Learning Relevance of Web Resources across Domains to make Recommendations

Julia Hoxha*
Karlsruhe Institute of Technology
Karlsruhe, Germany
Email: julia.hoxha@kit.edu

Peter Mika
Yahoo! Research
Diagonal 177, Barcelona, Spain
Email: pmika@yahoo-inc.com

Roi Blanco
Yahoo! Research
Diagonal 177, Barcelona, Spain
Email: roi@yahoo-inc.com

***Abstract*—Most traditional recommender systems focus on the objective of improving the accuracy of recommendations in a single domain. However, preferences of users may extend over multiple domains, especially in the Web where users often have browsing preferences that span across different sites, while being unaware of relevant resources on other sites.**

This work tackles the problem of recommending resources from various domains by exploiting the semantic content of these resources in combination with patterns of user browsing behavior. We overcome the lack of overlaps between domains by deriving connections based on the explored semantic content of Web resources. We present an approach that applies Support Vector Machines for learning the relevance of resources and predicting which ones are the most relevant to recommend to a user, given that the user is currently viewing a certain page.

In real-world datasets of semantically-enriched logs of user browsing behavior at multiple Web sites, we study the impact of structure in generating accurate recommendations and conduct experiments that demonstrate the effectiveness of our approach.

Keywords—cross-domain recommendations; hybrid semantic recommender; semantic logs; support vector machines;

I. Introduction

Recommender systems typically focus on items in a single domain, for example suggesting related books to a user who is currently viewing information about a book, or a list of movies when she is visiting a webpage about a film. The main objective in this context has been the improvement of the relevance of recommendations in one domain. Recently, there is a growing awareness [4], [3], [9] that users do not have a single interest and their needs span across different application areas. An emerging trend is the development of solutions that offer cross-domain recommendations.

In the Web, this task is more challenging than the traditional single domain recommender setup because there is a need to link items (resources in Web parlance) across Web sites. As such, we can expose visitors to novel resources by recommending pages belonging to various domains in terms of the type of information contained in the page.

Despite an increasing acknowledgment of the necessity for cross-domain recommendations, this research field is still new. A significant challenge in building such recommender systems is the small overlap between the users and items of different domains. In the Web context, the challenge is to understand the type of item(s) presented in a Web page and how they relate to items presented in other pages. In other words, there is a need for *semantic* approaches that can extract the *type* of items presented in a Web page and apply knowledge (an ontology) about the relations of items and item types.

To our knowledge, this work is the first to address the problem of generating recommendations of Web resources across domains using semantic information extracted from Web pages. We exploit ontological information to find relations between Web resources, establishing in this way bridges among domains. We present a new model for learning the relevance of resources and predicting which ones are the most relevant to recommend to a user, given that she is currently viewing a certain page. We evaluate our approach using real world usage data from navigation logs collected with a browser toolbar. We show that our method successfully exploits the semantics of resources and achieves performance beyond other classical recommender systems.

The contributions of this work are:
- a semantic recommendation approach, which exploits in novel ways the semantic structures of Web pages and their combination with usage patterns inherent in browsing logs.
- first real-world study on (1) leveraging user behavior at browsing *multiple* Web sites with structured markup data, and (2) the impact of structure in predicting relevant recommendations across various domains for Web users.

Outline. We present related work in Section II, then introduce in Section III the cross-domain recommendation framework. A relevance prediction approach is presented in Section IV. Section V presents the experiments done to evaluate our approach. We then draw conclusions in Section VI.

II. Related Work

Several approaches have been introduced to link semantic technologies and recommender systems. Some works [1], [2] uses semantic-based knowledge representations to describe user profiles, in order to make enhanced and understandable recommendations.The works in [7], [12] present a combination of collaborative and semantic features to generate recommendations. Similar to our approach, they combine the preferences of users implicit in their navigation through a sequence of Web pages with the semantic content of the pages. It is

*Work conducted during research stay at Yahoo! Research Barcelona.

demonstrated that an integrated approach yields significant improvements in terms of increasing accuracy of recommendations. Yet, in terms of semantic information, [7] considers only the attributes of the objects without addressing the structure of the data or taxonomy, and [12] uses a flat (matrix-based) representation of the semantic data, without capturing structure in the semantic similarity measures.

All these works provide recommendations in a single domain. Very recently, a few initiatives [4], [9] have emerged that investigate the task of cross-domain recommendations. Cremonesi et al. [3] offers a survey of these works. Ongoing work of Fernández-Tobías et al. [4] recently introduced a generic framework that, using DBpedia as basis, integrates knowledge from several domains to provide cross-domain recommendations. The framework shows the added-value of using the semantic information of items to link concepts from two domains. Yet, this work does not exploit the impact of usage-based features or the dynamics of past user behavior in determining items relevance. Another drawback is that an expert has to identify manually the semantic entities and relations of DBpedia, which can then be used to describe and link the domains of interest.

The majority of cross-domain recommendation approaches deal with collaborative filtering, missing the content-based features [5]. Moreover, there are no works that combine these two aspects, while also exploiting the structural representation of the content. A related approach is proposed by Loizou [9], which also uses a graph structure to represent relations between domains. In terms of semantics, the approach limits the mapping of an item to a page in Wikipedia, if it exists, otherwise free-form tagging is considered. Hence, the strategy fails to capture the full-fledged structure behind the items' content.

In our approach, we specifically cover this gap and exploit the impact of *semantic structures* in generating recommendations *across domains* in an open setting, without relying on a central knowledge base. We consider both usage-based and content-based features from structured data, showing how their combination improves the recommender results.

III. Cross-domain Recommendations

In a cross-site browsing scenario, our goal is to recommend to users the top-N pages that they might visit next. We consider a recommendation setting in which user and item profiles are distributed in multiple systems (domains). We adhere to the definition [17] of a domain as the set of similar items with the same characteristics that can be easily differentiated, e.g. movies, concerts, songs, news, artists, etc. Also, since the term "'category"' or "'type"' is sometimes used, we will use them interchangeably with "'domain"'.

Our goal is to suggest to a user resources of different types from across the Web, not just the website that the she is currently browsing. This scenario is illustrated in Figure 1: a user has previously browsed the website upcoming.yahoo.com, looking at information about the *Shaky Knees* festival in Atlanta, GA and two of its performers, the *Lumineers* and *The Antlers*. Subsequently, a user visits a page related to the same festival on the website eventful.com. Based on the history of the previous user, we might want to recommend the potentially relevant pages about the artists (different type from festival) in eventful.com or at upcoming.yahoo.com. Our recommendations need to be relevant, yet from various domains, in order to allow the user to discover new sources of information. We aim to develop a global recommender that suggests related items based on the current item the user is viewing and aggregate past behavior of users.

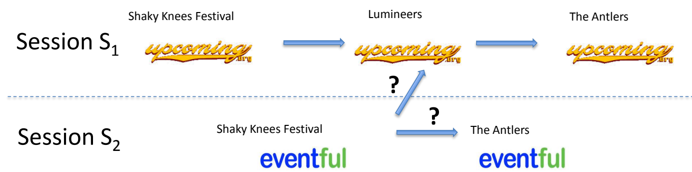

Fig. 1. Cross-domain recommendation task

In the following, we give the definition of our cross-domain recommendation task. Without loss of generality, we define the task when two domains are involved. Using the notation of [3], [5], let $\mathcal{U}_\mathcal{A}$, $\mathcal{U}_\mathcal{B}$ be the sets of users and $\mathcal{I}_\mathcal{A}$, $\mathcal{I}_\mathcal{B}$ be the sets of items with characteristics (user preferences and item attributes) in the domains $\mathcal{A}$ and $\mathcal{B}$, respectively. Our recommendation tasks is to *make joint recommendations* for items belonging to different domains, i.e., suggesting items in $\mathcal{I}_\mathcal{A} \cup \mathcal{I}_\mathcal{B}$ to users in $\mathcal{U}_\mathcal{A} \cup \mathcal{U}_\mathcal{B}$. There are various types of overlaps between domains that have been identified [3]. Our task is conducted in the setting where we might have user overlap among domains (i.e. a user browsing various domains), but no item overlaps (i.e. each domain has its own resources).

Since user and item profiles are distributed in multiple systems (domains), we have to establish a mechanism to link between such systems. To address the non-overlap situation, we have developed an approach that enables us to explore the content of Web pages and find semantic structures of resources across various domains. We use this information to build content-based relations between Web resources, which would be used as semantic bridges connecting different domains.

A. Semantic Enrichment

A crucial and novel element of our approach is that we exploit semantic information embedded in the content of the web pages, and combine this knowledge with the usage patterns inherent in the user browsing logs. Nowadays, it is increasingly common for web publishers to add metadata to the HTML content of their pages [11]. This enables search engines, Web crawlers, and browsers to extract, automatically-process and identify the type of entities and their attributes described in Web pages. Structured markup may be provided in different formats (e.g. RDFa, microdata, microformats) though most have a straight-forward translation to the RDF data model, the lingua franca of the Semantic Web. The annotations in Web pages may use different vocabularies supported by various consumers, for example the Open Graph Protocol,[1] supported by Facebook or schema.org markup [2] that is currently understood by both Bing, Google, Yahoo and Yandex.

The enrichment of usage logs with metadata is a complex process and is summarized here for brevity (see [6] for details).

[1]http://ogp.me/

[2]http://schema.org/

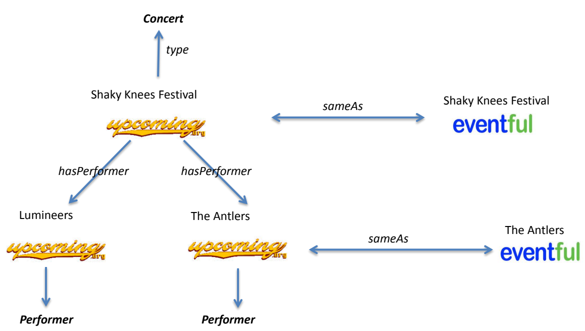

Fig. 2. Example of the Ontological Knowledge

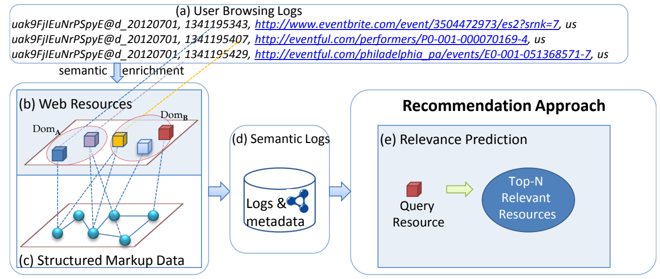

Fig. 3. Recommendation Approach

Our method is open-domain in that it does not exploit domain-specific heuristics. The process starts with the extraction of user browsing logs (tracked by a client toolbar) and segmenting logs in sessions, such that one session contains all requests of one user within a day. We then filter those sessions that include pages belonging to several sites of interest. For our experiments, we chose a set of sites[3] of events advertisement nature (concerts, conferences, etc.).

The next steps consist in identifying the set of unique pages in the filtered user logs and deploying metadata extraction and metadata analysis techniques. Different Web sites use different schemas to annotate their HTML elements. Hence, we semi-automatically align the concepts and relations among the schemas based on their respective semantics, in order to enable matching resources across different sites. The result is a reference ontology $\mathcal{O}$ with concepts and their semantic relations used for the annotation of resources across all sites.

Semantic Bridges. The resources are classified into different types, i.e. *classes* of the ontology (e.g. *Performer*, *Venue*) and are connected between each-other via semantic *relations* (e.g. *hasPerformer*, *hasVideo*). Figure 2 illustrates how the example user log in Figure 1 is enriched with ontological knowledge, for example that *Shaky Knees* is a festival related to two performers, and that the respective performer pages on Eventful and Upcoming represent the same real-world entities. We perform entity linking in order to identify pages that belong to different Web sites, but still semantically represent the same object in the real world (e.g. same performer, venue, etc.). This is modeled in $\mathcal{O}$ with the *sameAs* relations between resources of different sites. Entity linking is achieved by aligning the resources based on their semantic attributes and relations, then using Levenshtein distance for comparison.

In the following section, we describe how the captured ontological information is exploited to learn relevance between resources of various domains.

IV. Relevance Model

We consider sessions of user browsing logs, where a session S consists of a set of events e_j each representing a visit to a resource $r_j \in R'$ by a given user at a given time. A resource r_j may be linked to a resource in the ontology O. In the first step, we find a set $K \subset R'$ of resources that are *relevant* for a user to visit next, given r_i the current resource the user is viewing. The second step is to generate a set of recommendations $R \subseteq K$, s.t. $|R| = N$ (Fig. 3).

[3]eventful.com, eventbrite.com and upcoming.yahoo.com

Initially, we define resource *pair relevance* and *set relevance*, then propose an approach to estimate their values. We describe how we use the relevance values to generate inital recommendation set K.

Pair Relevance. Given two Web resources r_i and r_j, let pair relevance $P(rel|r_i, r_j)$ denote a value that captures the relevance of these resources to each-other. We give a probabilistic interpretation to the relevance values: they approximate the *likelihood* of r_j satisfying the user intent given the query resource r_i. In our case, $P(rel|r_i, r_j)$ is determined by a scoring function based on user access patterns and content of resources r_i and r_j. Pair relevance is an item-item similarity measure, thus the order in the pair is not important.

Set Relevance. We define the relevance of a set of recommendations as:

$$Rel(K|r_i) = 1 - \prod_{j \in K} (1 - P(rel|r_i, r_j)) \quad (1)$$

based on an *independence* assumption: given a query resource r_i, the conditional probabilities of two other resources satisfying the user are independent. The probability that the user will find none of two resources r_j and r_k relevant equals $(1 - P(rel|r_i, r_j))(1 - P(rel|r_i, r_k))$, s.t. $(1 - P(rel|r_i, r_j))$ is the probability that r_j fails to satisfy. The probability that the set K will all fail to satisfy equals its product, by the independence assumption. One minus that product equals the probability that some resource in the set will satisfy the user.

A. Relevance Learning and Prediction

We formulate the problem of estimating resource pair relevance $P(rel|r_i, r_j)$ as a binary classification task.

Learning Pair Relevance. We apply Support Vector Machines (SVMs), which are a established machine learning techniques for discriminative classification of high-dimensional sparse data. The task is to learn a decision function $f : \mathcal{R}^d \rightarrow Y$ based on an i.i.d training sample D_{train}=$\{(\mathbf{x_1}, y_1), (\mathbf{x_2}, y_2), ..., (\mathbf{x_n}, y_n)\}$, where each training example consists of a feature vector $\mathbf{x} \in \mathcal{R}^d$ and an output label $y \in \{-1, 1\}$. The learned function f is then used to predict the output label $sign(f(\mathbf{x_k}))$ for test example $\mathbf{x_k}$.

Our original input data consist of a set X of resource pairs. Each pair $x = \langle r_i, r_j \rangle \in X$ is initially mapped to a d-dimensional feature vector $\mathbf{x}$ via a function $\psi : X \rightarrow \mathcal{R}^d$. The output labels in $Y = \{-1, 1\}$ denote in our case the two classes:*non-relevant* and *relevant* resources in the pair.

Probability Estimates. For our relevance predictions, we are not just interested on hard decisions (labels), but rather the probability $P(rel|r_i, r_j)$ (Eq. 1). We formulate it as an estimate of the confidence in the correctness of the predicted label. It is defined as the class conditional posterior probability $P(y|\mathbf{x}) = P(y|\psi(x))$, i.e. the probability with which the feature vector $\mathbf{x}$ of pair $x = \langle r_i, r_j \rangle$ belongs to class y. We deploy Support Vector Machines (SVMs) as probabilistic models by calibrating the scores into an accurate class conditional posterior probability with the sigmoid function [13]:

$$P(rel|r_i, r_j) = P(y = 1|\psi(x = \langle r_i, r_j \rangle)) = \frac{1}{1 + exp^{(Af(x)+B)}} \tag{2}$$

fitted to the decision values of f, with parameters A and B estimated by minimizing the negative log likelihood of training data (using their labels and decision values) [8].

Predicting Relevant Resources. The approach allows us to learn a model, which we use to predict set K of resources that are relevant to a query resource r_i (i.e. have a pair relevance value above 0). We first derive pairs of resource r_i with other resources in the corpus, then apply the model learned with the SVMs to estimate the relevance $P(rel|r_i, r_j)$ for each pair.

Generating top-N Recommendations. We select the top-N resources from set K with the highest relevance to r_i, ordering by the predicted pair relevance values. This set composes the user recommendations for further Web navigation.

A crucial part of the prediction method is to define for the resource pairs the features (Sec. IV-B) that are effective in predicting an accurate relevance value. The novelty of this work is the introduction of features that exploit the semantic information of resources, in order to overcome the problem of lacking overlaps between domains.

B. Features

We engineer two groups of features: (1) content-based features that use the content of the resources, and (2) usage-based features, which exploit the information contained in the user logs. Some features especially captures the semantics of information in the ontology.

SEMANTICSIMILARITY: a measure estimated via a set spreading approach [15] using the structural information related to the Web resources. Our spreading approach (Fig. 4) appends to a resource description terms that are related to the original terms based on an ontology. This is the ontology $\mathcal{O}$ constructed in our semantic enrichment approach (Sec. III-A). The process starts with an initial set $RD_k = \{\langle t_k, w_k \rangle\}$

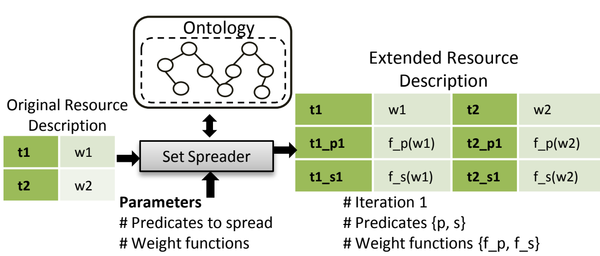

Fig. 4. Set Spreading Approach

for each resource description, where t is the semantic type (e.g. $Performer$, $Event$, etc.) of this resource k denoted by a concept in the ontology $\mathcal{O}$, and weight w_k denotes the importance of the concept term in describing the resource. Each resource description RD_k is then iteratively extended via spreading, utilizing the concepts and relations (predicates) in $\mathcal{O}$. The set spreading of an RD_k results in the resource description $RD'_k = \{\langle t_k, w_k \rangle, \langle t_k\_p, f\_p(w_k) \rangle\}$, which is extended by the term t_k_p related to t_k in $\mathcal{O}$ by the predicate p. We pre-defined a set of predicates to find, at each iteration, $related$ terms of the previous RDs. We use a simple function $f\_p(w_k) = 0.75 w_k$ to estimate weights, reducing them at each iteration.[4] The spreading process is terminated by predicates exhaustion. The final similarity of two resources is then the mean cosine similarity of their descriptions RD_i.

SHARETYPE: a binary value indicating if the pair (r_i,r_j) of resources have the same type (concept of the ontology $\mathcal{O}$).

SHARERELATION: a binary value indicating if resources of pair (r_i,r_j) share a relation (in the ontology $\mathcal{O}$) between them e.g. one resource is the event and the other resource is its venue, therefore sharing the relation *hasvenue*.

SYNTACTICSIMILARITY: the term vector similarity measure between any two Web resources, computed as the cosine angle between two vectors modeled out of the *bag-of-words* (BOW) representation of the HTML page of each resource:

$$sim_{syntactic}(r_i, r_j) = \frac{\mathbf{V}(r_i) \cdot \mathbf{V}(r_j)}{|\mathbf{V}(r_i)|\ |\mathbf{V}(r_j)|} \tag{3}$$

$\mathbf{V}(i)$ is a real-valued vector composed of the weights of terms found in the HTML content of resource r_i. The weights are computed using the TF-IDF weighting scheme [10]. As such, this feature entails only syntactic information.

We also define a group of session-based features, which are computed based on the user logs. As described earlier, we consider the implicit preference judgments of Web users captured in their interaction with the system (*clicks*). Since no explicit ratings are available, it is challenging to translate these preferences into measurable usage patterns in order to provide a collaborative-based filtering approach.

OBSERVEDRELEVANCEDEGREE: this measure captures observations from usage patterns in the sessions of browsing logs. We model the correspondence between resource usage counts and user interest as a heuristic mapping between the access patterns and the probability of relevance. We adapt the Expected Reciprocal Rank metric [16] for the setting of aggregated user sessions, introducing the scheme:

$$ORD(r_i, r_j) = \frac{2^{\mathrm{g(i,j)}}}{2^{\mathrm{g\_max}}} \tag{4}$$

$$\mathrm{g}(r_i,r_j) = \mathrm{n} \cdot \mathcal{F}(Freq(r_i, r_j)) \tag{5}$$

$$\mathcal{F}(Freq(r_i, r_j)) = \frac{|\{r_k \in S' | Freq_S(r_i, r_k) \le Freq_S(r_i, r_j)\}|}{|S'|} \tag{6}$$

where $S' \subseteq S$. The value $g(r_i, r_j)$ denotes the observed URL access frequencies in the overall user sessions. It is normalized to a common rating scale $[0, n]$, based on cumulative

[4]The less related the resources are via terms in $\mathcal{O}$, the smaller is their similarity degree at each iteration.

distribution function of $Freq(r_i, r_j)$ over the set of other URLs accessed in the same session with r_i and r_j, but co-located with r_i less frequently. The maximum relevance value is g_max.

CONDITIONALSIMILARITY: conditional probability of pair (r_i, r_j) occurring in the same session, given that resource r_i appears in that session:

$$sim_{conditional}(r_i, r_j) = \frac{Freq_S(r_i, r_j)}{Freq_S(r_i)} \quad (7)$$

SESSIONSIMILARITY: a binary value stating if the two resources in the pair (r_i, r_j) appear together in at least one user session from the set S.

$$sim_{session}(r_i, r_j) = \begin{cases} 1, & \text{if } Freq_S(r_i, r_j)¿0 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where $Freq_S(r_i, r_j)$ is the number of sessions in which resources r_i and r_j occur together.

V. EVALUATION

A. Experimental Setup

For our experiments, we have prepared new real-world datasets that combine user browsing logs at different Web sites with semantic metadata embedded in the Web pages.

1) Datasets: **User Logs Dataset.** The dataset of user logs (Table I) consists of browsing behavior data registered via the Yahoo! toolbar, which tracks HTTP requests upon agreement of users that have installed this toolbar. From a huge dataset, we extracted a sample of the user logs recorded within a five-week period. Our approach is open-domain, but to facilitate the evaluation of results we filtered logs from three different Web sites of event advertisement nature.

TABLE I. STATISTICS FOR THE DATASET OF USER LOGS

Logs period	01.Jul.2012 - 07.Aug.2012
Nr. users	494
Nr. logs	2244
Nr. unique URLs	1683
Nr. user sessions	526
Average session length	4.132 (10.31% sessions)
Mode session length	3 (38.86% sessions)
Min session length	2 (15.65% sessions)
Max session length	15 (0.18% sessions)
Nr. cross-site sessions	14.0 (2.58% sessions)

Structured Markup Dataset. The other dataset consists of structured markup data (metadata) extracted from the Web pages accessed in logs. These metadata are made available by the providers of the Web pages as annotations of the HTML elements.

Ground-truth Dataset. While clicks in the logs are signals of users preferences, we further acquired from human judges ground-truth values for evaluating resource relevance. Initially, we filtered the unique set of resources in the user logs, then extracted a large subset of these resources to pair up among each other. We did not perform a random extraction, rather selected resources by preserving a uniform distribution as in the original set, i.e. keeping the same proportion of resources from the different sites as in the logs. Table II shows the statistics of this *labeled dataset*. We showed the pairs (via a Web interface) to human judges, asking them for relevance feedback. They had to decide if the resources in the pair are relevant to each-other (i.e. if after visiting one resource, they would find the other relevant to view next).

TABLE II. STATISTICS OF THE LABELED DATASET

Nr. Resource Pairs	1230
Nr. unique Resources	387
Nr. Judges	13
Nr. Judgments/Pair	3
Inter-rater Agreement	80.2%
Non-Relevant Pairs	943 (76.67%)
Relevant Pairs	287 (23.33%)

2) Methodology and Evaluation Metrics: We performed two types of experiments for a thorough evaluation of our proposed approach, referred to as *Suadeo*[5]:
(1) We evaluate the relevance prediction approach using the ground truth dataset (Table II). This dataset is used for training a relevance model with SVMs and testing the results using 10-fold cross-validation. We use the metrics of Precision, Recall, F1, and Mean Absolute Error.
(2) We compare our recommendation approach to other baselines, applying the metrics of 1-call@N and precision P@N [14] on the ground-truth dataset. 1-call@N is the ratio of test queries for which we find at least one relevant item in their respective top-N recommendation list. The measure P@N reflects the ratio of the number of relevant resources in the top-N recommended resources.

B. Results of Relevance Prediction

We evaluate the performance of the SVM-based prediction approach applying different groups of features: usage-based, syntactic content-based, and semantic content-based only, as well as content-based and usage-based features combined. The results are illustrated in Table III.

We achieve a large improvement when using semantic features in addition to syntactic content-based features. This illustrates the advantage of capturing the structural information of the content when predicting relevant resources. The approach that combines both usage-based and semantic content-based features considerably improves the results, outperforming the others across all measures (i.e. higher precision, recall, ROC Area, and smaller errors).

C. Performance Comparison

We compare the performance of our recommendation approach with other baselines:

UBCF: a user-based collaborative filtering recommender [12] combined with content-related features that applies matrix factorization. It finds k similar items (neighbors) that are co-rated (or visited) by different users similarly. For a target item, predictions can be generated by taking an average of the target user's item ratings (or weights) on these neighbor items. Since we also deal with usage data, instead of rating we use an implicit binary weight associated to an item (i.e. Web resource) in a user session. In our experiments, we set k to 20. We extract resource pairs out of per-user recommendations: for each query resource, we find those users in the logs that have visited it, then generate the top-ranked recommendations for each user.

[5]Latin word for *advise*

TABLE III. Results of the relevance prediction approach using different sets of features

Features	Precision	Recall	F1	ROCArea	Mean Abs. Error	Root Rel. Sqr. Error(%)
Usage-based	84.1	82	77.7	60.3	0.2935	90.5532
Syntactic content-based	58.8	76.7	66.5	50.1	0.3578	99.9
Usage- & syntactic content-based	84.1	82	77.7	60.4	0.2935	90.5591
Semantic content-based	88.5	88.3	87.2	73.8	0.2075	76.0892
Usage- & semantic content-based	**89.9**△	**89.9**△	**89.3**△	**78.5**△	**0.1812**▽	**71.0287**▽

The final list consists of top-N recommendations ranked across all the filtered users.

IBCF: a classical item-based CF top-N recommendation algorithm.[6] As correlations of web resource similarity, we apply the tf-idf scores of their HTML content.

Suadeo: the proposed approach.

The recommendation performances of *Suadeo* and the baseline approaches in terms of 1-call@N and p@N are shown in Fig. 5. The following observations can be drawn: *Suadeo* outperforms all other methods in terms of P@N. Furthermore, there is an overall improvement over the other approaches in terms of P@N. These results corroborate that *Suadeo* achieves the goal of keeping relevant resources in the top-N recommendations.

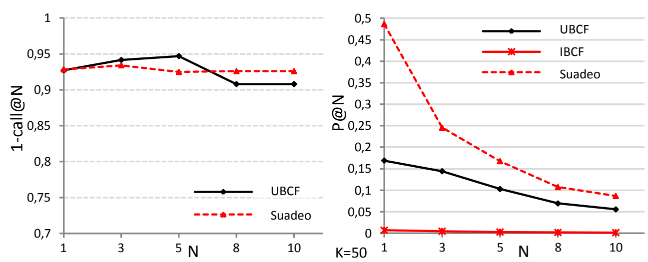

Fig. 5. Comparison of recommendations quality

In terms of 1-call@N, *UBCF* gives higher values for smaller N, and is outperformed by *Suadeo* for $N > 5$. These methods ensure to make at least one recommendation, among a few top-ranked resources, that is indeed relevant to the user. Whereas, *IBCF* recommender performs poorly with 1-call@N ≤ 0.125 for all N (for clarity not shown in graph). In conclusion, *Suadeo* contributes to providing valuable recommendations at top-N positions.

VI. Conclusions

In this work, we tackle the problem of generating user recommendations in an open-Web, cross-domain setting. We introduce a recommendation framework, whose novelty lies in the use of structural semantic information embedded in the content of the Web pages, and its combination with patterns of user browsing logs.

In cross-domain recommender systems, the expectation is that the generated recommendations may be less precise than those provided when considering only one domain. The advantage may not be the improved accuracy, but the added novelty that may offer users higher satisfaction and utility. However, this work presented an approach that is able to provide users with recommendations that are from different domains, yet highly relevant for the users. Through evaluation experiments on real-world datasets of semantically-enriched users logs, we showed the effectiveness of our approach and its superiority towards another traditional hybrid recommender system. An avenue of future work is to extend the framework with personalized recommendations that consider short-term and long-term preferences of users.

VII. Acknowledgments

The authors thank Pablo Castells (Universidad Autònoma de Madrid) for the helpful comments and feedback.

References

[1] I. Cantador, A. Bellogín, and P. Castells. A multilayer ontology-based hybrid recommendation model. *AI Commun.*, 2008.

[2] I. Cantador, P. Castells, and E. P. Superior. Extracting multilayered semantic communities of interest from ontology-based user profiles: Application to group modelling and hybrid recommendations. In *Computers in Human Behavior*, 2008.

[3] P. Cremonesi, A. Tripodi, and R. Turrin. Cross-domain recommender systems. In *ICDMW*, 2011.

[4] I. Fernández-Tobías, I. Cantador, M. Kaminskas, and F. Ricci. A generic semantic-based framework for cross-domain recommendation. In *HetRec*, 2011.

[5] I. Fernández-Tobías, I. Cantador, M. Kaminskas, and F. Ricci. Cross-domain recommender systems: A survey of the state of the art. In *CERI*, 2012.

[6] J. Hoxha. Semantic formalization of cross-site user browsing behavior. In *Research Technical Report, Archiv nr. 3025*. Institut AIFB, KIT, 2012. http://www.aifb.kit.edu/web/Techreport3025/en, 2012.

[7] X. Jin, Y. Zhou, and B. Mobasher. A maximum entropy web recommendation system: combining collaborative and content features. In *SIGKDD*, 2005.

[8] H.-T. Lin, C.-J. Lin, and R. C. Weng. A note on platt's probabilistic outputs for support vector machines. *Mach. Learn.*, 2007.

[9] A. Loizou. How to recommend music to film buffs: enabling the provision of recommendations from multiple domains. In *PhD thesis,University of Southampton*. Elsevier. In press, 2009.

[10] C. D. Manning, P. Raghavan, and H. Schutze. Introduction to information retrieval. *Information Retrieval*, 2010.

[11] P. Mika and T. Potter. Metadata statistics for a large web corpus. In *Proceedings of the Linked Data Workshop (LDOW) at the International World Wide Web Conference*, 2012.

[12] B. Mobasher, X. Jin, and Y. Zhou. Semantically enhanced collaborative filtering on the web. In *European Web Mining Forum*, 2003.

[13] J. Platt. Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In *Advances in Large Margin Classifiers*. 2000.

[14] Y. Shi, A. Karatzoglou, L. Baltrunas, M. Larson, N. Oliver, and A. Hanjalic. Climf: learning to maximize reciprocal rank with collaborative less-is-more filtering. In *RecSys*, 2012.

[15] R. Thiagarajan, G. Manjunath, and M. Stumptner. Computing semantic similarity using ontologies. In *Technical Report*. HP, 2008.

[16] S. Vargas and P. Castells. Rank and relevance in novelty and diversity metrics for recommender systems. In *RecSys*, 2011.

[17] P. Winoto and T. Tang. If you like the devil wears prada the book, will you also enjoy the devil wears prada the movie? a study of cross-domain recommendations. *New Generation Computing*, 26(3):209–225, 2008.

[6]Mahout implementation: http://mahout.apache.org/