UNIVERSITÀ
DEGLI STUDI
DI PALERMO

VASARI Project: Blended Recommendation for Cultural Heritage

Article

Accepted version

A. De Paola, A. Giammanco, G. Lo Re, M. Morana

In Proceedings of I-CiTies 2020 - 6th Italian Conference on ICT for Smart Cities And Communities

VASARI Project: Blended Recommendation for Cultural Heritage

Alessandra De Paola, Andrea Giammanco, Giuseppe Lo Re, Marco Morana
Universita' degli Studi di Palermo
Viale delle Scienze, edificio 6, Palermo, Italy
{alessandra.depaola, andrea.giammanco, giuseppe.lore, marco.morana}@unipa.it

***Abstract*—Supporting users in selecting the best works of art according to their interest is one of the most stimulating challenges in the field of *Cultural Heritage Promotion*. This task is made complex by several issues, such as the limited amount of time at users' disposal for cultural tours, the wide range of available artworks, and the heterogeneity and complexity of semantic features associated with them. In order to face such challenges and provide users with appropriate suggestions, we propose a blended recommendation approach that combines different techniques in order to capture latent relationships from wide sets of data associated with artworks.**

***Keywords*—Recommender Systems, Machine Learning, Smart Museums**

I. Introduction

Recommender Systems (RSs) play a pivotal role in improving users experience in Cultural Heritage scenarios [1] by supporting users in planning cultural tours which fully match their personal interests. These systems allow user to obtain a significant set of filtered suggestions, without being overwhelmed by the massive volume of information available from web sites and provided by travel guides.

The research laboratory of Artificial Intelligence and Distributed Systems of the University of Palermo conducts research in this field, by exploiting Artificial Intelligence techniques to innovate the fruition of cultural, artistic and historical heritage. In particular, the University of Palermo is partner, through the CINI, of the VASARI project (VAlorizzazione Smart del patrimonio ARtistico delle citta Italiane), which aims to create new ways to experience works of art and promote them, also by emphasizing smaller territorial realities through the use of intelligent Recommender Systems.

One of the most challenging issues related to the design of RSs, is the *cold start* problem occurring when new users interact with the system. Known users can be associated with a past history of interactions, which can be easily exploited to build a detailed profile of their interests, useful to provide new similar suggestions. New users, on the contrary, do not correspond to any known profile, thus it is very challenging to provide tailored suggestions.

In order to overcome such a problem, we propose to adopt a blended approach which combines different techniques, in order to design a system capable of dealing with multiple sources of heterogeneous information, by also leveraging users past behavior and current context, thus obtaining highly accurate suggestions for every class of users.

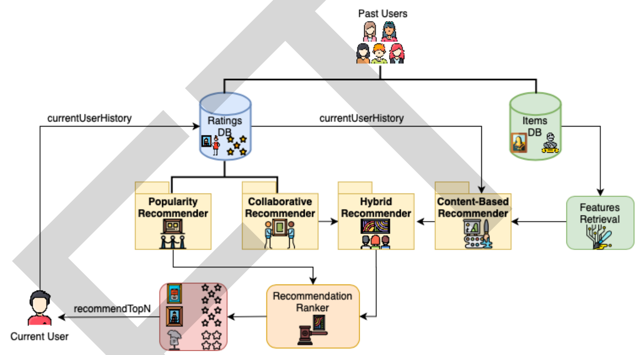

Fig. 1. Architecture of the proposed Recommender System.

II. Recommender System

In order to build a complex and complete Recommender System for cultural heritage, we propose to merge several algorithms according to four recommendation paradigms [1]:

- **Popularity based:** artworks with a large number of votes are suggested to users, considering as predicted rating for a specific artwork, the weighted average of its ratings;
- **Content based:** artworks are associated with semantic information, which is processed as feature vectors through machine learning predictive models;
- **Collaborative filtering:** latent relationships between artworks and users are learned and exploited to produce predicted ratings, by exploiting a cooperative approach;
- **Hybrid:** results from recommendation algorithms belonging to different paradigms (in our proposal, *Content based* and *Collaborative Filtering*) are combined into a single and more robust output.

The architecture of the proposed system is depicted in Figure 1. The Recommender modules produce lists of unordered couples $< item, predicted\_rating >$ according to the underlying recommendation paradigm; a *Recommendation Ranker* module sorts the recommendation lists coming from either the Popularity or Hybrid module, depending on the operating mode selected by the user o by the specific application.

As representative samples of each recommendation paradigm, we implemented 6 RSs:

1) a *popularity* based RS which exploits the IMDB rating formula to suggest objects widely appreciated by the community of users [1];

2) a *content* based RS which generates vectorial representation of the artworks by means of the *one hot encoding* in order to train a linear regressor [1], in particular, the *artist name*, *artist nationality*, *period*, *object type* and *caption* are interpreted as item features, and each artwork is represented as a binary vector, where 1s signal the presence of a certain feature in the artwork. We employ the *TF-IDF* score [2] to retrieve salient words from the verbose *caption* feature of the artworks;

3-5) two *collaborative filtering* approaches: *matrix factorization* [3], which captures the latent relationship between items and users by transforming the rating matrix into the dot product of two lower rank matrices by also considering users biases. *AutoRec* [4], in either the item-based and user-based variant, which extends the expressiveness of the *matrix factorization* model thanks to an autoencoder architecture with one single hidden layer of units;

6) a *hybrid* RS which properly merges the recommendations of *content* based and *collaborative filtering*. In particular, it realizes a weighted average of the two models' predicted ratings as follows: let w_{cb} be the weight for *content* based predictions, and w_{cf} be the weight for *collaborative filtering* predictions. Now, let us consider an item i, let $\widetilde{r_i}$ be the *content* based predicted rating for item i, and $\hat{r_i}$ be the *collaborative filtering* one. The final predicted rating for the *hybrid* RS is equal to $r_i = w_{cb} \cdot \widetilde{r_i} + w_{cf} \cdot \hat{r_i}$, where $w_{cb} + w_{cf} = 1$. In the simplest case, it is possible to choose $w_{cb} = w_{cf} = \frac{1}{2}$. However, a more intelligent approach would require a differential weight computation, in relation to the considered scenario. Thus, the values for weights w_{cb} and w_{cf} are adaptively computed in relation to the number of ratings currently available for the active user. In particular, let R_u be the number of ratings stored in the Ratings DB for user u, and let $\overline{R}$ be the average number of ratings for each user in the dataset. We compute:

$$w_{cb} = \begin{cases} 1 - \frac{R_u}{R_u + \overline{R}} & R_u \leq \overline{R} \\ \frac{1}{2} & R_u > \overline{R} \end{cases}$$

and $w_{cf} = 1 - w_{cb}$, thus implementing the following idea: the more ratings we have for the current user, and the more they quantitatively reach the average number of ratings expressed from other users, the more we can rely on a collaborative prediction for the ratings. When the current user has few ratings, the *hybrid* system will give more emphasis to *content* based recommendations, and then achieve an evenly bipartite combination with collaborative filtering when fully operational. In order to evaluate our proposal, we generated a synthetic dataset of explicit ratings for the artworks collection of the Minneapolis institute of art[1], following the guidelines proposed in [5]. We generated 100 users belonging to 10 different behavioral profiles, which essentially represent different artistic preferences in terms of *which* artworks are evaluated, and *how* they are evaluated. We trained and validated our RSs using the 5fold cross validation method. In Figure 2 we show some preliminar results, in particular, we show the *ROC* curves

[1]https://github.com/artsmia/collection

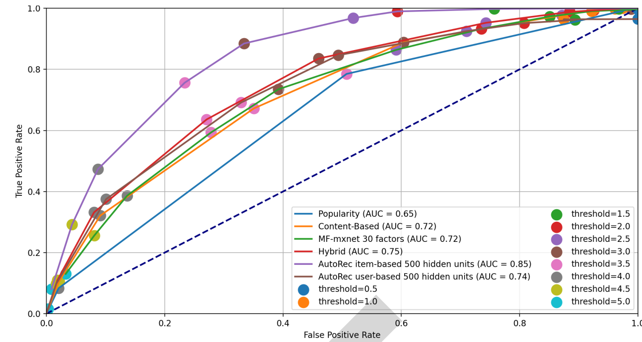

Fig. 2. ROC curves of the considered Recommender Systems against our synthetic ratings dataset set in Minneapolis Institute of art.

[6] for the Recommender algorithms underneath our system against our synthetic dataset, with different relevant/irrelevant rating decision thresholds. The *collaborative filtering* part of the *Hybrid* RS adopted in this simulation was based on the *matrix factorization* approach. However, results showed that higher values for the *Area Under the Curve* (AUC), which basically measures the discriminative power of a RS in terms of its ability to distinguish between relevant and irrelevant items, are obtained by the *AutoRec* item-based algorithm. Thus, we plan to exploit the more accurate suggestions coming from the *AutoRec* to improve the performances of the *Hybrid* system. Ontologies can also be used in order to overcome the heterogeneity of terms adopted by different users [7]. Moreover, results provided by the recommendation systems could also be refined according to the position of the user in relation to the work of art he is observing; to this aim, information coming from IoT sensors [8] could be exploited. Future work will also focus on a wider experimental analysis, and on the design of a graphical user interface for an easier fruition of the suggestions provided by the identified RSs. Moreover, the generality of the recommender system will allow to adopt the same framework to provide suggestions in a variety of application scenarios, such as smart campus [9], [10], [11] and other smart environments [12], [13]. Finally, we plan to investigate the use of reputation management mechanisms [14], [15] in order to detect and discard feedbacks provided by untrusted users that are interested in compromising the behavior of the system.

Acknowledgment

This research is partially funded by the Project VASARI of Italian MIUR (PNR 2015-2020, DD MIUR n. 2511).

References

[1] C. C. Aggarwal, *Recommender systems - The textbook*. Springer, 2016, vol. 1.

[2] J. Ramos *et al.*, "Using tf-idf to determine word relevance in document queries," in *Proceedings of the first instructional conference on machine learning*, vol. 242. Piscataway, NJ, 2003, pp. 133–142.

[3] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

[4] S. Sedhain, A. K. Menon, S. Sanner, and L. Xie, "Autorec: Autoencoders meet collaborative filtering," in *Proceedings of the 24th international conference on World Wide Web*, 2015, pp. 111–112.

[5] M. del Carmen Rodríguez-Hernández, S. Ilarri, R. Hermoso, and R. Trillo-Lado, "Datagencars: A generator of synthetic data for the evaluation of context-aware recommendation systems," *Pervasive and Mobile Computing*, vol. 38, pp. 516–541, 2017.

[6] P. Cremonesi, R. Turrin, E. Lentini, and M. Matteucci, "An evaluation methodology for collaborative recommender systems," in *2008 International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution*, 2008, pp. 224–231.

[7] P. Ferraro and G. Lo Re, "Designing ontology-driven recommender systems for tourism," in *Advances onto the Internet of Things*. Springer, 2014, pp. 339–352.

[8] S. Gaglio and G. L. Re, *Advances Onto the Internet of Things: How Ontologies Make the Internet of Things Meaningful*. Springer Publishing Company, Incorporated, 2014.

[9] S. Gaglio, G. Lo Re, M. Morana, and C. Ruocco, "Smart assistance for students and people living in a campus," in *2019 IEEE International Conference on Smart Computing (SMARTCOMP)*, June 2019, pp. 1–8.

[10] F. Concone, P. Ferraro, and G. Lo Re, "Towards a smart campus through participatory sensing," in *4th IEEE International Workshop on Sensors and Smart Cities (SSC 2018)*, Taormina, Italy, June 2018.

[11] V. Agate, F. Concone, and P. Ferraro, "Wip: Smart services for an augmented campus," in *The 4rd IEEE International Conference on Smart Computing (SMARTCOMP 2018)*, Taormina, Italy, June 2018.

[12] A. D. Paola, A. Giammanco, G. L. Re, and M. Morana, "Human mobility simulator for smart applications," in *2019 IEEE/ACM 23rd International Symposium on Distributed Simulation and Real Time Applications (DS-RT)*, 2019, pp. 1–8.

[13] A. de Paola, A. Giammanco, G. Lo Re, and G. Anastasi, "Detection of points of interest in a smart campus," in *2019 IEEE 5th International forum on Research and Technology for Society and Industry (RTSI)*, 2019, pp. 155–160.

[14] V. Agate, A. De Paola, G. Lo Re, and M. Morana, *A Simulation Framework for Evaluating Distributed Reputation Management Systems*. Cham: Springer International Publishing, 2016, pp. 247–254.

[15] V. Agate, A. De Paola, G. Lo Re, and M. Morana, "A platform for the evaluation of distributed reputation algorithms," in *2018 IEEE/ACM 22nd International Symposium on Distributed Simulation and Real Time Applications (DS-RT)*, Oct 2018, pp. 1–8.