

# Collaborative Filtering with Semantic Neighbour Discovery

Bruno Veloso<sup>1,2</sup>, Benedita Malheiro<sup>2,3</sup>, and Juan C. Burguillo<sup>1</sup>

<sup>1</sup> EET/UVigo – School of Telecommunication Engineering, University of Vigo, Spain

<sup>2</sup> INESC TEC, Porto, Portugal

<sup>3</sup> ISEP/IPP – School of Engineering, Polytechnic Institute of Porto, Portugal

**Abstract.** Nearest neighbour collaborative filtering (NNCF) algorithms are commonly used in multimedia recommender systems to suggest media items based on the ratings of users with similar preferences. However, the prediction accuracy of NNCF algorithms is affected by the reduced number of items – the subset of items co-rated by both users – typically used to determine the similarity between pairs of users. In this paper, we propose a different approach, which substantially enhances the accuracy of the neighbour selection process – a user-based CF (UbCF) with semantic neighbour discovery (SND). Our neighbour discovery methodology, which assesses pairs of users by taking into account all the items rated at least by one of the users instead of just the set of co-rated items, semantically enriches this enlarged set of items using linked data and, finally, applies the Collinearity and Proximity Similarity metric (CPS), which combines the cosine similarity with Chebyshev distance dissimilarity metric. We tested the proposed SND against the Pearson Correlation neighbour discovery algorithm off-line, using the HetRec data set, and the results show a clear improvement in terms of accuracy and execution time for the predicted recommendations.

**Keywords:** User-based collaborative filtering, semantic neighbour discovery, semantic enrichment

## 1 Introduction

Nowadays viewers are unable to search, find and choose their preferred content in real time due to the size of the search space both in terms of the number of offers and of the diversity of sources. Consequently, media content personalisation poses a significant challenge to media content distributors, *i.e.*, how to choose the most appropriate media content for each viewer from a vast list of sources and candidates per source. This information overload problem is typically addressed using Information Filtering technologies. A possible approach is to use Collaborative Filtering (CF), which relies on information retrieval and data mining techniques to provide recommendations based on the ratings of users with similar preferences. CF has three variants – user-based, item-based and tag-based – and two alternative implementations – memory-based, *e.g.*,  $k$ -Nearest Neighbour

( $k$ -NN), and model-based, *e.g.*, Singular Value Decomposition (SVD) [3]. Our research is focussed on a specific CF memory-based technique called user-based  $k$ -NN. This technique is influenced by several factors: (*i*) the similarity metrics used to determine the nearest neighbours; (*ii*) the sparsity of the existing user ratings, which emerge from the fact that each user usually rates only a very small percentage of the represented items; and (*iii*) the set of items used to determine the similarity between two users, which is based solely on co-rated items, *i.e.*, items rated by both users, ignoring all items rated by just one of the users. These weaknesses can lead to the selection of unrelated neighbours, resulting in the inability to provide interesting recommendations. Current nearest neighbour CF assessment methodologies are unable to identify these inefficiencies.

In this paper, we propose a new user-based  $k$ -NN filtering approach, which performs a semantic neighbour discovery (SND) using all items rated at least by one of the users, instead of just the subset of co-rated items, to select the most related neighbours and, thus, improve the accuracy of the final recommendations. For each user, it collects semantic data regarding all rated items, *e.g.*, the set of writers, actors, countries or directors, builds multiple vectors and applies the Collinearity and Proximity Similarity (CPS) metric, a similarity metric proposed by [16] to select the best neighbours. CPS combines the well known cosine similarity with Chebyshev distance dissimilarity, which determines the greatest differences between two related feature vectors.

Experiments were conducted with the HetRec 2011 data set, which contains movies, user ratings and tags, to evaluate the quality and accuracy of the recommendations produced. The metrics used were the Global Mean Absolute Error (GMAE), the Global Root Mean Square Error (GRMSE), the Global F1-measure (GF1) as well as the execution time. The comparison between the SND and the standard Pearson Correlation UbCF recommendations show a clear improvement in terms of GF1, comparable results in terms of GMAE and execution times related with the number of LOD features used.

The main contributions of this paper are: (*i*) the development of a new user-based  $k$ -NN collaborative filter supported by semantic neighbour discovery, *i.e.*, which uses linked data to improve the selection of the nearest neighbours; and (*ii*) the application of the Collinearity and Proximity Similarity (CPS) metric to  $k$ -NN.

The rest of the paper is organized as follows. Section 2 presents an overview of collaborative recommendation with semantic enrichment. Section 3 describes the proposed collaborative filter. Section 4 describes the data set, presents the adopted evaluation metrics as well as the tests, and discusses the results. Finally, Section 5 provides the conclusions and presents future development lines.

## 2 Semantically Enriched Collaborative Recommendation

Recommendation systems are valuable support tools for users or applications to find items of interest whenever the search space is too large. They are frequently used in personalisation and data retrieval tasks, and rely on different

techniques, *e.g.*, information filtering and semantic enrichment. The information filtering approaches for recommendation systems are organised in Content-based Filters (CbF), Collaborative Filters (CF) and Hybrid Filters (HF). This paper is focussed on  $k$ -NN CF.

## 2.1 Collaborative Filtering

CF algorithms recommend items by making predictions based on the history of items rated by the user. There are two main approaches: (i) memory-based algorithms, which calculate the similarity or correlation among users or items; and (ii) model-based algorithms, which learn predictive models. Collaborative filtering depends on the availability of the individual viewer records of watched, rated and/or tagged items. CF promotes diversity when dealing with different types of content and creates an effect of controlled serendipity. However, these filters suffer from: (i) data sparsity and scalability issues; (ii) the first-rater problem, *i.e.*, an item cannot be recommended without being rated by a user; (iii) the grey sheep problem, *i.e.*, users who never rate items; and (iv) the cold-start problem, *i.e.*, new users have no rated items. Memory and model-based approaches make predictions by analysing users (user-based), items (item-based) or tags (tag-based). Multiple examples can be found in the literature regarding both approaches. Whereas [10], [12] and [13] are instances of user, item and tag memory-based approaches, [14] and [17] are instances of user and item model-based approaches, respectively. This paper presents a UbCF with semantic enrichment.

## 2.2 Semantic Enrichment

Semantic enrichment is supported by public interconnected knowledge bases – the Linked Open Data (LOD) repositories. Together they are the fabric of the Semantic Web [2], where each *datum* is described and interlinked through Resource Description Framework (RDF) triplets. The subject-predicate-object triplets include uniform resource identifiers (URI) for resource identification and look up, using open standards. They provide a valuable source of information that can improve conventional recommender systems, if properly exploited [11]. Linked data can be used to enrich the profiles of users and items with new features, allowing the identification of new relationships. In this paper, we enrich the profiles of the users, using complementary features extracted from the LOD, and determine a multiple feature neighbour similarity, using CPS, to choose the nearest neighbours.

Martin *et al.* propose the usage of semantic reasoning techniques to improve the neighbour discovery. It is based on the products to recommend rather than on user similarity, explores hierarchical similarity [8]. The authors propose a pre-filtering approach limiting the neighbour search subspace at the level of the product class hierarchy, followed by the assessment of the ratings and semantic features to identify common interests. Our approach is more general since it is applicable, not only, to the complete set of LOD features, but also to any subset

of LOD features. Moreover, we rely on semantically enriched data to choose neighbours at the term level rather than at the concept level.

Kaveh *et al.* propose a weighted  $k$ -NN neighbour selection, taking into account the semantic distances between neighbours. The selection mechanism adjusts the weight of these distances to enhance or decrease the importance of the best and worst categories, respectively [6]. While this approach determines the similarity between users based on the discrimination of the relative importance between the categories, we equally consider the selected LOD properties. When compared with our approach, this category discrimination may lead to overspecialisation.

The hybrid film recommender designed by Kushwaha *et al.* [7] first uses semantic data from DBpedia to minimise the sparsity problem, then applies the Levenshtein distance to map films to DBpedia RDF links and, finally, determines the cosine similarity to find the best neighbours to generate predictions. While this recommender calculates the similarity between neighbours based exclusively on the features of the top co-rated items (5-star items), our proposal contemplates the features of all items rated by the users under analysis, including both liked and disliked items.

The collaborative recommender system reported by Martín-Vicente *et al.* uses semantic reasoning to address the matrix sparsity problem and improve the neighbour selection. User similarity is calculated using the relations and the number of hops between concepts represented in the system’s ontology [9]. While this proposal determines similarity between users at the concept level, we choose the nearest neighbours by calculating the similarity between users at the (semantically enriched feature) term level. Consequently, our approach is fully automated, *i.e.*, does not require any previous semantic annotation.

This related work review shows that semantic based neighbour discovery is becoming increasingly popular. Our SND approach, when compared with these works, does not require prior concept annotation as the semantic concept level approaches, uses the LOD for to enrich all items rated by any of the users under comparison with complementary features and determines user similarity with the CPS metric.

### 3 UbCF with Semantic Neighbour Discovery

User-based collaborative filtering systems suggest interesting items by relying on like-minded people called neighbours [1]. The quality of these NNCF depends largely on the similarity metrics used and, in particular, on the data used to determine the set of nearest neighbours. Typically, users have rated quite a small number of items and are compared on the basis of an even smaller subset – the set of co-rated items – regardless of the adopted similarity metric. To overcome this problem, we, rather than comparing two users based solely on the set of co-rated items (the intersection set), use all items rated by at least one of the users (the union set). Although this idea was first proposed by [15], we adopt a diverse approach herein called Semantic Neighbour Discovery (SND). Our UbCF with

SND, which is represented in Figure 1, encompasses three stages: (i) semantic enrichment; (ii) neighbour discovery; and (iii) item prediction.

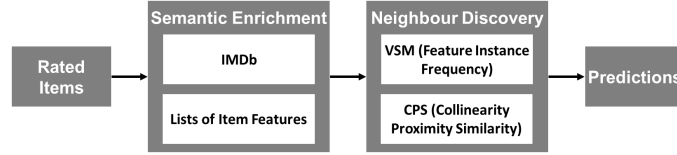


Fig. 1: Semantic UbCF architecture.

In the semantic enrichment stage the system collects, for each user and based on the full set of user rated items, multiple lists of semantic feature instances from the IMDb via the OMDb API<sup>1</sup>, *i.e.*, the lists of genres, actors, directors, writers, languages, countries, types and years of the user rated items. These lists are created beforehand off-line, using historical data and are updated on-line to include newly rated user items. Depending on the feature list, each element, representing a specific genre, actor, director, writer, type, year, language or country, holds the corresponding frequency count in the user rated items. During the neighbour discovery stage, the system, based on the semantic feature lists gathered, compares the active and current users to select the  $k$  users with the highest similarity to the active user. First, it creates a vector for each semantic feature and user with a dimension equal to the union of instances of the corresponding active and current user lists. Then, it calculates, for each semantic feature instance vector, the normalised Feature Instance Frequency (FIF) according to Equation 1:

$$FIF_{f,i,u} = \frac{freq_{f,i,u}}{\sum_i^n freq_{f,i,u}} \quad (1)$$

where  $freq_{f,i,u}$  is the frequency of the instance  $i$  of the semantic feature  $f$  in the items rated by the user  $u$  and  $n$  is the number of instances of the semantic feature  $f$  present in the items rated by the user  $u$ . The similarity between users is determined through the Collinearity and Proximity Similarity (CPS) using the Equation 2:

$$CPS_{a,c} = \beta \frac{\sum_{f=1}^m CS_{f,a,c}}{m} + (1 - \beta) \frac{\sum_{f=1}^m CDD_{f,a,c}}{m} \quad (2a)$$

$$CS_{f,a,c} = \frac{\sum_{i=1}^n \hat{A}_{f,i} \hat{C}_{f,i}}{\sqrt{\sum_{i=1}^n \hat{A}_{f,i}^2} \sqrt{\sum_{i=1}^n \hat{C}_{f,i}^2}} \quad (2b)$$

$$CDD_{f,a,c} = (1 - \max_i (|\hat{A}_{f,i} - \hat{C}_{f,i}|)) \quad (2c)$$

where  $CS_{f,a,c}$  is the cosine similarity between the active user  $a$  and the current user  $c$  regarding the semantic feature  $f$ ,  $CDD_{f,a,c}$  is the Chebyshev distance

<sup>1</sup> <http://omdbapi.com>

dissimilarity between the active user  $a$  and the current user  $c$  regarding the semantic feature  $f$ ,  $\hat{A}_f$  is the vector of the active user  $a$  regarding the semantic feature  $f$ ,  $\hat{C}_f$  is the vector of the active user  $c$  regarding the semantic feature  $f$ ,  $m$  is the number of semantic features used,  $i$  is the index of each semantic feature instance,  $n$  is the number of semantic feature instances and  $\beta$  is a linear combining parameter. We used  $\beta = 0.5$ , *i.e.*, attributed equal weights to the collinearity and the proximity metrics. In the final prediction stage, the algorithm computes the active user predictions using the Equation 3:

$$P_{a,i} = \bar{r}_u + \frac{\sum_{j=1}^k r_{j,i} CPS_{a,c}}{\sum_{j=1}^k |CPS_{a,c}|} \quad (3)$$

where  $P_{a,i}$  is the item  $i$  prediction for the active user  $a$ ,  $\bar{r}_u$  is the average rating for user  $u$ ,  $r_{j,i}$  is the rating attributed by user  $j$  to the movie  $i$ ,  $CPS_{a,c}$  is the similarity between the active user  $a$  and the current user  $c$  and  $k$  is the number of neighbours.

## 4 Tests and Results

The following subsections present the data set used, the evaluation metrics and the results obtained. The recommendation service holds two user-based collaborative filters with the following neighbour discovery algorithms: (i) Pearson correlation (PC); and (ii) semantic neighbour discovery (SND). Both algorithms select the best 50 neighbours ( $k = 50$ ). The SND algorithm can be configured to include, for each rated film, any combination of actor (A), country (C), director (D), genre (G), language (L), type (T), writer (W) and year (Y) features extracted from the LOD. The PC algorithm is the base algorithm for all comparisons. Our aim is to compare the accuracy, prediction errors and run time of the two neighbour discovery algorithms: SND and PC. The experiments were performed with an Intel i7-2600 3.4 GHz Central Processing Unit (CPU) and 16 GB DDR3 Random Access Memory (RAM) platform.

### 4.1 Data Set

The recommendation service was evaluated off-line with HetRec 2011 v. 2.0<sup>2</sup>. This data set was chosen due to lower data sparsity and data set size. It contains information about users (2113) and movies (10 197), including user movie ratings (average of 405 ratings per user) and movie tag assignments (average of 23 tags per user) together with the corresponding timestamps. In terms of semantic data, the data set covers 6404 actors, 106 countries, 4196 directors, 27 genres, 166 languages, 5 types, 16 471 writers and 97 years.

The data set was divided into the training set (80%) and the test set (20%), ensuring that, regardless of the number of ratings, each user has 80% of their

<sup>2</sup> <http://www.grouplens.org/datasets/hetrec-2011/>

ratings in the training set and the remaining 20 % are in the testing set. Since the data set used contains sparse information, only items that were actually rated by users are recommended. This way, the system increases the accuracy of the predictions, but, since the suggested items belong to the group of items seen by the user, the calculated precision value is always one [5].

## 4.2 Evaluation Metrics

The metrics for evaluating recommender systems can be divided into predictive and classification accuracy metrics.

**Predictive Accuracy Metrics** measure the error between the predicted rating and the rating assigned by the user. According to [4], there are two important metrics: the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The normalised *MAE* and *RMSE* values range from 0 (best) to 1 (worst). *RMSE*, when compared with *MAE*, emphasizes the largest errors.

The Global *MAE* (GMAE), which corresponds to the average *MAE* for all users, measures the average absolute deviation between the predicted rating and the rating assigned by each user. The Global *RMSE* (GRMSE), which is the average *RMSE* taking into account all users, measures the root mean square deviation between the predicted rating and the rating assigned by each user.

**Classification Accuracy Metrics** The *Recall*, *Precision* and *F1-measure* are classification accuracy metrics and, as such, can be used to measure the frequency with which a recommendation system provides correct recommendations. *F1-measure* is the harmonic mean of the *Precision* and *Recall*, where *Precision* is the number of relevant items recommended from the total number of items and *Recall* is the number of relevant items recommended from the total number of relevant items available. These metrics range from 0 (worst) to 1 (best).

The Global *F1-measure* (GF1), which corresponds to the average *F1-measure* for all users, was used to evaluate the performance of both algorithms.

## 4.3 Results

In this section, we present the set of tests performed to compare the accuracy, predicting errors and execution time of the two neighbour discovery algorithms: SND and PC. These analyses are done by calculating the *GF1*, the *GMAE* and the *GRMSE*, respectively. We expect SND to have improved accuracy – it uses all items rated by the users (not just the co-rated) and semantically enriches the items to identify additional similarities among users – and run time – at least in the cases where the volume of the LOD data used allows it. In terms of predicting errors, we expect SND to exhibit a similar behaviour.

The off-line performance was determined separately for the two neighbour discovery algorithms, using a 60 % similarity threshold. The training data set was submitted to each filter, and the individual user’s recommendations were

generated. Finally, these recommendations were evaluated using the test data set and calculating the corresponding global metrics.

Table 1 presents the results using different  $\beta$  values for the combination of CS and CDD in CPS. We can observe that  $\beta = 0.5$  provides improved average results in terms of time, prediction errors and good results in terms of accuracy.

Table 1: Average CPS Results for different  $\beta$  values

$\beta$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
GF1	0.739	0.747	0.752	0.756	<u>0.758</u>	0.761	0.762	0.763	0.764	<b>0.765</b>
GMAE	0.133	0.133	0.133	0.133	<b>0.133</b>	0.133	0.133	0.134	0.134	0.134
GRMSE	<b>0.170</b>	<b>0.170</b>	<b>0.170</b>	0.171	<u>0.171</u>	0.171	0.172	0.172	0.172	0.172
TIME (s)	616	614	609	612	<u>604</u>	<b>604</b>	614.71	620	626	621

Figure 2 compares the results of both algorithms – SND and PC – for all semantic feature combinations. While Figure 2a compares the *GF1* and the run

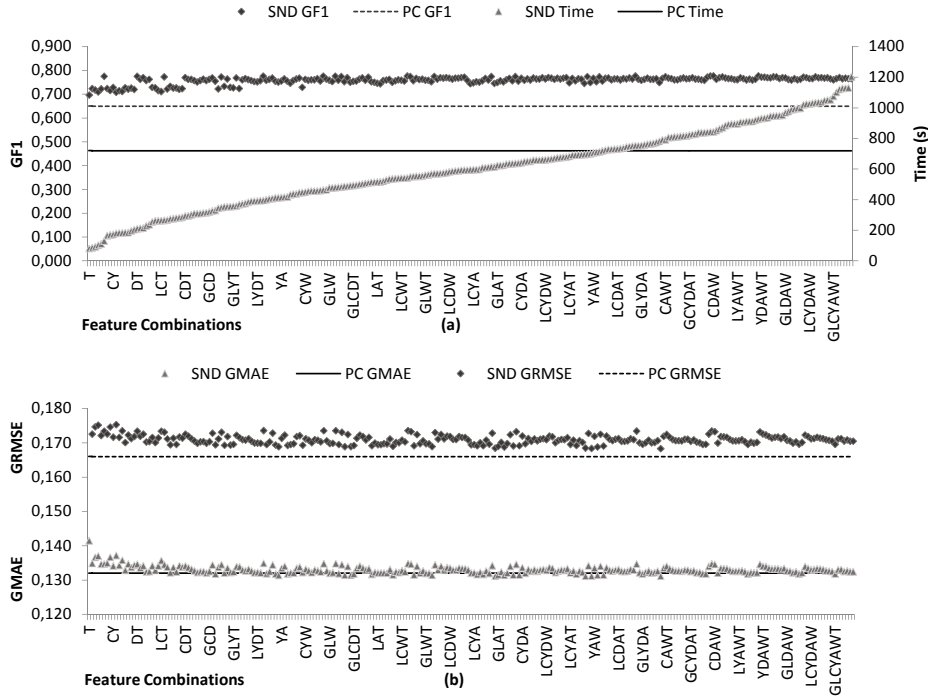


Fig. 2: SND *versus* PC: (a) GF1 and time; (b) GMAE and GRMSE

time for all semantic feature combinations, Figure 2b contrasts their normalised *GMAE* and *GRMSE*. The horizontal axis represents the 255 feature combinations involving the eight IMDb features – actor (A), country (C), director (D),



genre (G), language (L), type (T), writer (W) and year (Y). Figure 2a shows that SND is always more accurate – the black diamond data series *versus* the dashed black line – with an average improvement of 17 %. In terms of execution time, the SND algorithm – grey triangle data series – was faster than the base algorithm – the continuous black line – in 67 % of the cases. Figure 2b displays the negligible statistical errors that SND introduces. In average, the *GMAE* increases 1 % – grey triangle data series *versus* the continuous black line – and the *GRMSE* increases 3 % – black diamonds *versus* the dashed black line.

These results also show the relevance of selecting an appropriate set of LOD features for semantic enrichment. The analysis of the global metrics results for SND-GLCYDAWT, which corresponds to the version that uses all semantic features, and SND-DAT, which uses three features, illustrates this issue. Both combinations improve the recommendation accuracy 18 %, but SND-GLCYDAWT increases the run time 67 % while SND-DAT decreases the run time 25 %. In terms of the prediction errors, SND-DAT maintains the base algorithm errors and SND-GLCYDAWT maintains *GMAE* and increases *GRMSE* 3 %. In this case, the combination of director (D), actor (A) and type (T) provides better results than the combination of the eight IMDb features.

Figure 3 displays the top 8 results in terms of run time (a) and *GF1* (b). The horizontal axis represents the semantic feature combinations – actor (A), country (C), director (D), genre (G), language (L), type (T), writer (W) and year (Y).

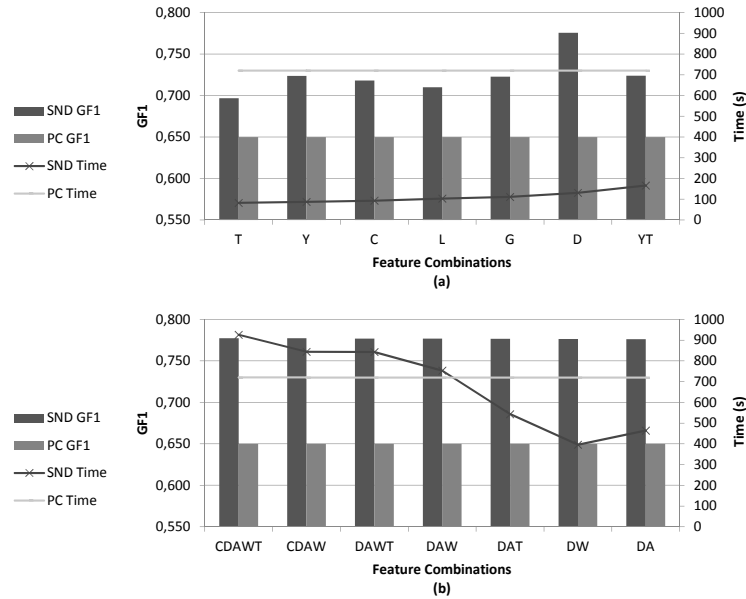


Fig. 3: Top 8 fastest (a) and most accurate (b) SND results.

We can observe in Figure 3a that the best feature in terms of run time and, then, *GF1* is director (D) since it reduces the run time 82 %, improves the recommendation accuracy 19 % and increases the recommendation errors by 2 % and 4 % for the *GMAE* and *GRMSE*.

SND-T, which corresponds to a semantic enrichment based solely on the type (T), reduces the run time by 89 %, improves the recommendation accuracy by 8 % and increases the recommendation errors by 7 % and 9 % for the *GMAE* and *GRMSE*. In Figure 3b, SND-DAW, which has a run time comparable to the base algorithm (5 % increase), improves the recommendation accuracy by 20 % and increases the recommendation errors by 2 % and 5 % for the *GMAE* and *GRMSE*, respectively.

As we had anticipated, the UbCF with SND generates more accurate predictions, is faster for the majority of the 255 feature combinations tested and introduces negligible prediction errors.

## 5 Conclusions

This paper describes an alternative neighbour discovery method for collaborative filters, which relies on semantic enrichment to expand the information available and explore additional relations among neighbours.

The SND algorithm, which uses all items rated by each user, and not only the co-rated items to discover the  $k$  nearest neighbours, shows considerable improvements in the quality of the recommendations in all SND feature combinations, when compared with the baseline algorithm (PC). The results obtained show that not all semantic features have the same impact and, thus, the set of semantic features to be used must be carefully selected. Our experiments showed that the best IMDb features are, for the HetRec data set, director (D), actor (A) and writer (W).

A straightforward implementation of the semantic neighbour discovery algorithm, *i.e.*, one that does not chose the semantic features to use, will face scalability and run time problems. These problems are, to some extent, shared with the baseline algorithm (PC). Our results show that it is possible to find a good compromise between the amount of semantic data used, the algorithm run time and the improvement of the recommendations accuracy. Nevertheless, in cases where the accuracy of the recommendations is crucial and the data set is huge, we recommend the adoption of resourceful execution platforms, *e.g.*, cloud computing platforms, to minimise the scalability problem together with the off-line pre-processing of semantic data enrichment prior to launching the on-line service. This approach allows the creation for all registered users of the corresponding data enrichment vectors (directors, actors, writers, genres, *etc.*) beforehand. Once the service is on-line, the UbCF with SND has to update the existing vector when new user feedback arrives and generate new semantic enrichment vector instances for new users.

As future work, we wish to apply this approach to item-based collaborative filters and assess the results to find the impact on the “new item problem”.

Furthermore, we intend to explore the use of other semantic sources to improve the subset of relevant semantic features.

## Acknowledgements

This work was partially financed by the ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme within project «POCI-01-0145-FEDER-006961» and by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) as part of project UID/EEA/50014/2013.

## References

1. Bellogín, A., Castells, P., Cantador, I.: Neighbor selection and weighting in user-based collaborative filtering: a performance prediction approach. *ACM Transactions on the Web (TWEB)* 8(2), 12 (2014)
2. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. *Scientific American* 284(5), 28–37 (2001)
3. Bobadilla, J., Ortega, F., Hernando, A., Gutiérrez, A.: Recommender systems survey. *Knowledge-Based Systems* 46, 109–132 (2013)
4. Herlocker, J.L., Konstan, J.A., Borchers, A., Riedl, J.: An algorithmic framework for performing collaborative filtering. In: *Proceedings of 22nd Annual international ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 230–237. ACM (1999)
5. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)* 22(1), 5–53 (2004)
6. Kaveh-Yazdy, F., Zare-Mirakabad, M.R., Xia, F.: A novel neighbor selection approach for knn: a physiological status prediction case study. In: *Proceedings of the 1st International Workshop on Context Discovery and Data Mining*. p. 2. ACM (2012)
7. Kushwaha, N., Vyas, O.: Semmovierec: extraction of semantic features of dbpedia for recommender system. In: *Proceedings of the 7th ACM India Computing Conference*. p. 13. ACM (2014)
8. Martín-Vicente, M.I., Gil-Solla, A., Ramos-Cabrera, M., Blanco-Fernández, Y., López-Nores, M.: A semantic approach to avoiding fake neighborhoods in collaborative recommendation of coupons through digital tv. *Consumer Electronics, IEEE Transactions on* 56(1), 54–62 (2010)
9. Martín-Vicente, M.I., Gil-Solla, A., Ramos-Cabrera, M., Pazos-Arias, J.J., Blanco-Fernández, Y., López-Nores, M.: A semantic approach to improve neighborhood formation in collaborative recommender systems. *Expert Systems with Applications* 41(17), 7776–7788 (2014)
10. Melville, P., Sindhvani, V.: Recommender systems. In: *Encyclopedia of machine learning*, pp. 829–838. Springer (2010)
11. Ostuni, V.C., Di Noia, T., Di Sciascio, E., Mirizzi, R.: Top-n recommendations from implicit feedback leveraging linked open data. In: *Proceedings of the 7th ACM Conference on Recommender Systems*. pp. 85–92. RecSys '13, ACM, New York, NY, USA (2013)

12. Papagelis, M., Plexousakis, D.: Qualitative analysis of user-based and item-based prediction algorithms for recommendation agents. *Engineering Applications of Artificial Intelligence* 18(7), 781–789 (2005)
13. Rey-López, M., Díaz-Redondo, R.P., Fernández-Vilas, A., Pazos-Arias, J.J.: T-learning 2.0: A personalised hybrid approach based on ontologies and folksonomies. In: *Computational Intelligence for Technology Enhanced Learning*, pp. 125–142. Springer (2010)
14. Sarwar, B.M., Karypis, G., Konstan, J., Riedl, J.: Recommender systems for large-scale e-commerce: Scalable neighborhood formation using clustering. In: *Proceedings of the Fifth International Conference on Computer and Information Technology*, 2002. vol. 1 (2002)
15. Symeonidis, P., Nanopoulos, A., Papadopoulos, A.N., Manolopoulos, Y.: Collaborative filtering: Fallacies and insights in measuring similarity. In: B. Berendt, A. Hotho, D. Mladenic, & G. Semeraro (Chairs), *Proceedings of the 17th European Conference on Machine Learning and 10th European Conference on Principles and the Practice of Knowledge Discovery in Databases Workshop on Web Mining*. pp. 56–67 (2006)
16. Veloso, B., Malheiro, B., Burguillo, J.C.: A multi-agent brokerage platform for media content recommendation. *Int. J. Appl. Math. Comput. Sci* 25(3) (2015)
17. Vozalis, M.G., Margaritis, K.G.: Applying svd on item-based filtering. In: *Proceedings of the Fifth International Conference on Intelligent Systems Design and Applications*, 2005. pp. 464–469. IEEE (2005)