HOPE (High Order Profile Expansion) for the new user problem on recommender systems

Diego Fernández
CITIC Research Center, University of A Coruña
A Coruña, Spain
diego.fernandez@udc.es

Vreixo Formoso
CITIC Research Center, University of A Coruña
A Coruña, Spain
vformoso@udc.es

Fidel Cacheda
CITIC Research Center, University of A Coruña
A Coruña, Spain
fidel.cacheda@udc.es

Victor Carneiro
CITIC Research Center, University of A Coruña
A Coruña, Spain
victor.carneiro@udc.es

ABSTRACT

The new user problem is well-known in recommender systems and refers to the difficulty to provide accurate recommendations to a user who has just arrived to the system or who has provided very few ratings. Profile Expansion techniques intend to increase the size of the user profile by obtaining information about user preferences in distinct ways and have already been used to expand the information provided to recommendation algorithms and improve results. In this work, discussed in more detail at [1], we present the High Order Profile Expansion techniques, which combine in different ways the Profile Expansion methods. The results show 110% improvement in precision over the algorithm without Profile Expansion, and 10% improvement over Profile Expansion techniques.

CCS CONCEPTS

• **Information systems** → **Information retrieval**; **Recommender systems**.

KEYWORDS

collaborative filtering, new user problem, profile expansion

1 INTRODUCTION

In this work we explore the Profile Expansion (PE) techniques [2], which were proposed to deal with the new user problem without involving the user. We propose a new set of methods called High Order Profile Expansion (HOPE) techniques, since instead of just using one PE technique, two or more techniques can be combined so that new and diverse ratings are added to user profiles with few ratings, which is discussed in more detail at [1]. In this work we show how the results obtained by PE techniques are improved by combining different PE techniques.

2 HIGH ORDER PROFILE EXPANSION

The idea of HOPE is to combine profile expansion techniques to take advantage of their benefits and minimize their errors. Using more than one algorithm to expand the user profile, the expansion will be more heterogeneous and, at the same time, still related to the initial user profile. The aim is also to expand the user profile in order to improve recommendations when facing the new user problem.

In our experiments, we will focus on Item Global (IG) and User Local (UL) PE alternatives because they represent the best PE techniques [2]. IG expand the user profile by searching in the system for the most similar items to those rated by the user, while UL only use the neighborhood computed for the main algorithm as a source of information to find new items. For the UL technique, different approaches are considered: most rated (MR) selects the most rated items in the neighborhood, local item neighbors (LN) uses the most similar items to those in the user profile and user-local clustering (LC) attempts to find similarities between items but taking into account only the ratings given by the local user neighborhood. These techniques will be combined according to two different proposals: serial and parallel.

In a serial process PE techniques are used one after the other. The kNN algorithm will use the doubly expanded profile to compute the recommendation list.

The parallel approach attempts to minimize errors in the item selection and rating prediction, being the initial user profile the input for all the expansion algorithms used. Since every expansion algorithm will obtain a new profile and the kNN technique needs a unique profile as input, it is necessary to combine the different profiles. Regarding the item selection, two alternatives have been considered: union and intersection.

As far as rating selection is concerned, when an item appears in more than one expanded profile, the final rating will be the mean of the ratings.

3 EXPERIMENTS

The experiments have been performed in a reduced version of the Netflix dataset, consisting of 8,362 items and 478,458 users, who have done 48,715,350 ratings. Regarding the methodology, we have randomly selected 1000 users for evaluation purposes. The new user problem has been simulated using a Given-N strategy.We focus on $N = 2$, since it is when the information is more scarce and HOPE techniques can be more useful. Moreover, we study the evolution of the algorithms according to N. From the remaining users, we have randomly chosen 90% of ratings for the training subset.

Regarding profile expansion size variations, we have used different l values (2, 5, 10 and 15) to check how the l combinations affect the results. Moreover, it is also interesting to know how the distinct algorithms evolve according to N, that is, according to the amount of information available. As baselines, the MR algorithm and the algorithm without PE are considered.

3.1 Serial

In serial combinations, two Profile Expansion techniques are used, one after the other. Best results were obtained with $l = 2$ and Table 1 shows P@5 and MAP values.

Table 1: P@5 and MAP for High Order Profile Expansion serial combinations, with $l = 2$. Significant improvements over baselines are highlighted.

Algorithms 1	Algorithms 2	$N = 2$ P@5	$N = 2$ MAP	$N = 10$ P@5	$N = 10$ MAP
LC	LC	0.153	0.026	0.146	0.034
	LN	0.147	0.026	0.148	0.036
	MR	0.156	0.025	0.153	0.036
	IG	0.149	**0.032**	0.147	**0.040**
	-	0.140	0.030	0.140	0.036
LN	LC	0.154	0.027	0.151	0.035
	LN	0.142	0.024	0.140	0.037
	MR	0.158	0.025	0.151	0.033
	IG	0.156	**0.032**	0.127	**0.041**
	-	0.139	0.029	0.141	0.039
MR	LC	**0.163**	0.026	0.153	0.034
	LN	0.152	0.025	0.146	0.032
	MR	0.155	0.024	0.151	0.034
	IG	0.160	0.030	0.137	0.039
	-	0.147	0.029	0.139	0.035
IG	LC	0.139	0.030	0.144	**0.040**
	LN	0.137	0.031	0.128	**0.041**
	MR	0.145	0.029	0.143	**0.041**
	IG	0.133	**0.035**	0.129	**0.045**
	-	0.122	0.035	0.118	**0.045**
No PE		0.078	0.029	0.140	0.036

Regarding the results, the LN algorithm works better when it is used as the first algorithm. Moreover, the IG algorithm performs better when it is the second algorithm, significantly improving the MAP values with respect to the baselines. The MR and LC algorithms depend on which algorithm is combined with them. These results make sense, because when the information about a user is scarce, global information is more prone to errors. However, when working with local information, although there is less information available, it tends to be related to the user profile.

3.2 Parallel

Table 2 shows P@5 and MAP results for union and intersection variants. Note that intersection alternative may lead to profiles with few items in common and therefore, profile expansion sizes have been increased with respect to the serial alternative, so that it is more probable that two different techniques have items in common.

While $l = 5$ is enough for obtaining good results when only UL algorithms are considered, a bigger profile value is required for IG techniques ($l = 100$). From Table 2, we can observe that MAP values are quite good, but P@5 results for the intersection variation do not improve those obtained with the serial alternative.

Table 2: P@5 and MAP for High Order Profile Expansion parallel alternatives. Significant improvements over baselines are highlighted.

Combinations	Algorithms 1	Algorithms 2	$N = 2$ P@5	$N = 2$ MAP	$N = 10$ P@5	$N = 10$ MAP
Intersection	IG	LC	0.128	0.029	0.148	0.038
		LN	0.139	0.028	0.141	0.039
		MR	0.133	0.027	0.143	0.038
	LC	LN	0.146	0.029	0.141	0.038
		MR	0.143	0.029	0.152	0.036
	LN	MR	0.147	0.028	0.141	0.038
Union	IG	LC	0.142	0.031	0.140	**0.041**
		LN	0.134	0.030	0.125	**0.040**
		MR	0.144	0.030	0.148	**0.040**
	LC	LN	0.137	0.025	0.147	0.034
		MR	0.144	0.027	0.152	0.034
	LN	MR	0.138	0.024	0.148	0.032
Simple PE	LC	-	0.140	0.030	0.140	0.036
	LN	-	0.139	0.029	0.141	0.039
	MR	-	0.147	0.029	0.139	0.035
	IG	-	0.122	0.035	0.118	0.045
No PE	-	-	0.078	0.029	0.140	0.036

On the other hand, union variation puts together the expansions obtained with the different algorithms. We have used $l = 2$ as profile expansion size. As shown in Table 2 the results do not improve those obtained with the serial alternative either. In fact, no combination gets a P@5 value better than MR baseline algorithm.

So, despite the fact that with the parallel alternatives the second expansion algorithm is not affected by the errors committed by the first one, the serial results are better because the second algorithm is provided with more information.

4 CONCLUSIONS

In this work, we present the HOPE techniques, in particular, serial and parallel alternatives. The experiments have shown how the serial alternative performs better than the parallel one. The reason is that in the serial alternative an algorithm is supplied with the information obtained by another algorithm. That does not happen with the parallel alternative, where the algorithms use the same information. As future work, we plan to combine more than two algorithms and use some other merging techniques for the parallel method.

ACKNOWLEDGMENTS

This research was supported by the by the Xunta de Galicia (Centro singular de investigacion de Galicia accreditation ED431G/01 2016-2019) and the European Union (European Regional Development Fund - ERDF).

REFERENCES

[1] Diego Fernández, Vreixo Formoso, Fidel Cacheda, and Victor Carneiro. 2019. High Order Profile Expansion to tackle the new user problem on recommender systems. *PloS one* 14, 11 (2019).

[2] Vreixo Formoso, Diego Fernández, Fidel Cacheda, and Victor Carneiro. 2013. Using profile expansion techniques to alleviate the new user problem. *Inf. Process. Manage.* 49, 3 (May 2013), 659–672. https://doi.org/10.1016/j.ipm.2012.07.005