Chapter 1
Introduction

With the emerging growth of the Internet, a huge amount of information is available to anyone. Even though everything *could* be accessed, the problem is to find relevant information. There are many examples where assistance is needed:

- **Online-Shopping:** Finding the right product within a huge catalogue is time-consuming for a user. Static online-shops organize their products within categories and hierarchies to facilitate browsing. Instead, personalized shops adapt their website to individual customers by factoring in their past actions. This helps the customer to find relevant products faster which leads to an increasing customer satisfaction, a higher purchasing rate and thus more profit. A successful example for such personalized recommender systems is Amazon[1].
- **Tagging:** Tagging is a popular technology in the Web 2.0. Tags allow the user to annotate items/ resources like songs, pictures, bookmarks, etc. with individual keywords. Tagging helps the user to organize his items and facilitate e.g. browsing and searching. But also in the process of tagging (that means annotating the 'right' keywords), assistance is important. Tag recommenders support the tagging process of a user by suggesting him tags that he is likely to use for an item.
- **Search Engines:** Web search is one of the most important tools of the Internet. It helps to find relevant information that is stored in the web. Typically, textual queries are used to search for web pages. The search engine returns a ranked list of pages that matches to the query. Most engines take the location of the user into account. Some engines adapt the results also to the individual user (Sun et al, 2005; Jeh and Widom, 2003).
- **Annotation:** Collaborative creation of content is another recent trend. The online encyclopedia Wikipedia[2] is the most famous example. All content is generated by the visitors and not by a small team of experts. Categorization of articles and links between articles are essential for browsing such large websites. But as the content is created by a large and diverse group of users, when editing content it is hard to find the right categories and links. Tools can help this process by suggesting categories or links.

[1] `http://www.amazon.com/`
[2] `http://en.wikipedia.org/`

S. Rendle: Context-Aware Ranking with Factorization Models, SCI 330, pp. 3–8.
springerlink.com

Table 1.1 Examples for scenarios of context-aware ranking

Scenario	Entity to rank	Context
Online Shopping	products	customer
Tagging	tags	user, bookmark/ song
Search Engine	web pages	search query, user
Wikipedia Annotation	categories	article

In all these examples, the problem is to rank *entities* given a *context*. Table 1.1 gives examples for the entity to rank and the context for the scenarios mentioned before.

In this work, we develop statistical methods that generate such context-aware rankings given observed data. Examples for this observed data are purchases in online shops, clicks on search result lists or already assigned categories for the wikipedia example. As we will discuss in detail in chapter 3, the problem setting is challenging and differs from standard machine learning settings:

1. Instead of classification or regression, we are interested in the less studied task of ranking. Moreover, unlike to the standard ranking literature, not a global ranking has to be found but the rankings should be *context-aware*. That means for each context another ranking is desired.
2. The observations are highly sparse and difficult to interpret. Sparseness means that for most data no observations have been monitored. For example a user has purchased only few products compared to the size of the whole catalogue. The interpretation of the non-bought products is difficult: Not having bought an item does not have to mean that the user dislikes this item but can also mean that he does not know it yet. This is especially crucial as we are mostly interested in ranking among the products a user has not bought yet.
3. The variables (e.g. user, product, web page, tag, ...) are defined over categorical domains with many levels. In contrast to real valued variables, we have no a priori knowledge about the space of these domains. E.g. one does not know a priori if two users are similar or not. This becomes even more crucial with the high sparsity mentioned above.

To solve these issues, we present a data interpretation that generates pairwise preferences for training. We develop a generic optimization criterion from the maximum a posteriori estimator of the pairwise interpretation. For learning model parameters, we present a gradient descent based optimization algorithm. We tackle the sparsity with factorization models that find latent representations of variable instances. These models allow to propagate information over variables, such that two 'similar' instances influence each other. Both the optimization framework and the proposed factorization models can solve problems of any number of modes. That means unlike traditional recommender systems that only work on two modes (e.g. user and item), our approach works also on problems with more modes (e.g. user, items, location, mood).

We provide case studies for item recommendation, tag recommendation and sequential set recommendation (item recommendation with time). Here, we compare our approach to state-of-the-art methods in each of these fields. It is important to note that our method of context-aware ranking is not limited to these applications. But within this book, we can only cover a fraction of all possible scenarios. We are confident that our method also works well in other applications and we are planing to continue this study in future work.

1.1 Overview

This book is organized in the three main parts of theory and application of context-aware ranking as well as general extensions.

1.1.1 Theory of Context-Aware Ranking

The first part deals with the general theory of context-aware ranking. In three chapters we discuss the problem setting, the optimization approach and the modelling with factorization models:

- **Chapter 3: Ranking from Incomplete Data**
  In this chapter, we develop a general method for context-aware ranking. We start with a detailed analysis of the observed incomplete data of a sparse relation over categorical variables. Then we formalize the task of context-aware ranking and show how pairwise preferences can be generated for obtaining training data. Afterwards we describe how a ranking relation can be expressed by a real valued function or by a tensor in the case of finite, categorical variables. Finally, we discuss evaluation metrics for ranking tasks.
- **Chapter 4: Learning Context-Aware Ranking**
  This chapter introduces the framework for Bayesian context-aware ranking (BCR). First, we derive the optimization criterion BCR-OPT which is the maximum a posteriori estimator. For optimizing models towards this criterion, we introduce the learning algorithm BCR-LEARN that is based on stochastic gradient descent with bootstrap sampling of training cases. We conclude with a comparison of BCR-OPT to other optimization criteria like element-wise losses and the area under the ROC curve.
- **Chapter 5: Factorization Models**
  As model, we propose factorization approaches which can model the latent interactions between variables. We discuss the Tucker decomposition (TD) and the PARAFAC model. From this we derive our pairwise interaction model (PITF) which is a special case of both TD and PARAFAC. We discuss the expressiveness and complexity of these models in detail.

1.1.2 Application of Context-Aware Ranking

In the second part, we apply our theory of context-aware ranking to three scenarios and compare our approaches to state-of-the-art methods within these fields.

- **Chapter 6: Item Recommendation**
  Item recommendation is the most well studied recommendation task for ranking. It is a two mode problem over users and items. Online shopping is an example for item recommendation. K-nearest neighbor and matrix factorization are the two most popular approaches for this task. We show how to apply our BCR optimization to both of these models. We compare BCR optimized models empirically to two state-of-the-art approaches: i.e. cosine-similarity and weighted least-square optimization.
- **Chapter 7: Tag Recommendation**
  Tag recommendation is a rather new field of study. Nevertheless it has already attracted a lot of research and many methods to solve this specific tasks have been proposed. In this chapter, we introduce the application of tag recommendation in detail and show how our context-aware ranking theory can be applied there. As indicated before, tag recommendation is a three mode problem over users, items and tags. We show empirically that our approach outperforms the current state-of-the-art methods including Folkrank and HOSVD both in runtime and quality. Furthermore our method was compared to many other approaches in the ECML/PKDD Discovery Challenge 2009 where we achieved the best quality.
- **Chapter 8: Sequential Set Recommendation**
  Time is a variable that is easy to track in almost all scenarios. In this chapter, we investigate the three mode problem of item recommendation with time. In comparison to the categorical variables that we have used so far, time is real valued. Thus, we have to treat it differently. Here, we examine sequences of shopping carts which reflects the sequential nature of time. For modelling, we use Markov chains. But instead of using standard chains, we introduce personalized (context-aware) chains – i.e. per user one chain. To overcome the sparsity problem, we factorize the chains. Empirically, we show that our new method of factorized personalized Markov chains outperforms both standard Markov chains and time-invariant factorization models.

1.1.3 Extensions

The third part covers two extensions that fall not directly under the theory of context-aware ranking.

- **Chapter 9: Time-variant Factorization Models**
  Instead of modelling the qualitative/ sequential aspect of time like in chapter 8, we model the time quantitatively inside the factors. Therefore, we make the Tucker decomposition time-variant by modelling each factor with a time-variant function. This function itself is factorized into basis functions and free parameters that should be estimated. Instead of choosing the basis functions fixed, we sample them from the observed data using a kernel approach. Finally, we evaluate time-variant models with Gaussian and exponential kernels on synthetic and real-world data sets. Note that this whole chapter is a general study of time-variant factors and not limited to context-aware ranking.

- **Chapter 10: One-Class Matrix Factorization**
  In the last chapter, we investigate a binary classification task over two modes where only one class is observed. Maximum margin matrix factorization (MMMF) is known to be a successful classifier for binary classification tasks over two modes but it is unclear how to apply it to one-class problems. Support vector machines (SVM) are another maximum-margin classifier that have already been applied for one-class problems. We transfer these ideas from one-class SVM to MMMF and propose one-class/ 1C-MMMF. Furthermore we extend it for cases where information about the prior class distribution is available to 1C-prior MMMF.

1.2 Contributions

The core contributions of this book are:

1. We develop a unified **theory of context-aware ranking** that subsumes several recommendation tasks including item, tag and context-aware recommendation.
2. **BCR optimization and learning** is proposed as a generic optimization framework.
3. **Factorization models** are used for modelling and we develop the **PITF** model for sparse problems.
4. **Factorizing Personalized Markov Chains (FPMC)** is introduced as an extension of Markov chains that also allows parameter estimation under sparsity.
5. We conduct **empirical studies** on the task of item recommendation, tag recommendation and sequential set recommendation.
6. **Time-aware factor models** are developed as a time-variant extension of general factorization models.
7. **One-class matrix factorization** with prior regularization is proposed to solve large scale problems with balanced classes.

1.3 Published Work

This book generalizes and builds on the following publications:

- Rendle and Schmidt-Thieme (2008), *Online-updating regularized kernel matrix factorization models for large-scale recommender systems*, in RecSys 08: Proceedings of the 2008 ACM conference on Recommender systems, ACM.
- Rendle, Marinho, Nanopoulos, and Schmidt-Thieme (2009), *Learning optimal ranking with tensor factorization for tag recommendation*, in KDD 09: Proceeding of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, New York, NY, USA.
- Rendle, Freudenthaler, Gantner, and Schmidt-Thieme (2009), *BPR: Bayesian personalized ranking from implicit feedback*, in Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI 2009).

- Rendle and Schmidt-Thieme (2009), *Factor models for tag recommendation in bibsonomy*, in Proceedings of the ECML-PKDD Discovery Challenge Workshop. **ECML/PKDD 2009 Best Discovery Challenge Award**
- Gantner, Freudenthaler, Rendle, and Schmidt-Thieme (2009), *Optimal ranking for video recommendation*, in Personalization in Media Delivery Platforms Workshop at the International ICST Conference on User Centric Media (PerMeD 2009).
- Rendle and Schmidt-Thieme (2010), *Pairwise interaction tensor factorization for personalized tag recommendation*, in Proceedings of the Third ACM International Conference on Web Search and Data Mining (WSDM 2010), ACM. **WSDM 2010 Best Student Paper Award**
- Rendle, Freudenthaler, and Schmidt-Thieme (2010), *Factorizing personalized markov chains for next-basket recommendation*, in WWW 10: Proceedings of the 19th international conference on World wide web, ACM, New York, NY, USA. **WWW 2010 Best Paper Award**

References

Gantner, Z., Freudenthaler, C., Rendle, S., Schmidt-Thieme, L.: Optimal ranking for video recommendation. In: Personalization in Media Delivery Platforms Workshop at the International ICST Conference on User Centric Media (PerMeD 2009) (2009)

Jeh, G., Widom, J.: Scaling personalized web search. In: WWW 2003: Proceedings of the 12th International Conference on World Wide Web, pp. 271–279. ACM, New York (2003)

Rendle, S., Schmidt-Thieme, L.: Online-updating regularized kernel matrix factorization models for large-scale recommender systems. In: RecSys 2008: Proceedings of the 2008 ACM Conference on Recommender Systems, pp. 251–258. ACM, New York (2008)

Rendle, S., Schmidt-Thieme, L.: Factor models for tag recommendation in bibsonomy. In: Proceedings of the ECML-PKDD Discovery Challenge Workshop (2009)

Rendle, S., Schmidt-Thieme, L.: Pairwise interaction tensor factorization for personalized tag recommendation. In: WSDM 2010: Proceedings of the third ACM International Conference on Web Search and Data Mining, pp. 81–90. ACM, New York (2010)

Rendle, S., Freudenthaler, C., Gantner, Z., Schmidt-Thieme, L.: BPR: Bayesian personalized ranking from implicit feedback. In: Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI 2009) (2009)

Rendle, S., Marinho, L.B., Nanopoulos, A., Schmidt-Thieme, L.: Learning optimal ranking with tensor factorization for tag recommendation. In: KDD 2009: Proceeding of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York (2009)

Rendle, S., Freudenthaler, C., Schmidt-Thieme, L.: Factorizing personalized markov chains for next-basket recommendation. In: WWW 2010: Proceedings of the 19th International Conference on World Wide Web, pp. 811–820. ACM, New York (2010)

Sun, J.T., Zeng, H.J., Liu, H., Lu, Y., Chen, Z.: Cubesvd: a novel approach to personalized web search. In: WWW 2005: Proceedings of the 14th International Conference on World Wide Web, pp. 382–390. ACM, New York (2005)