Editorial

Information Retrieval and Social Media Mining

María N. Moreno-García

Department of Computer Science and Automation, University of Salamanca, 37008 Salamanca, Spain; mmg@usal.es

The large amount of digital content available through web sites, social networks, streaming services, and other distribution media, allows more and more people to access virtually unlimited sources of information, products, and services. This enormous availability makes it very difficult for users to find what they are really interested in. Hence, the great current interest in developing personalized methods of information retrieval as well as reliable recommendation algorithms that help users to filter and discover what fits their preferences.

Social networks are a big source of data, from which valuable information can be extracted by means of datamining algorithms. Social media mining allows us to explore a wide range of aspects regarding users, communities, networks structures, information diffusion and so on, to be further exploited in multiple domains.

This Special Issue includes important contributions to the field of information retrieval and social media mining. Specifically, the articles published focus on three areas of research of great interest at the present: recommender systems, social media analysis, and sentiment analysis.

Collaborative Filtering (CF) is the approach most extensively used in recommender systems. It requires either explicit or implicit user ratings for items to be recommended. Then, recommendations provided to a user are based on the ratings of other users with similar preferences. Usually, each item is valued globally with a single rating; however, there are application domains in which different aspects of the items are rated. In these cases, multi-criteria recommendation models are required. Among them, one of the most recent and successful proposals is the utility-based multi-criteria recommendation approach, in which different utility functions can be used to model the value of an item from the perspective of a user. In this issue, an improvement of these models is presented in a proposal [1] that addresses user over-/under-expectations on items through penalty-enhanced models. These involve penalties in the range of $[-1, 1]$ for over-expectations and under-expectations that are added to the utility score and are learned in conjunction with expectations in the same optimization process used to generate the top-N recommendations by maximizing the normalized discounted cumulative gain.

Sometimes, collaborative filtering methods are combined with content-based approaches to solve some problems of the former and obtain more reliable recommendations. This combination is used in a cascade hybrid proposal for document recommendation presented in this issue [2]. A content-based method that makes use of document processing techniques and document metadata is applied first to provide an initial list of recommendations. It also uses a function that involves term frequency (*tf*) and inverse document frequency (*idf*) weights for document ranking. In a second step, collaborative filtering is used to re-rank the previous list.

Research on recommender systems also benefits from the intensive work currently being done in the field of deep-learning algorithms. Deep neural networks are being used to overcome some problems associated with matrix factorization methods since they are able to better represent complex relations between users and items. However, their use is justified if the complexity of the problem or the number of instances of the training set is high. This is the scenario of a paper in this Special Issue [3], in which a graph convolutional network (GCN) algorithm called PharmaSage is proposed for

providing pharmacy product cross-selling recommendations based on product feature information and sales data. The model was trained with a huge amount of real pharmaceutical data including almost a million products with complex properties and approximately 100 million sales transactions. This information is represented in a graph where each node represents a unique pharmacy product which also contains a vector encoding its descriptive data. Cross-selling for each pair of products is represented by undirected weighted edges between nodes. The GCN algorithm learns product embeddings by convolutions on aggregate neighborhood vectors. Finally, cosine similarity is applied to the output vectors to obtain recommendation scores.

Recommender systems are also one of the areas in which social data can be exploited to improve the reliability of recommendations. The incorporation of social functionalities in the recommender platforms has allowed their use in this domain. In [4], the concepts of trust and homophily derived from social structure are used to deal with the neighborhood bias of some CF recommendation methods which limits the number of items that can be recommended. Trust is derived from the friendship connections and is used to determine the degree of influence between users. Homophily is inferred from structural equivalence. This is a property often used to identify implicit communities in social networks. This is a way to capture the homophily concept since users belonging to the same community usually share interests and preferences. The similarities between users based on trust and homophily are used to extend the neighborhood of the active user and thus increase the number of potentially recommendable items.

Social media analysis is the focus of two articles in the Special Issue. One of them [5] presents a method for detecting significant events in social networks that can positively or negatively affect users. The changes in the user's followership network are used for event detection and are the base of a further analysis of the network dynamics. It is considered that an event for a given user takes place if the user experiences a follow burst or an unfollow burst in a time interval. To detect bursts, new follow/unfollow events are modeled as independent time series. Then, a time function representing the difference between the actual new follows/unfollows and the expected value for a given time is computed. A Personal Important Event (PIE) happens when the value of the function is higher than a threshold. The work also analyzes the evolution of the networks of users' followers and how the bursts caused by PIEs impact on the evolution.

The other paper focused on social media analysis presents a study about different aspects regarding the interrelationship of social media usage and perceived individual social capital [6]. A systematic procedure was applied to identify 80 scientific publications, which were analyzed in order to assess the measurement techniques used for evaluating social capital. Two operational techniques were identified. Additionally, the individual measurements items were explored to analyze future replication possibilities, resulting in no possibility of replication in an appreciable percentage of items. In the work, some consistencies and/or heterogeneity were detected in terms of operationalization, which can be useful for future studies.

In the research domains of information retrieval and social media mining, the application of language processing approaches to analyze sentiments is gaining increasing interest. In this context, the development of word embedding techniques based on deep learning have played an important role. In fact, word embedding is involved in a contribution to this issue [7], where sentiment analysis was performed for mining and summarizing opinions taking into account the context. The proposal, focused on news opinions, allows determining the relevance based not only on the text of the opinions, but also on the content of the news and its context. Topic detection from the opinion texts was performed by applying a hierarchical agglomerative clustering algorithm and using two different techniques to compute text similarity, with word embedding resulting as the best. The next steps are classifying the sentences according to the sentiment polarity and mapping topics and sentences. Finally, summary construction was provided after topic contextualization and sentence ranking were applied to news content. The topic was obtained by measuring the semantic similarity between the vocabulary associated with the topic and the news content.

We end this editorial by discussing another work that also addresses sentiment analysis [8]. In this case, the targets were questionnaire responses in telemonitoring programs to assist telemedicine patients. The aim was to monitor the adherence of patients to these programs from the sentiment polarity of their responses. The work presents the complete architecture of the system and also includes the collection and management of questionnaires. In addition, a new approach is introduced in the sentiment analysis that allows the monitoring of changes in patient's opinion across time through the repeated administration of a questionnaire. This is achieved by obtaining the polarity as a numerical value and modelling its sequence as a time series.

Conflicts of Interest: The author declares no conflict of interest. The funders had no role in the writing of the manuscript.

References

1. Zheng, Y. Penalty-Enhanced Utility-Based Multi-Criteria Recommendations. *Information* **2020**, *11*, 551. [CrossRef]
2. Borovič, M.; Ferme, M.; Brezovnik, J.; Majninger, S.; Kac, K.; Ojsteršek, M. Document Recommendations and Feedback Collection Analysis within the Slovenian Open-Access Infrastructure. *Information* **2020**, *11*, 497. [CrossRef]
3. Hell, F.; Taha, Y.; Hinz, G.; Heibei, S.; Müller, H.; Knoll, A. Graph Convolutional Neural Network for a Pharmacy Cross-Selling Recommender System. *Information* **2020**, *11*, 525. [CrossRef]
4. Sánchez-Moreno, D.; López, V.F.; Muñoz, M.D.; Sánchez, A.L.; Moreno, M.N. Exploiting the user social context to address neighborhood bias in collaborative filtering music recommender systems. *Information* **2020**, *11*, 439. [CrossRef]
5. Tang, T.; Hu, G. Detecting and Tracking Significant Events for Individuals on Twitter by Monitoring the Evolution of Twitter Followership Networks. *Information* **2020**, *11*, 450. [CrossRef]
6. Poecze, F.; Strauss, C. Social Capital on Social Media—Concepts, Measurement Techniques and Trends in Operationalization. *Information* **2020**, *11*, 515. [CrossRef]
7. Ramón-Hernández, A.; Simón-Cuevas, A.; García, M.M.; Arco, L.; Serrano-Guerrero, J. Towards Context-Aware Opinion Summarization for Monitoring Social Impact of News. *Information* **2020**, *11*, 535. [CrossRef]
8. Zucco, C.; Paglia, C.; Graziano, S.; Bella, S.; Cannataro, M. Sentiment Analysis and Text Mining of Questionnaires to Support Telemonitoring Programs. *Information* **2020**, *11*, 550. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.