A UNIFYING VIEW OF CONTEXTUAL ADVERTISING AND RECOMMENDER SYSTEMS

Giuliano Armano and Eloisa Vargiu
Dept. of Electrical and Electronic Engineering, University of Cagliari, Cagliari, Italy

Keywords: Contextual advertising, Recommender systems, Information Retrieval.

Abstract: From a general perspective, nothing prevents from viewing contextual advertising as a kind of Web recommendation, aimed at embedding into a Web page the most relevant textual ads available for it. In fact, the task of suggesting an advertising is a particular case of recommending an item (the advertising) to a user (the web page), and vice versa. We envision that bringing ideas from contextual advertising could help in building novel recommender systems with improved performance, and vice versa. To this end, in this paper, we propose a unifying view of contextual advertising and recommender systems. To this end, we suggest: *(i)* a way to build a recommender system inspired by a generic solution typically adopted to solve contextual advertising tasks and *(ii)* a way to realize a collaborative contextual advertising system a la mode of collaborative filtering.

1 INTRODUCTION

Let us note in advance that, in the literature, the term "context" is referred to "keywords used in search engines" in the area of contextual advertising, and to "events which modify the user behavior" in the area of recommender systems. In this paper we always adhere to the former interpretation. Therefore, we are not interested in context-aware recommender systems as in (Adomavicius and Tuzhilin, 2008) (Abbar et al., 2009) (Ramaswamy et al., 2009).

As discussed in (Broder et al., 2007), contextual advertising is an interplay of four players: (i) the *advertiser*, which provides the supply of ads; (ii) the *publisher*, which is the owner of the web pages on which the advertising is displayed; (iii) the *ad network*, which, as mediator between advertiser and publisher, is in charge of selecting the ads to put in the pages; and (iv) *users*, which visit the web pages of the publisher and interact with the ads. Similarly, a recommendation task may be described as an interplay of four players: (i) the *recommender*, which provides the supply of items to be recommended; (ii) the *publisher*, which is the owner of the web pages on which items are displayed for recommendation; (iii) the *recommender system*, which, as a mediator between recommender and publisher, is in charge of selecting the items to be recommended to a specific user; and (iv) *users*, which visit the web pages of the publisher/recommender and interact with the suggested items.

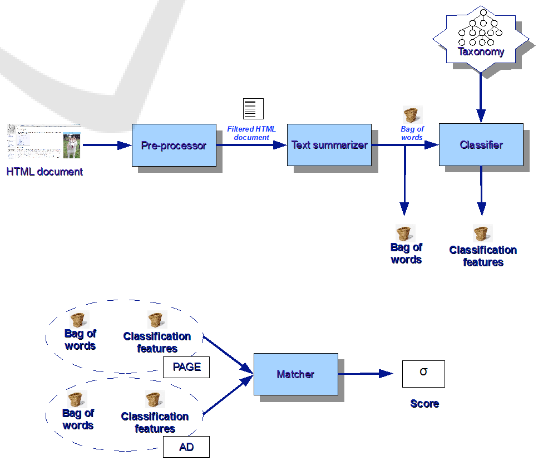

Figure 1: A contextual advertising system.

Although contextual advertising and recommender systems have been usually studied separately, they could be hypothesized as isomorphic structures, in which a task on one side corresponds to a task on the other side. For instance, the task of suggesting an advertising to a web page could be viewed as the task of recommending an item (the advertising) to a user (the web page), and vice versa.

Starting from this insight, in this paper we propose a unifying view presenting two novel approaches: a content-based recommender system devised *a la mode* of contextual advertising system and a collab-

Armano G. and Vargiu E..
A UNIFYING VIEW OF CONTEXTUAL ADVERTISING AND RECOMMENDER SYSTEMS.
DOI: 10.5220/0003097204630466
In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval* (KDIR-2010), pages 463-466
ISBN: 978-989-8425-28-7

orative contextual advertising devised *a la mode* of recommender system. To our best knowledge, this is the first attempt to combine the two approaches.

2 A RECOMMENDER SYSTEM *A LA MODE* OF CONTEXTUAL ADVERTISING

2.1 A Typical Contextual Advertising System

In our view, a generic system devoted to perform contextual advertising could be designed as depicted in Figure 1.

Pre-processor. Its main purpose is to transform an HTML document (a web page or an advertising) into an easy-to-process plain-text based document, while preserving important information. In particular, the main goal is to preserve the blocks of the original HTML document, while removing HTML tags and stop-words. Information about which phrases are part of the anchor text of the hypertext links could also be preserved.

Text summarizer. Text summarization techniques are divided into extractive and non-extractive. The input of a contextual advertiser being an HTML document, contextual advertising systems typically rely on the former. In particular, extraction-based techniques are applied to the blocks that form a web page –e.g., the title of the web page, the first paragraph, the paragraph with the highest title-word count. The text summarizer outputs a vector representation of the original HTML document, web page or advertising, in terms of bag of words (*BoW*).

Classifier. To alleviate possible harmful effects of summarization, both page excerpts and advertisings are classified according to a given taxonomy (Anagnostopoulos et al., 2007). The corresponding classification-based features (*CF*) are then used in conjunction with the original *BoW*.

Matcher. It suggests ads to the web page according to a similarity score based on both *BoW* and *CF*.

2.2 The Proposed Recommender System

As depicted in Figure 2, our proposal for building a recommender system involves two steps: user profiling and recommendation (Addis et al., 2010). Given a user, her/his profile is generated from the corresponding user history, i.e., from the set of documents s/he rated as relevant.

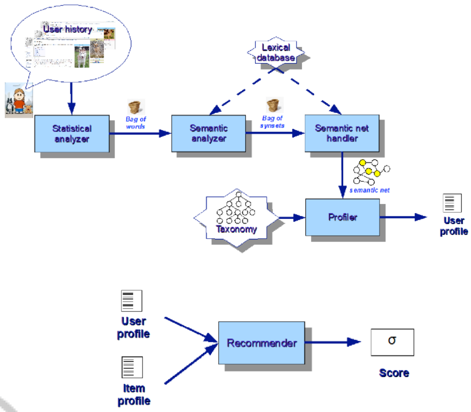

Figure 2: The proposed recommender system.

User Profiling

The user profiler is composed by four main modules: statistical document analyzer, semantic analyzer, semantic net handler, and profiler (Addis et al., 2009).

Statistical document analyzer. While analyzing documents rated as relevant by the user, this module is devoted to create the *BoW*, which collects all terms contained in the input documents, suitably weighted. The statistical document analyzer removes from the *BoW* all non-informative words such as prepositions, conjunctions, pronouns, and very common verbs using a stop-word list. Subsequently, it calculates the weight of each term adopting the TFIDF measure. The statistical document analyzer calculates an overall TFDIF considering all documents in the user history. Furthermore, the weights resulting from TFIDF undergo a cosine normalization. To reduce the dimensionality of the space, only the first N terms of the *BoW* are retained. The optimal value of N must be calculated experimentally. Hereinafter, the set of terms stored in the *BoW* will be called *features*. This module corresponds to the preprocessor and to the text summarizer adopted in the generic contextual advertising solution described previously.

Semantic words analyzer. This module creates the bag of synsets (*BoS*), which collects all synsets related to the selected features. To this end, the semantic document analyzer queries an online lexical database (e.g., WordNet (Miller, 1995)). After synset extraction, the semantic document analyzer assigns to each sysnset a weight according to the TFIDF of all related terms. This module corresponds to a text summarizer based on semantic information. In fact, a semantic approach can be also adopted in contextual advertising to improve the performances of the text summarization task.

Semantic net handler. This module aims to (i) build the semantic net from the *BoS* and (ii) extract its most relevant nodes. First, a semantic net is built in form of a graph, whose nodes are the synsets belonging to the *BoS* and whose edges are semantic relations between synsets. Four kinds of semantic relations are taken into account: *hyponymy* and its inverse (*hyperonymy*); *meronymy* and its inverse (*holonymy*). The semantic net handler is also in charge of pruning the network by dropping irrelevant nodes, identified according to their weight and to the number of connections with other nodes.

Profiler. This module is devoted to extract the user profile. To this end, it exploits a given taxonomy (e.g., WordNet Domains Hierarchy (Magnini and Cavagli, 2000)) and associates the proper category to each selected node. Considering the selected nodes, together with their weights, the profiler is able to identify the real interests of the user in terms of the given taxonomy. In particular, the user profile is represented as a set of pairs $\langle c_k, w_k \rangle$, where c_k is a category and w_k the corresponding weight in $[0,1]$. The semantic net handler and the profiler correspond to the classifier adopted in the generic contextual advertising solution previously described.

Recommendation

Once the user profile has been generated, the system can rank a new item i to determine whether it could be of interest for a specific user u. This can be done by measuring the distance between the vector-based representations of i and u, say $\vec{V}(i)$ and $\vec{V}(u)$. In particular, the textual information of an item i can be processed in a way similar to profile extraction: a set of categories of the given taxonomy with the corresponding relevance ratio are computed for the item, and the cosine distance between u and i is evaluated. Items obtaining a score greater than 0.5 are proposed to the user. It is easy to note that the recommender corresponds to the matcher adopted in the generic solution for contextual advertising, previously described.

3 A CONTEXTUAL ADVERTISING SYSTEM *A LA MODE* OF COLLABORATIVE FILTERING

3.1 A Typical Collaborative Recommender System

In our view, a generic system devoted to perform collaborative filtering could be designed as depicted in Figure 3.

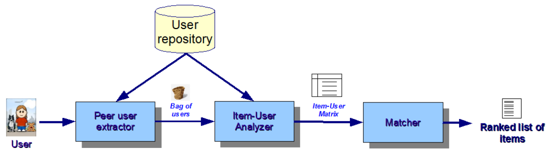

Figure 3: A collaborative recommender system.

The recommendation problem can be formulated as follows: let U be the set of all users and I be the set of all possible items that can be recommended (e.g., books, movies, and restaurants). Let f be a utility function that measures the usefulness of item i to user u, i.e., $f : U \times I \rightarrow R$, where R is a totally ordered set (e.g., non-negative integers or real numbers within a given range). Then, for each user $u \in U$, we want to choose the item $i' \in I$ that maximizes f. In recommender systems, f is typically represented by ratings and is initially defined only on the items previously rated by the users. For example, in an application for recommending books (e.g., Amazon.com), users initially rate a subset of books that they have read. This information is stored in the *user repository*, as sketched in Figure 3.

Peer user extractor. The main purpose of the peer user extractor is, given a user u, to detect her/his "peers". Peer users are other users that have similar preferences and tastes. The underlying idea is that only items that are most liked by the "peers" of user u would be recommended to her/him.

Item-user analyzer. This module is devoted to analyze the rates of the peers users and to build the corresponding user-item rating matrix, in which each row corresponds to a user (u and its peer users), each column corresponds to an item, and each cell corresponds to the rating given by that user to that item. Ratings are typically specified on the scale of 1 to 5.

Matcher. The main purpose of this module is to find, starting from the user-item rating matrix, the set of items to be recommended to u. Each corresponding score is calculated by taking into account the rates provided by users.

3.2 The Proposed Contextual Advertising System

Our idea of a collaborative contextual advertising system relies in suggesting ads to a web page p exploiting the "collaboration" of p with its peer pages. Figure 4 depicts the proposed high-level architecture.

Inlink extractor. This module is devoted to find, given a page p, the peer pages. In our opinion, suitable peer pages could be all the inlinks of p, i.e., all pages that link to p. First, this module creates the

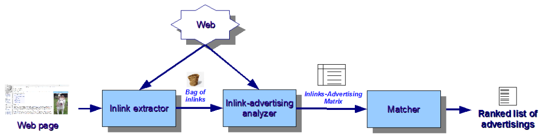

Figure 4: The proposed contextual advertising system.

bag of inlinks (*BoI*), which collects all the inlinks of a given page. To this end, Google AJAX Search API[1] or existing tools (such as Page Inlink Analyzer[2]) could be used. It is easy to note that this module corresponds to the peer user extractor previously described.

Item-advertising analyzer. First, this module parses all the extracted inlinks and, for each inlink i, extracts the corresponding list of ads weighting them according to the position in i. Then, the module builds the inlink-advertising matrix, whose generic element w_{ij} reports the weight for the inlink i and for the advertisement j. It is easy to note that this module corresponds to the peer user-item analyzer previously described.

Matcher. This module is devoted to suggest ads to the web page according to a similarity score. In principle, any similarity measure can be adopted: correlation, cosine-based, rated-based. This module corresponds to the matcher of the typical collaborative recommender system previously described.

4 CONCLUSIONS AND FUTURE DIRECTIONS

In this paper we proposed a unifying view of contextual advertising and recommender systems. To our best knowledge, this is the first attempt to combine these research fields.

As for future directions, we are currently setting up experiments to validate the content-based recommender system illustrated in Section 2. Furthermore, we are starting the implementation of the collaborative contextual advertising system illustrated in Section 3.

ACKNOWLEDGEMENTS

This work has been partially supported by Hoplo srl. We wish to thank, in particular, Ferdinando Licheri and Roberto Murgia for their useful suggestions.

[1]http://code.google.com/apis/ajaxsearch/
[2]http://ericmiraglia.com/inlink/

REFERENCES

Abbar, S., Bouzeghoub, M., and Lopez, S. (2009). Context-aware recommender systems: A service-oriented approach. In *3rd International Workshop on Personalized Access, Profile Management, and Context Awareness in Databases*.

Addis, A., Armano, G., Giuliani, A., and Vargiu, E. (2010). A recommender system based on a generic contextual advertising approach. In *Proceedings of ISCC'10: IEEE Symposium on Computers and Communications*, pages 859–861.

Addis, A., Armano, G., and Vargiu, E. (2009). Profiling users to perform contextual advertising. In *Proceedings of the 10th Workshop dagli Oggetti agli Agenti (WOA 2009)*.

Adomavicius, G. and Tuzhilin, A. (2008). Context-aware recommender systems. In *RecSys '08: Proceedings of the 2008 ACM conference on Recommender systems*, pages 335–336, New York, NY, USA. ACM.

Anagnostopoulos, A., Broder, A. Z., Gabrilovich, E., Josifovski, V., and Riedel, L. (2007). Just-in-time contextual advertising. In *CIKM '07: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 331–340, New York, NY, USA. ACM.

Broder, A., Fontoura, M., Josifovski, V., and Riedel, L. (2007). A semantic approach to contextual advertising. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 559–566, New York, NY, USA. ACM.

Magnini, B. and Cavagli, G. (2000). Integrating subject field codes into wordnet. In *Gavrilidou M., Crayannis G., Markantonatu S., Piperidis S. and Stainhaouer G. (Eds.)* Proceedings of LREC-2000, Second International Conference on Language Resources and Evaluation, pages 1413–1418.

Miller, G. A. (1995). Wordnet: A lexical database for english. *Commun. ACM*, 38(11):39–41.

Ramaswamy, L., Deepak, P., Polavarapu, R., Gunasekera, K., Garg, D., Visweswariah, K., and Kalyanaraman, S. (2009). Caesar: A context-aware, social recommender system for low-end mobile devices. *Mobile Data Management, IEEE International Conference on*, 0:338–347.