Flexible Adjustment of Feature Vector Correlations for Personalized Recommendations

Tsuyoshi Yamashita[1] and Kunitake Kaneko[1,2]

[1] Research Institute for Digital Media and Content, Keio University, Japan
[2] Faculty of Science and Technology, Keio University, Japan
{tullys,kaneko}@inl.ics.keio.ac.jp

Abstract. Personalized recommendation vectors are often correlated with the global feature vector, such as a popularity vector. Prior works tried to adjust the correlation using parameters. However, there are no parameter setting guidelines, and it is unclear whether the desired correlation is achieved. We propose CoCoA to generate the recommendation vector with a linear combination of the given source and global vectors. It ensures the cosine similarity between the recommendation and global vectors is a user-input value. A case study with a movie dataset observed CoCoA can flexibly control the effect of popular movies. Evaluations on three real-world graph datasets showed that only CoCoA achieved negative cosine similarity, where the globally important nodes are suppressed.

Keywords: personalization, recommendation, correlation adjustment

1 Introduction

Recently, personalized recommendations have become increasingly important due to the rapid growth of digital content. In content-based personalized recommendation, a recommender generates a recommendation vector that quantifies the importance of other content based on the user's interests. It is particularly helpful for new users without much user-specific history, as they are less likely to benefit from user-based recommendation. Graph-based recommendation is an example [1].

In the personalized recommendation, the recommendation vector tends to be correlated with global feature vectors, such as a popularity vector. Examples of such correlations include popularity bias and filter bubbles [2]. If the correlation is too strong, personalization will be compromised because popular content is recommended too much. On the other hand, if the correlation is too weak, accuracy will be sacrificed because minor and low-quality content may be recommended. Therefore, users need to adjust the correlation level according to the situation.

To mitigate the correlation, existing works introduced adjustment parameters [3–8]. However, these methods suffer from unclear parameter setting guidelines and a narrow adjustable range. Users require many trials to explore the parameters for the desired correlation. Additionally, the parameters to achieve the desired correlation may not exist.

This paper proposes CoCoA to compute recommendation vectors with a given source vector, while ensuring that the correlation between recommendation and global vectors is a user-input value $\in (-1, 1)$. In particular, CoCoA determines the recommendation vector with a linear combination between the source and global vectors. By mathematically determining the coefficient, the cosine similarity is ensured to be the user-input value. A case study with a movie rating dataset confirmed CoCoA can flexibly control the proportion of globally important movies included in the recommendation list. Evaluations using three real-world datasets showed CoCoA was the only method that achieved negative cosine similarity, which suppresses the globally important nodes.

Finally, the remainder of this paper is structured as follows. Sect. 2 discusses the related work. Sect. 3 provides the proposed method. Sect. 4 evaluates our method with real-world datasets. Sect. 5 concludes this paper.

2 Related Work

We organize the existing methods for adjusting correlations between feature vectors. In the following, we categorize them according to the timing and the quality of the adjustment.

First, in terms of the timing of adjustment, the existing methods are classified into pre-processing, in-processing, and post-processing. Pre-processing methods adjust the correlation before computing the recommendation vectors [3]. These methods need to reperform pre-processing to change the correlation level. Thus, they are not suitable for our applications. In-processing methods adjust the correlation while computing the recommendation vector. Some methods that use random walks on graphs control the probability of visiting the high degree nodes [4, 5]. Another method exploits the random walk lengths [6]. Post-processing methods compute the recommendation vector by manipulating the given source and global vectors. Some methods weight each element according to the value of the global vector [7, 8]. Our method falls under the post-processing category.

Second, in terms of the quality of adjustment, all existing methods do not provide the parameter setting guidelines. Thus, they suffer from heavy parameter tuning. In addition, the parameters for the desired correlation may not exist. On the other hand, our method clearly decides the parameter in terms of the cosine similarity between the recommendation and global vectors.

3 Proposed Method: CoCoA

This section introduces our method CoCoA. It guarantees that the cosine similarity between the computed recommendation vector and the user-input global vector (e.g., popularity vector) is an input value $\in (-1, 1)$.

3.1 Approach

We describe an approach to adjust the correlation between the recommendation vector $\hat{\boldsymbol{\pi}}$ and the global vector $\mathbf{g}$ using the input source vector $\boldsymbol{\pi}$. Here, the source

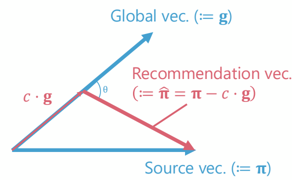


Fig. 1: An approach to adjust the cosine similarity using the linear combination.

vector is a vector computed by any recommendation algorithm, and the global vector is a feature vector defined for each dataset. The recommendation vector is computed by a linear combination between the source and global vectors. Therefore, the recommendation vector $\hat{\boldsymbol{\pi}}$ is computed as Eq. (1), where c is the coefficient of the linear combination. c affects the correlation between the recommendation vector and the global vector. In the following, the correlation is expressed by the cosine similarity.

$$\hat{\boldsymbol{\pi}} = \boldsymbol{\pi} - c \cdot \mathbf{g} \tag{1}$$

Fig. 1 illustrates the relationships among the recommendation vector $\hat{\boldsymbol{\pi}}$, the source vector $\boldsymbol{\pi}$, and the global vector $\mathbf{g}$. From Fig. 1, if the starting points of $\boldsymbol{\pi}$ and $\mathbf{g}$ are the same, the ending point of $\hat{\boldsymbol{\pi}}$ vector will match that of $\boldsymbol{\pi}$, and the starting point will exist on $\mathbf{g}$ or an extended line of it. The key point is the cosine similarity between $\hat{\boldsymbol{\pi}}$ and $\mathbf{g}$ changes monotonically over c. Moreover, there always exists a value of c that achieves any cosine similarity $\in (-1, 1)$. Thus, the existence of the unique c to meet the input cosine similarity is ensured.

3.2 Determination of the coefficient c

We describe how to decide the coefficient of the linear combination c such that the cosine similarity between $\hat{\boldsymbol{\pi}}$ and $\mathbf{g}$ is the user-input x. Note that in the following, we assume that $\|\boldsymbol{\pi}\| = \|\mathbf{g}\| = 1$ and $\boldsymbol{\pi} \neq \mathbf{g}$.

The cosine similarity $x \in (-1, 1)$ between $\hat{\boldsymbol{\pi}}$ and $\mathbf{g}$ is represented as Eq. (2), where k denotes the inner product between $\boldsymbol{\pi}$ and $\mathbf{g}$.

$$x = \frac{\hat{\boldsymbol{\pi}} \cdot \mathbf{g}}{\|\hat{\boldsymbol{\pi}}\|\|\mathbf{g}\|} = \frac{\boldsymbol{\pi} \cdot \mathbf{g} - c\|\mathbf{g}\|^2}{\|\boldsymbol{\pi} - c \cdot \mathbf{g}\|} = \frac{k - c}{\sqrt{1 - 2ck + c^2}} \tag{2}$$

Thus, Eq. (3) is introduced, showing the unique c for the cosine similarity x.

$$c = k - x\sqrt{\frac{1 - k^2}{1 - x^2}} \tag{3}$$

4 Evaluation

4.1 Settings

We evaluate the proposed method, CoCoA, using three real-world graph datasets listed in Table 1. In the following, we take graph datasets as the input data, the

Table 1: Datasets (K=10^3, M=10^6, B=10^9)

Name	Edge Type	Edge Weight	#nodes	#edges
MovieLens [10]	Undirected	Yes	418K	34M
LiveJournal [11]	Undirected	No	4.0M	35M
Twitter [12]	Directed	No	41.7M	1.5B

Personalized PageRank (PPR) vector [9] as the source vector, and the degree vector as the global vector. Here, the PPR vector quantifies the importance of each node with respect to user's interest nodes [9], and the degree vector quantifies the global importance of each node by the number of neighbors.

4.2 A case study using the movie rating dataset

We observe the high-ranking nodes in the recommendation vectors while changing the cosine similarities $\in [-0.5, 0, 0.5]$. We use MovieLens dataset [10], which contains 34 million ratings from about 330,000 users on about 87,000 movies. A graph is generated by treating user u and movie m as nodes and inserting an undirected edge (u, m) weighted by the rating. The personalized recommendation is achieved when the PPR seed node is set to the user's interest movie.

Table 2 shows the top 5 nodes of the recommendation vectors with respect to the "Avengers: Infinity War Part I". In Table 2, the movies in bold are based on American comics, which are directly related to the seed node. In Table 2, movies with high degrees are more likely to be ranked higher as the cosine similarity increases. In particular, in Table 2b, the influence of the degree is mitigated because the cosine similarity is set to 0. In this result, the top two movies are "Thor: Ragnarok" and "Avengers: Infinity War Part II", whose degree ranks are 468 and 535, respectively. They are produced by Marvel Studios, the same as the seed node. Moreover, Table 2a shows the result when the high-degree nodes are suppressed. High-degree movies such as "The Matrix" and "The Shawshank Redemption," which are highly ranked in Table 2c, are removed. Thus, the negative cosine similarity is effective in avoiding popular movies.

4.3 Comparison of the adjustable range of correlation

We evaluate existing correlation adjustment methods in terms of the adjustable range of the cosine similarity. Note that we exclude CoCoA from this evaluation, since it clearly achieves any cosine similarity. The compared methods were either in-processing or post-processing types, where we can change the correlation adjustment parameters for each query. As in-processing methods, we chose "α-based PPR [6]" and "Degree-based Weighting [4, 5]". As post-processing methods, we chose "Popularity Compensation[7]" and "Value Aware Ranking[8]". For each method, we measured the cosine similarities while changing the correlation adjustment parameters and recorded the maximum and minimum values.

Table 2: The top 5 nodes of the recommendation vectors.

(a) Cosine similarity:-0.5 ($c = 1.1528$)

Rank	Title	Degree Rank
1	**Thor: Ragnarok**	468
2	**Avengers: Infinity War Part II**	535
3	**The Dark Knight**	22
4	Inception	23
5	**Deadpool 2**	740

(b) Cosine similarity:0 ($c = 0.89324$)

Rank	Title	Degree Rank
1	**Thor: Ragnarok**	468
2	**Avengers: Infinity War Part II**	535
3	**The Dark Knight**	22
4	Inception	23
5	The Matrix	4

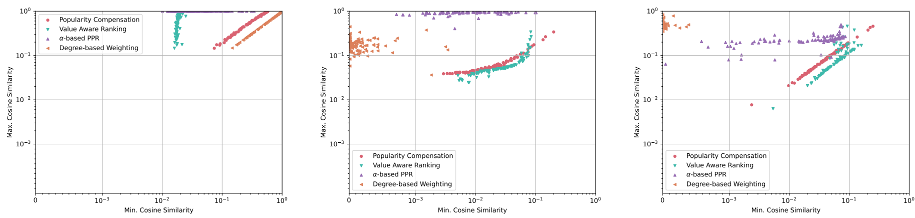
(c) Cosine similarity:0.5 ($c = 0.63367$)

Rank	Title	Degree Rank
1	The Matrix	4
2	The Shawshank Redemption	1
3	The Dark Knight	22
4	Inception	23
5	**Thor: Ragnarok**	468

Fig. 2 is a scatter plot of the maximum and minimum cosine similarities for each method. Each point represents the result for one seed node. Note that only positive regions are displayed, as the cosine similarity did not take negative values. According to Fig. 2, the results based on the cosine similarities are highly affected by the datasets and the seed nodes. For example, although "Degree-based Weighting" achieved cosine similarities below 0.001 in most settings, there were relatively many seed nodes with maximum values around 0.1. In the existing post-processing methods, the ranges were small. As a result, we confirmed the superiority of CoCoA, which achieves arbitrary cosine similarity regardless of the datasets and source vectors. Furthermore, CoCoA was the only method that achieved negative cosine similarities in our settings. Note that as mentioned in Sect. 4.2, there were situations where the negative cosine similarity was required.

5 Conclusion

Existing personalized recommendation methods cannot flexibly adjust the correlation between a recommendation vector and a global feature vector, such as a popularity vector. This paper proposes a post-processing method, CoCoA, to generate the recommendation vector whose cosine similarity to the global vector



(a) MovieLens **(b)** Live Journal **(c)** Twitter

Fig. 2: Comparison of the adjustable range of correlation.

is any input value $\in (-1, 1)$. In particular, CoCoA determines the recommendation vector with the linear combination of the source and global vectors. The cosine similarity is ensured by determining the coefficient of the linear combination based on mathematical considerations. A case study using the movie rating dataset showed changing the cosine similarity altered the presence of popular movies in the recommendation list. Furthermore, CoCoA was the only method that achieved negative cosine similarities, where the globally important nodes are suppressed. In future work, we will clarify the computational cost of CoCoA.

References

1. Park, S., Lee, W., Choe, B., Lee, S.G.: A Survey on Personalized PageRank Computation Algorithms. IEEE Access 7, 163049–163062 (2019)
2. Chen, J., et al.: Bias and Debias in Recommender System: A Survey and Future Directions. ACM TOIS 41(3), 1–39 (2023)
3. Abdollahpouri, H., Burke, R., Mobasher, B.: Controlling popularity bias in learning-to-rank recommendation. In: RecSys. pp. 42–46. Como Italy (2017)
4. Fujii, K., et al.: A Flexible Weighting Framework for Converting Relational Database to Hypergraphs. In: ICOIN. pp. 607–612. Chiang Mai, Thailand (2025)
5. Kloumann, I.M., Kleinberg, J.M.: Community membership identification from small seed sets. In: SIGKDD. pp. 1366–1375. New York, NY, USA (2014)
6. Yamashita, T., Kaneko, K.: Balancing global importance and source proximity for personalized recommendations using random walk length. In: Intell. Data Analysis. pp. 141–153. Konstanz, Germany (2025)
7. Zhu, Z., et al.: Popularity-opportunity bias in collaborative filtering. In: ACM WSDM. pp. 85–93. Israel and Online (2021)
8. Abdollahpouri, H., Burke, R., Mobasher, B.: Popularity-aware item weighting for long-tail recommendation (2016), https://arxiv.org/abs/1802.05382
9. Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank citation ranking: Bringing order to the web. Tech. rep., Stanford InfoLab (1998)
10. Harper, F.M., Konstan, J.A.: The MovieLens datasets: History and context. ACM TiiS 5(4), 1–19 (2015)
11. Leskovec, J., Krevl, A.: Stanford large network dataset collection, http://snap.stanford.edu/data
12. Kunegis, J.: KONECT: The Koblenz network collection. In: ACM WWW. pp. 1343–1350. Rio de Janeiro, Brazil (2013)