Visualization of Bias of Machine Learning for Content Recommendation

Ami Tochigi* — Ochanomizu University

Takayuki Itoh† — Ochanomizu University

Xiting Wang‡ — Microsoft Research Asia

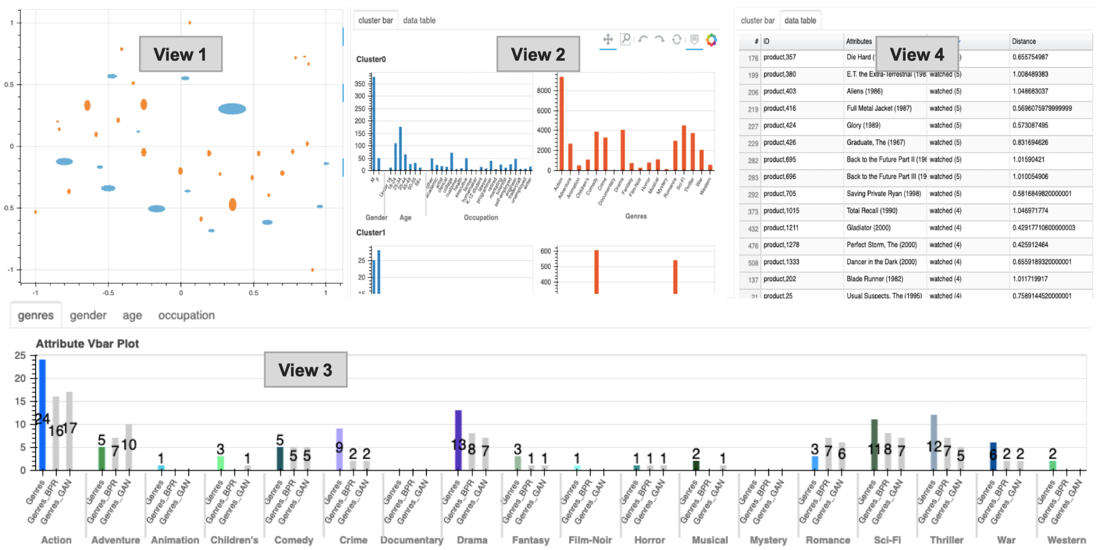

Figure 1: A snapshot of our system for visualization of biases of content recommendation. The system comparatively visualizes attributes of previously appreciated and currently recommended contents with the following four components. **View 1:** Scatterplot that displays the clustering results of contents and customers. **View 2:** Bar chart that displays the distribution of features of a user-specified cluster. **View 3:** Bar chart that displays features of a user-specified content or customer. **View 4:** Data table that displays the detailed information of a user-specified content or customer.

Abstract

Machine learning techniques have been applied to content recommendation systems. Meanwhile, the fairness/bias of machine learning has been an actively discussed issue. The bias of training datasets may cause such unfair or biased learning results. We developed a visualization system to assist the comparison between the training datasets and learning results and the discovery of the bias in the learning results. This visualization system has been applied to a movie recommendation system. This paper discusses how the bias exists in the movie recommendation results.

1 Introduction

Recommendation systems that suggest customers' favorite contents automatically have become widespread. Meanwhile, excessively personalized recommendation systems may cause biased recommendation results that contain limited ranges of contents. In recent recommendation systems, machine learning techniques [5] have been applied to recommendation engines in order to avoid the cold start problem and high loads caused by the increase in the number of contents and customers. This poster presents a visualization system for machine learning processes on recommendation systems. The presented study focuses on the bias of recommendations conducted by machine learning models. The system is useful while comparing statistics of appreciation and recommendation of the particular contents/customers and exploring how biased recommendations exist among the contents and customers.

*e-mail: tochigi.ami@is.ocha.ac.jp

†e-mail: itot@is.ocha.ac.jp

‡e-mail: xitwan@microsoft.com

2 Related Work

Visualization is a powerful tool for explainable recommendation systems, and actually, there have been several studies on visualization for recommendation systems. However, the field of study focusing on visualization of fairness and bias of recommendation systems. is a new field, and there have been few previous studies [1, 2]. They focused on general issues of fairness among people while operating the machine learning tasks. On the contrary, our study focuses on the bias of learning results for recommendation systems by visualizing the distribution of both people and contents.

3 Presented Visualization System

This section presents our visualization system that assists the comparison between the training datasets (the history of personal contents appreciation) and the recommendation results (a set of contents recommended to each customer) so that visualization users can verify the bias of the recommendation systems. We defined the visualization tasks as follows:

T1: Visualize the patterns of contents and customers, including the clusters of contents that are appreciated by similar customers, and the clusters of customers who appreciate similar contents.

T2: Find the clusters that recommendation results are much different from appreciation history that may cause unsatisfactory recommendations, or the clusters that contain diverse or outlier contents/customers that may cause biased recommendations.

T3: Observe the detailed statistics of attributes in a particular cluster of contents/customers so that users can find specific reasons for biased recommendations.

T4: Observe the detailed information of contents appreciated by a particular customer or customers who appreciated a particular content so that users can deeply discuss the cause of biased recommendations.

As shown in Figure 1, our implementation features four visualization components to compare the history of contents appreciation and recommendation results.

View 1 is a scatterplot that displays the clusters of contents and customers by applying a graph visualization technique [4]. **View 1** has the following two types of color codings, and users perform T1 and T2 by observing **View 1**. The scatterplot allows users zoom and filter operations so that they can focus on particular clusters.

Color coding 1 (Similarity with a user-specified node): When a user specifies a content node, the system calculates the cosine similarity of the feature vectors between other content nodes and the specified node and then assigns whose saturation is proportional to the similarity. The same applies to the case of customer nodes. We can observe how similar or dissimilar contents/customers distribute in the scatterplot.

Color coding 2 (Difference between appreciation and recommendation): The system calculates the cosine similarity between the features vectors of each content, or the feature vectors of each customer. Saturations assigned to the nodes are larger when the cosine similarity is smaller. We can observe which clusters contain many contents/clusters that have dissimilar recommendation results against the appreciations.

View 2 is a bar chart that displays the distribution of features of a user-specified cluster. Users observe the distributions of features of the clusters on **View 2** to perform **T3**. **View 3** is a bar graph that displays the features of user-specified contents and customers, and **View 4** is a data table that displays detailed information of user-specified contents and customers. Also, users can specify particular contents or customers by click operations and observe the detailed information of them on **View 3** and **View 4** to perform **T4**.

4 Example

4.1 Preprocessing

We had an experiment with a movie appreciation dataset [3]. The dataset contains 3,883 movies and 6,040 customers. Movies are categorized according to 18 genres. Customers have attributes including gender, age, and 21 types of occupations. We applied two machine learning models (BPR [5], BPR with GAN [6]) by consuming the movie appreciation dataset as a training dataset, and then conducted the recommendation to the customers of the dataset based on the learning results. We divided the customers into the training and test datasets, and then applied the training dataset to the machine learning process. Based on the learning results, we conducted the recommendation process and actually recommended 20 movies to the customers contained in the test dataset. Finally, we visualized 1,000 customers and 512 movies sampled from the test dataset.

4.2 Case Study

We visualized the dataset described above and explored it to discover the bias of recommendation results. Here, the blue bar charts in View 2 display the statistics of gender, age, and occupation of customers, while the orange bar charts display the statistics of genres of movies. View 3 has four tabs for gender, age, occupation, and genre. Users can selectively observe the differences in the statistics between appreciation and recommendation for each attribute. Colored bars in View 3 are statistics of appreciation while gray bars are statistics of recommendation.

We firstly observed View 1 with Color coding 2 and found several movie/customer clusters that contain movies or customers whose differences between appreciations and recommendations were large. many customers that have larger differences between appreciations and recommendations can be found in clusters in 1 and 5 on the scatter plot shown in Figure 2. These examples demonstrate that we can easily discover bias of recommendation results in particular clusters. The bar graphs in Figure 2 shows the statistics of customers in clusters 1 and 5. The left blue bar chart depicts that the number of female customers is larger than males in cluster 1. The right blue bar chart depicts that children (whose ages are less than 15) occupy a large part of cluster 5. This features are not present in the majority of the other clusters. These results suggest that the difference in recommendation results among the clusters is caused by the differences in attributes of customers.

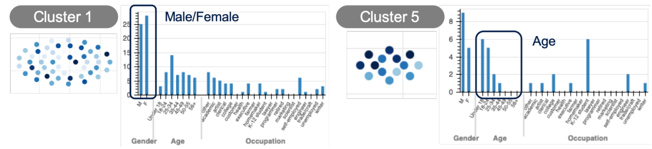

Figure 2: Customers in clusters 1 and 5 visualized by View 2.

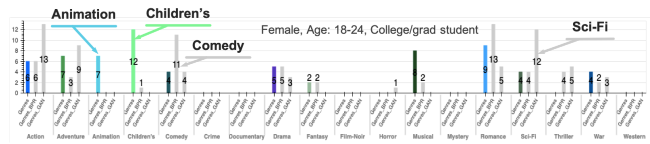

Figure 3: Genres of appreciated and recommended movies of particular customers belonging to clusters 5 visualized by View 3.

Next, we specified some customers in the clusters described above and visualized the differences of statistics of their appreciations and recommendations by View 3. Figure 3 shows the statistics of movies watched and movie genres recommended for customers belonging to cluster 5. The bar chart depicts that comedy and scientific fiction movies are well recommended to the corresponding customer (female, age between 18 and24, college/grad student) though she mainly appreciated animation and children's movies.

The above case study demonstrates a typical usage scenario of the presented visualization tool. We suppose users firstly overview the distribution and clusters of biased/unbiased customers and contents. Then, the users would focus on statistics of attributes of particular clusters, and/or explore the bias (the difference between appreciation and recommendation) of particular customers or contents while reading the detailed information of them shown on the data table.

5 Conclusion

This poster presented a visualization tool aiming the discovery of bias of machine learning results of the recommendation systems. This study supposes that a number of customers appreciate a number of contents and a recommendation system applies machine learning methods to form recommendation models from the history of contents appreciation of customers.

References

[1] Y. Ahn and Y. Lin, FairSight: Visual Analytics for Fairness in Decision Making, IEEE Transactions on Visualization and Computer Graphics, 26(1), 2020.

[2] A. Cabrera, et al., FairVis: Analytics for Discovering Intersectinal Bias in Machine Learning, IEEE Conference on Visual Analytics Science and Technology, 2019.

[3] F. M. Harper and J. A. Konstan. The Movielens datasets: History and context, ACM Transactions on Interactive Intelligent Systems (TiiS), 5(4), 2015.

[4] T. Itoh, K. Klein, Key-node-Separated Graph Clustering and Layout for Human Relationship Graph Visualization, IEEE Computer Graphics and Applications, 35(6), 30-40, 2015.

[5] S. Rendle, et al., BPR: Bayesian Personalized Ranking from Implicit Feedback, In Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, 452-461, 2009.

[6] I. J. Goodfellow, et al., Generative Adversarial Nets, In Proceedings of Neural Information Processing Systems (NIPS), 2672-2680, 2014.