Buon Appetito - Recommending Personalized Menus*

Michele Trevisiol
trevisiol@acm.org

Luca Chiarandini
luca.chiarandini@upf.edu

Ricardo Baeza-Yates
rbaeza@acm.org

Web Research Group
Dept. of Information and Communication Technologies
Universitat Pompeu Fabra, Barcelona, Spain

Yahoo Labs
Barcelona, Spain

ABSTRACT

This paper deals with the problem of *menu recommendation*, namely recommending menus that a person is likely to consume at a particular restaurant. We mine restaurant reviews to extract food words, we use sentiment analysis applied to each sentence in order to compute the individual food preferences. Then we extract frequent combination of dishes using a variation of the Apriori algorithm. Finally, we propose several recommender systems to provide suggestions of food items or entire menus, *i.e.* sets of dishes.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

Keywords

Recommender Systems, Food-word Recognition, Sentiment Analysis, Menu Recommendation.

1. INTRODUCTION

The quality of a restaurant is strongly related to the quality of the served food and drinks. In this work, we tackle the problem of *menu recommendation*, *i.e.*, the task of recommending good menus to people looking for a restaurant. In the following we will refer to "dishes" or "food items" for food and drinks indistinguishably and to "menu" as the set of food items a person is going to consume. Recommending menus to people may be useful in a number of ways. For example, it allows customers to pick the restaurants that offer dishes which are more in line with their tastes, or it may make them try new dishes which they have never tried before but are among the top-rated choices offered by the restaurant. In this work we use User Generated Content (UGC) to recommend menus and dishes to people. Analyzing and extracting the meaningful information from the user reviews of the places. We use Natural Language Processing (NLP) to extract information from public reviews of restaurants, performing Sentiment Analysis to estimate the user preference associated to each dish. We then build menus by finding sets of dishes that appear often together weighting them with the *positive* and *negative sentiments* previously extracted. Finally, we propose several recommender systems able to recommend dishes and menus and evaluate them quantitatively. Our results indicate that sentiment analysis at the sentence level helps to improve the quality of the recommendations. We are not aware of other research that exploit user reviews, extracting sentiments of each food item, in order to build profiles for people and restaurants to design a menu recommender system. Our approach is versatile as can be easily extended to recommend products or services of businesses based on the reviews of users, as is not strictly related to food items and restaurants.

*This work was done while the first two authors were PhD interns at Yahoo Labs Barcelona.

HT'14, September 1–4, 2014, Santiago, Chile.
ACM 978-1-4503-2954-5/14/09.
ACM http://dx.doi.org/10.1145/2631775.2631784 .

Related Work Previous work about place recommendation within Yelp, are based on hidden factors [6], extraction of subtopics [4], or optimization for mobile devices [1]. A different problem related to reviews is extracting keywords in order to improve readability [10]. The majority of related work is about recipe recommendation (*e.g.*, [3, 9, 8, 5]) where the ingredients or the recipes are usually explicitly rated by users. However they deal with explicit user ratings and require a dataset of recipes' feedback. The studies done on recipes are very different from our, and they cannot be applied on a general textual review service. Our work differs from previous on basic aspects, first we use User Generated Content, which includes a degree of complexity due to noise and language use. We do not rely on structured databases of recipes, in which a complete knowledge is available, but deal with fragmented information extracted solely from the user reviews of restaurants. Moreover, we are dealing with the problem of menu recommendation not in an abstract way, as a standalone problem, but contextualized to the user environment. This means that we are not just recommending food to people but rather recommending dishes *at a particular restaurant*. The recommendation task is constrained by these factors but the result is more related to everyday life applications.

2. ANALYSIS

In this paper, we use a collection of user's reviews of places on Yelp, a popular online urban guide that contains a list of business (*e.g.*, restaurants, coffee shops) and allows people to write reviews about their experiences at each place. Reviews also contain a rating, expressed as number of "stars" (from 1 to 5, 1 being the worst and 5 being the best rating). In total, the dataset contains $11,537$ businesses, $43,873$ users, and $229,907$ reviews. In order to reduce noise and sparsity, we split a review in sentences, we remove stopwords in each sentence, and we lemmatize all the words. In order to capture the most popular food in reviews, we use three publicly available sources[1] building a dictionary of around $9,000$ food items. We then extract the food words from each review (at the sentence-level) obtaining an average of 4.7 food words per review, meaning that users tend to comment more than one dish. Sentiment analysis is the use of NLP techniques to identify subjective information from text. We are interested in understanding the *polarity* of a text, *i.e.*, the amount in which it is positive or negative, and to do that we use the LIWC 2007 [7] dictionary of sentimentally-annotated words. Given a text, we define the sentiment score with $S = \frac{p-n}{p+n}$, where p is the number of positive words, and n is the number of negative words. Moreover, we compare the sentiments extracted by the entire review with the sentiments extracted on each sentence of the same review. We observe that in the case of sentiment of sentences, there is a majority of positive and negative, while review sentiments are more mixed (*i.e.*, $-1 < S < +1$). Indeed, mixed sentiments occur only in 12% of the sentences, against 57% in the case of review-level sentiment analysis. Our experiments shown that splitting by sentence allows us to get a more precise, clean, and localized characterization of the text.

From Dishes to Menus. A menu is a set of dishes which are served during the same meal, and we are interested in detecting menus that people often like. We apply an extension of the Apriori algorithm [2] that deals with *fuzzy* sets, sets whose elements have degrees of membership that in our case are represented by the sentiments extracted from the sentences. Note that for each review, the membership of each food word is the averaged sentiment that it receives in each sentence, normalized to fall into the interval $[0, 1]$.

3. RECOMMENDATION EXPERIMENT

We implement 3 recommender systems and a simple baseline which predict which food items will be present in each given new review. In order to have a meaningful dataset, we consider only users with an average number of food items per review greater or equal to the global average (4.7). We pick 80% of the data as train set (from which we build the user and place profile to train the recommenders), and the remaining as test set. Our evaluation is done on the top K food items (@K) where K represents the number of food items for each test review. All the approaches are based on user and restaurant profiles built on the training data, the first one on the dishes frequently consumed by the user, whereas the latter one on the dishes frequently ordered in the restaurant. Based on the findings showed in Section 2, we propose the following algorithms: (a) `avg-sent`, return the most frequent positive food items that belong to both the user and restaurant profiles, but discarding food items that have a sentiment smaller than the average sentiment in either the user or the restaurant profile; (b) `user-words`, a standard collaborative-filtering recommender that estimates the similarity of the users' tastes in order to rate any food item that belong to the restaurant (the items are weighted with the positive sentiments); (c) `menu-words`, uses the frequent and good menus extracted with the Fuzzy Apriori algorithm, in this case the profiles are built by sets of food words (item sets), the menus. Finally we implement also a baseline, `zero-sent`, that does not consider the sentiments and simply recommend the most frequent items that belong to both the user and the restaurant profiles.

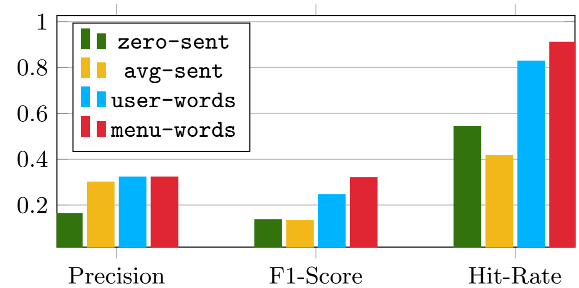

Figure 1: Performance of the sentiment-based recommender systems among different evaluation metrics.

Results Discussion. The results of the recommender systems (Figure 1) are evaluated with Hit-Rate in addition of Precision and F1-Score, in order to highlight the coverage of the algorithms. At a first glance the sentiment-based approaches outperform the baseline showing that sentiment analysis boosts the precision of the recommended food items, thus increasing the probability of recommending dishes that the user will actually order at the restaurant. However, the `full-sent` has significantly worse performance in term of F1-Score and Hit-Rate compared to the `*-words` algorithms, this is due because its strong filtering that might reduces at zero the items to recommend. Among the last two algorithms, the results show interesting insights. Even if `user-words` is slightly better in term of Precision, it decreases in term of Recall and F1-Score. Whereas the recall and precision of `menu-words` are very close to each other meaning that including the most frequent and good good menus considerably boosts the overall accuracy. Moreover in terms of Hit-Rate, `menu-words` is showing the biggest coverage, with around 90% of the users receiving at least one relevant recommendation, whereas `user-words` is slightly smaller.

4. CONCLUSIONS

We introduced the novel problem of menu recommendation for restaurants, describing for the first time how to build a sentiment-based menu recommender system. We use the Fuzzy Apriori algorithm to extract frequent and good food items. We compared different algorithms showing how the sentiment analysis and the menu lead to increase the performance in term of Precision, F1-Score and Hit-Rate. Further work includes inverting the recommendation task such as to recommend restaurants given the dishes a person may want to eat, or even, to recommend the most appealing menus to the restaurants based on the reviews.

Acknowledgments

This work was partially funded by Grant TIN2012-38741 (Understanding Social Media: An Integrated Data Mining Approach) of the Ministry of Economy and Competitiveness of Spain.

[1] *WordNet*: (`wordnet.princeton.edu`), *Oregon State University Food Glossary* (`food.oregonstate.edu`), and *BBC Food*: (`www.bbc.co.uk/food`).

5. REFERENCES

[1] C. Biancalana, A. Flamini, F. Gasparetti, A. Micarelli, S. Millevolte, and G. Sansonetti. Enhancing traditional local search recommendations with context-awareness. In *User Modeling, Adaption and Personalization*, pages 335–340. Springer, 2011.

[2] M. Delgado, N. Marín, D. Sánchez, and M.-A. Vila. Fuzzy association rules: general model and applications. *Fuzzy Systems, IEEE Transactions on*, 11(2):214–225, 2003.

[3] J. Freyne and S. Berkovsky. Recommending Food : Reasoning on Recipes and Ingredients. *User Modeling, Adaptation, and Personalization*, 6075:381–386, 2010.

[4] J. Huang, S. Rogers, and E. Joo. Improving restaurants by extracting subtopics from yelp reviews. In *iConference*, pages 1–5, Germany, 2014. Round One Yelp Data Challenge Winners, 2013.

[5] F.-f. Kuo, C.-T. Li, M.-K. Shan, and S.-y. Lee. Intelligent menu planning: Recommending Set of Recipes by Ingredients. In *Proceedings of the ACM multimedia 2012 workshop on Multimedia for cooking and eating activities - CEA '12*, page 1, New York, New York, USA, 2012. ACM Press.

[6] J. McAuley and J. Leskovec. Hidden factors and hidden topics. In *Proceedings of the 7th ACM conference on Recommender systems - RecSys '13*, pages 165–172, New York, New York, USA, 2013. ACM Press.

[7] J. W. Pennebaker, M. E. Francis, and R. J. Booth. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, page 71, 2001.

[8] C.-y. Teng, Y.-r. Lin, and L. A. Adamic. Recipe recommendation using ingredient networks. In *Proceedings of the 3rd Annual ACM Web Science Conference on - WebSci '12*, pages 298–307, New York, New York, USA, 2012. ACM Press.

[9] M. Ueda, M. Takahata, and S. Nakajima. User âĂŹ s Food Preference Extraction for Personalized Cooking Recipe Recommendation. In *Proc. of the Second Workshop on Semantic Personalized Information Management: Retrieval and Recommendation*, 2011.

[10] J. Wang. *Clustered Layout Word Cloud for User Generated Online Reviews*. PhD thesis, Virginia Polytechnic Institute and State University, 2012.