Information Processing and Management 51 (2015) 695–717

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

An investigation on the serendipity problem in recommender systems

Marco de Gemmis *, Pasquale Lops, Giovanni Semeraro, Cataldo Musto

Department of Computer Science, University of Bari Aldo Moro, Via E. Orabona 4, I-70125 Bari, Italy

ARTICLE INFO

Article history:
Received 3 August 2014
Received in revised form 9 June 2015
Accepted 15 June 2015
Available online 29 June 2015

Keywords:
Recommender systems
Serendipity problem
Knowledge representation
Spreading activation
Affective feedback
Facial expressions

ABSTRACT

Recommender systems are filters which suggest items or information that might be interesting to users. These systems analyze the past behavior of a user, build her profile that stores information about her interests, and exploit that profile to find potentially interesting items. The main limitation of this approach is that it may provide accurate but likely obvious suggestions, since recommended items are similar to those the user already knows. In this paper we investigate this issue, known as *overspecialization* or *serendipity problem*, by proposing a strategy that fosters the suggestion of surprisingly interesting items the user might not have otherwise discovered.

The proposed strategy enriches a graph-based recommendation algorithm with background knowledge that allows the system to deeply understand the items it deals with. The hypothesis is that the infused knowledge could help to discover hidden correlations among items that go beyond simple feature similarity and therefore promote non-obvious suggestions. Two evaluations are performed to validate this hypothesis: an in vitro experiment on a subset of the HETREC2011-MOVIELENS-2K dataset, and a preliminary user study. Those evaluations show that the proposed strategy actually promotes non-obvious suggestions, by narrowing the accuracy loss.

1. The filter bubble and the serendipity problem

In the book "*The Filter Bubble: What the Internet Is Hiding from You*", Eli Pariser argues that Internet is limiting our horizons (Parisier, 2011). He worries that personalized filters, such as Google search or Facebook delivery of news from our friends, create individual universes of information for each of us, in which we are fed only with information we are familiar with and that confirms our beliefs. These filters are opaque, that is to say, we do not know what is being hidden from us, and may be dangerous because they threaten to deprive us from *serendipitous* encounters that spark creativity, innovation, and the democratic exchange of ideas. Similar observations have been previously made by Gori and Witten (2005) and extensively developed in their book "Web Dragons, Inside the Myths of Search Engine Technology" (Witten, Gori, & Numerico, 2006), where the metaphor of search engines as modern *dragons* or gatekeepers of a treasure is justified by the fact that "the immense treasure they guard is society's repository of knowledge" and all of us accept dragons as mediators when having access to that treasure. But most of us do not know how those dragons work, and all of us (probably the search engines' creators, either) are not able to explain the reason why a specific web page ranked first when we issued a query. This gives

* Corresponding author.
E-mail addresses: marco.degemmis@uniba.it (M. de Gemmis), pasquale.lops@uniba.it (P. Lops), giovanni.semeraro@uniba.it (G. Semeraro), cataldo.musto@uniba.it (C. Musto).

http://dx.doi.org/10.1016/j.ipm.2015.06.008
0306-4573/

Fig. 1. IMDb suggestions for the movie *Star Trek into Darkness*.

rise to the so called *bubble of Web visibility*, where people who want to promote visibility of a Web site fight against heuristics adopted by most popular search engines, whose details and biases are closely guarded trade secrets.

Also recommender systems, which suggest to users items or information they might be interested in (Ricci et al., 2011), give their contribution to the filter bubble (Kamishima, Akaho, Asoh, & Sakuma, 2012; Resnick, Konstan, & Jameson, 2011). These systems analyze a user's past behavior, maybe find others who have a similar history, and use that information to provide suggestions. For example, if you tell the Internet Movie Database (IMDb)[1] that you like the movie *Star Trek into Darkness*, it will suggest movies liked by other people who liked that movie ("People who liked this also liked…" in Fig. 1), most of whom are probably science-fiction fans. Furthermore, one of those recommendations is a movie of the same saga, which is likely to be already known to the user. The user will be provided with items within her existing range of interests and her tendency towards a certain behavior is reinforced by creating a self-referential loop. This drawback is known as *overspecialization* or *serendipity problem* (McNee, Riedl, & Konstan, 2006), and stems from the fact that the goal of the system is to find items that best match the model of user preferences in order to improve accuracy, regardless of the actual usefulness of the suggestions. The importance of taking into account factors, other than accuracy, which contribute to the perceived quality of recommendations is emphasized in recent research (Castells, Wang, Lara, & Zhang, 2011; Hurley & Zhang, 2011; Zhang, Séaghdha, Quercia, & Jambor, 2012). One of these factors is serendipity, that can be seen as the experience of receiving unexpected suggestions helping the user to find surprisingly interesting items she might not have otherwise discovered, or that would have been really hard to discover (Herlocker, Konstan, Terveen, & Riedl, 2004). Serendipity has been recognized as a goal that often conflicts with

[1] www.imdb.com.