

Report on the Workshop on Social Media for Personalization And Search (SoMePeAS)

Ludovico Boratto
Eurecat
ludovico.boratto@acm.org

Andreas Kaltenbrunner
Eurecat
kaltenbrunner@gmail.com

Giovanni Stilo
Sapienza Università di Roma
stilo@di.uniroma1.it

Abstract

The 1st Workshop on Social Media for Personalization And Search was held on April 9, 2017 in conjunction with the 39th European Conference on Information Retrieval (ECIR 2017). The scientific program included paper presentations, posters, and a final discussion. The keynote was delivered by Dr. Fabrizio Silvestri, who provided an overview of how recent advances in the use of embeddings can be effectively employed in advertisement and recommendation.

1 Introduction

On April 9, 2017, the 1st Workshop on Social Media for Personalization And Search was held in Aberdeen, in conjunction with the 39th European Conference on Information Retrieval (ECIR 2017). The workshop was organized by the Digital Humanities unit at Eurecat (Spain) and by the Department of Computer Science at the Sapienza University of Rome (Italy). The aim of the workshop was to collect novel ideas in the use of social media to mine user behavior for personalization and search technologies, to provide a common ground for researchers working in this area.

The workshop had more than 30 participants. The keynote speaker was Dr. Fabrizio Silvestri from Facebook. The scientific program included paper presentations, posters, and a final discussion. The papers covered topics that go from the use of neural feature embeddings to predict user response in real-time bidding, over doctor recommendation in a health social network, to an approach to recommend to journalists aspects of a news item that are still uncovered. The posters analyzed real-world phenomena, such as the perception of ISIS by the Dutch on Twitter, and what are the smoking cessation causes according to what the users post on social media and online forums.

The event concluded with a discussion session to briefly summarize the outcomes of the workshop.

2 Summary of the keynote talk

“On The Use of Embeddings in Search and Personalization Applications”, by Fabrizio Silvestri. The talk of the keynote speaker opened with a summary of Google’s word2vec (a technology that allows to create a vector representation of the words in a text corpus) and of the two main architectures that can be employed to use it (i.e., *skipgram* and *CBOV*). The talk then moved to three novel approaches that use the embeddings in industrial applications. The first of them aims at creating a matching between queries and ads, to allow advertisers to bid on relevant queries, which is based on the idea of semantic embeddings; the embeddings are learned using a large dataset consisting of user browsing sessions. The second application aims at performing query rewriting, based on a query embedding algorithm, which jointly models query content as well as its context within a search session; this allows semantically similar queries to be mapped into vectors close in the embedding space, and to perform the query expansion as a k-nearest neighbor search. The final application aims at predicting what is the next mobile app that a user is going to open; the objective is achieved by first defining a set of basic and session features that are designed to capture the sequential correlation between different actions involving mobile apps, then, based on these features, a Parallel TAN model is built to efficiently solve the app prediction problem. The take-home message of the talk was that the embeddings can be effectively employed to solve different types of problems, but they should be trained on the specific application in which they are going to be used.

3 Summary of the papers

“Recommending doctors and health facilities in the HealthNet Social Network”, by Fedelucio Narducci, Cataldo Musto, Marco Polignano, Marco de Gemmis, Pasquale Lops, and Giovanni Semeraro. This work presented *HealthNet* (HN), a social network that helps patients meeting the best doctor for their health condition. The recommender system component suggests to a user patients similar to her, and generates suggestions about doctors and hospitals that best match her patient profile. This is achieved thanks to a semantic matching, able to compute a similarity also between patients which do not share the exact same condition. A preliminary study aimed at obtaining comments and suggestions on an alpha version of HN from three focus groups of patients, practitioners, and health organizations. The outcome of this preliminary analysis showed that all the groups were very enthusiastic about HN.

“Neural Feature Embedding for User Response Prediction in Real-Time Bidding (RTB)”, by Enno Shioji and Masayuki Arai. In the paper, the authors apply a novel neural feature embedding technique to predict the probability of user response prediction in Real-Time Bidding (RTB), using a real-world dataset consisting of ad impressions. A large amount of user web history is used to learn high quality feature representations, which are then used to predict (rare) user responses. A comparison with a commonly used model in the industry showed that the technique can outperform the baseline method, especially when labeled data is scarce and when the feature sparsity problem is most acute.

“Capturing users’ information and communication needs for the press officers”, by Giovanni Stilo, Christian Morbidoni, Alessandro Cucchiarelli, and Paola Velardi. This paper presented a methodology for a recommender system able to suggest to a journalist, for a given event, the aspects still uncovered in news articles in which the readers’ interest focuses. The approach characterizes an event according to the echo it had in online news sources and associates it with the corresponding readers’ communicative and informative patterns, detected through the analysis of Twitter and Wikipedia, respectively. The methodology finally aligns the results of this analysis from the temporal perspective and identifies as recommendations the concepts that emerge as topics of interest on Twitter and Wikipedia, and are not covered in the published news articles.

4 Summary of the posters

“ISIS in the Eyes of Dutch”, by Bas Hendrikse, Mena B. Habib, and Maurice van Keulen. The first poster aimed at analyzing the perception of the Dutch about ISIS on Twitter. The authors employed Text classification, Topic Modeling, and visualization tools to filter the tweets, extract ten topics about ISIS on which Dutch people tweeted about, and display the most used words in these tweets thanks to a word cloud. Results showed that the main topics people tweeted about were: Dutch politics, ISIS in the media, attacks in Iraq and Syria, ISIS and the Islam, other terrorist organizations, the name of ISIS, refugees, a demonstration for ISIS in the Hague, and support for and against ISIS from countries. Regarding the perception of the users, results show that people are not scared or angry, but mostly disappointed in the actions of ISIS. Indeed, most of the tweets are not about their expression of a certain emotion, but often just make a statement of what happened.

“Smoking Cessation Causes: Contrasting Evidence from Social Media and Online Forums”, by Farooghian and M. Oussalah. The authors employ a system that gathers stories from *iCanQuit*, an online service provided by Australia cancer institute NSW. Moreover, using the Twitter Streaming API, geolocated tweets related to smoking cessation within UK region have been collected for three months. An automated classifier has been employed to identify the four most dominant categories in Health field for each document of blogs or Twitter dataset. The automated classifier highlighted four categories: weight loss, mental health, addiction, and support group. The analysis revealed that both blogs and Twitter datasets agree that the dominant source of smoking cessation is related to weight loss (shape body appearance), or body-look, while a support group, which includes any clinician support, has little impact on the smokers’ quit motivation. A result that may be precious for the health authorities.

5 Conclusions

Overall, the 1st Workshop on Social Media for Personalization And Search was a success, both in terms of number of participants and of interests that emerged during the presentations and the final discussion, creating new relationships and novel ideas in this area. Tentative plans to organize the 2nd workshop next year were formed, along with plans to have a special issue in a journal on these topics, which will soon be organized.