

TipMe: Personalized advertising and aspect-based opinion mining for users and businesses

Dimitris Proios

Department of Informatics and
Telematics

Harokopio University of Athens
Athens, Greece

Magdalini Eirinaki

Computer Engineering
Department

San Jose State University, CA,
USA

Iraklis Varlamis

Department of Informatics and
Telematics

Harokopio University of Athens
Athens, Greece



- Introduction
- Related Work
- Proposed Model
 - Recommender Workflow
 - Aspect and opinion Extraction
 - Aspect Aggregation
 - Recommendation engine
- Aspect Aggregation
- Recommendation engine
- Implementation
- Prototype Application
- Future Work



Introduction

Problem on the User-side

Techniques of “targeted advertising” reach selected consumers based on their demographics, psychographics and behaviour (e.g. purchase history).

Promotions are purely content- or context-based.

In the end a user has to spend a lot of time reading reviews.



Introduction

Problem on the Business-Side

To act proactively and position their products where there is still place in the market. There is a need to understand:

- i) how potential clients are grouped,
- ii) what are the specific interests of each group,
- iii) how competitive products are to each other,
- iv) what are their pros/cons and impact to the end-users.



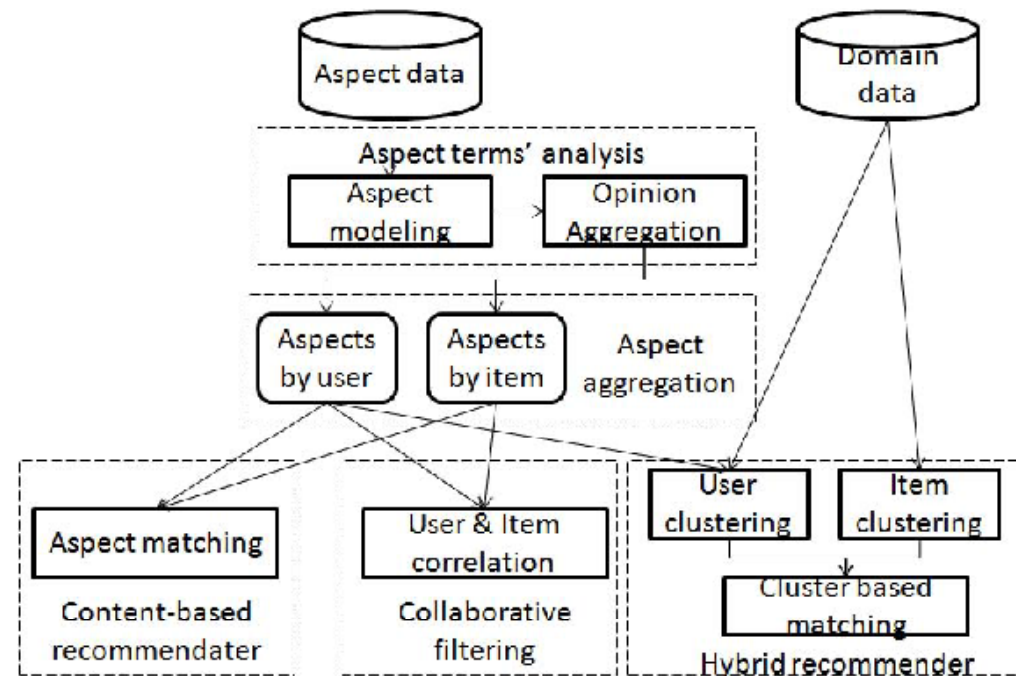
Introduction

Online data can help both:

- Users through getting recommendations of products/services according to their needs.

- Businesses by taking feedback on their weak points

& advertising their strong points to the right users.



Related Work

State of the art ad-targeting systems combine variations of the following elements:

- Location,
- demographics,
- interests (as implied by groups joined/apps used in a social network application, keywords used in searches, etc.),
- cookies (of previously visited websites), (device used, web page visited, etc.)



Related Work

Several interesting works on aspect identification and sentiment polarity of reviews.

The adaption of aspects and opinion scores for recommendation purposes is only a recent trend.

Enrichment of the dataset through sentiment extraction, gives additive value to current ad-targeting systems.



Related Work

In our work we try to bridge together:

- aspect based opinion
- mining and recommendation systems areas
- intuitive visualizations of the knowledge extracted

It's a two-way tool:

- A personalized recommender system for the end-user
- A market analysis tool for the businesses,
overviewing pros&cons,
compare with the competition
and 'drill-down' to the opinions of the market-share.



Proposed Model

The idea is to identify the main aspects and associated values a user is interested in.

Example:

A user expressed an opinion about a business. He mentions that a restaurant had a “**tasteless white wine**, our system based on this comment will provide the user :

A. With more aspects of the business using other users’ reviews.

So we propose these proposed aspects of this specific business to the user. This action can also happen prior to the review combined with geographical location data (check-ins).

B. With alternatives, based on the positive/negative aspects as that were derived from the reviews of all other users with similar interests.

This way we and target multiple aspects that fit the user’s profile.



Proposed Model

Similarly we assume that the target of the service is also the business that received the specific review.

The system will provide business owners with the aspects that users commented about their business.

By this way, business owners will be aware of their '**strengths**' and '**weaknesses**' and target their advertisement campaigns to the appropriate users.

Provide business owners an overview of aspects provided by users mentioned in reviews of their business category.

This way they will be able to identify '**opportunities**' and '**Threats**'



Recommender workflow

Each review r in the set of reviews R
and it is a text provided by a user for a business.

$$r_k = \langle u_i, b_j, text_k \rangle : u_i \in U, b_i \in B$$

LEXICAL ANALYSIS

“aspect” -> Set(Nouns | Verbs)

Usually aspects are accompanied by adjectives or adverbs which are located in the same lexical clause and provide a positive, neutral or negative orientation.

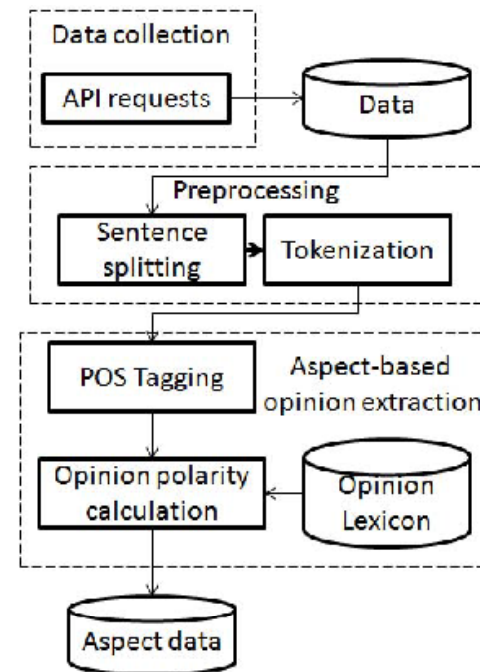
e.g. “reluctant service”, “the worst food”, etc.



LEXICAL ANALYSIS

The textual analysis of the review leads to the detection of aspects with associated opinion values.

$$r_k = \langle u_i, b_j, \langle a_l, s_l, v_l \rangle \rangle$$



Aspect and opinion extraction

First step in aspect-based analysis:
Identify and cluster the most interesting aspects to “dominant aspects”, using topic-modelling or hierarchical clustering Techniques.

The result is a mapping between the dominant aspects DA, where da is one of the M total number of aspects.

$$r_k = \langle u_i, b_j, \langle da_m, s_m, v_m \rangle_{m=1..M} \rangle$$

Aspect aggregation

By aggregating the aspects we get a set of triplets:

$$interests_{u_i} = \langle da_m, s_{im}, v_{im} \rangle_{m=1..M}$$

Similarly, we define the “impression” about a business b as:

$$impression_{b_j} = \langle da_m, s_{jm}, v_{jm} \rangle_{m=1..M}$$



Recommendation engine

Aspect matching:

Based on the aggregated knowledge, we can simply search items that match user's interests.

Aspect-based collaborative filtering

The general task of collaborative filtering in our case can be defined as follows:

Given a set of users,

a set of businesses ,

a matrix of users' interests,

the task is to predict the rating of a user for a business .

(using SVD, FunkSVD or SVD++ algorithms)



Implementation

To test the prototype of our application we used a dataset provided by Yelp, a platform with social-network structure made for businesses, containing more than 40.000 businesses and 250.000 users.

We restricted our analysis on the Tips. The reasons behind this decision are:

- typically provide positive or negative opinions,
- Most of the time a tip judge only one or two aspects of the targeted business, focusing on the most important ones and
 - it's faster to process tips than full reviews.



Implementation

We kept information about categories of businesses and we identified users by Yelp user_id.

We parsed every tip with the Stanford Core NLP parser which generates a part-of speech annotated text.

Example:

For the tip, “Great food, huge portions and a gift shop and showers.” The result is The following:

JJ NN JJ NNS CC DT NN NN CC NNS .
 Great food, huge portions and a gift shop and showers.

```
(ROOT
  (NP (JJ Great) (NN food)) (, ,)
  (NP (JJ huge) (NNS portions)) (CC and)
  (NP (DT a) (NN gift) (NN shop)) (CC and)
  (NP (NNS showers)) (. .)
)
```



Implementation

Next we processed the POS annotated tips and found the polarity of the opinion they convey.

We checked if the nouns, verbs or adjectives of a tip were opinionated words contained in a combined lexicon. (“AFFIN-111” and “Harvard General Inquirer”)

If any sentiment words were spotted inside a sentence, the respective polarity score is added to the aspect of the tip.

```

(ROOT
  (NP (JJ Great) (NN food)) (, ,)
  (NP (JJ huge) (NNS portions)) (CC and)
  (NP (DT a) (NN gift) (NN shop)) (CC and)
  (NP (NNS showers)) (. .)
)

```

Diagram illustrating the implementation of sentiment analysis on a sentence. The sentence is: "Great food, huge portions, a gift shop and showers." The words are annotated with their part-of-speech (POS) tags and sentiment scores. The scores are: "Great" (+0.8), "food" (0), "huge" (+0.1), "portions" (0), "a" (0), "gift" (0), "shop" (0), and "showers" (0). The words "Great", "huge", "gift", and "shop" are highlighted in green, while "showers" is highlighted in grey.

Prototype Application

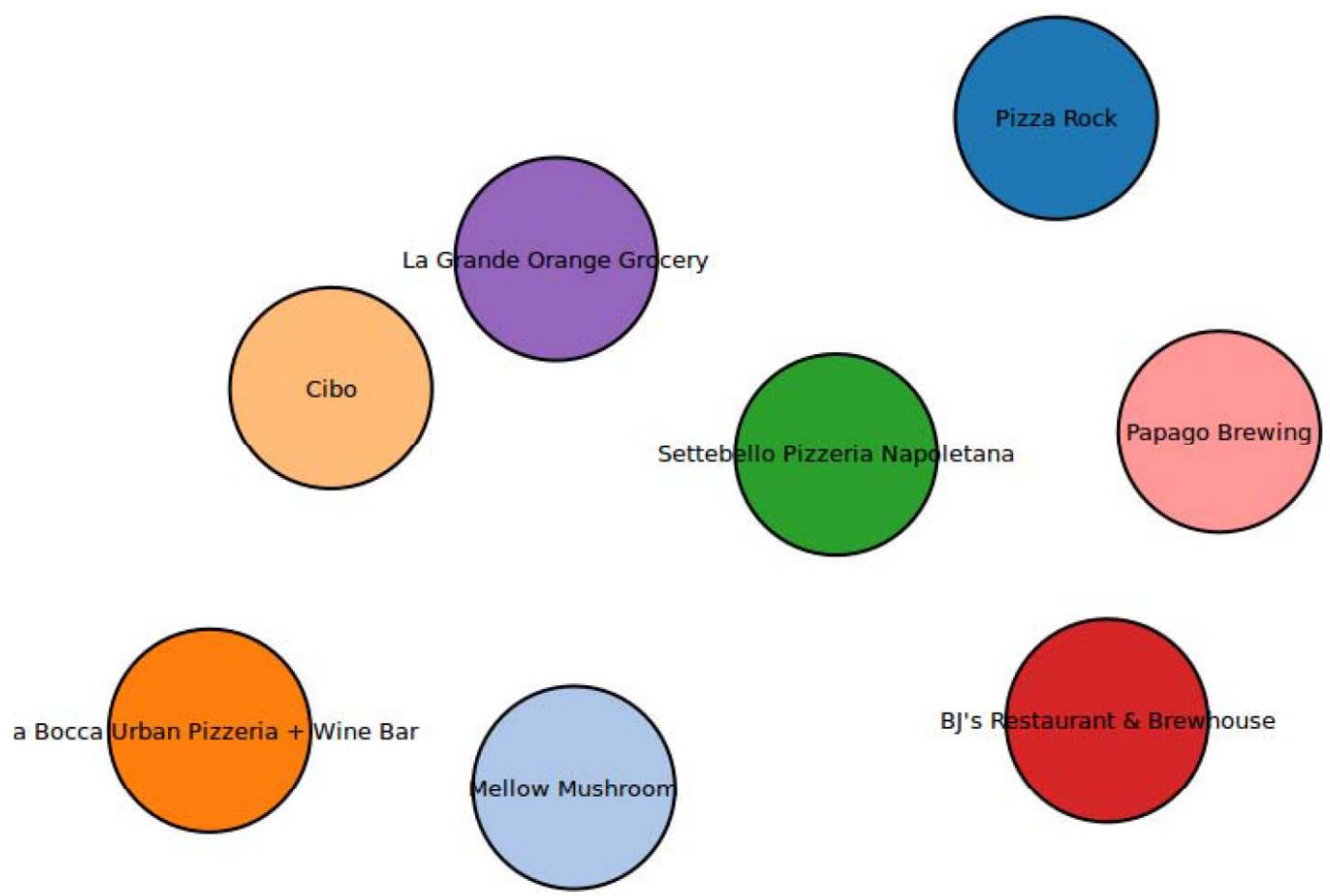
For visualization purposes we used the d3.js.

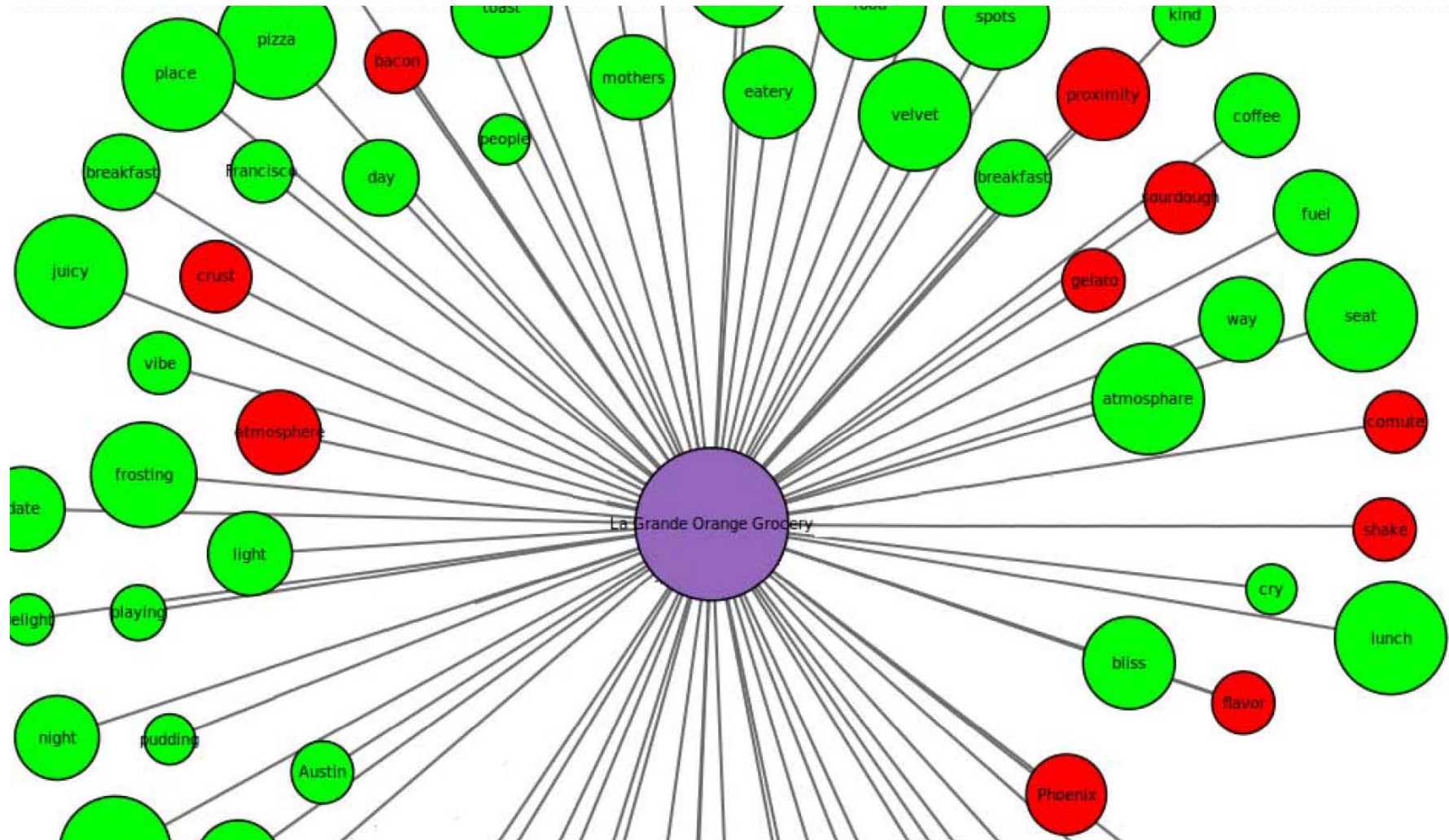
The prototype application builds on a graph-based visualization

Which provide services on both clients and business owners.

We implemented three different scenarios to demonstrate the usability of the tool.







Future work

Part of our future work we plan to focus on the improvement of aspect extraction and aspect-based opinion mining performance.

Broader lexicons (eg. containing emoticons) will improve significantly the performance.

We also intend to consolidate synonym words by incorporating a thesaurus and explore the effectiveness of aspect aggregations into more high-level categories.

We intend to use live feed from multiple APIs (Yelp, Foursquare, Google+, Twitter, Facebook) to combine multi-source information, as well as provide this service to the public.

