Recommendation in Social Media Networks

Flora Amato[+], Vincenzo Moscato[+*], Antonio Picariello[+*], Giancarlo Sperlí[+]
DIETI
[+]University of Naples ''Federico II"
Via Claudio 21, 80125, Naples, Italy
[*] CINI - ITEM National Lab
Complesso Universitario Monte Santangelo, 80125, Naples, italy
{flora.amato, vmoscato, antonio.picariello@, giancarlo.sperli}@unina.it

***Abstract*—In this paper, we propose and describe a novel recommender system for big data applications that provides recommendations on the base of the interactions among users and generated multimedia contents in one or more social media networks, leveraging a collaborative and user-centered approach. Preliminary experiments using data of several online social networks show how our approach obtains very promising results.**

I. Introduction

In the age of *Big Data* a large amount of information - data coming from social media networks, digital libraries, news archives, travel websites, e-commerce portals, and so on - is now widely available to a large population of users. In the last decade, *Recommender Systems* [1] have been introduced to facilitate the browsing of such collections, assisting users to find "what they need" within this ocean of information.

Generally, recommender systems help people in retrieving information that match their preferences or needs by recommending products or services from a large number of candidates, and support people in making decisions in various contexts: what items to buy, which photo or movie to watch, which music to listen, what travels to do, or even who they can invite to their social network, just to make some examples.

Formally, a recommender system deals with a set of *users* $U = \{u_1 \dots, u_m\}$ and a set of *items* $O = \{o_1, \dots, o_n\}$. For each pair (u_i, o_j), a recommender can compute a *score* (or a *rank*) $r_{i,j}$ that measures the expected interest of user u_i in item o_j (or the expected utility of item o_j for user u_i), using a *knowledge base* and a *ranking* algorithm that generally could consider different combinations of the following characteristics: (i) user preferences and past behavior, (ii) preferences and behavior of the user community, (iii) items' features and how they can match user preferences, (iv) user feedbacks, (v) context information and how recommendations can change together with the context.

In the last years, several recommendation techniques have been proposed in the literature for supporting several kinds of applications in *heterogeneous social networks*. They exploit information extracted by the network to improve the accuracy of recommendation and provide new types of suggestions. As an example, they can recommend not only items but also groups, friends, events, tags, etc. to users, using particular algorithms that can be classified in three different classes.

The first one is composed by a set of algorithms leveraging a graph as data model, in which nodes and edges represent respectively single social entities and the relationships among them. Jiang et al.[2] propose a modified version of *Random Walk* on a star-structured heterogeneous network, centered on social domain which is connected with other domains, to select transferable items in auxiliary domains, bridge cross-domain knowledge with the social domain, and accurately predict user-item links in a target domain.

The algorithms that try to learn user and item preferences, through factorization of the related rating matrices, compose the second class. Jamali et al.[3] propose *Heteromf*, a context-dependent matrix factorization model taking in account jointly two types of latent factors: general latent factors for every entity type and context-dependent latent factors for each context in which entities are involved.

The third class is constituted by algorithms that model the interactions among users and items as "latent topics", enriching the recommendation process with other kind of information. To such purposes, Yuan et al.[4] propose a probabilistic model to exploit the different types of information (such as spatial, temporal and activity data) extracted from social networks to infer behavioral models for users. In turn, Pham et al. [5] propose an algorithm able to measure the influence among different types of entities, with the final aim to achieve higher recommendation accuracy and to understand the role of each entity in recommendation process.

From the other hand, if we take into account the architectural point of view, the last generation of recommender systems is usually composed by one or more of the following components[6].

- A *pre-filtering* module that selects for each user u_i a subset $O_i^c \subset O$ containing items that are good candidates to be recommended; such items usually match user preferences and needs.
- A *ranking* module that assigns w.r.t. user u_i a rank $r_{i,j}$ to each candidate item o_j in O_i^c using the well-known recommendation techniques (i.e., *content-based*, *collaborative filtering* and *hybrid* approaches) that can exploit in several ways items' features and users'preferences, feedbacks (in the majority of cases in terms of *ratings*) moods (by opinion or sentiment analysis) and behavior [1].

978-1-5090-6549-3/17 $31.00 © 2017 IEEE
DOI 10.1109/BigMM.2017.55

IEEE computer society

- A *post-filtering* module that dynamically excludes, for each user u_i, some items from the recommendations' list; in this way, a new set $O_i^f \subseteq O_i^c$ is obtained on the base of user feedbacks and other contextual information (such as data coming from the interactions between the user and the application). Eventually, depending on applications, recommended items can be arranged in *groups* according to additional constraints.

In this paper, we propose a novel *collaborative* and *user-centered* approach that provides *social* recommendations on the base of the all different kinds of interactions among users and generated contents in one or more social media networks.

Thus, in our approach several aspects related to users - i.e., *preferences* (usually coded in the shape of items' metadata), *opinions* (textual comments to which it is possible to associate a particular sentiment), *behavior* (in the majority of cases logs of past items' observations and actions made by users in the social environment), *feedbacks* (usually expressed in the form of ratings) - are considered and integrated together with *items' features* and *context information* within a general and unique recommendation framework that can support different social applications using proper customizations (e.g., recommendation of news, photos, movies, etc.).

The paper is organized as follows. Section 2 describes the proposed strategy for recommendation and provides a recommender system overview with some implementation details. Section 3 reports preliminary experimental results and Section 4 gives some concluding remarks and discusses future work.

II. The Recommendation Framework

The basic idea behind our proposal is that when a user is browsing a particular items' collection, the recommender system: (i) determines a set of useful *candidate* items on the base of user actual needs and preferences (*pre-filtering stage*); (ii) opportunely assigns to these items a rank, previously computed exploiting items' intrinsic features, users' past behaviors, and also using users' opinions and feedbacks (*ranking stage*); (iii) dynamically, when a user "selects" as interesting one or more of the candidate items, determines the list of most suitable items (*post-filtering stage*), also considering other context information in the shape of constraints on items' features.

A. Pre-filtering Stage using user preferences

In the *pre-filtering* stage, our aim is to select for a given user u_h a subset $O_h^c \subset O$ containing items that are good "candidates" to be recommended: such items usually have to match some (static) user preferences and (dynamic) needs. Each item subjected to recommendation may be represented in different and heterogeneous feature spaces. For instance, a photo may be described by a set of metadata as title, set of tags, description, etc.

The first step consists in clustering together "similar" items, where the similarity should consider all (or subsets of) the different spaces of features.

To this goal, we employ *high-order star-structured co-clustering* techniques - that some of the authors have adopted in previous work [7], [8] - to address the problem of heterogeneous data filtering, where a user is represented as a set of vectors in the same feature spaces describing the items.

B. Ranking Stage using user behavior and items similarity

In our model we assume the existence of a finite set of *Action Symbols* (S) coding all the possible "interactions" among a set of *Users* (U) and a set of *Objects* (O) in a Social Media Network, which can be properly captured during several browsing sessions using log information.

In such a context, we can consider different examples of actions: users' reactions or comments to published contents (e.g., a post or a photo), user visualization or rating of a given content, and so on.

We introduce the following definitions.

Definition 2.1 (Log tuple): A *log tuple* can be defined by $l = (s, u, o, \lambda_1, \cdots, \lambda_k)$, where $s \in S$, $u \in U$, $o \in O$ and $\lambda_1, \cdots, \lambda_k$, are particular attributes (e.g., timestamp, type of reaction, text and tags of a comment, etc.) used to describe a particular action.

Definition 2.2 (Log): A *Log* (L) is a finite sequence of log tuples.

Intuitively, a log tuple corresponds to an observation of $l.s$ performed by the user $l.u$ on a given object $l.o$ along with the associated attributes of the observation $\lambda_1, \cdots, \lambda_k$. By convention, if action a_2 occurs after a_1 in a log, then the action a_2 occurred temporally after a_1.

Starting from a log, it is possible to model interactions among users and the generated multimedia content in social medial networks as a particular labeled graph as in the following definition.

Definition 2.3 (User-Content Social Graph): A *User-Content Social Graph* is a couple (G, γ), where: $G = (O, E)$ is a *directed graph*; $\gamma : E \rightarrow \{pattern, sim\} \times R^+$ is a *labeling function* that associates each edge in $E \subseteq O \times O$ with a pair (t, w), t being the type of the edge which can assume two enumerative values ($pattern$ and $similarity$) and w is the weight of the edge.

In our model, we list two different cases: (i) a *pattern label* for an edge (o_j, o_i) denotes the fact that an item o_i was accessed immediately after an item o_j by any user in a given browser session and, in this case, the weight w_j^i is the number of times o_i was accessed immediately after o_j; (ii) a *similarity label* for an edge (o_j, o_i) denotes the fact that an item o_i is similar to o_j and, in this case, the weight w_j^i is the "similarity" between the two items. Thus, a link from o_j to o_i indicates that part of the importance of o_j is transferred to o_i.

A user-content social graph can be *local* if it is related to a given user or *global* if it concerns an entire social community.

Leveraging log information in the graph building process, we have to choose for each kind of social media network:

- the list of "consecutive" actions in the log with the related attributes that can instantiate an edge, as an example:
  - a user visualized/published two objects in consecutive temporal instants of the same browsing session,
  - a user provided two positive reactions or comments to two different objective in successive times of the same browsing session;
  - a user marked two objects as "favorite" in consecutive temporal instants of the same browsing session,
  - etc.
- the particular attributes (e.g. tags, keywords or other relevant information extracted from a text) describing user actions on different objects that can be used in the similarity computation together with object metadata and low-level features of multimedia contents.

Figure 1 describes an example of log and the related global user-content social graph.

On the base of the user-content social graph, each object can be opportunely *ranked*.

Definition 2.4 (Recommendation Grade): Given an item $o_i \in O$, its *recommendation grade* $\rho(o_i)$ is defined as follows:

$$\rho(o_i) = \sum_{o_j \in P_G(o_i)} \hat{w}_{ij} \cdot \rho(o_j) \quad (1)$$

where $P_G(o_i) = \{o_j \in O | (o_j, o_i) \in E\}$ is the set of predecessors of o_i in G, and $\hat{w}_{ij}$ is the normalized weight of the edge from o_j to o_i. We note that for each $o_j \in O$ $\sum_{o_i \in S_G(o_j)} \hat{w}_{ij} = 1$ must hold, where $S_G(o_j) = \{o_i \in O | (o_j, o_i) \in E\}$ is the set of successors of o_j in G.

In [9], it has been shown that the ranking vector $R = [\rho(o_1) \ldots \rho(o_n)]^T$ of all the items can be computed as the solution to the equation $R = C \cdot R$, where $C = \{\hat{w_{ij}}\}$ is an ad-hoc matrix that defines how the importance of each item is transferred to other items.

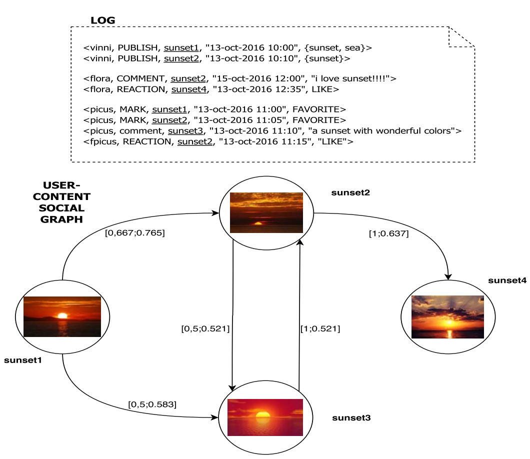

Figure 1: Log with the related User-Content Social Graph.

The matrix can be seen as a linear combination of: (i) a *local matrix* $A_h = \{a_{ij}^h\}$ for each user u_h, where its generic element a_{ij}^l is defined as the ratio of the number of times item o_i has been accessed by user u_h immediately after o_j to the number of times any item in O has been accessed by u_h immediately after o_j; (ii) a *global matrix* $A = \{a_{ij}\}$, where its generic element a_{ij} is defined as the ratio of the number of times item o_i has been accessed by any user immediately after o_j to the number of times any item in O has been accessed immediately after o_j; (iii) a *similarity matrix* $B = \{b_{ij}\}$ such that b_{ij} denotes the similarity between two items o_i and o_j (e.g., a semantic relatedness[10], [11] based on a given taxonomy using some high-level features values, eventually combined with a low-level features comparison in the case of multimedia data).

The successive step is to compute *customized* rankings for each individual user. In this case, we can rewrite previous equation considering the ranking for each user as $R_h = C \cdot R_h$, where R_h is the vector of preference grades, customized for a user u_h considering only items in the related O_h^c. In [9], it has been demonstrated that C, under certain assumptions and transformations, is a real square matrix having positive elements, with a unique largest real eigenvalue and the corresponding eigenvector has strictly positive components, and the equation can be solved using *Power Method*. Rank can be finally refined using user attached sentiments and ratings [6].

C. Post-Filtering Stage using context information

We have introduced a *post-filtering* method for generating the final set of "real" candidates for recommendation using *context* information. The context is represented by means of the well-known *key-value* model [12] using as dimensions some of the different feature spaces related to items. In our system, context features can be expressed either directly using some *target items* (e.g. objects that have positively captured user attention) or specifying the related values in the shape of *constraints* that recommended items have to satisfy.

Assume that a user u_h is currently interested in a target item o_j. We can define the set of candidate recommendations as follows:

$$O_{h,j}^f = \bigcup_{k=1}^{M} \{o_i \in O_h^c \,|\, a_{ij}^k > 0\} \cup \{o_i \in NNQ(o_j, O_h^c)\} \quad (2)$$

The set of candidates includes the items that have been accessed by at least one user within k steps from o_j, with k between 1 and M, and the items that are most similar to o_j according to the results of a *Nearest Neighbor Query* ($NNQ(o_j, O_h^c)$) functionality[1]. The ranked list of recommendations is then generated by ranking the items in $O_{h,j}^f$, for each item o_j selected as interesting by user u_h, using the ranking vector R_h thus obtaining the final set O_h^f. Finally, for each user all the items that do not respect possible context constraints are removed from the final list.

[1]Note that a positive element a_{ij}^k of A^k indicates that o_i was accessed exactly k steps after o_j at least once.

III. Preliminary Experiments

We designed and implemented a first prototype of a recommender system that builds and manages a user-content social graph, providing the basic facilities for ranking computation and querying and data filtering, using data from several online social networks. In particular, our recommender systems was realized using the Apache Spark engine based on the Hadoop technological stack: the log files are stored on HDFS while user, context and object descriptions are managed by Cassandra. Co-clustering, ranking and post-filtering techniques were implemented on the top Spark machine learning and graph analysis libraries and leveraging SPARK SQL facilities; in turn, multimedia similarities were computed using Windsurf library [13].

In this paper, we report several preliminary experimental results aiming at evaluating the effectiveness of the introduced ranking method (*Social Ranking (SR)*) with respect to a human ground truth.

We used a subset of the *Yahoo Flickr Creative Commons 100 Million Data* (YFCC100M)[2] multimedia collection, provided by Yahoo in 2014. In particular, we exploited users' social interactions (friendships, tags, publishing, comments, favorites) with the related multimedia data (images retrieved using Flickr API[3] about animal, landscape and nature domains) to build our user-content graph[4].

In order to evaluate the performance of the proposed algorithm, we defined a human-generated ranking (representing the unique gold standard), asking a group of about 20 students to rank the results of different queries with respect to the relevance of the retrieved images, in terms of topics and multimedia content. We have used two classes of ranking methods: *Popularity* (PR)[14] - based on popularity of images computed by a linear combination of number of favorites, comments and likes - and *Collaborative* (CR)[15] - based on the user feedbacks and interactions among users.

Table I shows the results of ranking comparison using the *Kendall's Tau* (τ) and *Spearman's Rank Correlation* (ρ) coefficients obtained by comparing in pairs the following methods: Social Ranking (SR), Human Ranking (HR), Popularity Ranking (PR) and Collaborative Ranking (CR).

It is possible to note that our ranking (that considers both popularity and collaborative aspects) presents the most similar behavior with respect to the human ground truth.

IV. Conclusion and Future Works

In this paper we proposed a novel collaborative and user-centered approach exploiting the interactions among users and generated contents in one or more social media networks.

Moreover, preliminary experiments on ranking effectiveness show how our approach provides promising and interesting results.

[2]https://webscope.sandbox.yahoo.com

[3]https://www.flickr.com/services/api

[4]An edge is created when a user visualized/published two objects in consecutive temporal instants of the same browsing session, or when she/he marked two objects as "favorite" in consecutive temporal instants of the same browsing session.

	τ	ρ
SR - HR	0,80	0,91
PR - HR	0,64	0,71
CR - HR	0,78	0,88
SR - PR	0,61	0,67
SR - CR	0,78	0,85
PR - CR	0,63	0,88

Table I: Ranking comparison (Social Ranking (SR), Human Ranking (HR), Popularity Ranking (PR) and Collaborative Ranking (CR)).

Future works will be devoted to exploit our ranking approach for effectively supporting multimedia recommendation in heterogeneous social media networks, providing more detailed experimental results and a comparison with the most diffused approaches for social recommendation with respect to different datasets.

References

[1] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds., *Recommender Systems Handbook*. Springer, 2011.

[2] C.-D. T. Knowledge, "Social recommendation with cross-domain transferable knowledge," *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, vol. 27, no. 11, 2015.

[3] M. Jamali and L. Lakshmanan, "Heteromf: recommendation in heterogeneous information networks using context dependent factor models," in *Proceedings of the 22nd international conference on World Wide Web*. ACM, 2013, pp. 643–654.

[4] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann, "Who, where, when and what: discover spatio-temporal topics for twitter users," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2013, pp. 605–613.

[5] T.-A. N. Pham, X. Li, G. Cong, and Z. Zhang, "A general recommendation model for heterogeneous networks," *IEEE Transactions on Knowledge & Data Engineering*, no. 12, pp. 3140–3153, 2016.

[6] F. Colace, M. D. Santo, L. Greco, V. Moscato, and A. Picariello, "A collaborative user-centered framework for recommending items in online social networks," *Computers in Human Behavior*, vol. 51, pp. 694–704, 2015.

[7] D. Ienco, C. Robardet, R. G. Pensa, and R. Meo, "Parameter-less co-clustering for star-structured heterogeneous data," *Data Mining and Knowledge Discovery*, vol. 26, no. 2, pp. 217–254, 2013.

[8] I. Bartolini, V. Moscato, R. G. Pensa, A. Penta, A. Picariello, C. Sansone, and M. L. Sapino, "Recommending multimedia visiting paths in cultural heritage applications," *Multimedia Tools Appl.*, vol. 75, no. 7, pp. 3813–3842, 2016.

[9] M. Albanese, A. d'Acierno, V. Moscato, F. Persia, and A. Picariello, "A multimedia recommender system," *ACM Trans. Internet Techn.*, vol. 13, no. 1, p. 3, 2013.

[10] F. Amato, A. Mazzeo, V. Moscato, and A. Picariello, "A framework for semantic interoperability over the cloud," in *Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference on*. IEEE, 2013, pp. 1259–1264.

[11] F. Amato, A. De Santo, V. Moscato, F. Persia, A. Picariello, and S. R. Poccia, "Partitioning of ontologies driven by a structure-based approach," in *Semantic Computing (ICSC), 2015 IEEE International Conference on*. IEEE, 2015, pp. 320–323.

[12] G. Adomavicius, R. Sankaranarayanan, S. Sen, and A. Tuzhilin, "Incorporating contextual information in recommender systems using a multidimensional approach," *ACM Transactions on Information Systems (TOIS)*, vol. 23, no. 1, pp. 103–145, 2005.

[13] I. Bartolini and M. Patella, "Multimedia queries in digital libraries," in *Data Management in Pervasive Systems*. Springer, 2015, pp. 311–325.

[14] H. Steck, "Item popularity and recommendation accuracy," in *Proceedings of the fifth ACM conference on Recommender systems*. ACM, 2011, pp. 125–132.

[15] P. Resnick and H. R. Varian, "Recommender systems," *Communications of the ACM*, vol. 40, no. 3, pp. 56–58, 1997.