Socially-aware Video Recommendation using Users' Profiles and Crowdsourced Annotations

Marco Bertini, Alberto Del Bimbo, Andrea Ferracani, Francesco Gelli, Daniele Maddaluno, Daniele Pezzatini*
Università degli Studi di Firenze - MICC
marco.bertini, alberto.delbimbo, andrea.ferracani, daniele.pezzatini@unifi.it

ABSTRACT

The recent explosive growth in the use of social networks has raised the question of how to meet the emerging demand for services that address the interests of the users. In this paper we show how considering homophily in social networks can improve video recommendation, using inferred user profiles and modeling users' interests. We propose a socially-aware framework for video commenting, sharing and interest discovery that combines recommendation algorithms, clustering techniques, tools for video tagging and evaluation of tag semantic relatedness. The system allows to connect to friends, curate a personal profile and get video recommendations through a social network.

Categories and Subject Descriptors

H.3.5 [**Information Storage and Retrieval**]: Online Information Services; H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Algorithms, Design, Experimentation

Keywords

Social video tagging; social video recommendation

1. INTRODUCTION

Nowadays one of the main challenges of a social network is targeting and personalization of services and items to help users to choose from a wide variety of alternatives deriving from the huge amount of user-generated content in the network. Another challenge is to guide users to create and update their public profiles as means to motivate users to be engaged with the systems. Social networks and commercial sites like Facebook, Netflix, YouTube and Digg already propose to users contents of interest based on past users' interactions (e.g. in the YouTube personal home page), past user ratings, watching preferences and items metadata (Netflix suggestions of movies), or profiles of similar users (e.g. the recommendation system of Digg news), exploiting collaborative filtering (CF) and collective intelligence algorithms. Some of these networks also feature public profiles, like the Facebook timeline, which collects users activity and offers the possibility to manually curate personal pages. However, all these social networks address the issue of recommendations and targeted services considering only a few discrete user activities like voting or tagging.

The capability of recommending relevant videos to targeted users can help them to find the most relevant content according to their recurrent viewings or preferences. As shown in [21], recommendation is a powerful force in driving users to watch other videos, much more than direct searching for new videos. Typically the approaches presented in the scientific literature are based on textual analysis of the metadata that accompany a video, possibly complemented by some multimedia content analysis. In [19] the multimodal relevance between two video documents is expressed as the combination of textual, visual, and aural relevance, and relevance feedback is used to automatically adjust intra-weights within each modality and inter-weights among different modalities by users' click-through data. The YouTube recommendation system, described in [6] uses two broad classes of data: *i)* content data, such as the raw video streams and video metadata such as title, description, etc., and *ii)* user activity data, either explicit (e.g. video rating/liking) and implicit (watching a video for a long time). The videos that a user is likely to watch after having watched a given seed video are considered as related and are selected using association rule mining (co-visitation counts). In [16] user activity (video play) is used to create a user profile, by accumulating the tags used by the video uploaders to describe their videos. The user profiles are then used to compute the set of recommended videos based on their tags. In [12] it has been proposed a weighted fusion of different modalities based on user activity (e.g. fast forward/rewind) during the video play. The use of social features to augment the evaluation of video similarity based on its content has been proposed in [4]. In [7] it has been proposed a system that merges video content and social networks to gather semantic metadata to describe interaction, usage and opinion of video content. Social similarity of videos, expressed as popularity distribution across social circles has been used to improve video recommendation in [10]. In [1], the social network of a user is exploited to create his initial user profile and to suggest him videos based on those tagged by his friends using a tool for temporal video commenting; video

*Authors listed in alphabetical order

SAM'13, October 21, 2013, Barcelona, Spain.

http://dx.doi.org/10.1145/2509916.2509924.

tagging is also used to update a list of user's interests that are mapped on semantic categories using DBPedia.

In this paper we introduce a socially-aware framework that combines a profiling module (through online social network, i.e. Facebook) and user's activity analysis, like semantic tagging, to generate targeted services for the users of the social network. User profiles are generated semi-automatically, reducing the effort required to the users for their creation, and are exploited to perform recommendation of videos, users and resources of interest. The general idea is to improve and propagate interests and connections through the network (social influence) and to exploit a known tendency of social networking, called homophily [11] in sociology, according to which contacts between similar people occurs more frequently than among dissimilar people. In fact, information contained in interest networks and in friendship networks is highly correlated and homophily can be used to achieve higher performance in interest targeting and friendship prediction, as shown in [20]. In this regard, the main contribution of this work is the use of users' interest similarity, estimated from semi-automatically created user profiles, to support video recommendation as well as interest targeting and friendship prediction. By computing similarity between users considering their interest profiles and not only their voting activity, we are also able to face the common "cold start" issue for recommenders. In fact, classic user-based CF algorithms have the problem to make recommendations to novel users who have not yet expressed preferences on any item in the network.

This paper is organized as it follows: Sect. 2 describes the components of the system; Sect. 3 evaluates the proposed framework; conclusions are drawn in Sect. 4, along with a description of future work.

2. THE SYSTEM

The system[1] allows users to connect to friends, comment and semantically annotate videos at frame level, share and browse videos, interests and users through recommendation, clustering and semantic similarity, and to create and curate personal profiles of interests in a semi-automatic way. In this regard it can be classified as an hybrid recommender system combining a collaborative filtering approach with content-based filtering techniques. As noted in [3], the collaborative system approach in social networks needs to be extended, because the nature of these networks is somehow bilateral and also users have to be considered as 'items' that participate in social interactions. In this regard, the recommendation system proposes, in the personal user home page, not only videos but also these most similar users.

The system consists of three main parts that are closely interconnected: *i)* a recommendation engine of videos and similar users, viewable in the personal home page of the social network (similarly to the YouTube home page), *ii)* a user profiling engine for automatic creation of public profiles of interests, *iii)* a clustering engine that is responsible of the categorization of resources in the network and also, through the aid of semantic distances, makes recommendations and suggestions of resources that match those interests. The profile can then be edited and refined by users over time in a semi-automatic way.

Users' similarity is computed not only on the traditional paradigm of a model built from the past behavior of the user (considering numerical rates and "implicit" rates from viewing activity) or from similar decisions made by other users, but also on data that can be defined as having an implicit or explicit semantic (i.e. semantics extracted from users' comments or explicitly declared in user profiles); in particular, it is taken into account information extracted from Facebook (e.g. page 'likes'), users' activity in the system (comments and annotations) and information provided manually (categories of interest) or automatically computed (semantic analysis and categorization of annotations).

The modeling of user interests is carried out with three strategies: initially, in the "cold start" scenario, the system tries to extract some observable properties from users' Facebook profiles, then it uses algorithms to compute a user's neighborhood based on profile similarity, exploiting collaborative filtering techniques and implicit interest indicators, and finally it clusters resources extracted from the network activity, assigning the classifications of these clusters to the users' personal profile of interests. In this way the system integrates a collaborative filtering and a content based filtering approach and the user interaction paradigm is combined with featured-based recommendation algorithms, binding the neighborhood based recommendations to the network social graph, as suggested in [2, 8, 18]. The workflow of the system is shown in Fig. 1.

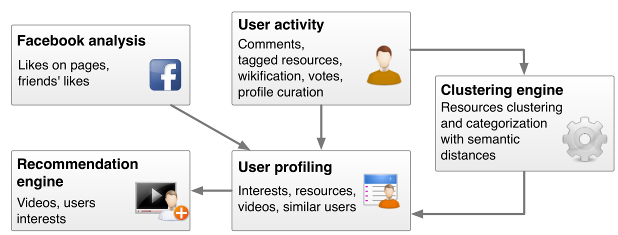

Figure 1: System workflow.

Crowdsourced video tagging and concept extraction.

The key data through which the system extracts semantic information about users is constituted by their comments on video frames and by any resource automatically detected or manually tagged. In [9] it has been showed that tag-based recommenders have a better interest ratio than people-based recommenders, and recommenders that combine both elements are even better. In this regard the system features an automatic extraction of semantic entities and provides also a widget for manual semantic tagging which allows to add, at video frame level, Wikipedia and Facebook resources within the text of the comments (Fig. 2). The interface has been designed to make the process similar to that of commenting and tagging photos on a social network like Facebook, following a procedure that is familiar to users. Semantic tags are automatically detected within comments using Named Entity Detection based on rules and gazetteer [5] and with a 'wikification' procedure that identifies Wikipedia entities in text comments [14]; during 'wikification', the system selects one sense, when more senses are available for a term on Wikipedia, considering the best semantic relatedness based on Wikipedia article internal links structure [13]. These semantic tags are used to represent the video topics, and their association to users' interests is computed in real-time when users post new comments, to update their personal profiles.

[1]available at: http://fiona.micc.unifi.it/intime

Clustering and categorization of resources is then refined by Mahout scheduled jobs, as described in the following.

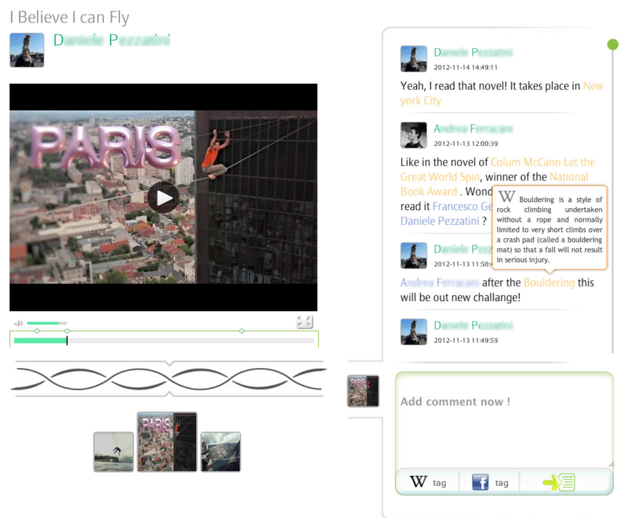

Figure 2: The annotation widget: users can comment at video frame level and add resources both from Facebook and Wikipedia. Automatically detected semantic entities in the comments are highlighted.

Recommendations, profile creation and curation.

Recommendation of videos is content-based, computed evaluating users' semantic profiles similarity. Profiles are created on the basis of the resources that have been tagged or extracted automatically from user comments on video frames. All the extracted resources are represented in the network by their corresponding Wikipedia page text document. Resources are vectorized using the TF-IDF algorithm and clustered with Fuzzy K-Means using a semantic distance. This distance is computed using the Wikipedia Link-based Measure (WLM), that takes into account the similarity of two Wikipedia articles with respect to the comparison of their incoming and outgoing links [13]. The clustered resources are associated to a two-levels taxonomy of interests (12 categories in the first level and 50 in the second, such as *Music* and *Rock music*, inspired by the taxonomy of Vimeo) by computing a weighted average of the semantic distances of the k resources closest to the centroid with respect to the items of the taxonomy. Each item of the taxonomy has its own corresponding Wikipedia article, used to compute the semantic distance.

Profiles generation is obtained matching users' interest and labelled cluster of resources, and is refined and updated using collaborative filtering techniques (implicit, e.g. click through, and explicit profile curation). The initial profile is based on personal data obtained from the Facebook Graph API: the system analyses user 'likes' on Facebook pages, extracting their Facebook categories, or the likes of the friends of the user, in case the user has no 'likes' activity. Then the categories extracted from Facebook are compared with the items of the system taxonomy computing a semantic distance between the Facebook category and each taxonomy item, using the WLM measure.

Video suggestions are computed considering and comparing the profiles of interest of the most similar users in the network. For this a memory-based algorithm is adopted which uses the map/reduce paradigm implemented in Apache Mahout on an Hadoop cluster in order to handle the scalability issue that such approach usually presents.

Users are described using a vector that contains the percentage of interest for all the categories of the system. Percentages of interest are normalized counting user's Facebook likes on categories in the "cold start scenario" and refined later considering the number of the resources added by users for each interest while curating their public profile. To this end, the system features a user profile view presenting automatically extracted interests and, for each of them, a carousel of suggested resources that users can drag and promote in their public profile. This vector of weighted categories is used to compute user similarity and to determine a user neighborhood inside the network. Once the neighborhood n is defined, the recommendation is generated by ranking items using the preferences expressed by users in n. Formally, the proposed system modifies the user-based recommendation algorithm, in that it uses the similarity of user interests to select the items on which recommendation is computed (Alg. 1).

given a user u **foreach** *other user* w **do**
  compute an interests similarity s between u and w.
  create a neighborhood n containing top k users, ranked by similarity.
end
foreach *item* i *that some user in* n *has a preference for, but that* u *has no preference for yet* **do**
  foreach *other user* v *in* n *that has a preference for* i **do**
    compute a similarity s between u and v.
    incorporate v's preference for i, weighted by s, into a running average.
  end
end
return *the top items, ranked by weighted average*

Algorithm 1: Algorithm for user similarity recommendation based on users' profile of interest. Different users' similarity measures have been tested in the experiments.

3. EXPERIMENTAL RESULTS

The recommendation process can be viewed as a prediction problem: the system should be able to predict the user's level of interest in specific items, such as videos or other users, and rank these according to their predicted values [15]. In order to evaluate the accuracy of the prediction, we can extract a percentage of the collected data, represented by users ratings on videos, and use them as test data, that is not used to train the recommendation system. The recommender engine produces rating predictions for the missing test data, that are compared to the actual values in order to evaluate the accuracy. The performance is evaluated using Root Mean Square Error (RMSE), since this is the most commonly used for this task. The more accurately the recommender predicts users rating, the lower the RSME will result.

Our dataset is composed by 138 videos and 51 users (with 152 expressed preferences - 23 ratings of 1, 16 of 2, 24 of 3, 45 of 4 and 44 of 5 - sparsity level= $1 - \frac{expr.\ ratings}{users \times videos} = 0.978$). User interests profiles are represented by 383 resources that are organized in 15 main categories. User's ratings on videos vary on a 1 to 5 scale vote system. We chose to select 90% of our data-set as training set, and to perform an evaluation on the remaining 10% of the data, using a repeated random sub-sampling validation (1000 iterations). In the first experiment, we tested three different distance measures between users: Euclidean distance, Pearson correlation and Log-Likelihood, in terms of RMSE between predicted value of rating and real user rating. As shown in Fig. 3 *left)*, results show that Pearson correlation tends to behave badly when small amounts of data are involved; Euclidean distance provides best results, therefore we adopted such distance metric in our system. All the tested similarity measures obtain approximately the same accuracy when the number of considered neighbors grows.

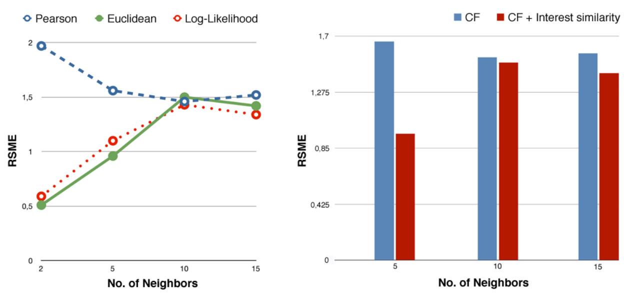

Figure 3: *left)* RMSE comparison between different similarity measures; *right)* RMSE comparison between Mahout user-based recommendation (CF) and the proposed recommender algorithm (CF + Interest similarity).

In the second experiment we compared our recommendation algorithm with the user-based Collective Filtering algorithm [17] that does not consider interest profile similarity. We evaluated the RMSE of both approaches, computing the values at different sizes of the neighborhood. The results are shown in Fig. 3 *right)*. We can observe that the number of users adopted for the neighborhood in our approach does actually affect the quality of prediction. The baseline user-based algorithm obtains its best results when the neighborhood size is 10, even if variations are small. The proposed algorithm always performs significantly better, in particular when a small number of neighbors is involved. For instance, if we consider the top 5 similar users of the population to produce the recommendation, our system produce RMSE of 0.96, while the classical CF algorithm produce an RMSE of 1.66. When the size of the neighborhood grows, the two approaches tend to give similar results, although our proposed solution performs better.

4. CONCLUSIONS

In this paper we presented a socially-aware framework that exploits homophily, represented using users' interests profiles, to improve videos and resources recommendations. We have shown that it is possible to use a combination of algorithms for the computation of semantic distances (WML) with classical clustering techniques for user profiling and expansion of knowledge and we have provided a variation to the classical algorithms of user similarity based on ratings in favor of a neighborhood computed on users profiles. An evaluation of the proposed recommender system has shown improvements with respect to traditional collaborative filtering approaches.

Future work will focus on: *i)* integration of a system to extract salient video frames to be presented to users, in order to increase comments and manual tagging, *ii)* use of tag propagation techniques applied to users' semantic annotations.

Acknowledgments. The authors thank Daniele Daveri Niederwinkler for his contribution to the design of the graphical user interface, and Tiberio Uricchio for his technical help on Hadoop/Mahout.

References

[1] M. Bertini, A. Del Bimbo, A. Ferracani, and D. Pezzatini. A social network for video annotation and discovery based on semantic profiling. In *Proc. of WWW*, 2012.

[2] P. Bonhard and M. A. Sasse. Knowing me, knowing you - using profiles and social networking to improve recommender systems. *BT Technology Journal*, 25(3), 2006.

[3] X. Cai, M. Bain, A. Krzywicki, W. Wobcke, Y. S. Kim, P. Compton, and A. Mahidadia. Collaborative filtering for people to people recommendation in social networks. In *Proc. of ACAI*, 2010.

[4] M. Cherubini, R. de Oliveira, and N. Oliver. Understanding near-duplicate videos: a user-centric approach. In *Proc. of ACM MM*, 2009.

[5] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In *Proc. of ACL*, 2002.

[6] J. Davidson, B. Liebald, J. Liu, P. Nandy, T. Van Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston, and D. Sampath. The YouTube video recommendation system. In *Proc. of ACM RecSys*, 2010.

[7] S. Davis, I. Burnett, and C. Ritz. Using social networking and collections to enable video semantics acquisition. *IEEE MultiMedia*, 16(4):52–61, 2009.

[8] J. Golbeck. Generating predictive movie recommendations from trust in social networks. In *Proc. of ICTM*, 2006.

[9] I. Guy, N. Zwerdling, I. Ronen, D. Carmel, and E. Uziel. Social media recommendation based on people and tags. In *Proc. of ACM SIGIR*, 2010.

[10] X. Ma, H. Wang, H. Li, J. Liu, and H. Jiang. Exploring sharing patterns for video recommendation on YouTube-like social media. *Multimedia Systems*, pages 1–17, 2013.

[11] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 2001.

[12] T. Mei, B. Yang, X.-S. Hua, and S. Li. Contextual video recommendation by multimodal relevance and user feedback. *ACM TOIS*, 29(2), 2011.

[13] D. Milne and I. H. Witten. An effective, low-cost measure of semantic relatedness obtained from Wikipedia links. In *Proc. of AAAI WIKIAI*, 2008.

[14] D. Milne and I. H. Witten. Learning to link with Wikipedia. In *Proc. of ACM CIKM*, 2008.

[15] B. Mobasher. Data mining for web personalization. *The adaptive web*, pages 90–135, 2007.

[16] J. Park, S.-J. Lee, S.-J. Lee, K. Kim, B.-S. Chung, and Y.-K. Lee. Online video recommendation through tag-cloud aggregation. *IEEE Multimedia*, 18(1), 2011.

[17] S. Schelter and S. Owen. Collaborative filtering with Apache Mahout. In *Proc. of ACM RecSys Challenge*, 2012.

[18] R. R. Sinha and K. Swearingen. Comparing recommendations made by online systems and friends. In *Proc. of Delos-NSF WPRSDL*, 2001.

[19] B. Yang, T. Mei, X.-S. Hua, L. Yang, S.-Q. Yang, and M. Li. Online video recommendation based on multimodal fusion and relevance feedback. In *Proc. of ACM CIVR*, 2007.

[20] S.-H. Yang, B. Long, A. Smola, N. Sadagopan, Z. Zheng, and H. Zha. Like like alike: joint friendship and interest propagation in social networks. In *Proc. of WWW*, 2011.

[21] R. Zhou, S. Khemmarat, and L. Gao. The impact of YouTube recommendation system on video views. In *Proc. of ACM SIGCOMM IMC*, 2010.