is a complex task. The research literature stresses the complexity of potential context definitions (Ingwersen et al. 2004, Ruthven 2004). However, many models and implementations restrain context to the similarity of information needs to a given profile (Mandl & Womser-Hacker 2000, Belkin et al. 2004, Ma & Goharian 2005). These models can be called content dependent. The complexity of the notion of context shows that content dependent definitions can only capture a fraction of the task. It is obvious that for the same content (query and documents) different contexts with different optimal solutions exist. These different solutions can be achieved better by content independent context models.

We propose a model which allows the integration of context in a potentially content independent manner. For that endeavor, we rely on an extension of the MIMOR model for the individualization of information retrieval results.

In MIMOR, several black box matching functions are combined into a linear combination committee machine which reflects the user's vague individual cognitive concepts expressed in relevance feedback decisions. An extension based on the soft computing paradigm couples the relevance feedback processor and the matching function into a unified retrieval system. User relevance feedback can be considered to be the best technique to improve the results of information retrieval systems. However, it requires considerable effort by the user and it is not often done. One reason is that users do not see the benefit and that relevance feedback is often used only for improving the current query and is subsequently lost.

In previous research, we have developed the MIMOR system, which stores user relevance feedback decisions to foster the long-term optimization of a information retrieval system. MIMOR is based on the fusion of several information retrieval systems (Womser-Hacker 1997). The influence of systems is based upon their previous performance for the user measured by the relevance feedback. The fusion is currently implemented as a weighted linear combination of the individual systems and has been successfully applied to both mono-lingual as well multi-lingual retrieval tasks (Hackl et al. 2002). MIMOR has several advantages:

- Individual systems do not need to be optimized and can be regarded as black boxes
- The fusion as linear combination is transparent for users
- User relevance feedback is well exploited and stored for long-term use
- Fusion of several systems leads to good performance

MIMOR has been extended to improve performance at early stages when little relevance feedback is available and consequently performance is low. The system integrates the relevance feedback data of all users into a public model which dominates the retrieval process at the beginning of usage. When more relevance decisions of an individual user are present, his individual (private) model will gain strength and dominate the results. In this paper, we extend MIMOR such that it also considers the context of the user. This can be done in a similar way the individualization was integrated into the model.

The remainder of the paper is organized as follows. Section 2 reviews the basics of fusion and relevance feedback systems. Section 3 repeats the design of the MIMOR system and presents the extension to model context. Section 4 concludes the paper.

2 Fusion and Relevance Feedback in Information Retrieval

Fusion and relevance feedback are considered as main factors to improve effectiveness in information retrieval. Whereas fusion methods combine different perspectives of document representation and query-document-matching, relevance feedback is based on the individual perspective of the user. There were two main directions of research on relevance feedback: first, to re-weight terms automatically depending on their distribution over relevant and non-relevant documents, second, to provide a term expansion within the initial query (Harman 1992a). Later on the information retrieval community paid attention to more detailed aspects, namely the relation between modification of weights and query expansion, the selection of further query terms and the effectiveness of the number of iterations etc.

Fusion methods delegate a task to different systems and integrate each result returned into one final result. For information retrieval tasks, this means the integration of different probabilities for the relevance of a document. In a non-formal view, fusion connects different perspectives of representing and indexing the content and the topical aspects of a document. These approaches are also referred to as poly-representation (Ingwersen 1994) and have also contributed to the development of elaborate context models (Ingwersen et al. 2004).

Fusion in information retrieval has been inspired by results of the TREC evaluation campaign (Voorhees & Harman 1997, 1999, 2001). TREC is a large scale evaluation forum for information retrieval research which provides a uniform testbed for many retrieval tasks. The data comprises test collections, queries or topics, relevance judgements and the evaluation methodology. Researchers may use the TREC data to achieve comparability to others.

Experiments within TREC have shown that the results of similarly well performing information retrieval systems often differ. This means that while the systems find the same percentage of relevant documents, the overlap between their results is sometimes low (Voorhees & Harman 1997). Therefore, fusion seems to be a promising approach and has been applied to text retrieval (Bartell et al. 1994; Fox & Shaw 1994, Lee 1995, Voorhees et al. 1995, McCabe et al. 1999, Savoy & Rasolofo 2000).

Fusion research aims at finding out which retrieval or indexing methods should be combined, which committee machine architecture should be used and which features of collections indicate that a fusion might lead to positive results.

In experimental systems, the methods to be fused are applied to the same collection. However, fusion can treat collections without overlap as well. A collection may be split into artificial sub-collections which are handled by a retrieval system (Savoy & Rasolofo 2000). In such a case, the goal of the fusion can be regarded as an attempt to derive knowledge about which collection leads to good results.

3 MIMOR: A Learning Model for Fusion

The results of the TREC conferences and other empirical studies have shown that relevance feedback may be an effective technique to improve retrieval quality (Harman 1992a, 1992b). In our opinion, powerful learning methods for information retrieval need to extend the range of relevance feedback effects beyond the modification of the query in order to achieve long-term adaptation to the subjective point of view of the user. The mere change of the query often results in improved quality, however, the information is lost after the current session.

3.1 The MIMOR Model

The complexity of information retrieval led to the idea of combining components which gained positive assessments within the long development process of information retrieval. Moreover, we assumed that the adequate selection and adaptation of useful functionality is the key for success. The challenge was to gain the optimal combination within special contexts. The basic idea of the MIMOR model (Womser-Hacker 1997) is to optimize its quality through learning from user feedback. The task of the users is to provide relevance feedback about the retrieved documents which is used to judge the collaboration of retrieval systems within the MIMOR fusion system.

Consequently, the framework MIMOR does not rely on changes to the document or the query representation when processing relevance feedback information. Instead, it focuses on the central aspect of a retrieval function, the calculation of the similarity between document and query. Like other fusion methods, MIMOR accepts the results from individual retrieval systems as black boxes. These results are fused by a linear combination which is stored during many sessions although more complex fusion methods can also be applied. The vector of weights for the fusion over all contributing systems can be seen as the MIMOR model at a certain point in time.

The weights for the systems experience are modified through learning. They adapt according to relevance feedback information provided by users and create a long-term model for future use. That way, MIMOR learns which systems were successful in the past and therefore in the training data. In MIMOR, the following formula gives the retrieval status value (RSV) for a document. Arguments are the normalized RSV of the fused systems and their weights.

$$(1) \qquad RSV_{MIMOR}(doc) = \frac{\sum_{system=1}^{N} (\omega_{system}\ RSV_{system}(doc))}{N}$$

The learning features are central in MIMOR. The weight of the linear combination of each information retrieval system is adapted according to the success of the system measured by the relevance feedback of the users. A system which gave a high RSV to a document which received positive relevance feedback should contribute more intensely to the final result. The following formula (2) enables such a learning process by iteratively increasing the weight of positively evaluated systems. It shows how to calculate the weight change for a system weight based on relevance feedback and the RSV assigned to a document. After the modification, the weights of all systems need to be normalized.

$$(2) \qquad \begin{array}{l} \Delta\omega_{system} = \varepsilon\ RF_{user}(doc)\ RSV_{system}(doc) \\ \varepsilon \quad learning \quad rate \in [0;1] \\ RF_{user} \quad relevance \quad feedback \in [-1;1] \end{array}$$

From a machine learning perspective MIMOR can be interpreted as a committee machine, where the experts are retrieval systems. They represent the cognitive structures of the authors of the documents and the users who are seeking for information (see fig. 2). MIMOR uses a linearly weighted combination of the results. The learning algorithm for these weights guarantees that experts who were successful in the past acquire a higher influence. More complex dynamic forms of committee machines may be applied as well.

The parameters for the linear combination could also be learned by a nonlinear learning algorithm like neural networks or support vector machines. The features and properties of the individual retrieval engines can serve as input jointly with the RSVs calculated. MIMOR has also been applied to individualized text categorization (Mandl & Womser-Hacker 2001).

3.2 User Model in MIMOR

MIMOR is designed for a multi-user environment which is typical for companies. During use, MIMOR develops a user model for each individual person using the system. Unlike other user models in information retrieval, MIMOR adapts the core of an information retrieval system and applies the user's decisions to the calculation of the RSV. A MIMOR model for each person can be introduced leading to optimal user models. However, the training of MIMOR requires a substantial number of relevance feedback values forcing the user to submit many decisions. Another disadvantage is common to all inductive and incremental learning algorithms. The occurrence of outliers in the initial learning phase may cause unstable learning behavior. Long training time or a degradation of the retrieval behavior may be the consequences.

Both problems can be avoided by introducing separate private and public models. The private model contains a user-specific MIMOR model optimized by all the individual relevance feedback decisions. One public model is trained with all decisions of all users of the system. The public MIMOR model is therefore optimized but not individualized. It represents a consensus or compromise between all users and their individual differences in relevance assessment. It should be applied in the absence of an individual model, especially for any user beginning to work with MIMOR. Both models are combined linearly in order to reach a final result each having a weight associated with it. Over a period of time, the beginner will collect a significant number of relevance judgements and will eventually reach a fully individualized and saturated model. During this process, the public model will lose its influence while the importance of the private model grows (Mandl & Womser-Hacker 2001).

A parameter p is increased from zero to one while the user collects his decisions. We define p a a function of the number of relevance judgements provided by a user.

$$p_{user}(j) = \frac{1}{1 + e^{-(j-a)\lambda}} \qquad (3)$$

j *number of relevance judgements*

The following figure 1 shows a plot of the function for $\lambda = 0.1$ The optimal setting for λ has to be determined by experiments.

The models can be seen as a set of weights over all contributing systems:

- private model: ($\omega_{\text{private, A}}$; $\omega_{\text{private, B}}$; $\omega_{\text{private, C}}$; … ; $\omega_{\text{private, N}}$)
- public model: ($\omega_{\text{public, A}}$; $\omega_{\text{public, B}}$; $\omega_{\text{public, C}}$; … ; $\omega_{\text{public, N}}$)

The parameter p controls the weight of the private model in the final result. The inverse weight is given to the public model. The parameter p needs to be increased during the use of a retrieval system until the individual user model is saturated. The following formula reflects the modifications for a personalized MIMOR model integrating the private and the public model where q is the query and d a document:

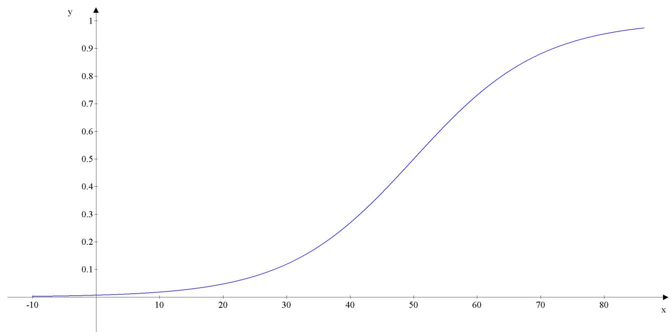

Fig 1: Function to Adapt Private Model Weight

(4) $$RSV_{PP}(q,d) = \frac{\sum_{i=1}^{N}((p\,\omega_{private,i} + (1-p)\,\omega_{public,i})\,RSV_i(q,d))}{N}$$

$$RSV_i(q,d) \quad \textit{the result of retrieval system } i$$

The distinction between private and public model can be further refined introducing group models at intermediate levels.

MIMOR learns the subjective relevance judgements of users or user groups. It models these opinions by changing the influence of several black box methods. The combination of these systems is optimized until the users´ cognitive decisions about relevance are adequately mirrored by the MIMOR model.

3.3 MIMOR and Context Adaptation

The context adatptation applies a similar algorithm as the individualization in MIMOR. Each user can be potentially in many different contexts. Consider different tasks at a company like customer research, proposal development or project work. In each of these contexts every individual may have his own preferences expressed by potential relevant documents. This preference is captured by collecting the user relevance judgements for the individual. In the same manner, all relevance judgements given by different users in the same context are assumed to capture the preferences for that context.

As a results, a MIMOR model for each context is established. Again, the models can be seen as a set of weights over all information retrieval engines:

- C: context model: ($\omega_{\text{context, A}}$; $\omega_{\text{context, B}}$; $\omega_{\text{context, C}}$; … ; $\omega_{\text{context, N}}$)
- G: global model: ($\omega_{\text{global, A}}$; $\omega_{\text{global, B}}$; $\omega_{\text{global, C}}$; … ; $\omega_{\text{global, N}}$)

We assume that just like an individual user model, each context model is not well established when a new context arises or develops. As a consequence, we assume that a global or context free model is built from all context models. This can be done by summarizing all relevance judgements given in all contexts. The final result considers both the global and context model.

$$(5) \quad RSV_{CG}(q,d) = f(\text{ context model}, \text{global model}, \text{q}, \text{d})$$

A parameter c governs the strength of the context model opposed to the global model. The weight of the global model is defined as 1-c. We assume that the fusion of the two models is realized as summation:

$$(6) \quad RSV_{CG}(q,d) = \frac{\sum_{i=1}^{N}((c\,\omega_{context,i} + (1-c)\,\omega_{global,i})\,RSV_i(q,d))}{N}$$

Again the modification of the parameter c can be modeled as a function of the number of relevance judgements available for that context.

$$(7) \quad p_{context}(j) = \frac{1}{1+e^{-(d-a)\lambda}}$$
$$d \quad number\ \ of\ \ relevance\ \ judgements$$

In a setting where individualization as well as context adaptation is desired, our model can integrate both. The final result is basically a fusion of four models which can be carried out in different ways:

$$(8) \quad \begin{aligned} &RSV(q,d) = f(\text{ context model }, \text{global model, private model, public model, q, d}) \\ &RSV\ (q,d) = f(RSV_{CG}(q,d), RSV_{PP}(q,d)) \end{aligned}$$

We propose a model with a emphasis on transparency which sums the two factors.

$$(9) \quad RSV\ (q,d) = \frac{RSV_{CG}(q,d) + RSV_{PP}(q,d)}{2}$$

The public and the global model may be identical but they do not necessarily have to be. In some cases, context and individual models can be represented differently and then no requirements are made about their values.

In addition to the individualization model we introduce a new feature in the learning algorithm. In cases where negative relevance feedback outweighs the positive judgements of the user, the feedback can be applied to penalize not only the retrieval systems assigning high values to the documents considered irrelevant. Furthermore, we propose that this information is also used to penalize the strength of the specific model, being it the individual or the context model. This may lead to a faster approximation of an optimal solution. We assume that negative relevance judgement is communicated to our system by a value smaller than zero and positive judgements as positive numbers between zero and one.

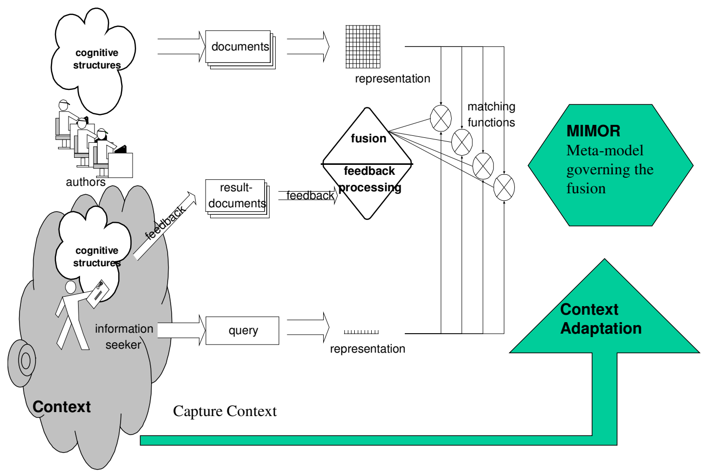

Fig. 2. Context adaptation in MIMOR as integration of relevance feedback processing and fusion

4 Conclusion and Future Work

We presented a model for context adaptation of an information retrieval system. The model is developed by extending a model for individualization which has previously been tested at the Cross Language Evaluation Forum (CLEF). The resulting model integrates both individualization as well as context adaptation.

The modified MIMOR model presented here needs to be evaluated in order to assess its effectiveness. A framework for evaluating context in information retrieval is outlined by Belkin et al. 2004. Context is modeled by some terms or keywords and thus by content. However, individualization is still hard to evaluate. We intend to extract a basis for evaluating individualization of retrieval effectiveness from the jurors relevance assessments from CLEF. For some tasks, the same document in different languages is judged by the several jurors. These judgements differ and can be interpreted as individual opinions about the relevance of the same document.

References

1. Bartell, B.; Cottrell, G, and R. Belew (1994): Automatic Combination of Multiple Retrieval Systems. In: Proc 17th Annual Intl ACM SIGIR Conf on Research and Development in Information Retrieval. Dublin, Ireland, pp. 173-181.
2. Belkin, N.J.; Muresan, G.; Zhang, X.-M. (2004): Using User's Context for IR Personalization. In: SIGIR 2004 Workshop on Information Retrieval in Context (IRiX). http://ir.dcs.gla.ac.uk/context/ pp. 23-25
3. Crestani, Fabio (1994): Comparing Neural and Probabilistic Relevance Feedback in an Interactive Retrieval System. In: IEEE Intl Conference on Neural Networks, vol. 5, Orlando. pp 3426-3430.
4. Fox, E.; Shaw, J. (1994): Combination of Multiple Searches'. The Second Text Retrieval Conference. (TREC-2), NIST Special Publication, pp. 243-252.
5. Fuhr, N. and C. Buckley: 1991, 'A Probabilistic Learning Approach for Document Indexing'. ACM Transactions on Information Systems 9 (3), 223-248.
6. Hackl, R., Kölle, R., Mandl, T. and C. Womser-Hacker (2002): Domain Specific Retrieval Experiments at the University of Hildesheim with the MIMOR System. In: Cross Language Evaluation Forum: Results of the CLEF 2002 Evaluation Campaign. Working Notes CLEF-Workshop, Rome, pp. 241-244.
7. Harman, Donna (1992a): Relevance Feedback Revisited. In: 15th Proc. Annual Intl. ACM SIGIR Conf on Research and Development in Information Retrieval. New York, USA, pp. 1-10.
8. Harman, D Donna (1992b): Relevance Feedback and other Query Modification Techniques. In: Frakes, W.B. and R. Baeza-Yates (eds.): Information Retrieval. Data Structures & Algorithms. Englewood Cliffs, NJ, pp. 241-263.
9. Harman, Donna (1994): 'Overview of the Third Text REtrieval Conference (TREC-3)'. The Third Text Retrieval Conference (TREC-3), NIST Special Publication, pp. 1-20.
10. Ingwersen, Peter: (1994); Polyrepresentation of Information Needs and Semantic Entities. Elements of a Cognitive Theory for Information Retrieval Interaction. In: 17th Annual Intl. ACM SIGIR Conference on Research and Development in Information Retrieval, New York, USA, pp. 101-110.
11. Ingwersen, Peter; Ruthven, Ian; Järvelin, Kalervo (2004): IR in Context - the Conceptions of Context. In: SIGIR 2004 Workshop on Information Retrieval in Context (IRiX). http://ir.dcs.gla.ac.uk/context/ pp. 6-9
12. Iyer, R., Lewis, D., Schapire, R., Singer, Y. and A. Singhal: 2000, 'Boosting for Document Routing'. Ninth ACM Conf on Information and Knowledge Management (CIKM). Washington, USA, pp. 70-77.
13. Lee, J. H. (1995) Combining Multiple Evidence from Different Properties of Weighting Schemes. In: Proc 18th Annual Intl ACM SIGIR Conf on Research and Development in Information Retrieval. Seattle, WA, USA, pp. 180-188.
14. Ma, Li; Goharian, Nazli (2005): Query length Impact on Misuse Detection in Information Retrieval Systems. In: Proc ACM SAC Symposium on Applied Computing (SAC). Information Access and Retrieval (IAR) Track. Santa Fe, New Mexico, USA. March 13.-17. pp. 1070-1075.
15. Mandl, Thomas; Womser-Hacker, Christa (2000): Ein adaptives Information Retrieval Modell für Digitale Bibliotheken'. Informationskompetenz - Basiskompetenz in der Informationsgesellschaft. Proceedings 7. Intl. Symp. für Informationswissenschaft (ISI). pp. 1-16.
16. Mandl, Thomas; Womser-Hacker, Christa (2001): Fusion approaches for mappings between heterogeneous ontologies'. In: Proc. 5th European Conference on Research and

Advanced Technology for Digital Libraries (ECDL) Darmstadt, Germany, LNCS. pp. 83-94.
17. McCabe, M.C., Chowdhury, A., Grossmann, D. and O. Frieder (1999): A Unified Framework for Fusion of Information Retrieval Approaches. In: Eighth ACM Conf on Information and Knowledge Management (CIKM), Kansas City, USA, pp. 330-334
18. Ruthven, Ian (2004): "and this set of words represents the user's context …" IR in Context - the Conceptions of Context. In: SIGIR 2004 Workshop on Information Retrieval in Context (IRiX). http://ir.dcs.gla.ac.uk/context/ pp. 10
19. Savoy, J. and Y. Rasolofo (2000): Report on the TREC-9 Experiment: Link-based Retrieval and Distributed Collections. In: The Ninth Text Retrieval Conference (TREC-9). pp 579-588.
20. Voorhees, E.; Gupta, N.; Johnson-Laird, B. (1995): Learning Collection Fusion Strategies. In: 18th Annual Intl ACM SIGIR Conf on Research and Development in Information Retrieval. Seattle, WA, USA, pp. 172-179.
21. Voorhees, Ellen; Harman, Donna (1997): Overview of the Sixth Text REtrieval Conference (TREC-6). The Sixth Text Retrieval Conference (TREC-6), pp 1-24.
22. Voorhees, Ellen; Harman, Donna (1999): The Eighth Text Retrieval Conference (TREC-8). NIST Special Publication. http://trec.nist.gov/pubs/trec8/t8_proceedings.htm.
23. Voorhees, Ellen; Harman, Donna (2001): The Tenth Text Retrieval Conference (TREC-10). NIST Special Publication. http://trec.nist.gov/pubs/trec10/t10_proceedings.htm.
24. Womser-Hacker, Christa (1997): Das MIMOR-Modell. Mehrfachindexierung zur dynamischen Methoden-Objekt-Relationierung im Information Retrieval. Habilitation thesis. University of Regensburg, Germany, Information Science.