Chapter 11
Conclusion

In this book, we have studied multi-mode prediction problems. The focus was on problem settings with large categorical domains and high sparsity. Due to the sparsity and usually high imbalance, we are not interested in classification but in ranking the entities of one of the modes. Instead of creating one global ranking, the rankings should be context-aware – i.e. we create many rankings that depend on a given context. Important applications for this setting are recommender systems. There exist several recommender tasks, two of the most well studied ones are personalization and tag recommendation. Our developed method of context-aware ranking subsumes both of them and includes also other settings like time-awareness. Moreover other well-known applications like web search or multi-label classification (e.g. annotation, wikipedia categorization) can be seen as an instance of context-aware ranking.

Based on these ideas, we have introduced the theory of context-aware ranking. From a Bayesian analysis, we have derived the optimization criterion BCR-OPT which is the general MAP estimator of a parametrized model. The learning algorithm BCR-LEARN is a maximization procedure for BCR-OPT. As it is based on stochastic gradient descent, any model that can be expressed as a differentiable, non-recursive function with a finite set of parameters can be optimized. The bootstrapping approach with the proposed drawing schemes makes it applicable even in cases where the number of ranking triples is huge. We have demonstrated the usefulness of BCR on the task of item recommendation, tag recommendation and sequential set recommendation. Here, we have applied BCR to a wide variety of models including several factorization models (matrix factorization, TD, PARAFAC, PITF), k-nearest-neighbour and Markov chains. Throughout these experiments, BCR optimization has outperformed other state-of-the-art approaches like weighted regularized least-square in quality. These results indicate that choosing the right optimization criterion is important. Furthermore, it shows that BCR is generic and a good choice for many applications.

With respect to modelling, we have focused on factorization models. For two-mode problems, matrix factorization models are known to generate high quality predictions (e.g. for regression or binary classification). In this book, we extended

S. Rendle: Context-Aware Ranking with Factorization Models, SCI 330, pp. 173–176.
springerlink.com

these factorization approaches to multi-mode problems. The Tucker decomposition and the PARAFAC model are such multi-mode factorization models. We have discussed their strength and limitations in detail showing that TD results in slow runtime for our multi-mode settings. Furthermore our empirical results indicate that optimizing model parameters for both TD and PARAFAC with standard Gaussian priors (aka ridge regression) can result in bad prediction quality. Our assumption is that in sparse settings, both of these models are too expressive and lack an a priori structure. To solve this, we propose the more restricted model PITF that explicitly models pairwise-interactions. We have shown that both TD and PARAFAC subsume this model. But in our evaluation, PITF outperforms both TD and PARAFAC. Thus our empirical results indicate that restricting the expressiveness and predefining a structure on TD models makes sense in sparse settings.

General factorization models like TD assume finite variable domains. For handling infinite domains like time, we present two extensions. The first one is a Markov chain, where sequential pattern can be found. Here, we extend the general Markov chain with personalization and secondly we model the transition cube with a factorization model (e.g. TD/PARAFAC/PITF). We have shown that this model subsumes both standard MCs and the standard non-time-aware factorization models. The second kind of time-variance we investigate, is variance within factors. Each factor is modelled time-dependent by decomposing it into basis functions and free parameters. We have shown, how these basis functions can be generated/ sampled from the observed data using a kernel approach.

11.1 Summary of Contributions

In total, the contribution of this book are:

1. **Theory of Context-Aware Ranking**
   We introduce context-aware ranking. This subsumes many important tasks like item recommendation and tag recommendation. Although each of these tasks has attracted a lot of research, they are usually studied isolated of each other. Our work brings them together which allows to transfer results from one domain to the other.
2. **BCR optimization and learning**
   We develop the Bayesian Context-aware Ranking (BCR) method that consists of a new optimization criterion BCR-OPT and a learning algorithm BCR-LEARN. BCR-OPT is the MAP estimator of the model parameters given context-aware ranking constraints that are derived from sparse observations. BCR-LEARN is a generic learning algorithm for context-aware ranking that is based on stochastic gradient descent with bootstrap sampling.
3. **Factorization Models (PITF)**
   For modelling the multi-mode data, we suggest to use factorization models based on Tucker decomposition. We discuss the practical limitations of general Tucker decomposition in terms of runtime and regularization. To solve this, we have

developed the pairwise interaction model PITF and shown that for ranking problems, the model complexity is linear.

4. **Factorizing Personalized Markov Chains**
   Moreover, we have extended Markov chain models by personalization and factorization. Personalization allows each user to have an individual Markov chain – i.e. own transition probabilities. By factorizing transition cubes, we solve the problem of sparsity in the data. That means, information propagates over the whole transition cube and the estimation gets more reliable than MLE with full parametrized models.
5. **Empirical Studies & Applications**
   We have applied our theory of context-aware ranking, the BCR optimization method and the factorization models to several applications: (1) online shopping, movie rental and online TV (Gantner et al, 2009) (item recommendation), (2) bookmark and music tagging (tag recommendation) and (3) sequential basket recommendation. In all of these applications, our proposed models have shown to outperform current state-of-the art methods. Furthermore our method has won the tag recommender challenge of the 2009 ECML/PKDD Discovery Challenge.
6. **Time-aware Factor Model**
   For handling time variance, we have extended the general factorization model with time-variant factors. This is done by decomposing each factor into a set of basis functions and free parameters. We provide a method for generating the set of basis functions from the observed data using a kernel assumption.
7. **One-Class Matrix Factorization**
   We extend the binary Maximum Margin Matrix Factorization classifier to handle one-class problems. This is done by transferring ideas from one-class support vector machines to matrix factorization. We show that the optimization of 1C-MMMF is invariant to the size of the bias and thus can be kept constant. We extend this model by prior regularization to 1C-PMMMF which allows to take assumptions about the class distribution prior into account.

11.2 Future Directions

Besides the directions that have been mentioned for specific tasks throughout this work, we see three major directions of future work:

- **Multirelational Prediction**
  The settings of this work can be seen as single relational. That means that there is one relation (e.g. a customer buys a product) and this relation should be predicted – i.e. the instances of one variable should be ranked given the others. But often additional information is available like information about products (price, category, ...) or information about customers (gender, age, ...). Especially in sparse settings, this additional information might be helpful to create better predictions of the target relation. There is already much work (Tso and Schmidt-Thieme, 2006; Agarwal and Chen, 2009; Tso-Sutter et al, 2008) on special cases (e.g. attribute-aware or tag-aware recommender systems) but only limited one for the

general case (Lin et al, 2009). Thus, one direction of future work is to develop a generic method for multi-relational settings and to integrate it in our work on context-aware ranking.

- **Regularization**
  As the empirical results for tag recommendation have shown, the restricted PITF is able to outperform PARAFAC and Tucker decomposition. But both PARAFAC and TD subsume PITF, so they should be able to generate at least as good predictions as PITF. We have argued that the reason might be the regularization together with the sparsity. Thus, a very interesting topic would be to investigate other regularization methods than 0-based Gaussian priors. Solving this problem would help to build better factorization models where the structure of the model is chosen data-dependent.
- **Applications**
  In this book we could only investigate some applications of context-aware ranking. Due to the success of our approach in our selected applications, we assume that it might also improve results on other tasks. Examples for further applications are: (1) personalized web-search, where web pages are suggested for a user given a query, (2) context-aware advertising, where ads are ranked based on a context like user, web page, last actions, (3) annotation like wikipedia categorization or (4) multi-label classification in general.

References

Agarwal, D., Chen, B.C.: Regression-based latent factor models. In: KDD 2009: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 19–28. ACM, New York (2009)

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)

Lin, Y.R., Sun, J., Castro, P., Konuru, R., Sundaram, H., Kelliher, A.: Metafac: community discovery via relational hypergraph factorization. In: KDD 2009: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 527–536. ACM, New York (2009)

Tso, K., Schmidt-Thieme, L.: Evaluation of attribute-aware recommender system algorithms on data with varying characteristics. In: Ng, W.-K., Kitsuregawa, M., Li, J., Chang, K. (eds.) PAKDD 2006. LNCS (LNAI), vol. 3918, pp. 831–840. Springer, Heidelberg (2006)

Tso-Sutter, K., Marinho, L., Schmidt-Thieme, L.: Tag-aware recommender systems by fusion of collaborative filtering algorithms. In: Proceedings of 23rd Annual ACM Symposium on Applied Computing (SAC 2008), Fortaleza, Brazil (to appear) (2008)