DualDiv: Diversifying Items and Explanation Styles in Explainable Hybrid Recommendation

Kosetsu Tsukuda and Masataka Goto
National Institute of Advanced Industrial Science and Technology (AIST), Japan
{k.tsukuda, m.goto}@aist.go.jp

ABSTRACT

In recommender systems, item diversification and explainable recommendations improve users' satisfaction. Unlike traditional explainable recommendations that display a single explanation for each item, explainable hybrid recommendations display multiple explanations for each item and are, therefore, more beneficial for users. When multiple explanations are displayed, one problem is that similar sets of explanation styles (ESs) such as user-based, item-based, and popularity-based may be displayed for similar items. Although item diversification has been studied well, the question of how to diversify the ESs remains underexplored. In this paper, we propose a method for diversifying ESs and a framework, called DualDiv, that recommends items by diversifying both the items and the ESs. Our experimental results show that DualDiv can increase the diversity of the items and the ESs without largely reducing the recommendation accuracy.

CCS CONCEPTS

• **Information systems Recommender systems**.

KEYWORDS

recommender systems, explainable recommendations, diversity

ACM Reference Format:

Kosetsu Tsukuda and Masataka Goto. 2019. DualDiv: Diversifying Items and Explanation Styles in Explainable Hybrid Recommendation. In *Thirteenth ACM Conference on Recommender Systems (RecSys '19), September 16–20, 2019, Copenhagen, Denmark.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3298689.3347063

1 INTRODUCTION

Most studies dealing with recommender systems (RSs) have traditionally put priority on recommendation accuracy [1] where users' preferences toward items are often measured based on their histories of item consumption [4]. However, it has been recognized that the diversity of items in the recommended list is also an important factor for increasing users' satisfaction with RSs because diversified recommended items allow users to browse a wider range of item types [9, 18]. Recommendation diversification aims to generate a ranked list of items in which items are dissimilar to each other but nonetheless relevant to the user's preference [3, 14, 18].

To improve the persuasiveness and user satisfaction of RSs, it is also useful to provide explanations so that users can understand why each item is recommended [13]; such approach is called explainable recommendation (ER) [16]. In ERs, a concrete explanation is generated from the *explanation style (ES)*. An ES can be user-based (*e.g.*, "this item is recommended because users similar to you like it."), popularity-based (*e.g.*, "this item is recommended because it is popular among users."), *etc.* Although most studies related to ERs had focused on a single ES, where a single explanation is displayed for each item [6, 11, 12, 15], it has been reported that users prefer multiple ESs, where many explanations are displayed for each recommended item [8]. A recommendation that can deal with multiple ESs is called *explainable hybrid recommendation* [8].

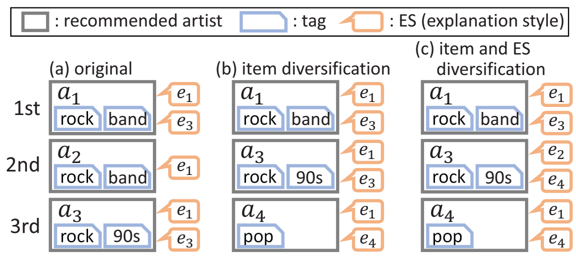

Figure 1: DualDiv framework concept.

Because of such usefulness of item diversification and ERs, multiple explanations in item-diversified recommendations could be beneficial to users. However, there are some problems associated with explanations. A study by Kouki *et al.* [8] has reported that subjective persuasiveness of each ES is different from one user to another. To cope with such difference, displaying all possible explanations generated from all ESs, say seven ESs, is not appropriate: it has been reported that giving too many explanations leads to the user's information overload, so displaying three or four explanations is usually sufficient. However, it is difficult for a RS to know in advance which ESs are preferred by each user. Moreover, even when preferred ESs are known for a specific user, it may be subject to change depending on the domain (*e.g.*, songs and books) and context (*e.g.*, time and user's intent). Therefore, it may not be ideal to fix three or four ESs for all items.

In light of the above, we propose a new framework *DualDiv* for diversifying both the items and the ESs. We explain intuitively DualDiv by using toy examples in Fig. 1 which shows three results of artist recommendations to a user. List (a) represents the original recommendation results where, for each recommended artist, tags attached to an artist as well as ESs are displayed. In list (a), because the user likes rock music, top three artists are rock-related. To make the recommendation results more attractive to the user, the list is diversified in terms of tags, and list (b) is generated. Now, a pop artist a_4 is recommended instead of a_2. However, for the two rock artists a_1 and a_3, the same ESs e_1 (*e.g.*, user-based ES) and e_3 (*e.g.*, social-based ES) are displayed. If the user does not think these ESs are persuasive, she may select neither of the recommended

RecSys '19, September 16–20, 2019, Copenhagen, Denmark

ACM ISBN 978-1-4503-6243-6/19/09...$15.00
https://doi.org/10.1145/3298689.3347063

artists. To solve this problem, DualDiv also diversifies the ESs as shown in list (c): the ESs e_2 (*e.g.*, popularity-based ES) and e_4 (*e.g.*, item-based ES) are displayed for a_3 because e_1 and e_3 are displayed for a_1. Note that in list (c), only the ESs are diversified, while the order of the artists is the same as in list (b). If the user thinks e_2 is persuasive, she would select a_3 and listens to a_3's songs. Thus, DualDiv aims to increase the probability of a user accepting at least one recommended item by diversifying both the items and the ESs.

Our main contributions are as follows. (1) We propose a framework DualDiv that recommends items by diversifying both the items and the ESs. To the best of our knowledge, this is the first study to propose the concept of ES diversification. DualDiv is a general framework and can be applied to any domain of item recommendations. (2) We propose a method for diversifying ESs that avoids displaying the same ESs for similar items as much as possible. Our method can also diversify ESs among all recommended items regardless of similarities between the items, by changing the parameter. (3) Our experiments in the context of artist recommendations show that DualDiv can diversify artists and ESs without largely reducing recommendation accuracy in terms of Recall.

2 DUALDIV

Given a target user, DualDiv generates a list of items by diversifying both the items and the ESs as follows. We first generate recommendations with personalized explanations for each item based on the method of Kouki *et al.* [8] (Section 2.2). We then generate an item-diversified list by using the method of Dou *et al.* [5] (Section 2.3). Because ESs cannot be diversified just by combining these two methods, in Section 2.4, we present DualDiv core method for diversifying the ESs. In this paper, we develop DualDiv in the domain of artist recommendations by using the data of Last.fm.

2.1 Notation

Let U and A denote the sets of users and artists, respectively. For each artist $a \in A$, a set of tags attached to a is defined as T_a, and each tag $t \in T_a$ has a relevance score with respect to a that is represented by $trel(a, t)$. For $u \in U$, $A_u \subset A$ represents a set of artists preferred by u, and $F_u \subset U$ represents a set of u's friends on Last.fm. Moreover, let $A_{pop} \subset A$ denote a set of top 1,000 popular artists on Last.fm obtained by Last.fm API. Given the target user u, the set of candidate artists for a recommendation is defined as $C = \bigcup_{u' \in F_u} A_{u'} \cup A_{pop} \setminus A_u$ [8]. Therefore, our goal is to generate a personalized ranked list of artists from C.

2.2 Explainable Hybrid Recommendation

Our first step is to generate recommendations with personalized explanations. To this end, we adopt a hybrid recommender system, called HyPER [7]. HyPER makes use of probabilistic soft logic (PSL) [2] that enables HyPER to develop recommendation models through a set of template rules in first-order logic syntax. HyPER automatically learns to balance the different input signals and computes the probability that the user accepts each item.

To compute the probability that user u listens to the recommended artist $a \in C$, we use the following eight rules [8].

Rule 1: $\textsc{SimUsers}_{CF}(u, u') \wedge \textsc{Prefers}(u', a) \Rightarrow \textsc{Listens}(u, a)$
Rule 2: $\textsc{SimArtists}_{CF}(a, a') \wedge \textsc{Prefers}(u, a') \Rightarrow \textsc{Listens}(u, a)$
Rule 3: $\textsc{SimArtists}_{content}(a, a') \wedge \textsc{Prefers}(u, a') \Rightarrow \textsc{Listens}(u, a)$
Rule 4: $\textsc{SimArtists}_{last.fm}(a, a') \wedge \textsc{Prefers}(u, a') \Rightarrow \textsc{Listens}(u, a)$

Table 1: Example explanations generated from ESs: artist "The Beatles" is recommended to a user.

ES	We recommend *The Beatles* because:
e_1	Last.fm users *Anna*, *Tom*, and *John* with whom you share similar music tastes, listens to *The Beatles*.
e_2	People who listen to your preferred artists *Bob Dylan* and *Johnny Cash* also listen to *The Beatles*.
e_3	*The Beatles* has similar tags with your preferred artist *The Kinks*
e_4	According to Last.fm, artist *The Beatles* is similar to your preferred artist *Queen*.
e_5	Artist *The Beatles* is tagged with *60s* and *britpop* that are in your profile.
e_6	Your friend *Lisa* likes artist *The Beatles*.
e_7	Artist *The Beatles* is very popular in the Last.fm database with 3.67 M listeners and 516.17 M playcounts.

Rule 5: $\textsc{Tag}(a, t) \wedge \textsc{Tag}(a', t) \wedge \textsc{Prefers}(u, a') \Rightarrow \textsc{Listens}(u, a)$
Rule 6: $\textsc{SimFriends}(u, u') \wedge \textsc{Prefers}(u', a) \Rightarrow \textsc{Listens}(u, a)$
Rule 7: $\textsc{PopularArtist}(a) \Rightarrow \textsc{Listens}(u, a)$
Rule 8: $\neg\textsc{Listens}(u, a)$

In all rules, the atom $\textsc{Listens}(u, a)$ represents the inferred probability that user u will listen to a, while the remaining atoms are binaries. Rule 1 captures the intuition that similar users like similar artists. Similarity between two users is computed based on the set of common artists they prefer. $\textsc{SimUsers}_{CF}(u, u')$ is 1 iff u' is included in the top 20 similar users of u. Rule 2, 3, and 4 capture the intuition that a user listens to similar artists. In Rule 2, similarity between two artists is computed by the set of common users who prefer both artists, while in Rule 3, similarity is computed according to the set of common tags attached to the artists. Similarity in Rule 4 is obtained through Last.fm API. In all cases, artists in the top 20 similar artists of a have values of 1 for $\textsc{SimArtists}$. Rule 5 states that a user listens to artists with the same tag. Rule 6 captures the intuition that friends share similar tastes in music and listen to the same artists. Rule 7 implies that a user tends to listen to popular artists. Finally, Rule 8 models a general belief that a user will not listen to an artist. Due to space limitation, we do not give the details of the similarity computation methods. We refer readers to Kouki *et al.* [8] for more details.

After HyPER computes $\textsc{Listens}(u, a)$ for all artists in C, we generate a ranked list of the artists by sorting them in descending order of $\textsc{Listens}(u, a)$. For each artist, we also create personalized explanations. In HyPER, each rule has an ES and explanations are generated from the ESs. Example explanations for seven ESs are shown in Table 1, where ES e_i corresponds to Rule i. Kouki *et al.* [8] have reported that the order of average subjective persuasiveness of the seven ESs is $e_4 \to e_2 \to e_3 \to e_7 \to e_5 \to e_6 \to e_1$. They have also reported that displaying three or four explanations for each artist is sufficient. Following this insight, in this paper, we display three explanations for each recommended artist. Given target user u, let R^{exp} and E_a^c denote the ranked list of all artists in C and the sets of all ESs created for $a \in R^{exp}$, respectively.

2.3 Item Diversification

Having obtained R^{exp}, our next step is to diversify the artists in R^{exp}. In this paper, we generate the artist-diversified recommendation list based on the diversification method proposed by Dou *et al.* [5], which is a general form of the xQuAD framework [10].

Their method [5] employs a greedy algorithm which iteratively selects items and generates a diversified ranking list. Let S_n denote the top n artists selected so far. The $n+1$th artist is given by:

$$a_{n+1} = \underset{a \in R^{exp} \setminus S_n}{\arg\max} \left[\alpha \cdot \textsc{Listens}(u, a) + (1-\alpha) \cdot \Phi(a, S_n, T)\right],$$

where $\Phi(a, S_n, T)$ computes the diversity of a and $T = \bigcup_{a \in C} T_a$; specifically, $\Phi(a, S_n, T)$ represents a tag richness score of a given the set S_n. α is a parameter that controls the tradeoff between relevance (*i.e.*, $\textsc{Listens}(u, a)$) and diversity. The term $\Phi(a, S_n, T)$ is decomposed as $\Phi(a, S_n, T) = \sum_{t \in T} w_t \cdot ntrel(a, t) \cdot \phi(t, S_n)$, where w_t is the importance of tag t for u. Following Santos *et al.* [10], we consider all tags as being equally important: $w_t = \frac{1}{|T|}$. In turn, $ntrel(a, t)$ represents the normalized relevance of artist a with respect to tag t and is given by $ntrel(a, t) = \frac{trel(a, t)}{\sum_{t' \in T_a} trel(a, t')}$. Finally, $\phi(t, S_n)$ is the discounted importance of tag t given S_n, where $\phi(t, S_n) = 1$ if $n = 0$ and $\phi(t, S_n) = \prod_{a' \in S_n}[1 - ntrel(a', t)]$ otherwise. Based on the above iterative process, we re-rank all artists in R^{exp}; let R^{div} denote the ranked list of diversified artists.

2.4 Explanation Style Diversification

Having obtained R^{div}, our final step is to diversify ESs for each artist. Note that in this step, the order of artists in R^{div} does not change. One simple approach to select three ESs for each artist is by selecting them according to the order of average subjective persuasiveness of ESs mentioned in Section 2.2. However, in this approach, similar sets of ESs may be often displayed for recommended artists. To solve this problem, we propose a method that also employs a greedy algorithm and generates a diversified ranked list of ESs for each artist. Our method is based on the following criterion: similar sets of ESs should not be displayed for similar artists. Our method is applied for each artist in top-to-bottom order of R^{div}. Given $n+1$th ranked artist $a_{n+1} \in R^{div}$, our goal is to select three ESs, which are denoted by $E^s_{a_{n+1}}$, from $E^c_{a_{n+1}}$ created in Section 2.2. Note that when we create $E^s_{a_{n+1}}$, $E^s_{a_i}$ for a_i ($1 \le i \le n$) has already been created. We create $E^s_{a_{n+1}}$ based on (1) if $e \in E^c_{a_{n+1}}$ is included in $E^s_{a_i}$ and (2) if a_{n+1} is similar to a_i.

Fig. 2 shows an example process of selecting mth ($1 \le m \le 3$) ES for a_3 (*i.e.*, $n = 2$), where a_3 is similar to a_1 and dissimilar to a_2. When $E^c_{a_3}$ consists of five ESs, e_2 is selected first because it is the only ES not included in either of $E^s_{a_1}$ and $E^s_{a_2}$. Then e_6 is selected because it is included in only $E^s_{a_2}$ of dissimilar artist a_2. Finally, all of the remaining ESs (*i.e.*, e_1, e_4, and e_7) are included in $E^s_{a_1}$ of similar artist a_1; but e_4 and e_7 are also included in $E^s_{a_2}$. Therefore, to increase the overall diversity of ESs, e_1 is selected.

Formally, the mth ES $e \in E^c_{a_{n+1}}$ is given by:

$$e = \underset{e' \in E^c_{a_{n+1}} \setminus E^s_{a_{n+1}}}{\arg\max} \left[\Psi(e', a_{n+1})\right]. \quad (1)$$

$\Psi(e', a_{n+1})$ represents the diversity of e' and is decomposed by:

$$\Psi(e', a_{n+1}) = \sum_{1 \le i \le n} \left(1 - \delta(e' \in E^s_{a_i})\right) \cdot sim(a_{n+1}, a_i),$$

where $\delta(x)$ is 1 when x is true and 0 otherwise. $sim(a_{n+1}, a_i)$ represents similarity between a_{n+1} and a_i, and is given by $sim(a_{n+1}, a_i) = \|T_{a_{n+1}} \cap T_{a_i}\| / \|T_{a_{n+1}} \cup T_{a_i}\|$. Although Eq. 1 considers duplications of ESs between similar artists, those between dissimilar artists are not taken into account. Therefore, if all artists in R^{div} are dissimilar, almost the same sets of ESs may be displayed for all artists. To solve this problem, we consider the ES duplications between two arbitrary artists and expand Eq. 1 as follow:

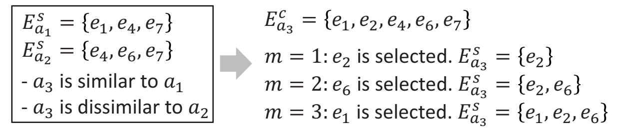

Figure 2: An example process to select three ESs $E^s_{a_3}$ for a_3.

$$e = \underset{e' \in E^c_{a_{n+1}} \setminus E^s_{a_{n+1}}}{\arg\max} \left[\beta \cdot \Psi(e', a_{n+1}) + (1-\beta) \cdot \Omega(e', a_{n+1})\right], \quad (2)$$

where $\Omega(e', a_{n+1}) = \sum_{1 \le i \le n} \left(1 - \delta(e' \in E^s_{a_i})\right)$ and β is a parameter. If e' is already selected for many high-ranking artists in R^{div}, the importance of e' decreases regardless of similarity. When more than one ESs have the same score in Eq. 2, the ES having the highest order of average subjective persuasiveness is selected [1].

3 EXPERIMENTS

This section reports on the evaluation of our proposed framework DualDiv using the Last.fm dataset.

3.1 Evaluation Setup

[Dataset] As target users, we randomly selected 300 users of American nationality from Last.fm dataset [17]. By using the Last.fm API, we collected the following data: top 20 friends and top 25 preferred artists for each target user, top 20 preferred artists for each friend, and top 20 tags for each artist. In summary, our dataset consisted of 5,789 users, 29,435 artists, and 31,969 tags. For each target user, five preferred artists were randomly sampled as test data and added to C (described in Section 2.1), and the remaining 20 artists were used as training data.

[Comparisons] In addition to our proposed framework DualDiv that diversifies both the items and the ESs, we use the following three comparisons.

- **NonDiv**: this method diversifies neither artists nor ESs and corresponds to Fig. 1 (a). The recommendation list R^{exp} generated in Section 2.2 is used for this method. Regarding ESs, we select three ESs from E^c_a in the order of average subjective persuasiveness.
- **ItemDiv**: this method, which corresponds to Fig. 1 (b), diversifies only artists and uses the recommendation list R^{div} generated in Section 2.3. The ESs for each artist are selected in the same way as in NonDiv.
- **RandDiv**: this method also uses R^{div} for artist ranking but randomly selects three ESs for each artist. This is also a kind of ES diversification because, unlike in NonDiv and ItemDiv, there are no specific priorities among ESs.

For ItemDiv, RnadDiv, and DualDiv, the value of α in Section 2.3 is set to 0.75, while β in Section 2.4 is set to 0.5 for DualDiv.

[Evaluation Metrics] We evaluate the recommendation accuracy in terms of Recall, where Recall@k is defined by the ratio of artists in the user's test data that are correctly included in R^{div}_k, which is the top k artists in R^{div}. Regarding artist diversity, we use AILD (artist ILD, where ILD stands for intra-list diversity [18]),

[1]When $|E^c_{a_{n+1}}| < 3$, the number of selected ESs is equal to $|E^c_{a_{n+1}}|$.

Table 4: Example results for a user of NonDiv, ItemDiv, and DualDiv. Note: see Table 1 for example explanations generated from each ES. (Abbreviations: The Human League (THL), Orchestral Manoeuvres in the Dark (OMD), Backstreet Boys (BB))

	NonDiv			ItemDiv			DualDiv		
	Recall@5=0.4, AILD@5=0.558 EILD@5=0.350, WEILD@5=0.137			Recall@5=0.2, AILD@5=0.701 EILD@5=0.400, WEILD@5=0.0768			Recall@5=0.2, AILD@5=0.701 EILD@5=0.700, WEILD@5=0.207		
Rank	Artist	Tags	ESs (E_a^s)	Artist	Tags	ESs (E_a^s)	Artist	Tags	ESs (E_a^s)
1	THL	electronic, 80s	e_2, e_3, e_4	THL	electronic, 80s	e_2, e_3, e_4	THL	electronic, 80s	e_2, e_3, e_4
2	Anything Box	electronic, 80s	e_3, e_4, e_5	Anything Box	electronic, 80s	e_3, e_4, e_5	Anything Box	electronic, 80s	e_3, e_4, e_5
3	OMD	electronic, 80s	e_2, e_3, e_4	OMD	electronic, 80s	e_2, e_3, e_4	OMD	electronic, 80s	e_2, e_5, e_6
4	Dead or Alive	electronic, 80s	e_2, e_3, e_4	BB	pop, 90s	e_3, e_4, e_7	BB	pop, 90s	e_4, e_5, e_7
5	Book of Love	electronic, 80s	e_2, e_4, e_5	*NSYNC	pop, 90s	e_3, e_4, e_7	*NSYNC	pop, 90s	e_3, e_5, e_7

Table 2: Comparison results.

	NonDiv	ItemDiv	RandDiv	DualDiv
Recall@5	**0.115**		0.111	
Recall@10	**0.190**		0.173	
Recall@20	**0.311**		0.275	
AILD@5	0.680		**0.753**	
AILD@10	0.715		**0.776**	
AILD@20	0.745		**0.804**	
EILD@5	0.391	0.433	0.644	**0.731**
EILD@10	0.454	0.480	0.632	**0.699**
EILD@20	0.517	0.537	0.628	**0.686**
WEILD@5	0.117	0.0985	0.159	**0.184**
WEILD@10	0.122	0.100	0.141	**0.157**
WEILD@20	0.128	0.0993	0.122	**0.135**

which is computed as follows:

$$\text{AILD@}k = \frac{\sum_{a_i \in R_k^{div}} \sum_{a_j \in R_k^{div} \setminus \{a_i\}} dist(a_i, a_j)}{k(k-1)}, \quad (3)$$

where $dist(a_i, a_j)$ represents the distance between two artists and is computed by $dist(a_i, a_j) = 1 - sim(a_i, a_j)$ (see Section 2.4 about the definition of $sim(a_i, a_j)$). The high value of AILD indicates that artists in the recommendation list are diversified well. Similarly, we use EILD (explanation style ILD) to evaluate the ES diversity. EILD@k is also computed as in Eq. 3, but $dist(a_i, a_j)$ is given by $dist(a_i, a_j) = 1 - \|E_{a_i}^s \cap E_{a_j}^s\| / \|E_{a_i}^s \cup E_{a_j}^s\|$. EILD puts equal weights between arbitrary pairs of artists. In contrast, we also evaluate the ES diversity by using WEILD (weighted EILD), which is given by:

$$\text{WEILD@}k = \frac{\sum_{a_i \in R_k^{div}} \sum_{a_j \in R_k^{div} \setminus \{a_i\}} sim(a_i, a_j) \cdot dist(a_i, a_j)}{k(k-1)}.$$

$dist(a_i, a_j)$ is the same as in EILD. By using WEILD, we can evaluate if dissimilar sets of ESs are selected between similar artists. Regarding NonDiv, R_k^{exp} is used instead of R_k^{div} for all metrics.

3.2 Results

Table 2 shows comparison results. Note that ItemDiv, RandDiv, and DualDiv have the same values of Recall and AILD because RandDiv and DualDiv use the same artist ranking as ItemDiv. We can see that these three methods can improve AILD without largely reducing Recall compared to NonDiv. In addition, in terms of EILD, DualDiv significantly outperforms the other three methods. Although RandDiv also outperforms NonDiv and ItemDiv, when there are many, say seven, candidate ESs for an artist (*i.e.*, $|E_a^c| = 7$), RandDiv may be able to randomly select ESs that are not overlapping with those of similar artists. However, when there are, for example, only four candidate ESs, it rarely happens that the appropriate three ESs are selected by chance. In contrast, DualDiv can greedily select the three most appropriate ESs one by one regardless of the number of candidate ESs. Hence, DualDiv shows better performance than RandDiv. In ItemDiv, because the artists are diversified, the scores of $sim(a_i, a_j)$ tend to be lower than in NonDiv; this leads to the lower WEILD than NonDiv. Nonetheless DualDiv, which uses the same artist ranking with ItemDiv, outperforms NonDiv. These results indicate that DualDiv can successfully select dissimilar sets of ESs, especially between similar artists.

Table 3: Effects of parameter β in DualDiv.

	$\beta = 0$	$\beta = 0.25$	$\beta = 0.5$	$\beta = 0.75$	$\beta = 1$
EILD@5	0.734	0.731	0.731	0.730	0.725
WEILD@5	0.179	0.184	0.184	0.184	0.184

Next, we evaluate the effects of parameter β in DualDiv. When β is small, DualDiv aims to reduce duplications of ESs among all artists. When β is large, it aims to reduce duplications especially between similar artists. In Table 3, as expected, EILD is high for small β while WEILD is high for large β. Therefore, we can tune DualDiv by changing the value of β depending on the type of ES diversification we want to obtain.

Finally, Table 4 shows top five recommendation results of NonDiv, ItemDiv, and DualDiv for a user, where only characteristic tags for each artist are listed. Due to space limitation, explanations generated from each ES are not shown. In terms of tags, all artists in NonDiv are related to "electronic" and "80s," while in DualDiv, two artists related to "pop" and "90s" are included. This enables the user to browse a wide range of artists. In terms of ESs, DualDiv shows six kinds of ESs, while NonDiv and ItemDiv shows four and five kinds of ESs, respectively. Therefore, even if this user puts the highest priority on e_6, she can find at least one artist explained based on that ES. In addition, unlike ItemDiv, DualDiv can select more diverse sets of ESs for similar artists (*i.e.*, five ESs for artists ranked at 1-3 and four ESs for artists ranked at 4-5). This leads to a higher value of WEILD and increases the probability with which the user accepts at least one of the recommended similar artists.

4 CONCLUSION

This paper proposed DualDiv, a framework for recommending items by diversifying both the items and the ESs. Our experimental results using Last.fm dataset showed that DualDiv was able to diversify the artists and ESs without largely reducing Recall. As future work, we plan to develop DualDiv in other domains such as movies and sightseeing spots, because DualDiv as a general framework can be applied to any domain of item recommendations.

[Acknowledgments] This work was supported by JST ACCEL Grant Number JPMJAC1602, Japan.

REFERENCES

[1] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.

[2] S. H. Bach, M. Broecheler, B. Huang, and L. Getoor. Hinge-loss markov random fields and probabilistic soft logic. *The Journal of Machine Learning Research*, 18(1):3846–3912, 2017.

[3] F. Belém, R. Santos, J. Almeida, and M. Gonçalves. Topic diversity in tag recommendation. In *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys '13, pages 141–148, 2013.

[4] R. Burke. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4):331–370, 2002.

[5] Z. Dou, S. Hu, K. Chen, R. Song, and J. R. Wen. Multi-dimensional search result diversification. In *Proceedings of the 4th ACM International Conference on Web Search and Data Mining*, WSDM '11, pages 475–484, 2011.

[6] J. L. Herlocker, J. A. Konstan, and J. Riedl. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, CSCW '00, pages 241–250, 2000.

[7] P. Kouki, S. Fakhraei, J. Foulds, M. Eirinaki, and L. Getoor. HyPER: A flexible and extensible probabilistic framework for hybrid recommender systems. In *Proceedings of the 9th ACM Conference on Recommender Systems*, RecSys '15, pages 99–106, 2015.

[8] P. Kouki, J. Schaffer, J. Pujara, J. O'Donovan, and L. Getoor. Personalized explanations for hybrid recommender systems. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, IUI '19, pages 379–390, 2019.

[9] P. Pu, L. Chen, and R. Hu. A user-centric evaluation framework for recommender systems. In *Proceedings of the 5th ACM Conference on Recommender Systems*, RecSys '11, pages 157–164, 2011.

[10] R. L. Santos, C. Macdonald, and I. Ounis. Exploiting query reformulations for web search result diversification. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 881–890, 2010.

[11] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web*, WWW '01, pages 285–295, 2001.

[12] A. Sharma and D. Cosley. Do social explanations work?: Studying and modeling the effects of social explanations in recommender systems. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13, pages 1133–1144, 2013.

[13] N. Tintarev. Explanations of recommendations. In *Proceedings of the 1st ACM Conference on Recommender Systems*, RecSys '07, pages 203–206, 2007.

[14] S. Vargas, L. Baltrunas, A. Karatzoglou, and P. Castells. Coverage, redundancy and size-awareness in genre diversity for recommender systems. In *Proceedings of the 8th ACM Conference on Recommender Systems*, RecSys '14, pages 209–216, 2014.

[15] J. Vig, S. Sen, and J. Riedl. Tagsplanations: Explaining recommendations using tags. In *Proceedings of the 14th International Conference on Intelligent User Interfaces*, IUI '09, pages 47–56, 2009.

[16] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, and S. Ma. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '14, pages 83–92, 2014.

[17] Y. Zhang, J. Tang, Z. Yang, J. Pei, and P. S. Yu. Cosnet: Connecting heterogeneous social networks with local and global consistency. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, pages 1485–1494, 2015.

[18] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. In *Proceedings of the 14th International Conference on World Wide Web*, WWW '05, pages 22–32, 2005.