Modelling Psychological Needs for User-dependent Contextual Suggestion

Di Xu, Jamie Callan

Language Technologies Institute, School of Computer Science, Carnegie Mellon University
Pittsburgh, PA, USA
dix@cs.cmu.edu, callan@cs.cmu.edu

Abstract

This paper presents our approach for the Contextual Suggestion track of 2014 Text REtrieval Conference (TREC). The task aims to provide recommendations on points of interests (POI) for various kinds of users under different contexts. This becomes challenging due to the limited amount of training data provided by TREC and the demanding constraints for a suggestion to be judged as relevant. Our approach does not deviate from existing Machine Learning based methods in principle, but sticks closely to the defined relevance judgement criteria, by focusing primarily on modelling users' preferences on POI categories, and investigating upon their psychological expectations on the textual descriptions of the POIs. The latter is considered as our novelty in this work. Support Vector Regression was used for suggestion ranking, an ad-hoc web information extractor was used to collect POI descriptions, and a description evaluation mechanism was engaged to select proper POI descriptions subject to the nature of the POIs. Our results suggest that our methods are effective in obtaining satisfying user-specific POI rankings and generating descriptions that meet users' psychological expectations.

Index Terms: Information retrieval, information Extraction, intelligent information systems, support vector machines, text mining, machine learning

1. Introduction

With the advancement of mobile and wireless communication technologies, the popularization of location based recommender systems (LBRSs) has taken place in the past couple of years. Yelp, Google Places, Foursquare, etc. are typical examples of LBRSs. These systems by far have been successful, and they do not require very sophisticated ranking mechanisms because simple features such as the number of good ratings on a particular point of interest (POI) are pretty strong indications of a good suggestion to most users, as the majority of users share the common patterns of interests. These facts have, however, also lead to the less efforts made on generating user-dependent LBRSs.

A user dependent LBRSs should not provide suggestions to a user without considering its prior interests on such kind of POIs, even though the suggestion may be well endorsed in a global sense. The preferences regarding entertainments may vary from different user groups. Different cultural background, social economic status, personalities, gender, ages, etc. all play an important role in shaping one's preference towards different POIs. For example, a female user is likely to prefer shopping as opposed to a male user.

The 2014 Text REtrieval Conference (TREC) Contextual Suggestion track investigates search techniques for complex information needs that are highly dependent on context and user interests, as suggested by the guidelines. In this task, a set of geographical regions are provided as contexts, along with a set of user profiles. Each user profile consists of rankings of sets of provided POIs in several contexts. The goal is to learn from the provided context-specific information for each user, and provide a ranked list of POIs for each user under a new context. In addition, the URL and a textual description of each recommended POI are needed along with the title of the POI.

The majority of the systems presented in the past collect POI information from well-known recommender services such as Yelp, Google Places, Trip Advisor, Foursqure, etc. Some systems exploit the open web to a greater extent by performing Information Extraction (IE) directly from hub pages that already contain a list of promising POIs that are publicly desired. In the work done by Luo & Yang [1], a Wiki Travel homepage for a target city is first located, and then POI names are extracted using a variety of heuristics. The collected POI names are then reinvestigated for more information such as URL, category, description and geographical location.

Many existing approaches adopt an Information Retrieval (IR) based framework for retrieving and ranking POI. For example, George et al. [2] crawled and indexed web pages of certain levels of interests, and in their work, user profiles are analyzed to generate queries that reflect users' preferences, and then structured queries are generated to address important fields such as title and anchor text.

In addition to traditional Rocchio-like query generation approaches, a more principled approach it to explicitly extract features from the POIs, user profile and context information. This allows the traditional Vector Space models to be used, and also motivates machine learning based methods such as ranking by Support Vector Machines (SVMRank). Yang & Fang [3] used both positively and negatively rated POIs in the profiles to score new POIs based on their distances to the positive and negative POIs. Later, Jiang & He [4] used linear regression to further include other features to generate a final ranking.

Unfortunately, only a limited amount of context-specific user profiles are provided by TREC, making it difficult to train more sophisticated ranking models. In fact, this makes practical sense because recommender systems in real life often face the same issue of insufficient training data to generate user-specific models. Still, it is noticeable that the user profiles provided this year intentionally span two different contexts: Chicago, Illinois and Santa Fe, New Mexico, which is an indication that the recommender system is expected to make context-dependent suggestions.

We have observed two major trends through the past works: (1) feature extractions and the use of machine learning algorithms are replacing traditional IR based methods; (2) more attentions are being given to finding POIs that are of particular interests to different user groups, rather than relying on a univer-

sal background model. We will follow these trends and propose our machine learning based approach, which focuses on making reasonable and practical assumptions to model user-specific preference patterns and learning from users' general psychological expectations towards POI descriptions conditioned on their categories.

2. Task Formulation

The task in general is to provide a user with recommendations about the entertainments available under a particular context subject to the person's preferences. We are provided with a number of user profiles, and each user profile consists of ratings made by that user on different POIs under different contexts. In detail, each user profile has 50 rated POIs in Chicago, Illinois and 50 in Santa Fe, New Mexico. Our goal is to provide for each user 50 such suggestions for additional 50 contexts. Each suggestion consists of the title, description and URL of the POI. We formulate the task into the following subtasks:

Web crawling and information extraction is to collect POI information including title, description, URL and other features from the web. This is discussed in section 3.2 and 3.6.

Category-based user preference modelling is to learn the preference of each user towards each POI category from their profiles, as discussed in section 3.4. We use the mean and variance of the user's ratings on each particular POI category to model the user's preference.

User and context-specific POI ranking is to put together useful features introduced in section 3.3 for each user-category-context triple. SVM based ranking mechanism is used for scoring and ranking. This will be discussed in detail in section 3.5

Category based description selection as described in section 3.7 is to find the best description text among those collected during the description crawling process described in section 3.6. Text mining techniques are attempted to collect features that reflects the nature of each chunk of description text. For each category, a Regression Tree trained on user profiles will be used to score the alternative descriptions and select the most promising one. We consider this to be a novelty in our paper.

3. System Descriptions

3.1. Important Assumptions

We assume (1) the preferences of the users remain the same as what they used to be at the time when their profiles were generated. This assumption should hold as introduced in the TREC 2014 Contextual Suggestion Guidelines that during the evaluation the same users who created the training user profiles will be summoned to rate suggested the POIs in different contexts.

We assume that (2) most of the time users directly infer the category from the title, and (3) sometimes from both the title and a brief glance of the description if the title appears ambiguous, and (4) if they have to look into the description, it triggers a two-stage process where in the first stage the category is inferred, then in the second stage reflections and judgements are made. Further, we assume (5) a user will not actively pursue or positively rate a POI unless its category is of interests.

The above assumptions make practical sense because in real life most POI titles contain keywords that allow users to directly

infer their categories. Should the title be ambiguous, a quick glance at some keywords in the description will help them identify the category. For instance, if a person does not recognize the POI title "Carnegie Mellon", a glance at the first sentence of the description will help him target keywords such as "University", thus he immediately refers the category as school without looking further. As for whether the description text is interesting by itself, he may look further down the description to spot information such as histories and specialties of the university. According to Cognitive Psychology [5], peoples' judgements are biased by their preference on the categories, such as prior knowledge, past experience, personal attitudes and many other subjective views, leading them to rate a POI by comparing it with their mental prototypes or relating it to past experiences. For example, mercenary people are unlikely to be interested in luxury stores where price tags contradicts to their money saving principles; males are less likely to be interested in clothing stores since wandering among aisles of clothes is simply boring to them; and people studying sciences are less interested in art or history museums due to knowledge limitations. Notice that these are broad generalizations that are true for the majority of people, and my not hold for particular exceptions. But such generalizations are important as they provide prior knowledge that is useful with the absence of other evidences, which is typical in our case and many other real-life applications.

Additionally, under the context of recommender systems where dozens of alternatives are available, if the category of a POI is deemed not interesting, users are not likely to bother wasting time looking in detail, but to resort to other alternatives of interested categories.

These are fundamental modelling assumptions, as our POI ranking model does not consider features from the description text or the website, because we regard being interested in its category the prior condition for a user to be interested in a POI. Apart from that, our description selection model assumes that (6) if user is interested in a POI based on its category inferred from its title or a glance at its description, his expectations on POI descriptions are guided by the category. We will elaborate more in section 3.7.

In terms of geographical relevance, as introduced in the guidelines, any POI with 5 hours driving distance from the provided coordinate is considered appropriate. Therefore we will not make discriminations on POIs that are within the acceptable distance to the provided coordinates of each context, based on their geographical proximities.

3.2. POI Collection

We used **Google Places API** primarily to collect POI titles and other information. We decided not to rely on other POI search engines such as Yelp or Trip Advisor because we focused on POI category analysis, which requires a consistent labelling mechanism. Yelp provides two-level categorical information for most POIs, while Google Places provides significantly fewer number of categories. We chose Google Places because our training data is very limited and fewer numbers of categories suggests less training data needed to train the model.

This step is crucial to the recall performance of the system, and beyond this stage no extra POIs will be considered. In particular, **Nearby Search** and **Text Search** APIs are used to obtain the titles and references of the POIs.

To perform Nearby Search, one needs to define the POI categories to be searched over. We have defined a list of POI categories used by Google Places API, including amusement

park, aquarium, art gallery, bakery, bar, book store, bowling alley, cafe, casino, church, city hall, department store, food, library, mosque, movie theater, museum, neighborhood, night club, park, place of worship, restaurant, RV park, shopping mall, stadium, synagogue, university and zoo. The search radius is set to 320,000 meters which is calculated by assuming an average driving speed of 40 mph in 5 hours.

To perform Text Search, we enumerated based on common sense a list of general queries from the composition {"popular", "interesting", "amazing"} × {"attractions", "entertainments"}, which motivates the search engine to return well known POIs that are within the search radius. Moreover, we have analyzed the Wiki Travel Pages for all the contexts and automatically extracted name entities in bold face that are likely to be POI titles. These titles are then searched over through the Google Places Text Search API to obtain more POIs in each context.

The results obtained from both Nearby Search and Text Search are merged to produce a final pool of POI references.

3.3. POI Feature Extraction

The collected POI references are stable IDs used to obtain detailed information of the POIs via the **Place Details** API. The API provides a certain amount of information regarding the POI, including title, address, categories, user ratings, etc. In order to support user preference modelling and ranking, features need to be developed based on those information. In particular, we have considered the following features for each POI.

- the number of user ratings warped by a power function
- the first moment (mean) of user ratings
- the second moment (variance) of user ratings
- the number of uploaded photos
- the number of associated categories
- the core category based on inverted document frequency (idf) heuristics

These features will be used during the POI ranking along with features extracted for a particular user and context.

We chose to warp the number of user ratings by a power function because in the places that are less visited by tourists, such Santa Fe, New Mexico, restaurants often receives higher numbers of ratings than the actual tourist destinations. This is because most restaurants constantly receives comments from local residents, and since Santa Fe is not a popular tourist destination comparing to many others, its true tourist destinations may receive fewer comments comparing to local restaurants. To correct this problem and let the restaurants to be less dominating due to their higher number of received user comments, a power function is used to warp this number. We have set the warping factor to be 0.3 for contexts that are similar to Santa Fe.

Idf based heuristics are used to select the core category for those POIs that have multiple categories on record by Google. For example, a restaurant may be co-labelled as cafe and establishment. This is because a cafe is a restaurant in most cases and almost all POIs that profit from customers are establishments. To infer the most representative functionality of a POI, using idf is a reasonable solution because the more frequent a category is labelled, the more general this category may be, and vice versa. Here, the category "cafe" appears far less frequently than the other two labels on record. Therefore the core category of this POI is regarded as "cafe".

3.4. User Preference Modelling

The concept of user preference is open-ended. Commonly it refers to a user's levels of interests towards different POI categories. But there is more than one data-driven way of estimating the preferences.

One approach will be consider the relative frequency of the number of good or poor ratings given by a user on a particular category with respect to the user's total number of good or poor ratings. But this method suffers from serious bias as the sampled data is seriously unbalanced regarding POI categories. Therefore the estimation obtained with this method will be significantly biased or even inconclusive for some categories.

Another approach will be considering the mean rating of the POIs of a particular category made by the same user. This approach becomes problematic if a user rates the POIs in an unstable way. For example, a user who is not interested in restaurants may rate all instances as neutral, while another user who specializes in eating may give high rating to some restaurants, but also give poor ratings to other restaurants. Therefore the mean rating may not be informative in these scenarios. To resolve such issues, higher order moments need to be used.

We chose to use the mean, the second and third momentums of ratings made by the target user on POIs of a particular category as features of his preference pattern towards that category. If a rated POI has multiple categories, its rating will be used not only when computing user preference statistics on its core category, but also for all other associated categories.

When it comes to pattern mining, a common approach is to perform clustering on the data points and learn models for each cluster. The intuition is that by merging patterns are similar to each other as one cluster, fewer models are needed to model the patterns, and thus more training data will be available for parameter estimations with lower variances. Motivated by this, we have also tried clustering users based on their preferences encoded by the above features. Unfortunately, we have observed no obvious formation of distinguishable clusters, as illustrated in figure 1. That is why we chose not to perform clustering on user profiles, although intuitively it sounds reasonable. Therefore, we will take statistics from each user profile and model the preference of the user independently.
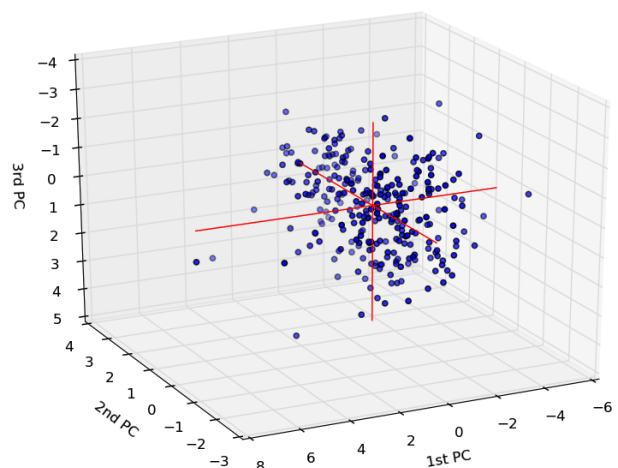

Figure 1: Principled Component (PC) Analysis on Category Preferences for the 299 provided user profiles

3.5. User and Context Specific POI Ranking

So far, we have features for each POI and features that reflect a user's preference pattern on particular POI categories, in order to perform context and user dependent POI ranking, the only thing left is to design features to distinguish different contexts.

In the user profiles, unfortunately, only two different contexts are provided, therefore we could not conduct extensive studies or design sophisticated models on how the nature of the context affects user's preferences over various POIs. Intuitively, in metropolis such as New York, Los Angeles and Chicago, a person tends to visit man-made landscapes such as landmarks, museums, amusement parks and famous restaurants, while in smaller cities with relatively lower population, a person tends to visit some natural landscape such as national parks, natural reserves and resorts. Based on this weak assumption, we adopt a binary feature to simply distinguish small cities from metropolises.

Now that we have features for each POI, user and context, but to train a regression model, we need to determine from the user profiles the training label of each POI-user-context triple. That label should be a rating that best reflects the user's preference on the POI category. However, what we are provided with are user's ratings on the description and the website for each POI, which are indirect reflections of their preference on the categories.

Interestingly, we have observed that most of the time a user's rating on the description is no worse than his rating on the website, but sometimes the rating on the website may be higher. In order to determine a rating that accurately reflects one's general preference towards POI categories, we will pick the maximum value of the two ratings as a best guess. Based on the assumptions that when a user gives lower website rating than the description rating, it implies the user has a clear idea of the category of the POI, only the website does not look impressive to him; however, when the situation is reversed, it implies the user failed to infer the category from the description, possibly due to low quality gibberish texts, but managed to infer the category from the website and restore his interests.

With all the features and training labels mentioned above, we have constructed a ν-Support Vector Regression model [6] to model each user's interest in a particular POI under a particular context, knowing his profile. The underlying implementation is based on libSVM [7], with minor modifications. The model will be evaluated by applying it directly to rank the POIs and computing the resulting precision at the top 20 ranked POIs. Therefore, our ranking system is technically precision-oriented.

The training has been conducted over the 299 user profiles regarding 100 POIs located in 2 different contexts. Degrees of interests are addressed by duplicating the instances where the user is "strongly interested" or "strongly disinterested". To avoid over-fitting as our data is not large enough, we have used linear kernel with leave-1-out cross validation, and selected a relatively large regularization coefficient for a wide margin.

3.6. Description Crawling

Extracting description text from the web is ad-hoc by nature, because the homepages of the POIs are highly diversified in terms of structure and organization, and the occasionally we also need descriptions available from third-party sources such as Wikipedia, and customer reviews from Yelp. Therefore, one may find multiple chunks of descriptive text for the same POI, but address different aspects of it. Unfortunately, the qualities of the texts are not consistent, such that there is no unique source that consistently provides the best descriptive text of all POIs.

An ad-hoc and highly engineered homepage spider has been developed to extract textual information from the homepage of POIs. The program will identify chunks of text on a web page and also navigate to tabs that are labelled such as "about us", "history", "mission", etc. which are likely to contain descriptive text addressing the nature of the POI. On average, the spider finds about 3 chunks of text from a homepage. The number of chunks the spider crawls is affected by the design of the homepage, and typically the number of tabs available on the page. Apart from the homepage of a POI, Wikipedia also contains introductory information regarding some popular POIs. Google, Bing and Yahoo Local Search occasionally provides such introduction crawled from Wikipedia and customer reviews from Yelp and Google+ along with the normal search results on a particular POI. Such introductions and reviews are also collected by the web crawler. The average length of the crawled descriptions is close to 500 characters long.

Now that for each POI we have a pool of descriptions that are of different styles and qualities. The motivation of collecting more than one chunks of descriptive text for a POI is to allow us to find more appropriate ones among others to ensure the quality of the description text, which is an important part of the evaluation.

3.7. Category Based Description Selection

Recall that we have made the assumption that users' expectation of a description is affected by the categories that the POI belongs to. For example, given a museum or cultural district, one may expect narrative description texts that address the history of the place; and given a famous restaurant or shopping center, one may expect persuasive text that address how wonderful and cheap the foods or products are. On the other hand, the reverse may also be true. A user is more likely to be interested in a restaurant or shopping center if he is informed about how delicious the foods are or how cheap the products sale; and one may not be convinced if the description about a museum or cultural district lacks literal seriousness as they should have.

Furthermore, we have also formed groups of categories. This is because although different categories may lead to different user expectations on the description text, many of them do not differ much from each other. For example, the desired description texts for zoos and aquariums may not differ significantly, and similarly for museums and galleries. Therefore, one may consider amusement park, restaurant, shopping center, night club as one group, while museum, historical district, place of worship, library as another group, and park, zoo, aquarium, natural reserve as yet another group. We have manually distributed all existing POI categories into 3 groups. One group is about culture, history and art; one group is about money spending events; and the last one is about nature and wild life. The benefit of grouping categories is that it brings more training data for each group, because we have only a limited number of labelled category-description pairs for model training, and their corresponding POIs may not span all categories to be studied.

The labelled training data were generate by randomly selecting 150 POIs from all the top 50 POIs returned by the POI ranking system for all the contexts, and labelling their descriptions texts manually on a scale of -1 to 1, with -1 as unsatisfying, 0 as mediocre, and 1 as satisfying.

Text Mining techniques have been adopted in selecting features from the description texts that reflects their characteristics. In particular we have extracted some heuristic features,

such as the number of exclamatory marks and question marks, the longest possible timespan indicated by four-digit numbers range from 1500 to 2014, the number of dollar signs, and the length of the description text. Moreover, we have also used the Stanford Parser [8] to inspect the grammatical correctness of the sentences in the text, and calculate the proportion of grammatically correct sentences. In addition, we have also used the Stanford Log-linear Part-Of Speech (POS) Tagger [9] to extract features, such as the proportion of nouns (NN+NNS), proper nouns (NNPS/NNP), cardinal numbers (CD), adjectives (JJ), comparative adjectives (JJR) and superlative adjectives (JJS), etc.. On top of that, we have also gathered all adjectives and adverbs and manually crafted a set of general commendatory terms such as "excellent", "awesome", "delicious", "exciting", and a set of general derogatory terms such as "awful", "bad", "horrible", "boring". Based on that we have also calculated the proportion of commendatory terms and derogatory terms in each description, during the process we also consider negations that may appear before commendatory and derogatory terms. If a commendatory term is preceded by a negation, we will count it as a half of a derogatory term.

Three Regression Trees were trained independently using Scikit-learn [10] with the manually labelled description texts mentioned above, one for each category group. A tree takes a description whose POI's category belongs to the tree's designated category group, and outputs a predicted level of satisfactoriness. For a particular POI, all of its collected description texts were run through the tree and ranked by their scores, and the top ranked description was selected.

4. Evaluation Results and Analysis

Table 1 shows the evaluation results of our submission "dixlticmu" and the average of the medians of all submissions across 299 POI-user pairs. We have also performed statistical significance analysis using Paired t-Test. The results suggest that our user and context specific POI ranking system generally performs better than the average. In particular, our system is significantly better in terms of precision (prec@5) with strong statistical evidences, and Time-Based Gain (TBG) with some evidence, and is slightly better in terms of Mean Reciprocal Recall (MRR). This indicates that our system is able to return more POIs of interest, with proper description text and URL. This is expected as we trained our user-specific POI ranking SVM based on prec@5, which also explains why our MRR performance is not as good as precision, since reciprocal recall was not investigated or optimized in our system.

Unfortunately, at this point we do not have a direct feedback on how our POI description selection system works comparing to others. Yet based on our assumptions, we can still infer that the description selection system has managed to provide reasonably well description texts for the returned POIs. This is because if the top ranked POIs deviates from a user's preference, we assume the user will not bother giving the description text a high rate since the rating reflects his level of interests. Therefore if the performance of the user and context specific POI ranking system is horrible, we do not have evidence on the quality of the description texts. However, since our POI ranking model has shown to be working well, we therefore have some evidence to claim that the description selection system is working reasonably well. Of course, it is also possible that the POI ranking model works even better than it appeared to be, due to degradation caused by a poorer description selection system.

Motivated by the released evaluation results for each con-

	prec@5	MRR	TBG
avg. medians	0.3491	0.535	1.3685
dixlticmu	0.3906	0.5431	1.4828
Advantage	+12%	+1.5%	+8.4%
P-value	0.0011	0.3986	0.1254

Table 1: Evaluation results released by TREC.

text, we have conducted a study to see if our recommender system performs better or worse for certain contexts than others in terms of prec@5. In particular, we have found that for our system performed more than 0.1 absolute better than the averaged median for Boise, Walla Walla, College Station, Bloomington, Portland, Redding, San Diego, Virginia Beach, Yuma, Clarksville, Buffalo, Sacramento, Anchorage, Honolulu and Lawton. And for Erie, Lancaster, Kalamazoo, Homosassa Springs, Toledo, Albuquerque and Kansas City, our system performed more than 0.1 absolute worse than median. The above locations have been labelled in figure 2.
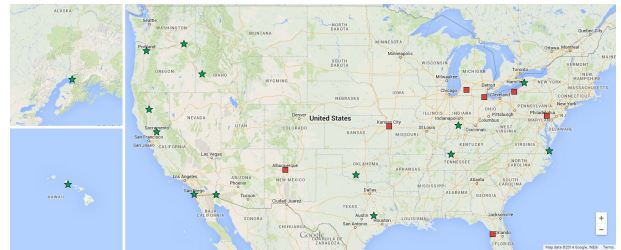

Figure 2: The distribution of contexts (cities) for which our recommender system performs more than 0.1 better (green stars) or worse (red cubes) than averaged median in terms of prec@5. The map was provided by Map data ©2014 Google.

Unfortunately, we cannot conclude from these observations that our system is better or worse on contexts in terms of the city scale, economic status, population density, climate, geology or even physical distance to the border. It is noticeable that most contexts that our system performed poorer are close to the Great Lakes region, but this information alone does not make much sense to us. Therefore, from this analysis we can claim with enough evidence that our system performs equally well for contexts of different scale, economic status, population density and geographical attributes, which may be considered as another merit of our system.

5. Conclusions

In this paper we have summarized our recommender system for the TREC 2014 Contextual Suggestion task. Our system consists of a web information crawling and extraction module that prepare the resources for making suggestions, a SVM based POI ranking module, and a Regression Tree based POI description selection module. We have also made some practical assumptions that have made our models simpler and more efficient. Generally speaking, our approach focused on using Machine Learning approaches to model user's needs and expectations from a psychological perspective.

The evaluation results suggest that our precision oriented system is competent in terms of prec@5 and TBG. They also suggest that the user preference models were successful in capturing particular interests of most users, which also implies that

the underlying features are useful. Although the results are not directly indicative of the performance of our category based description selection method, but at least we can tell that it is reasonably good, otherwise the results could have been much worse than the median. More importantly, the results give us some evidence to claim that our human Psychology motivated approach works well in general.

In future, we will focus more on improving MRR performance of our system, and also on engineering better features for user preference modelling and description selection. We will also focus on incorporating more POI information from multiple sources, and adopt a more fine-grained POI categorization schema for more accurate preference modelling. We are also interested in developing a recounting system that extracts information from the web that are align with users' preferences, and automatically generates description text accordingly.

6. Acknowledgements

7. References

[1] J. Luo and H. Yang, "Boosting venue page rankings for contextual retrieval–georgetown at trec 2013 contextual suggestion track," Georgetown University, Tech. Rep., 2013.

[2] G. Drosatos, G. Stamatelatos, A. Arampatzis, and P. S. Efraimidis, "Duth at trec 2013 contextual suggestion track," Democritus University of Thrace and Athena Research and Innovation Center, Tech. Rep., 2013.

[3] P. Yang and H. Fang, "An exploraton of ranking-based strategy for contextual suggestion," University of Delaware, Tech. Rep., 2012.

[4] M. Jiang and D. He, "Pitt at trec 2013 contextual suggestion track," University of Pittsburgh, Tech. Rep., 2013.

[5] L. W. Barsalou, *Cognitive psychology: An overview for cognitive scientists*. Psychology Press, 2014.

[6] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural computation*, vol. 12, no. 5, pp. 1207–1245, 2000.

[7] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[8] R. Socher, J. Bauer, C. D. Manning, and A. Y. Ng, "Parsing with compositional vector grammars," in *In Proceedings of the ACL conference*. Citeseer, 2013.

[9] K. Toutanova, D. Klein, C. D. Manning, and Y. Singer, "Feature-rich part-of-speech tagging with a cyclic dependency network," in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*. Association for Computational Linguistics, 2003, pp. 173–180.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *The Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.