

Diversifying Contextual Suggestions from Location-based Social Networks

M-Dyaa Albakour, Romain Deveaud, Craig Macdonald and Iadh Ounis
University of Glasgow, UK
firstname.lastname@glasgow.ac.uk

ABSTRACT

In this paper, we study the emerging Information Retrieval (IR) task of contextual suggestion in location-based social networks. The aim of this task is to make personalised recommendations of venues for entertainments or activities whilst visiting a city, by appropriately representing the context of the user, such as their location and personal interests. Instead of only representing the specific low-level interests of a user, our approach is driven by estimates of the high-level categories of venues that the user may be interested in. Moreover, we argue that an effective model for contextual suggestion should not only promote the categories that the user is interested in, but it should also be capable of eliminating redundancy by diversifying the recommended venues in the sense that they should cover various categories of interest to the given user. Therefore, we adapt web search result diversification approaches to the task of contextual suggestion. For categorising the venues, we use the category classifications employed by location-based social networks such as FourSquare, urban guides such as Yelp, and a large collection of web pages, the ClueWeb12 corpus, to build a textual classifier that is capable of predicting the category distribution for a certain venue given its web page. We thoroughly evaluate our approach using the TREC 2013 Contextual Suggestion track. We conduct a number of experiments where we consider venues from the closed environments of both FourSquare and Yelp, and the general web using the ClueWeb12 corpus. Our empirical results suggest that category diversification consistently improves the effectiveness of the recommendation model over a reasonable baseline that only considers the similarity between the user's profile and venue. The results also give insights on the effectiveness of our approach with different types of users.

1. INTRODUCTION

The emergence of Location-based Social Networks (LBSNs), such as FourSquare and Facebook Places, offers enormous information that can be exploited to address localised information needs of citizens in smart urban spaces [15]. Allan *et al.* identified emerging information seeking scenarios in these smart spaces where the user may express no explicit query (*zero-query*) and an Information Retrieval (IR) system is required to entertain the user by sug-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

IIIX '14, August 26 - 29 2014, Regensburg, Germany.

Copyright 2014 ACM978-1-4503-2976-7/14/08 ...\$15.00.

<http://dx.doi.org/10.1145/2637002.2637018>.

gesting venues to visit that are relevant to the context of the user (e.g. the location and the known user profile on the LBSN) [2]. The recent introduction of the Contextual Suggestion track in the TREC evaluation campaign signified the importance of addressing this type of information need [8]. The Contextual Suggestion track introduces a new IR personalisation task where the challenge lies in accurately modelling the user preferences as indicated in their profiles on the LBSNs in order to provide relevant venues to visit that match their taste. Existing state-of-the-art approaches use the description of the user's preferred venues and reviews of the venues on LBSNs to build a user profile and suggest venues based on their similarity with the constructed profile [26]. Such approaches represent the direct low-level preferences of a user, and may not sufficiently deal with the ambiguity implied by the zero-query challenge in this task, which may require modelling higher level interests of the user. For example, if a user has shown interests in various museums in the past, it might be inferred that she is generally interested in arts. Therefore, it would be appropriate to recommend art galleries and cultural centres to this user. Collaborative filtering approaches can be employed to infer higher-level user's interests by mining preferences of similar users on the LBSN [19]. However, they require detailed profiles from a large number of users, which may not be available. Moreover, the aforementioned state-of-the-art approaches do not consider redundancy in recommending many venues of the same category that may exist in one area. For the user in the previous example, if she is visiting an area with plenty of museums, it is desirable to avoid the redundancy of recommending only museums by diversifying the recommendations. Instead, if that user has shown interests in other types of venues, such as bars, but to a less extent than museums, it would still be appropriate to also recommend bars to her.

Diversification approaches have been studied in web IR to deal with the ambiguity of queries and to reduce the redundancy of retrieved documents [1, 22, 23]. Similarly, and since a zero-query is ambiguous by definition, we argue that diversity can also deal with the ambiguity involved in contextual suggestion, as it avoids redundancy and may enhance the available choices of venues for the user. Recent approaches for diversity in web search, e.g. [22], identify explicit aspects of an ambiguous query and then rerank the search results to maximise the probability of covering all aspects in the top results, whilst reducing redundancy. We propose to adapt the notion of query aspects to contextual suggestion by extracting the high-level categories of venues that the user is interested in, based on their low-level interests as identified in their venue preferences on the LBSN. Considering the high-level categories seems natural since it provides an approximation of the low-level user interests in the information domain of venues world-wide, which is potentially very large and sparse. Our proposed approach infers the probability distribution of high-level venue categories of interest from

the user preferences. Furthermore, using this probability distribution, we present an adaptation of a state-of-the-art diversification approach in IR, namely the xQuAD framework [22], to the contextual suggestion problem. With this adaptation, we produce a diversified ranking of venues that balances between matching the user’s low-level interests, from their social profile on the LBSN, while also covering the inferred high-level venue categories from the user preferences. In our approach, the venue categories can be obtained directly from the category classifications in location-based social networks such as FourSquare and Urban guides such as Yelp, which are considered to be high quality sources. We further propose a text classifier that is capable of predicting the category of a venue from its public profile (i.e. an HTML page) on the web. This is done by using a large corpus of web pages, the ClueWeb12 corpus, and thousands of venues collected from FourSquare and Yelp. The availability of such a classifier allows our approach to be capable of suggesting venues from various sources that include any LBSN or the web in general. Indeed, our approach is generalisable beyond any specific LBSN and is not limited by a social network setting, as opposed for example to the aforementioned collaborative filtering approaches [3, 19, 27], which are limited by the information available about the users’ preferences within a single social network.

Using the TREC 2013 Contextual Suggestion track, we empirically evaluate our diversification approach for contextual suggestion using two different venue categorisations. The track serves as a user study, since it involved crowd-sourced users who judged personalised suggestions [9]. We show that diversification consistently improves the effectiveness of the suggestions over a reasonable baseline that only considers the similarity between the user’s profile and the venue. In our evaluation, we consider venues from Yelp and the FourSquare LBSN. Moreover, to show the generality of our approach beyond a specific LBSN, we also consider an evaluation where we suggest venues from a subset of the ClueWeb12 corpus.

As a summary, the contribution of the paper is twofold. Firstly, we devise and thoroughly evaluate a diversification approach for contextual suggestion that alleviates the potential redundancies of recommended venues in a personalised manner. Secondly, we develop a venue category prediction approach, which ensures that our diversification is applicable beyond a specific LBSN.

The remainder of the paper is structured as follows. We first discuss the related work in Section 2. Then, we introduce our general model for personalisation to provide contextual suggestion in Section 3. Following that, in Section 4, we introduce our adaptation of the web IR diversification approach to the contextual suggestion problem. In Section 5, we propose an approach for venue category prediction. We then describe our experimental setup for evaluation and discuss the results of our experiments in Section 6. Finally, we discuss the conclusions stemming from these results in Section 7.

2. RELATED WORK

A recent study [15] has identified a variety of emerging localised information needs that citizens often have in smart urban spaces. Among these information needs are the contextual suggestion of venues or activities to users without explicitly specifying keyword queries [2]. With the emergence of LBSNs such as FourSquare and Facebook Places, and the growing interest in smart cities [6], the IR task of contextual suggestion is becoming more popular and several approaches have been proposed in the literature. Two streams of approaches can be recognised in the literature, namely, standard IR personalisation approaches and collaborative filtering approaches, which we further discuss below.

The first stream of approaches for contextual suggestion rely on traditional IR personalisation techniques [12], which require modelling the interests of the user, i.e. building some sort of a user

profile that can be then used to provide more tailored results to the user’s needs [4]. In social media platforms, modelling the interests of the user to provide personalised search has been performed from explicit user’s interactions which include the content of the posts and the messages they share, or the pages on the web that they have previously “bookmarked”, e.g. [18]. Similarly, for contextual suggestion in LBSNs, the user’s low-level interests has been modelled using the description of venues or locations they are interested in (e.g. language models describing venues they like). For example, Kurashima *et al.* [16] proposed a probabilistic topic modelling approach to infer the next visit of a user by mining historical logs of GPS. Hu *et al.* [13] employs a similar approach but they use text and locations extracted from social media (e.g. Twitter). Yang *et al.* [26] incorporates opinions provided by previous users into the constructed language model of the venues. The aforementioned personalisation approaches suffer from two drawbacks. First, they are not capable of inferring the higher level interests of a user. For example, if a user has shown interests in many museums in the past (by rating them highly), a cultural centre may be deemed irrelevant to the user by a language modelling (LM) approach, since its description may have a different language model compared to the language model of museums rated highly by the user. However, the user may still consider it as relevant as it satisfies her interests in arts in general. The second drawback is that the suggestions will always be biased towards the types of venues that are rated highly by the user, which introduces redundancy. Our proposed approach tackles these drawbacks by both modelling the user’s high level categories of interests and diversifying the recommendations according to these interests.

The second stream of work in contextual suggestion relies on collaborative filtering approaches that perform recommendation based on the venues that similar users have liked in the past within the LBSN, e.g. [3, 19, 27]. These approaches do not require any knowledge about the description of the venues and can tackle the first drawback identified above, but they require large past feedback from a lot of users to accurately infer the user’s interests. In other words, they may suffer from the sparsity of user profiles in LBSNs. To tackle sparsity, Noulas *et al.* [19] used random walk with a collaborative filtering approach based on latent space models and computed a variety of similarity criteria with venue’s visit frequencies on the LBSN. Bao *et al.* [3] tackled the sparsity issues by identifying candidate local experts in one area and matching those to a given user. Nevertheless, collaborative filtering approaches are still restricted to users within a single LBSN.

In this work, we aim to adapt the first stream of contextual suggestion approaches and address their drawbacks. In particular, we aim to both identify the higher level user’s interests and eliminate the redundancy in the top recommended venues by adapting a state-of-the-art diversification approach for web search to the problem of contextual suggestion. Our work is related to recent approaches for diversity in recommender systems e.g. [24, 29]. Zhang and Hurley [29] introduces the notion of user profile partitioning using clustering algorithms. The diversity is then achieved by generating recommendations from each constructed partition and then combining them uniformly in the final list of recommendations. Vargas and Castells [24] introduce the notion of sub-profiles within the user profile using available categorisation of the recommended items’ domain. They achieve diversity by combining recommendations from each generated sub-profile and by taking into account the importance of each sub-profile in the final list of recommendations. Our approach is similar to [24], in the sense that we aim estimate the higher-level interests of the users in the various venue categories and use that as a prior knowledge for the diversification approach. However, unlike [24], which relies on collaborative fil-

tering within a LBSN, our proposed approach can be generalised beyond a single LBSN as we show later.

3. A MODEL FOR CONTEXTUAL SUGGESTION

In this section, we describe our general approach for providing personalised venue recommendations given that we know the explicit interests of the user. First, we formalise the problem of contextual suggestion. Then, we describe our language modelling approach for contextual suggestion.

3.1 Problem Formulation

Firstly, we assume that users have expressed preferences on venues that they are interested in. These preferences are expressed as ratings that the user provided on previous venues they have visited or read about indicating the degree to which they like these venues. These ratings may take many forms, e.g. integer rating scales such as 0-5 stars, binary ratings (like/dislike), or even unary rating such as ‘has visited before’. This is akin to user ratings of information items, such as movies, books, or clothes in e-commerce platforms, which are used in recommender systems [10]. Considering a universe of venues V , from a *venue source* such as a LBSN (e.g. FourSquare), the user’s preferences are specified by the ratings $a(u, v_i)$ given by the user u to a subset of venues $V_u \subset V$. For each venue $v_i \in V_u$, $a(u, v_i)$ is a numeric score that reflects the intensity of the user’s interest in that venue. The aim of our contextual suggestion model is to produce a ranked list of recommended venues for a certain context, specified by the user, her current location and/or the time. In this paper, we simplify the contextual suggestion problem by not considering the time as part of the context.

3.2 A Language Modelling Approach for Contextual Suggestion

For a context specified by the user u and their location l_u , we consider a set of *candidate* venues, a subset of venues $V_l \subset V$ that are close enough to the user’s location l_u . For example, we can limit the candidate venues to be those within a certain distance from the user’s location. This is a simplified approach that is supported by a recent study, which shows that users are more likely to visit venues close to their current location [27]. In future work, we may consider a distance decay factor to extend this simplified filtering of venues. The candidate venues are ranked by a recommendation score $r(u, v)$; $v \in V_l$ to produce the ranked list of venues that can be suggested to the user.

To estimate the recommendation score for a contextual personalised ranking of the venues, we adopt a content-based approach for recommendation, where we build a model describing the contents of both the user profile and the venues to be suggested to the user. To model the user’s interests, we adapt a common modelling approach for personalisation [12], where we construct two profiles that represent the user preferences: (i) a positive profile and (ii) a negative profile. The *positive* profile describes the venues that the user likes, i.e. the venues for which the user has explicitly provided positive feedback. Formally, these venues are denoted by $\mathcal{V}_u^+ = \{v \in V_u : a(u, v) \geq \eta^+\}$; where η^+ represents a threshold over which the rating $a(u, v)$ is considered positive (e.g. above 3 stars). On the other hand, the *negative* profile describes the venues that the user dislikes, i.e. the venues for which the user has explicitly provided negative feedback. Formally, these venues are denoted by $\mathcal{V}_u^- = \{v \in V_u : a(u, v) \leq \eta^-\}$; where η^- represents a threshold below which the rating $a(u, v)$ is considered negative (e.g. below 2 stars).

Using the constructed profiles of the user, we estimate the rec-

ommendation score $r(u, v)$ by estimating the similarity between the user’s liked venues \mathcal{V}_u^+ and a given candidate venue v and combining this with the dissimilarity between the user’s disliked venues \mathcal{V}_u^- and the candidate venue v . To achieve this, we use a language modelling (LM) approach where we construct a unigram language model for the description of the candidate venue, denoted by θ_v , and each of the profiles, denoted by $\theta_{\mathcal{V}_u^+}$ and $\theta_{\mathcal{V}_u^-}$ respectively. Following this, we rank the candidate venues using the following recommendation score:

$$r(u, v) = \alpha \cdot KL(\theta_v || \theta_{\mathcal{V}_u^+}) - (1 - \alpha) \cdot KL(\theta_v || \theta_{\mathcal{V}_u^-}) \quad (1)$$

where $KL(\theta_v || \theta_{\mathcal{V}_u^+})$, and $KL(\theta_v || \theta_{\mathcal{V}_u^-})$ calculate the Kullback Liebler (KL) divergence between the language model of the candidate venue and that of the positive and negative profiles respectively, $0 \leq \alpha \leq 1$ is a parameter that regulates the contribution of each divergence component to the final score. To build the language model of a venue θ_v , we use the home page of the venue on the LBSN or on the web as a textual representation. The home page of the venue contains information about the venue such as the venue’s name, the description and the category of the venue (e.g. Food or Shopping). In addition, it incorporates a social aspect in the form of comments provided by users, as in FourSquare and Yelp. To build the language models of the user’s profiles $\theta_{\mathcal{V}_u^+}$ and $\theta_{\mathcal{V}_u^-}$, we use a concatenation of the home pages’ contents for all the venues in the profile. Finally, we use Dirichlet’s prior smoothing [28] when estimating the KL divergences in Equation (1).

4. CATEGORY DIVERSIFICATION FOR CONTEXTUAL SUGGESTION

The ranking of venues produced by the language modelling approach of Equation (1) in the previous section considers only the low-level interests of the user represented in the language models of the positive and the negative profiles. Moreover, this ranking may result in redundancy by recommending many venues of the same category. Even if these venues are of interest to the user, she may still want to consider suggestions of other categories of interest. We aim to introduce diversity in the contextual suggestions to cover the venue categories of interest to the user. Our intuition is that diversity can deal with the ambiguity of the zero-query in contextual suggestion by approximating user’s interests into high-level venue categories and by eliminating potential redundancy of recommending venues of the same category. Our approach can still ensure that such diversification is performed in a personalised manner whereby we estimate the personal interest of the user in each of the various venue categories. In this section, we describe how we adapt a state-of-the-art approach for web search result diversification to the contextual suggestion problem.

Effective web search result diversification are based on an explicit representation of query aspects, such as IA-Select [1] and xQuAD [22], where the coverage of these aspects are estimated for each document in the retrieved ranked list of results in order to produce a diversified ranked list. Those approaches have been shown to outperform prior non-explicit content-based diversification approaches such as MMR [23]. We propose to adapt an explicit diversification approach where we consider the notion of query aspects to be equivalent to the high-level venue categories underlying the user’s profile as described in their previous ratings. Furthermore, we propose to develop our adaptations using the xQuAD framework, since it is amongst the top performing diversification methods in four successive web search evaluation campaigns [23].

The xQuAD framework uses a greedy algorithm that re-ranks documents returned by a baseline retrieval model to balance between relevance and diversity. Given an ambiguous query q and an

initial ranking R produced for this query, xQuAD builds a new ranking S by iteratively selecting the highest scored documents from R , according to a probabilistic mixture model:

$$(1 - \lambda) \cdot P(d|q) + \lambda \cdot P(d, \bar{S}|q) \quad (2)$$

where $P(d|q)$ is the relevance component of xQuAD after a document likelihood retrieval model, and $P(d, \bar{S}|q)$ is the diversity component of xQuAD that quantifies the likelihood of observing this document but not the documents already in S , which were selected in previous iterations of the algorithm. In our adaptation, we equate the user profile u (the representation of the information needs in the contextual suggestion task) to the search query q , the venues to documents, and the venue categories to query aspects. Therefore, the objective function of the xQuAD framework for contextual suggestion can be written as:

$$(1 - \lambda) \cdot P(v|u) + \lambda \cdot P(v, \bar{S}|u) \quad (3)$$

where $P(v, \bar{S}|u)$ is the diversity component that quantifies the likelihood of observing the venue v but not the venues already in S , and $P(v|u)$ denotes the probability that the venue v is of interest to the user and can be estimated using the ranking scores in Equation (1). More precisely, we estimate this probability as follows:

$$P(v|u) \propto \frac{r(u, v)}{\sum_{v_i \in R} r(u, v_i)} \quad (4)$$

where R is an initial ranking of venues produced by the language modelling approach.

As for the diversity component $P(v, \bar{S}|u)$, Santos *et al.* [22] defined it as a marginalisation of the probabilities over an explicit set of query aspects (which in our case are the categories of venues) as follows:

$$P(v, \bar{S}|u) = \sum_{c_i \in \mathcal{C}} \left[P(c_i|u) P(v|c_i) \prod_{v_j \in S} (1 - P(v_j|c_i)) \right] \quad (5)$$

where:

- \mathcal{C} is the set of all venue categories considered. This set is uniform for all users, and can be obtained from the categorisations employed by LBSNs such as FourSquare and Yelp;
- $P(c_i|u)$ defines the *category importance*, the probability that the user u is interested in the category $c_i \in \mathcal{C}$. As we show below, we devise two different approaches for estimating this probability; and
- $P(v|c_i)$ is the *category coverage*, which estimates the venue's coverage of the category in question. It equates to the document's coverage of a query aspect and is typically calculated using the approach in [22], where the coverage is based on the relevance estimation of documents for a query aspect. In our case, it is based on the category of the venue in the LBSN or the predicted venue category that we estimate below.

To estimate the probability distribution $P(c_i|u)$, the probability of a preference towards a certain venue category c for a given user u , we devise two different approaches. The first approach aims to achieve a *non-personalised* diversification, whereas the second approach aims to achieve a *personalised* diversification.

With *non-personalised* diversification, the probability $P(c_i|u)$ is uniform for all users and all categories. In particular, we estimate this probability as follows:

$$\forall c_i \in \mathcal{C} : P(c_i|u) = \frac{1}{|\mathcal{C}|} \quad (6)$$

For *personalised* diversification, estimating the probability distribution $P(c_i|u)$ involves inferring a distribution of categories of interest for the user using her preferences encapsulated in both user's profiles \mathcal{V}^+ and \mathcal{V}^- . We estimate this probability as a marginalisation of probabilities over all the venues in the original ranking of venues R , generated by the language modelling approach. In this marginalisation, we assume the obvious independence of the venue category from the user and we compute it as follows:

$$P(c_i|u) \propto \sum_{v \in R} P(c_i|v) \cdot P(v|u) \quad (7)$$

where $P(v|u)$ can be estimated using Equation (4) from both user's profiles \mathcal{V}^+ and \mathcal{V}^- ; and $P(c_i|v)$ is the probability that the venue v corresponds to the category c_i . This probability can be obtained in two different ways:

- $P(c_i|v)$ can be directly obtained from the categorisation of venues in the LBSN. For example, if the venue v is classified in the LBSN as a shopping venue then $P('shopping'|v) = 1$ and $P(c_x|v) = 0$ for all other categories $c_x \in \mathcal{C}$.
- If the venue source is not a LBSN, e.g. the web, then the estimation of $P(c_i|v)$ is more challenging. In the following section, we develop a category prediction approach that estimates the probability $P(c_i|v)$ using a web document representing the venue v , such as its home page on the public web. This makes our approach applicable to venues beyond a single LBSN.

5. VENUE CATEGORY PREDICTION

Touristic guides and LBSNs provide a high-level categorisation of their venues to aid users in browsing the available places to visit. The category information available in such guides or networks can be considered as accurate information about the categories that apply to a venue (e.g. Food or Shopping). However, in some cases, venues may not be available in a certain LBSN. Therefore, we need to be able to predict the category of the venue. In fact, this would be the case when we consider the entire web as a source of venues for recommendation.

In this section, we develop an approach for predicting the venue category given a web document that represents the venue. We propose to use a large collection of web documents to be able to classify any web document that may represent a venue into the right venue category. In other words, for a web document d that represents a venue v , e.g. its home page, our classifier would be able to estimate a category probability distribution $P(c_i|v)$ for all the categories $c_i \in \mathcal{C}$.

Our approach is inspired by [5], which maps Facebook likes to DMOZ categories using a supervised machine learning approach that builds a textual classifier trained with top web search results. Similarly, we start from ground truth data obtained from a large number of venues in a LBSN. Using these venues, we can retrieve, from a web collection, a number of web documents that either represent the venue (from the web site of the venue) or mention the venue. We make the assumption that these documents represent the actual category of the venue in the LBSN. These documents can be then used to train a textual classifier with a supervised machine learning approach to be able to predict the category probability distribution for a web document representing a venue of an unknown category.

The process of training the textual classifier has the following stages:

- The first stage is *sampling venues* from a LBSN that can be used for the training process. Formally, we consider a subset

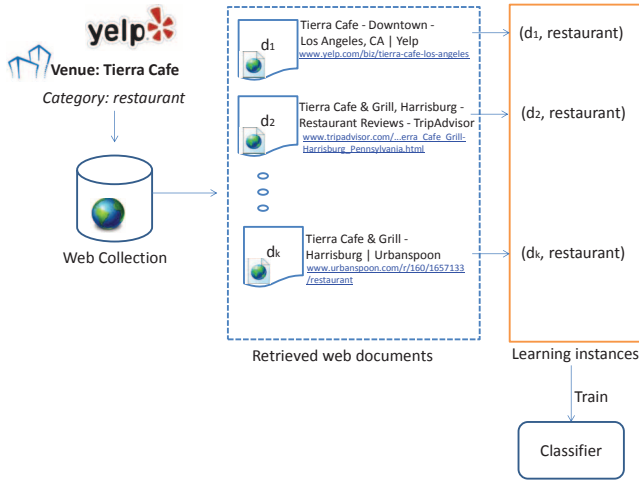


Figure 1: An example of the training process of the classifier

of venues $V_{sample} \subset V$ obtained from a LBSN as the sample venues where the categorisation’s information are available.

- The second stage is *retrieving web documents* that are relevant for each venue in the sample in order to create learning instances for the classifier. For each venue $v \in V_{sample}$ of category c_v , we retrieve a set of web documents D_v and we consider each web document $d_i \in D_v$ to be labelled with c_v . The assumption is that the content of those documents will be representative of the venue v and therefore they are also representative of the category c_v . To retrieve the documents D_v , we use the title of the venue t_v as a query and consider the top k retrieved documents for this query from a large web collection to represent the set D_v . For retrieval, we explore using a standard retrieval model, such as BM25, as well as state-of-the-art learning-to-rank approaches as we show later. For each document $d_i \in D_v$, we use the pairs (d_i, c_v) as a learning example for the classifier where the document d_i is mapped to the category c_v .
- The last stage is *applying the supervised machine learning* using the learning instances extracted in the previous stage. The supervised machine learning algorithm uses the terms of the documents as features for learning.

In Figure 1, we show a concrete example of the training process. For the venue “Tierra Cafe” in the sample venues V_{sample} , we retrieve, from a web collection, k documents that are relevant to the venue using “Tierra Cafe” as a query. From each document retrieved, a learning instance is created where the venue category associated to the document is considered to be “restaurant”, which is the original category of the sample venue “Tierra Cafe”. Once the training is completed, we have a textual classifier that can estimate the category probability distribution $P(c_i|v)$ for a venue v given the web page that represents it. Figure 2 shows a concrete example of the category prediction using our text classifier after being trained using the process above.

Our proposed approach for category prediction is evaluated in the experiments in the coming section, before evaluating our category diversification for contextual suggestion.

6. EXPERIMENTS

In this section, we aim to thoroughly evaluate our approach for contextual suggestion. Since the category prediction is a required

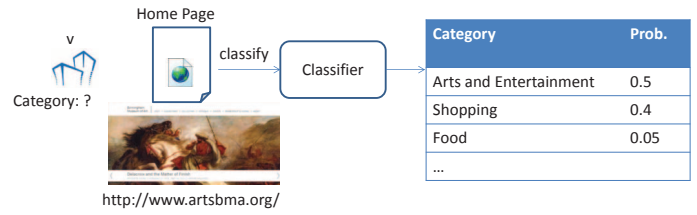


Figure 2: An example of predicting the category of the venue represented with a web page

component for the diversification to work, we first validate our category prediction approach. Following this evaluation and using the conclusions stemming from it, we then conduct a number of experiments to assess the quality of the contextual suggestions produced with the diversification approach and compare it to our LM baseline introduced in Section 3. The remainder of the section is structured as follows. We first identify the concrete research questions we aim to answer by conducting our experiments. Then, we present our experimental setup highlighting the various datasets used. The evaluation of the category predictions is then presented, followed by the evaluation of the diversified contextual suggestions and an analysis of the results.

6.1 Research Questions

In our experiments, we aim to answer the following research questions:

- **RQ1:** How effective is our classification approach for predicting the category of a venue given its web page? What is the best setting under which the classifier produce the most accurate predictions? (Section 6.3)
- **RQ2:** Can our diversification approach improve the quality of contextual suggestion over a reasonable baseline model? (Section 6.4)

Next, we discuss our experimental setup and describe how the above questions can be answered.

6.2 Experimental Setup

In this section, we first describe the dataset of the TREC 2013 Contextual Suggestion track used to evaluate our diversification approach. Then, we describe the venue sources we used to identify candidate venues for our model. Moreover, we present the venue category sets considered in our diversification method. Finally, we describe the experimental setup for venue category prediction.

6.2.1 Dataset

We evaluate our contextual suggestion model using the TREC 2013 Contextual Suggestion track [9]. The dataset used in this track consists of user profiles and geographical contexts. In addition, it consists of 50 sample venues within the city of Philadelphia, PA in the USA. These sample venues are used to construct the users’ profiles. Each user profile consists of a 5 point rating scale (from 0 to 4) explicitly provided by the user to web pages and descriptions of the sample venues. These ratings are similar to those usually available in LBSNs. The geographical contexts are 50 selected city centres in the USA and the task is to provide contextual suggestions for pairs of users and city centres. The judgments collected are a total of 223 pairs of user/city centre pairs with contextual suggestions of venues judged from 115 unique users in 36 unique geographical contexts (city centres). The task also involves providing a URL of the venues from either the open web or a large collection of web

documents, the ClueWeb12 corpus, together with a short description of the suggestion. Both the descriptions and the web pages are rated by the actual users but since the descriptions are free text provided by the participating groups in the track, it is not possible to reuse the judgments of those descriptions. Therefore, we evaluate our contextual suggestions using only the judgments provided for the URLs or the ClueWeb12 identifiers of the suggestions. We also note that the Contextual Suggestion track does not use the traditional pooling approach that TREC uses for relevance assessment but instead asks users to judge the top 5 suggestions produced by each submitted run individually. Hence, there are cases where URLs or ClueWeb12 identifiers are judged more than once for the same user/city centre pair, but with different judgments. For these URLs, we aggregate the various judgments available into a single one by taking the highest rating available.

6.2.2 Venue Sources

The participating groups in the TREC 2013 Contextual Suggestion track used a variety of sources for venues, such as Google Places, FourSquare, Yelp and ClueWeb12 CS, with varying success [9]. Therefore, to show the generality of our approach, we consider a number of these sources in our experiments. In particular, to generate the candidate venues that can be ranked for a certain user profile in a given city centre, we collect venues from the FourSquare LBSN and Yelp, which is both an urban guide and a LBSN. Using the FourSquare API,¹ we collected all the venues for the 50 city centres of the track in a 4 km² grid centred at those locations. Similarly, using the Yelp API,² we collected venues for the same grids. The total number of venues collected from FourSquare and Yelp is 60,212 and 7,096 venues respectively. For each collected venue, we have the title of the venue, its category according to the corresponding LBSN, its web page and other information. As an alternative venue source, we also consider candidate venues from the ClueWeb12 corpus where we are not restricted to a certain LBSN. In particular, as candidate venues, we use the web pages in a subset of ClueWeb12, the ClueWeb12 CS, created specifically for the track [9]. ClueWeb12 CS contains 30,144 documents where each corresponds to one of the geographical contexts (the city centres) in the track and may potentially represent a candidate venue suitable for the contextual suggestion problem.

6.2.3 Venue Categories

In order to show the generality of our diversification approach beyond specific LBSNs, in our experiments we consider a variety of combinations of the source for candidate venues and the set of venue categories for diversification. Table 1 summarises these combinations. For the first two, denoted by (I1, I2), we consider venues from only a specific LBSN (FourSquare and Yelp respectively) and use the categories of those venues that are available in the corresponding LBSN. Both FourSquare and Yelp classify venues according to their own hierarchy of categories, of which we only consider the top level. For the last two settings (w1, w2), we use an open setting where we are not restricted to a certain LBSN but instead we can suggest venues from a snapshot of the web. In this case, we use the ClueWeb12 CS dataset as a source of candidate venues. We also employ our venue category prediction (c.f. Section 5) to estimate the category distribution of a venue given a web page from the ClueWeb12 CS, where we train the classifier using the top level categories of FourSquare (w1) or Yelp (w2).

The category sets used in our experiments for the diversification approach and the category prediction are the top level categories

¹<http://developer.foursquare.com>

²<http://www.yelp.com/developers/>

Table 1: Experiments Specification

Setting	Venue Sources	Categories
I1	Specific LBSN (FourSquare)	FourSquare
I2	Specific LBSN (Yelp)	Yelp
w1	Web Subset (ClueWeb12 CS)	FourSquare
w2	Web Subset (ClueWeb12 CS)	Yelp

Table 2: Top-level categories used for diversification and category prediction

FourSquare	Yelp
Arts and Entertainment	Arts and Entertainment
Food	Beauty & Spas
Nightlife Spots	Education
Great Outdoors & Services	Food
Shops	Hotels & Travel
Travel & Transport	Religious Organisations
	Shopping
	Sports & Leisure
	Restaurants
	Nightlife

of FourSquare and Yelp, which are listed in Table 2. Yelp has a total of 22 top-level categories in its hierarchy; we excluded 12 of those that are not suitable for contextual suggestion. Examples of categories excluded include ‘Property’, ‘Public Services & Government’ and ‘Local Services’. FourSquare has a total of 9 top-level categories of which we excluded 3 that are not suitable for contextual suggestion, namely ‘Residences’, ‘Professional’ and ‘College and Universities’. Both of the hierarchies have finer-grained categorisation under the top categories. For example, different types of cuisines under ‘Food’ and different types of sport under ‘Sport & Leisure’. In future work, we will consider finer-grained categories in lower levels of the category hierarchy. For example, the types of cuisines that a user likes.

6.2.4 Category Prediction Setup

In Section 5, we presented our approach for venue category prediction that involves 3 stages for training a text classifier. First, for the sampling stage, we randomly selected 5000 FourSquare venues and 5000 Yelp venues from the collected venues used in our combinations I1 and I2. Next, for the second stage of retrieving example documents (web pages) corresponding to the training venues, we used the full ClueWeb12 dataset, which is a large web collection of around 733 million English web pages. We index the collection using the Terrier platform [20] without stemming and without removal of stopwords. To retrieve documents corresponding to the training venues, we experimented with 3 different retrieval approaches, namely ranking documents by BM25 alone, and by using two learning-to-rank approaches. In particular, we use AFS [17] – which creates a learned model that linearly combines feature values – and also the state-of-the-art LambdaMART learning-to-rank technique [11, 25]³, which creates a learned model based on regression trees. To train the learning-to-rank techniques, we use 200 queries from the the TREC Web tracks 2009-2012, randomly split into training and validation sets, so as to prevent overfitting. For applying learning to rank, we used 63 different features that includes document weighting models, term-dependence proximity models, hyperlink analysis (inlink counts and URL lengths), and content quality (e.g. spam classification). Finally, for the last stage of applying the supervised machine learning for text classification using

³<http://code.google.com/p/jforests/>

Table 3: F-1 scores of the combinations of four supervised machine learning algorithms used for text classification with three different retrieval models and top 10 documents $k = 10$. Training and testing are done using with a 5-fold cross validation. Best performances are highlighted in bold.

	Yelp				FourSquare			
	Naive Bayes	J48	Random Forests	SVM	Naive Bayes	J48	Random Forests	SVM
BM25	0.314	0.463	0.474	0.444	0.282	0.480	0.483	0.440
AFS	0.304	0.496	0.476	0.456	0.346	0.485	0.487	0.454
LambdaMART	0.236	0.477	0.475	0.424	0.336	0.491	0.511	0.451

the documents’ terms as features, we experimented with 4 different approaches, namely Naive Bayes, SVM [14], Random Forests [7] and Decision Trees (J48) [21].

6.3 Evaluation of Category Prediction

In this section, we aim to answer our first research question RQ1. The aim is to assess the quality of venue category prediction with our approach and to identify the best setting under which the category prediction performs. The venue category prediction is important to perform the diversification of contextual suggestion, and will therefore help us later in answering our second research question RQ2. We conduct a number of experiments where we assess the accuracy of the classification of web pages (documents from ClueWeb12) using a 5-fold cross validation. We aim to validate our approach and to identify the most accurate combination for retrieving web documents that correspond to the sample venues (the second stage of the training in Section 5) and the supervised machine learning approach for our text classifier (the third stage of the training in Section 5). For this, and as described in Section 6.2, we applied three different retrieval approaches: the BM25 retrieval model, and the two learning-to-rank approaches AFS and LambdaMART. We selected the top 10 retrieved documents ($k = 10$) for each venue to use them as examples for training the classifier. We report, in Table 3, the F-1 accuracy scores of the four supervised machine learning approaches used for the text classification and described in Section 6.2.

First, we can observe from Table 3, that both J48 and Random Forests are the top performing supervised machine learning approaches with comparable performances on both used category sets (top-level FourSquare and Yelp categories) and across the different retrieval approaches. For example, in the case of the Yelp categories, the best accuracy is achieved when using J48 and the AFS retrieval approach (F-1=0.496), and with FourSquare categories, the best accuracy is achieved when using Random Forests and LambdaMART (F-1=0.511). We can also observe that the performance is consistent on both LBSNs, which suggests the robustness of our approach. It is also worth noting that with a higher number of categories of venues (the Yelp categories), the classifier is still producing consistent results.

Next, we examine whether a better retrieval model improves the accuracy of venue category prediction. The learning-to-rank approaches (LambdaMART and AFS) are known to be more effective than the BM25 model in web retrieval [11, 17, 25] and therefore we aim to examine if the category prediction improves when using those learning-to-rank approaches as opposed to using the BM25 baseline model. In Table 3, and when comparing the second and the third row with the first row, we observe that overall using the learning-to-rank approaches for retrieval (AFS and LambdaMART), results in increased classification accuracy. This is again consistent for both category sets (from both LBSNs). The only exception when BM25 results in better classification accuracy than both learning-to-rank retrieval approaches is the case where

Table 4: F-1 scores of the classification accuracy with Random Forests using the LambdaMART retrieval model and different number of top documents to train the classifier (k). Training and testing are done using with a 5-fold cross validation. Best performances are highlighted in bold.

k	3	5	10	20	30	40
Foursquare	0.396	0.438	0.511	0.541	0.593	0.614
Yelp	0.410	0.442	0.475	0.542	0.578	0.589

Naive Bayes is used as a supervised machine learning approach using the Yelp categories (i.e. the first column of Table 3). However, since we observe a much lower accuracy for Naive Bayes, we can still generally conclude that a better retrieval approach improves the accuracy of an appropriate supervised machine learning approach to predict the category of a venue using its web page. This again validates the soundness of our approach and justifies using a state-of-the-art web retrieval approaches for identifying web pages to train the text classifier. Overall, the achieved F-1 accuracy score of 0.511 is reasonable for our task, since the task is challenging where some venues may actually belong to multiple categories.

Finally, in this evaluation, we study the effect of k : the number of top documents used to train the classifier. To do this, we use the best performing combination of the supervised machine learning approach and the retrieval model, which is the Random Forests and the LambdaMART learnt model. We vary the number of top documents used for training each time and report the F-1 score using 5-fold cross validation in Table 4. The results suggest that there is trend of an increasing classification accuracy with larger values of k , which suggests that considering more documents retrieved in response to a query containing the venue’s name informs the classifier better about the category that the venue belongs to. Hence, it results in an overall better classification accuracy.

To summarise our findings for research question RQ1, we can conclude that the accuracy of our classification approach for predicting the category of the venue is reasonable and it increases especially when using an effective retrieval model (learning-to-rank with LambdaMART) and a large number of representative documents of a venue as training data.

6.4 Evaluation of Diversification for Contextual Suggestion

In this section, we address our second research question RQ2, where we evaluate the effectiveness of our diversification approach for contextual suggestion (c.f. Section 4) and compare it to our LM model as a baseline (c.f. Section 3). As discussed in the experimental setup and illustrated in Table 1, we conduct a number of experiments where we consider venues from two LBSNs and also the ClueWeb12 to show the generality of our approach beyond a specific LBSN. For our LM baseline, we use equal weights for the positive and negative profiles, i.e. $\alpha = 0.5$. For the diversification approach, we rerank the suggestions of the similarity baseline and

Table 5: Evaluation results of the personalised diversification approach for contextual suggestion using the TREC 2013 Contextual Suggestion track. Best performances are highlighted in bold. For the diversification approach, the percentage of increase or decrease over the LM baseline is reported in parenthesis. All metrics are estimated using the judgments of URLs and ClueWeb12 identifiers and ignoring the judgments of descriptions since they are not usable (c.f. Section 6.2.1). Under this setup, TREC 2013 median of P@5 is 0.374 for all runs, and 0.125 for runs that use the ClueWeb12 CS dataset.

	Venue Sources	Method	Categories	P @ 3	P @ 5	MRR	% j@5
1	FourSquare	LM baseline	–	0.442	0.360	0.624	67.98%
2	FourSquare	personalised diversification	FourSquare	0.462 (+4.5%)	0.385 (+6.9%)	0.640 (+2.5%)	68.43%
3	FourSquare	non-personalised diversification	FourSquare	0.431 (-2.4%)	0.354 (-1.6%)	0.620 (-0.6%)	63.94%
4	Yelp	LM baseline	–	0.073	0.062	0.167	11.13%
5	Yelp	personalised diversification	Yelp	0.078 (+6.40%)	0.064 (+2%)	0.172 (+2.9%)	10.88%
6	Yelp	non-personalised diversification	Yelp	0.074 (+1.36%)	0.062 (0.00%)	0.171 (+2.33%)	10.86%
7	ClueWeb12 CS	LM baseline	–	0.088	0.081	0.212	26.78%
8	ClueWeb12 CS	personalised diversification	FourSquare	0.097 (+10.17%)	0.088 (+8.89%)	0.223 (+4.47%)	28.10%
9	ClueWeb12 CS	personalised diversification	Yelp	0.095 (+7.72%)	0.089 (+10.00%)	0.217 (+2.42%)	26.60%
10	ClueWeb12 CS	non-personalised diversification	FourSquare	0.083 (-5.68%)	0.082(+1.23%)	0.202 (-4.71%)	27.22%
11	ClueWeb12 CS	non-personalised diversification	Yelp	0.079 (-10.22%)	0.081 (0.00%)	0.205 (-3.30%)	27.04%

set the λ parameter of the xQuAD framework to 0.5, where the relevance and diversity components have equal weights. In future work, we aim to conduct a further study to choose the trade-off parameter, e.g. using supervised learning. For example, this parameter may be sensitive to different types of users or different types of locations. We experimented with the two different approaches, described in Section 4, to estimate the category probability distribution: (i) the *personalised diversification* approach where the category probability distribution is inferred from the positive and negative profiles constructed for the user and calculated using Equation (7), and (ii) the *non-personalised diversification* approach where all categories have equal importance for all users and the category probability distribution is calculated using Equation (6). When using the ClueWeb12 CS, a subset of ClueWeb12, as source of candidate venues, we use our category prediction approach to produce the category probability distribution that is necessary to the diversification approach and consider two category sets as illustrated in Table 1 (settings w1 and w2). We employ the top performing classification approach as identified from the conclusions of the experiments in the previous section. In particular, we use LambdaMART for retrieval with $k = 40$ documents (the second stage for training in Section 5) and Random Forests as a supervised machine learning algorithm (the third stage for training in Section 5).

In Table 5, we report the results for our LM baseline approach and the diversification approaches (personalised and non-personalised) using the various combinations of venue and category sources of Table 1. We report two official TREC measures of the Contextual Suggestion, namely P@5 and MRR, and also P@3. As discussed in Section 6.2, these measures are estimated from the relevance of the venue’s web page only (URLs or ClueWeb12 identifiers). Indeed, in the Contextual Suggestion track, judgments are also provided for a description generated by each run for each suggestion. However, it is not possible to reuse the relevance judgments for the descriptions since they are free text. Moreover, the last column in Table 5 reports the percentage of judgments available for the recommended venue at rank 5 in each case.

The first part of Table 5 (rows 1-3) reports the results when the candidate venues are selected from our FourSquare crawl. We observe that the personalised diversification approach does increase the precision of the contextual suggestion (the second row). Especially at higher cut-off points, the precision is markedly improved (+6.9% for P@5 vs. +4.5% for P@3). This is interesting as it shows that diversification approach has a higher chance to cover the cat-

egories of interest to the user. Moreover, the performance of the baseline that the diversification is improving on is reasonable, since its P@5 is around the median of all runs in the TREC 2013 Contextual Suggestion track (0.374). From the third row, we observe that the non-personalised diversification degrades the performance of the contextual suggestion. This is expected since the task is a personalised one, but the degrade is marginal, which highlights the power of our diversification approach that balances between relevance to the user’s interests and the overall diversity of categories.

The second part of Table 5 (rows 4-6) reports the evaluation measures when we consider venues from the Yelp LBSN. We observe a similar pattern where the personalised diversification approach improves on the LM baseline for the various effectiveness measures (e.g. +2.9% for MRR). However, in this case, the non-personalised diversification marginally improves two of the effectiveness measures but to a less extent than the personalised diversification (e.g. +1.36% for P@3). Overall, the performance when using Yelp degrades when compared to the performance achieved when using FourSquare. The degrade in performance can be explained by the figures in the last column of Table 5 where the percentage of judged venues dropped from around 68% to 11%.

The last part of Table 5 (rows 7-11) reports the results when we consider suggestions from the ClueWeb12 CS dataset. As before, we observe a consistent improvement of the personalised diversification approach over the LM baseline for the various measures. Moreover, we observe an overall degrade in performance of the non-personalised diversification against the LM approach, but the degrade is marginal and in one case we observe a slight improvement (+1.23% for P@5 when using the FourSquare categories). The results here also show that using either of the Yelp or FourSquare category sets for personalised diversification improves the effectiveness of contextual suggestion. Finally, we observe low precision scores when using ClueWeb12 CS as a source of candidate venues. However, the performance is comparable with the median of all runs in the track that use the same source of venues (median of P@5 for those runs is 0.125). Moreover, the weaker performance can be further explained by the low figure of the percentage of venues judged reported in the last column of Table 5.

As a summary and in answer to our research question RQ2, we can conclude that the personalised diversification approach we have proposed is indeed useful for contextual suggestion and can consistently improve the precision of recommendation over an LM baseline. In contrast, the non-personalised diversification generally re-

sults in a less effective performance, which is expected since the task is a personalised one. However, it is interesting to observe that the degradations in performance are only marginal, which indicates that diversification in general is a promising approach for contextual suggestion. Next, we perform further analysis on our diversification approach to investigate its performance with regards to different types of users.

6.5 Analysis of the Diversification Results

The previous experiments suggest that when the personalised diversification approach is applied, the overall effectiveness of the recommendation is improved. In Figure 3, we show a concrete example from our experiments on how personalised diversification of contextual suggestion works. In particular, we illustrate a user profile defined by explicit ratings of venues and next to it we list the top 5 recommendations using our LM baseline and our personalised diversification approach. We can observe that with diversification, new categories of venues are pushed to the top 5 list, for which the LM baseline has only presented historical places and museums. Now, we aim to establish the contribution of the diversity to the observed improvement in performance bearing in mind that our approach achieves personalisation by balancing between the similarity and the diversity of categories that the user likes. In order to quantify the contribution of diversity to the recommendation effectiveness, we first aim to identify the different types of users in terms of the variety of category interests. Then, we establish whether our diversification approach works for the different types of users in providing a recommendation list that is diverse and that represents their categories of interests. Finally, we aim to quantify if there is any correlation between the improvement of recommendation effectiveness and diversity.

Users are different in terms of the variety of their interests. Some users may have few categories of venues that they would like to visit, e.g. they only like bars. Others may have a variety of equal interests to different categories of venues. To measure this variation, we can estimate the entropy of category probability distribution for the given user denoted by $H(P(c_i|u))$. Users with higher values of $H(P(c_i|u))$ have more diverse interests, where users with lower values of this entropy has fewer categories of interests.

To establish whether our diversification approach works for the different types of users in providing a recommendation that is diverse to match their category of interests, we devise the following method. Given the original ranking R and the diversified ranking S , we quantify the category probability distribution $Q_1(c_i)$ from the original ranking R and the category probability distribution $Q_2(c_i)$ from the diversified ranking S . Ultimately, the diversification approach should produce a category probability distribution that is closer to original distribution of the user $P(c_i|u)$. This is quantified using the cross-entropy between the category probability distribution of the user and the inferred one from the ranking $H(P, Q)$. Therefore, we can check if the diversification approach does produce a more representative ranking of the user interests if $H(P, Q_2)$ is less than $H(P, Q_1)$, i.e. when the following is true:

$$\Delta H(P, Q) = H(P, Q_2) - H(P, Q_1) \leq 0 \tag{8}$$

We quantified this difference for the top 50 users in terms of category diversity (high entropy $H(P(c_i|u))$) and plot the results for the w2 setting of Tabel 1 in Figure 4. We see that in most cases (86% of the cases), the difference is negative and the diversification approach succeeds in providing venues that match the diversity of category interests of those users. For the users in the bottom of the list, i.e. the least 50 users in terms of category diversity (low entropy $H(P(c_i|u))$), we plot the differences in Figure 5, we observe

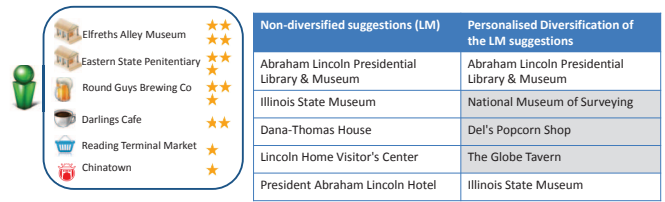


Figure 3: An example of a user profile and two ranked list of contextual suggestions for that user in the city of Springfield

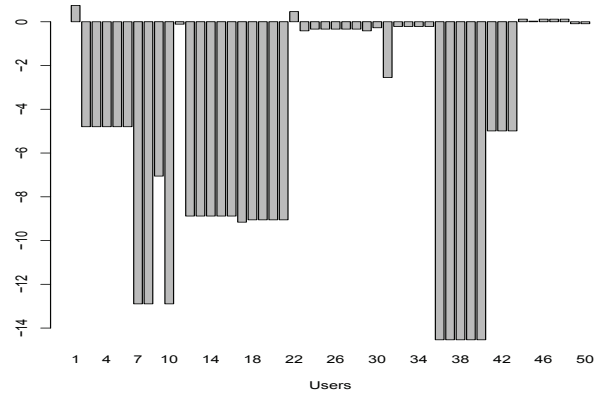


Figure 4: $\Delta H(P, Q)$ for the top 50 users ranked by entropy of category distribution $H(P(c_i|u))$

a different picture, where in most cases, the difference between the two rankings is minimal as the diversification has still produced a ranking that represents the category distribution of the user’s interests. However, in some cases, the original ranking was better (in 30% of the cases $\Delta H(P, Q) \leq 0$). This validates our diversification approach and shows that it is capable of producing a ranking of recommendation that better matches the category interests of the user and it shows that it works particularly well for users with high variety of interests.

Finally, we aim to find if there is any correlation between the improvement in the recommendation effectiveness and the diversity performance. For that we estimate the correlation between the improvement in precision $\Delta(P@5)$ and the differences $\Delta H(P, Q)$. We observe that there is no strong correlation between the two, but since the effectiveness measure does not take diversity into account, we accept this observation and we leave for future work an evaluation that takes diversity into account when measuring the effectiveness of contextual suggestion.

7. CONCLUSION

In this paper, we have developed an approach for diversifying the categories of contextual suggestion in LBSNs. Our approach can promote the categories that the user is interested in, and it is also capable of eliminating redundancy of the recommended venues in the sense that they should cover various categories of interest to the given user. Our approach generalises beyond a specific LBSN as we have devised a venue category prediction by building a textual classifier using a large collection of web pages, the ClueWeb12 dataset. Our thorough empirical evaluation suggests that diversity can improve the effectiveness of contextual suggestion over a LM baseline that only considers the similarity between the user’s pro-

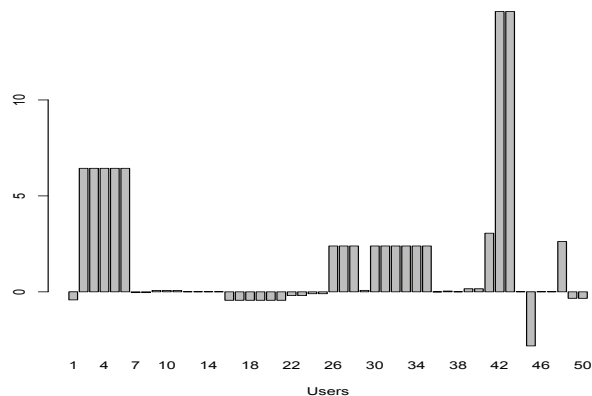


Figure 5: $\Delta H(P, Q)$ for the least 50 users ranked by entropy of category distribution $H(P(c_i|u))$

file and venue. In particular, the diversification improves the effectiveness of contextual suggestion when we take into account the category preferences of the user, i.e. personalised diversification. The results also give insights on the effectiveness of our approach with different types of the users. Our analysis shows that users with higher variety of interests benefits most from diversification. Finally, there is plenty of room for future work. For example, we aim to devise a more comprehensive evaluation procedure to measure diversity in contextual suggestion, since the current measures do not take diversity into account.

Acknowledgments

This work has been carried out in the scope of the EC co-funded project SMART (FP7-287583).

8. REFERENCES

[1] R. Agrawal, S. Gollapudi, A. Halverson, and S. Jeong. Diversifying Search Results. In *Proceedings of WSDM*, 2009.

[2] J. Allan, B. W. Croft, A. Moffat, and M. Sanderson. Frontiers, Challenges, and Opportunities for Information Retrieval: Report from SWIRL 2012 the Second Strategic Workshop on Information Retrieval in Lorne. *SIGIR Forum*, 46(1):2–32, May 2012.

[3] J. Bao, Y. Zheng, and M. F. Mokbel. Location-based and Preference-aware Recommendation using Sparse Geo-social Networking Data. In *Proceedings of SIGSPATIAL*, 2012.

[4] N. J. Belkin. Some(what) Grand Challenges for Information Retrieval. *SIGIR Forum*, 42(1):47–54, June 2008.

[5] B. Bi, M. Shokouhi, M. Kosinski, and T. Graepel. Inferring the Demographics of Search Users: Social Data Meets Search Queries. In *Proceedings of WWW*, 2013.

[6] V. Bicer and V. Lopez. Searching in the City of Knowledge: Challenges and Recent Developments. In *Proceedings of SIGIR*, 2013.

[7] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[8] A. Dean-Hall, C.L.A. Clarke, J. Kamps, and P. Thomas. Overview of the TREC 2012 Contextual Suggestion Track. In *Proceedings of TREC*, 2012.

[9] A. Dean-Hall, C. L. A. Clarke, J. Kamps, P. Thomas, N. Simone, and E. Vorhees. Overview of the TREC 2013 Contextual Suggestion Track. In *Proceedings of TREC*, 2013.

[10] M. D. Ekstrand, J. T. Riedl, and J. A. Konstan. *Collaborative Filtering Recommender Systems*. Now Publishers Inc, 2011.

[11] Y. Ganjisaffar, R. Caruana, and C. Lopes. Bagging Gradient-boosted Trees for High Precision, Low Variance Ranking Models. In *Proceedings of SIGIR*, 2011.

[12] R. Ghorab, D. Zhou, A. O’Connor, and V. Wade. Personalised Information Retrieval: Survey and Classification. *User Modeling and User-Adapted Interaction*, pages 1–63, 2012.

[13] B. Hu and M. Ester. Spatial Topic Modeling in Online Social Media for Location Recommendation. In *Proceedings of RecSys*, 2013.

[14] T. Joachims. A Statistical Learning Model of Text Classification for Support Vector Machines. In *Proceedings of SIGIR*, 2001.

[15] H. Kukka, V. Kostakos, T. Ojala, J. Ylipulli, T. Suopajarvi, M. Jurmu, and S. Hosio. This is Not Classified: Everyday Information Seeking and Encountering in Smart Urban Spaces. *Personal & Ubiquitous Computing*, 17(1), 2013.

[16] T. Kurashima, T. Iwata, T. Hoshide, N. Takaya, and K. Fujimura. Geo Topic Model: Joint Modeling of User’s Activity Area and Interests for Location Recommendation. In *Proceedings of WSDM*, 2013

[17] D. Metzler. Automatic Feature Selection in the Markov Random Field Model for Information Retrieval. In *Proceedings of CIKM*, 2007.

[18] M. G. Noll and C. Meinel. Web search Personalization via Social Bookmarking and Tagging. In *Proceedings of ISWC’07*, 2007.

[19] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo. A Random Walk around the City: New Venue Recommendation in Location-based Social Networks. In *Proceedings of PASSAT*, 2012.

[20] I. Ounis, G. Amati, V. Plachouras, B. He, C. Macdonald, and D. Johnson. Terrier Information Retrieval Platform. In *Advances in Information Retrieval*, pages 517–519. Springer Berlin Heidelberg, 2005.

[21] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., 1993.

[22] R. L. T. Santos, C. Macdonald, and I. Ounis. Exploiting Query Reformulations for Web Search Result Diversification. In *Proceedings of WWW*, 2010.

[23] R.L.T. Santos, C. Macdonald, and I. Ounis. On the Role of Novelty for Search Result Diversification. *Information Retrieval*, 15(5):478–502, 2012.

[24] S. Vargas and P. Castells. Exploiting the Diversity of User Preferences for Recommendation. In *Proceedings of OAIR*, 2013.

[25] Q. Wu, C. J. C. Burges, K. M. Svore, and J. Gao. Ranking, Boosting, and Model Aaptation. Technical Report MSR-TR-2008-109, Microsoft, 2008.

[26] P. Yang and H. Fang. Opinion-based User Profile Modeling for Contextual Suggestions. In *Proceedings of ICTIR*, 2013.

[27] M. Ye, P. Yin, W. Lee, and D.-L. Lee. Exploiting Geographical Influence for Collaborative Point-of-interest Recommendation. In *Proceedings of SIGIR*, 2011.

[28] C. Zhai and J. Lafferty. A Study of Smoothing Methods for Language Models Applied to Information Retrieval. *ACM Transactions on Information Systems (TOIS)*, 22(2), 2004.

[29] M. Zhang and N. Hurley. Novel Item Recommendation by User Profile Partitioning. In *Proceedings of WI-IAT*, 2009.