Review on Movie Recommender Systems

Shikha Sharma[1] *Lakshika Bhardwaj*[1] *Komal Goel*[1] *Saurabh Srivastava*[2]
[1]*B.Tech Student,* [2]*Faculty*
Department of Information Technology
ABES, Engineering College, Ghaziabad, India
[1]*shikha105789@gmail.com* [1]*bhardwajlakshika@gmail.com*
[1]*kittu93komal@gmail.com* [2]*saurabh.srivastava@abes.ac.in*

Abstract:

Recommender systems are a subclass of information filtering system that seek to predict the 'rating' or 'preference' that user would give to an item.

Recommender systems research is being slowed by the difficulty of replicating and comparing research results. Published research uses various experimental methodologies and metrics that are difficult to compare. It also often fails to sufficiently document the details of proposed algorithms or the evaluations employed. Researchers waste time implementing well-known algorithms the new implementations may miss key details from the original algorithm or its subsequent refinements. When proposing new algorithms, researchers should compare them against finely-tuned implementations of the leading prior algorithms using state-of-the-art evaluation methodologies. With few exceptions, published algorithmic improvements in our field should be accompanied by working code in a standard framework, including test harnesses to reproduce the described results. To that end, we present the design and freely distributable source code of LensKit, a flexible platform for reproducible recommender systems research. Lens Kit provides carefully tuned implementations of the leading collaborative filtering algorithms, APIs for common recommender system use cases, and an evaluation framework for performing reproducible offline evaluations of algorithms.

Keywords: LensKit, Recommender systems, GroupLens

I. INTRODUCTION:

Rating prediction is the elementary task performed with rating-based data. Given a particular item and user, the goal is to predict the user's true rating for the item in question. Early work on rating prediction focused on neighborhood-based methods such as the GroupLens algorithm. Personalized recommendations can be generated for any user by first predicting ratings for all items the user has not rated, and recommending items with the highest predicted ratings. The capability to predict ratings has other interesting applications. Rating predictions can beingincorporated with content-based scores to create a preference augmented search procedure. Rating prediction also facilitates an active approach to collaborative filtering using expected value of information. In such a framework the predicted rating of each item is interpreted as its expected utility to the user Recommender evaluation is also not handled consistently between publications. Even within the same basic structure, there are many evaluation methods and metrics that have been used. Important details are often omitted or unclear, compounding the difficulty of comparing results between papers or lines of work.

.LensKit is intended to provide a robust, extensible basis for research and education in recommender systems. It is suitable for deployment in research-scale recommender systems applications, developing and testing new algorithms, experimenting with new evaluation strategies, and classroom use. Its code and surrounding documentation also serves as documentation of what is necessary to take recommender algorithms from the mathematical descriptions in the research literature to actual working code.

II. PAPER WISE CRITICAL REVIEWS

Paper	**Methods and Technologies Used**	**Strength**
[1]Rethinking the Recommender Research Ecosystem: Reproducibility, Openness, and LensKit	Took lensKit data ,used SVD	Demonstrated the utility of LensKit
[2]Modeling User Rating Profiles For Collaborative Filtering	This model combines the intuitive appeal of the multinomial mixture and aspect models, with the strong high level generative semantics of LDA and mPCA.	The URP model performs significantly better than the other methods, and obtains the lowest prediction error
[3]Latent Semantic Analysis for Multiple-Type Interrelated Data Objects	Pearson method which uses correlation coefficient to find the nearest neighbors for an active user	Based on the mutual reinforcement principle as used in the traditional LSA,M-LSA identifies the most salient concepts among all the co-occurrence data and represents each object in a unified semantic spa
[4]Fast Maximum Margin Matrix Factorization for Collaborative Prediction	Took lensKit data, factor models as feature learning, low-norm factorization under maximum margin matrix factorization.	In this work, they have shown that it is possible to “scale-up” MMMF to large problems

III. PAPER WISE RESULTS DISCUSSION

[i]Rethinking the Recommender Research Ecosystem: Reproducibility, Openness, and LensKit	In this paper, they have presented LensKit, an open source recommender systems toolkit intended to foster more open recommender systems research. LensKit is a flexible platform to enable researchers to easily implement and evaluate recommender algorithms. LensKit can also serve as a production recommender for small to medium scale lab or field trials of recommender interfaces, which will benefit from the finely-tuned implementations of the leading algorithms.
[ii]Modeling User Rating Profiles For Collaborative Filtering	In this paper they have presented the URP model for rating-based collaborativefiltering. Their model combines the intuitive appeal of the multinomial mixture and aspect models, with the strong high level generative semantics of LDA and mPCA.
[iii]Latent Semantic Analysis for Multiple-Type Interrelated Data Objects	They compare their result with the standard SVM on the email-word matrix and LSA based method. They set α = 0.3. It is clear that M-LSA can outperform the baseline, while LSA cannot. When they set the number of dimensions to 50, M-LSA achieves 0.803 on micro-F1.

[iv]Fast Maximum Margin Matrix Factorization for Collaborative Prediction	In this work, we have shown that it is possible to "scale-up" MMMF to large problems. We used gradient descent on *U*, *V* and θ to find an approximate minimum to the MMMF objective. Although $J(U, V, \theta)$ is not convex, an empirical analysis indicated that local minima are, at worst, rare.

IV. CONCLUSION

It is currently difficult to reproduce and extend research results and algorithmic developments in recommender systems .A culture of open code and reproducible experiments in common frameworks will move the recommender systems community forward by fostering increased reuse and extension and making it easier to compare against prior results. LensKit is a flexible platform to enable researchers to easily implement and evaluate recommender algorithms. LensKit can also serve as a production recommender for small to medium scale lab or field trials of recommender interfaces, which will benefit from the finely-tuned implementations of the leading algorithms.

We have performed experiments demonstrating that the LensKit implementations meet or beat the canonical implementations of the best-known algorithms. We have further demonstrated that the platform is a valuable tool for addressing open questions in the field. Development on LensKit is ongoing. Also the URP method has the capability to operate directly on ratings data, and to efficiently predict all missing ratings in a user profile. This means URP can be applied to recommendation, as well as many other tasks based on rating prediction.. For online applications where it is impractical to refit the model each time a rating is supplied by a user, the result of interest is strong generalization performance. The aspect model cannot be applied in a principled manner in such a scenario, and we see that URP outperforms the other methods by a significant margin. Based on the mutual reinforcement principle as used in the traditional LSA, M-LSA identifies the most salient concepts among all the co-occurrence data and represents each object in a unified semantic space. M-LSA is general and covers several variants of LSA as special cases. All the experiments showed that M-LSA is effective in utilizing all the information on a multiple-type graph. M-LSA can be applied to any co-occurrence data involving multiple types of objects, thus has potentially many applications in multiple domains.

V. FUTURE WORK

Support for content-based and hybrid recommender systems. The current algorithms in LensKit are all collaborative filtering algorithms, but the API and infrastructure are designed to support more general types of recommenders as well. We will be implementing example hybrid content/collaborative algorithms to demonstrate this capability. A framework for learning parameters through cross-validation, many algorithms have features, such as learning rates and smoothing terms that may be learned through experimentation. A general framework will be provided to learn parameters, including an evaluation methodology to accurately measure the performance of the resulting automatically-tuned algorithm. Additional evaluation strategies in addition to the train-test prediction evaluation, temporal evaluation methods will be provided, such as time-averaged RMSE and Profile MAE, and recommendation list evaluations .We can also extend LensKit with recommender throughput evaluation strategies. Our community is very successful today, but we believe the recommender systems ecosystem approach to research offers an opportunity for even greater success in the future .The research community will benefit from increased publication of working implementations, easier access to the state of the art, and a library of reproducible evaluations on publicly-available data.

ACKNOWLEDGMENT

We gratefully acknowledge our mentor Saurabh Srivastava(Assistant Professor, Department Of Information Technology, ABES Engineering College,INDIA) for his constant motivation and support.

REFERENCES

[1] Rethinking the Recommender Research Ecosystem:Reproducibility, Openness, and Lens Kit Michael D. Ekstrand, Michael Ludwig, Joseph A. Konstan, and John T. RiedlGroupLensResearchUniversity of MinnesotaDepartment of Computer Science{ekstrand,mludwig,konstan,riedl}@cs.umn.edu

[2]A Big-Data Perspective on AI: Newton, Merton, and Analytics IntelligenceFei-Yue Wang, State Key Laboratory of Management and Controlfor Complex Systems 2012

[3]Business Intelligence from Social Media A Study from the VAST Box Office ChallengeYafeng Lu, Feng Wang, and Ross Maciejewski■ Arizona State University2014.

[4]Clouds forScalableBigDataAnalyticsDomenico Talia University of alabria, Italy2013.[5] Using In-Memory Analytics to Quickly Crunch Big DataLee Garber 2012.

[6] Big Data: New Opportunities and New Challenges, Katina Michael, University of WollongongKeith W. Miller, University of Missouri–St. Louis.

[7]Latent Semantic Indexing: An overviewINFOSYS 240 Spring 2000 Final Paper Barbara Rosario.

[8] Mining Social Media Data for Understanding Students' Learning Experiences in Chen, Student

Member, IEEE, Mihaela Vorvoreanu, and Krishna Madhavan.

About Authors

Shikha Sharma is a final year student of Btech in ABES Engineering College at Uttar Pradesh Technical University. She has done her final year project on "Movie Recommender System". She has written and sent the research paper on Movie Recommender System to get published. Her research interests include Big data technology.

Lakshika Bhardwaj is a final year student in ABES Engineering College at Uttar Pradesh Technical University. She has done her major project of Btech on "Movie Recommender System".

Komal Goel is a final year student in ABES Engineering College at Uttar Pradesh Technical University. She has done her major project of Btech on "Movie Recommender System". Her area of interests include web development.

Saurabh Kumar Srivastava is working as Assistant Professor in the department of Information Technology at ABES, Engineering College Ghaziabad (U.P.). He obtained his M-Tech (Computer Engineering) from Shobhit University and B.Tech(C.S.) from Uttar Pradesh Technical University, Lucknow (U.P.).

He has been in teaching since 7+ years. During his teaching he has coordinated several Technical fests and National/ International Conferences at Institute level. He has attended several seminars, workshops and conferences at various levels. His area of research includes Data mining, information retrieval, soft computing, and Web Technology.