RECOMENDAÇÃO DE REVIEWS PARA DONOS DE PONTOS DE INTERESSE

THIAGO RESENDE PEREIRA PRADO

RECOMENDAÇÃO DE REVIEWS PARA DONOS DE PONTOS DE INTERESSE

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

Orientadora: Mirella Moura Moro

Belo Horizonte

Março de 2017

THIAGO RESENDE PEREIRA PRADO

REVIEW RECOMMENDATION FOR POINTS OF INTEREST'S OWNERS

Dissertation presented to the Graduate Program in Ciência da Computação of the Universidade Federal de Minas Gerais in partial fulfillment of the requirements for the degree of Master in Ciência da Computação.

Advisor: Mirella Moura Moro

Belo Horizonte

March 2017

Ficha catalográfica elaborada pela Biblioteca do ICEx - UFMG

Prado, Thiago Resende Pereira.

P896r Review recommendation for points of interest's owners
/ Thiago Resende Pereira Prado. – Belo Horizonte, 2017.
xxiv, 48f.: il.; 29 cm.

Dissertação (mestrado) - Universidade Federal de Minas Gerais – Departamento de Ciência da Computação.

Orientadora: Mirella Moura Moro

1. Computação – Teses. 2. Sistemas de recomendação. 3. Recuperação da informação. 4.Redes sociais on-line. I. Orientador. II. Título.

CDU 519.6*73(043)

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FOLHA DE APROVAÇÃO

Review recommendation for points of interest's owners

THIAGO RESENDE PEREIRA PRADO

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

PROFA. MIRELLA MOURA MORO - Orientadora
Departamento de Ciência da Computação - UFMG

PROF. ANA PAULA COUTO DA SILVA
Departamento de Ciência da Computação - UFMG

Giseli Lopes
PROFA. GISELI RABELLO LOPES
Departamento de Ciência da Computaçao - UFRJ

Belo Horizonte, 31 de março de 2017.

To those who inspired it...

Acknowledgments

I would like to thank all the people who contributed in some way to the work described in this thesis. First and foremost, I thank my academic advisor, Professor Mirella M. Moro, for accepting me into her group. During my tenure, she contributed to a rewarding graduate school experience by giving me intellectual freedom in my work, engaging me in new ideas, and demanding a high quality of work in all my endeavors. Additionally, I would like to thank all the friends I made in the CS+X Laboratory over these last years, who made the lab a fun and cool environment to work on and from whom I learned so much. I also thank CAPES and CNPq for funding my scientific endeavors.

Finally, I would like to acknowledge friends and family who supported me during my time here. First and foremost I would like to thank Mom, Dad and Helena for their constant love and support. I am lucky to have met André Bueno, Luis Fernando, Pericles and Pedro Lopes who made my time here at UFMG a lot more fun.

"Let me tell you something you already know. The world ain't all sunshine and rainbows. It's a very mean and nasty place and I don't care how tough you are it will beat you to your knees and keep you there permanently if you let it. You, me, or nobody is gonna hit as hard as life. But it ain't about how hard ya hit. It's about how hard you can get it and keep moving forward. How much you can take and keep moving forward. That's how winning is done! Now if you know what you're worth then go out and get what you're worth. But ya gotta be willing to take the hits, and not pointing fingers saying you ain't where you wanna be because of him, or her, or anybody! Cowards do that and that ain't you! You're better than that! But until you start believing in yourself, ya ain't gonna have a life."

(Sylvester Stallone, Rocky Balboa)

Resumo

Revisões online tornaram-se uma maneira poderosa para os usuários disponibilizarem suas opiniões para todos. Tais revisões são extremamente valiosas para os consumidores quando eles estão procurando informações antes de adquirir um produto ou serviço. Em uma tentativa de ajudar os consumidores a identificar as melhores revisões, muitos sites permitem que os usuários votem se uma revisão é útil. Então, os algoritmos de recomendação são utilizados para facilitar a tarefa do usuário encontrar revisões que possam ser de seu interesse. Em resumo, tais aplicativos de revisão online geralmente recomendam as revisões mais úteis para um cliente ler. No que diz respeito a revisões de locais (ou pontos de interesse), para o proprietário (ou administrador, gerente, etc.) do estabelecimento, é importante ter uma maneira rápida e confiável de identificar as revisões com informações relevantes para melhorar os serviços ofertados, uma vez que lidam com volumes potencialmente grandes de dados (muitos usuários escrevendo muitos comentários para muitos itens). Neste trabalho, introduzimos um novo problema: identificar a utilidade de uma revisão para o proprietário de um estabelecimento. Portanto, propomos criar a classificação das revisões de acordo com sua relevância para a tomada de decisões, ou seja, direcionado aos proprietários e não aos clientes. O ranking proposto considera aspectos e sentimentos presentes nas revisões. Finalmente, a nossa avaliação experimental considera um *ground truth* (construída com base na opinião de especialistas) e um *baseline* (considera a similaridade entre revisão do usuário e a respectiva resposta do estabelecimento) também propostos neste trabalho e mostra que a nossa solução está muito próxima da classificação ideal.

Palavras-chave: Recomendação de review, Redes sociais.

Abstract

Online reviews have become a powerful way for users to make their opinions available to everyone. Such reviews are extremely valuable for consumers when they are looking for information before acquiring a product or service. In an attempt to help consumers identify helpful reviews, many sites allow users to vote if a review is useful. Then, recommendation algorithms come to the rescue in matching reviews to the consumers who are reading them. Such online review applications usually recommend the most useful reviews for consumers to read. Regarding reviews of points of interest, for the establishments owner (or administrator, manager, etc), it is important to have a fast and reliable way to identify the reviews with relevant information for improving the services provided, once they deal with potentially big volumes of data (many users writing many reviews for many items). In this work, we introduce a new problem: identifying the helpfulness of a review for the owner of an establishment. Therefore, we propose creating a ranking of reviews according to their relevance for decision making, i.e. targeting the owners and not clients. The proposed ranking considers aspects described and sentiments present in the reviews. Finally, our experimental evaluation considers a ground truth (constructed by experts opinion) and a baseline (considering the similarity between review and its respective answer provided by the establishment owner) also proposed in this dissertation and show that our solution is very close to the ideal ranking.

Keywords: Review Recommendation, Social Network.

List of Figures

2.1 Foursquare search interface: characteristics of the use 6
2.2 Spotsetter recommendation interface by pulling data from multiple sources such as social networks and review websites 7
2.3 Collaborative filtering example 8
2.4 Content-based filtering example 8
2.5 Tripadvisor review vote 9
2.6 Review Recommender System: given a reader and an item (e.g., a notebook) with reviews written by authors, the system retrieves a personalized ranking of reviews. 10
2.7 Review: Fantastic location and terrible room service. 11
2.8 Review: Delicious breakfast and terrible room service 12
2.9 Good bedroom size and terrible room service 12

3.1 POIView - POIs opinion visualization, allowing to detect main areas according to location, seasons, checkins or categories 15

4.1 Review with positive and negative aspects (total score: 0.21) 22
4.2 Review with only positive aspects (total score: 0.85) 22
4.3 Review with only negative aspects (total score: -0.645) 23
4.4 Aspect scores along a year for the same hotel 24

5.1 Ranking evaluation steps 30
5.2 A real example of a POI owner review answer 31
5.3 Web interface used by experts to evaluate reviews 33
5.4 Experts age group and gender distribution 34
5.5 Hotels average evaluation score by experts in a 5.0 scale 35
5.6 Algorithm evaluation scenarios 37
5.7 Baseline validation against the ground truth 39

5.8 Cumulative Distribution Function on nDCG for OwnerView against the baseline 39
5.9 Review: generic answer from hotel 40

List of Tables

5.1 Filtered reviews in pre-processing step 30
5.2 Dataset statistics with number of aspects in one review, number of positive/negative reviews and answers. 32
5.3 Ground Truth statistics with score, number of evaluations and percentage representation. 34

Contents

Acknowledgments **xi**

Resumo **xv**

Abstract **xvii**

List of Figures **xix**

List of Tables **xxi**

1 Introduction **1**

2 Basic Concepts and Problem Definition **5**

2.1 Location Based Social Networks 5
2.2 Recommendation Systems 6
2.3 Online Review Systems 8
2.4 Review Recommendation 10
2.5 Our Problem Definition 10

3 Background and Related Work **13**

3.1 Sentiment Analysis 13
3.2 Spam, Fake Review Detection 16
3.3 Influence of Reviews on Sales 17
3.4 Predicting Review Helpfulness/Quality 18

4 Identifying Helpful Reviews for POI's Owners **21**

4.1 Aspects Extraction and Sentiment Analysis 21
4.2 Calculating Aspect Weight 23
4.3 Review Score 26

5 Experimental Evaluation **29**
5.1 Data and Pre-processing 29
5.2 Ground Truth 32
5.3 Methods for Comparison 34
5.4 Evaluation Metrics 36
5.5 Experimental results 36

6 Conclusion and Future Work **41**

Bibliography **43**

Chapter 1

Introduction

Online reviews have become a powerful way for users to make their opinions available to everyone [Askalidis and Malthouse, 2016; Chatterjee, 2001; Duan et al., 2008]. Indeed, the number of online reviews at many specialized websites, such as TripAdvisor and Yelp[1], has been growing significantly. Such reviews are extremely valuable for consumers when they are looking for information before acquiring a product or service. Analyzing review textual data (which represents the thoughts and communication between users) enables to understand the public needs and concerns about what constitutes valuable information from academic, marketing or policy-making perspectives. For example, Lacic et al. aim to predict the satisfaction of air travelers by analyzing their reviews [Lacic et al., 2016].

One specific scenario is when the items being reviewed are locations. Specially, location based social networks, such as TripAdvisor[2] and FourSquare[3], are now important tools for users to choose hotels, restaurants and tourist attractions – locations known as *points of interest* (POI). The contents of these social networks are generated by users, thus providing access to the opinions of many individuals. A user may: contribute with an opinion, evaluate a POI by a rating or indicating whether liked or not, and write a review. Then, the problem with big volumes of data persists: TripAdvisor handles more than 500 million reviews and opinions from 7 million places in more than 135,000 destinations[4]; whereas FourSquare handles data of more than 93 million places[5]. With so many reviews, how can a user find a proper one?

In an attempt to help consumers identify useful reviews, many sites allow users

[1] Yelp: `http://www.yelp.com`
[2] TripAdvisor: `http://www.TripAdvisor.com`
[3] FourSquare: `https://www.FourSquare.com`
[4] TripAdvisor Fact Sheet: `https://tripadvisor.mediaroom.com/US`, May 2017
[5] FourSquare About Us: `https://FourSquare.com/about`, May 2017

to vote if a review is helpful. While most websites just show the percentage of positive votes or the average of received votes, some of them provide the grade that each user gave to a review. However, this evaluation tends to be sparse with many reviews without any feedback [O'Mahony and Smyth, 2009]. This problem is due to the rich-get-richer effect, in which reviews on top tend to receive more feedback, while recent reviews are rarely read [Liu et al., 2007]. Even if the grades given to reviews are too sparse to help users identify relevant reviews, they can provide important data to create a model to automatically predict the quality of a review [Jo and Oh, 2011].

Regarding each POI review, several aspects may affect the users while writing them, including: noise level, quality of products or services, weather, season and existing expectations. In this context, identifying and managing these factors can provide customers and owners with valuable information through the interpretation of large amounts of reviews [Maroun et al., 2016]. Nonetheless, how to manage is different for each role. Specifically, for the establishment's owner (or administrator, manager, etc), it is important to have a fast and reliable way to identify the reviews with relevant information for improving the services provided. Then for the client, it is important to identify reviews with details about the place that will help to decide where to go, eat, visit or stay.

Indeed, consumer online reviews have become a major factor in business reputation and brand image due to the popularity of TripAdvisor, Yelp and online review websites. A negative review can really damage the reputation of a business. The problem is so serious that an industry of *reputation management* has arisen: companies, such as *Reviewsthatstick*[6], attempt to remove or hide bad reviews such that more favorable content is found when potential customers look for products and services.

Furthermore, data from online reviews have become a major source for different types of research. For example, considering different contexts, Krawczyk and Xiang [2016] use a text analysis approach to create perceptual maps from the most frequent terms used in a data set collected from an online travel agency; and Law et al. [2017] assess the effectiveness of sentiment analysis for dishwasher defect discovery. Then, considering only data from hotel reviews, Hu et al. [2017] propose a novel multi-text summarization technique for identifying the top-k most informative sentences of hotel reviews. Then, Lappas et al. [2016] formalize the visibility of a hotel to the customer based on evaluating its vulnerability to fake review attacks; Ye et al. [2009] empirically investigate the impact of online consumer-generated reviews on hotel room sales; and Phillips et al. [2009] study review valence (positive vs. negative reviews), hotel familiar-

[6]ReviewsThatStick: `http://reviewsthatstick.com`

ity (well-known vs. lesser-known hotels), and reviewer expertise (expert vs. non-expert reviewers) as independent factors and show that on average, exposure to online reviews enhances hotel consideration by consumers. Such previous works consider data from online reviews for tackling an interesting problem in a unique context.

In this dissertation, we introduce a new problem: identifying the helpfulness of a review for the *owner* of an establishment. The relevance of a review to the establishment differs from the relevance to the customer because now the goal is not to help decide which product to buy or which place to go to. For example, considering two reviews: (a) complains about the distance to the nearby train station; and (b) complains about the hotel staff. A traveler may find (a) more important than (b), but surely the hotel owner will be more concerned about (b), as not much can be done about (a). Following such an example, the focus here is to identify comments on important issues, especially those regarding the establishment, which can be improved to increase customer satisfaction and help in making strategic and administrative decisions.

Also, providing a calculated average of the received rates (grades) or a summary of all comments to the establishment owner is not enough, became: (i) the average is just a number for a global view of a set of aggregated reviews; (ii) the summary will potentially present complains and compliments for all aspects at the same time; (iii) and none of them provides a way for the owner to answer individual critics. In other words, a major goal here is to point out *textual* and *individual* reviews that qualify existing problems, so that the owner can properly answer and address them.

Hence, we propose creating a ranking of reviews according to their relevance for POI decision making, i.e. targeting the owners and not clients. The proposed ranking considers aspects described and sentiments present in the reviews. In order to evaluate our ranking, we build a ground truth dataset formed by expert opinions on the relevance of a set of reviews to a set of POI owners. Likewise, as there is no similar work to use as baseline, we build one from scratch based on the similarity between reviews and their answers. Overall, our main contributions are: a ground-truth dataset constructed by experts; a baseline built by considering the hypothesis that there is a relation between the review and its answer from the POI owner; a new algorithm to recommend reviews for POIs *owners*, called *OwnerView*, and a through experimental evaluation on a big volume of data.

The rest of the work is organized as follows. Chapter 2 overviews basic concepts, and Chapter 3 goes over related work. In Chapter 4, we describe how to measure the helpfulness of a review. Then, Chapter 5 presents our experimental evaluation and its results, and Chapter 6 concludes this work.

Chapter 2

Basic Concepts and Problem Definition

We first introduce basic concepts, both for generic and review recommendation, and then state the problem.

2.1 Location Based Social Networks

Location based social networks (LBSN) are web platforms that reflect the social networking structures of real world [Zheng, 2012]. In recent years, the study of LBSNs has attracted attention because they consider interaction information among users along with their geographic location for a period of time. Such information is useful for developing applications such as recommendation systems of places, reviews, travel planning, among others [O'Mahony and Smyth, 2009; Zheng et al., 2011].

Foursquare[1] is one of the most popular LBSN. It recommends points of interest based on user location and opinions (such as hotels and restaurants). Figure 2.1 shows an example of Foursquare use, with locations in a big map and their individual comments in a box. Also, Foursquare lets friends know where a user is. It applies gamification[2] strategies to motivate users for collecting points, prize badges, and, eventually, coupons.

After the success of Foursquare, other companies were motivated to create platforms to improve user experience, such as Spotsetter[3]. Spotsetter aims to combine

[1]Foursquare: `http://www.foursquare.com`

[2]Gamification: is the strategy of interaction between people and companies based on the offer of incentives that stimulate the engagement of the public with the brands in a playful way

[3]Spotsetter: `http://www.spotsetter.com`

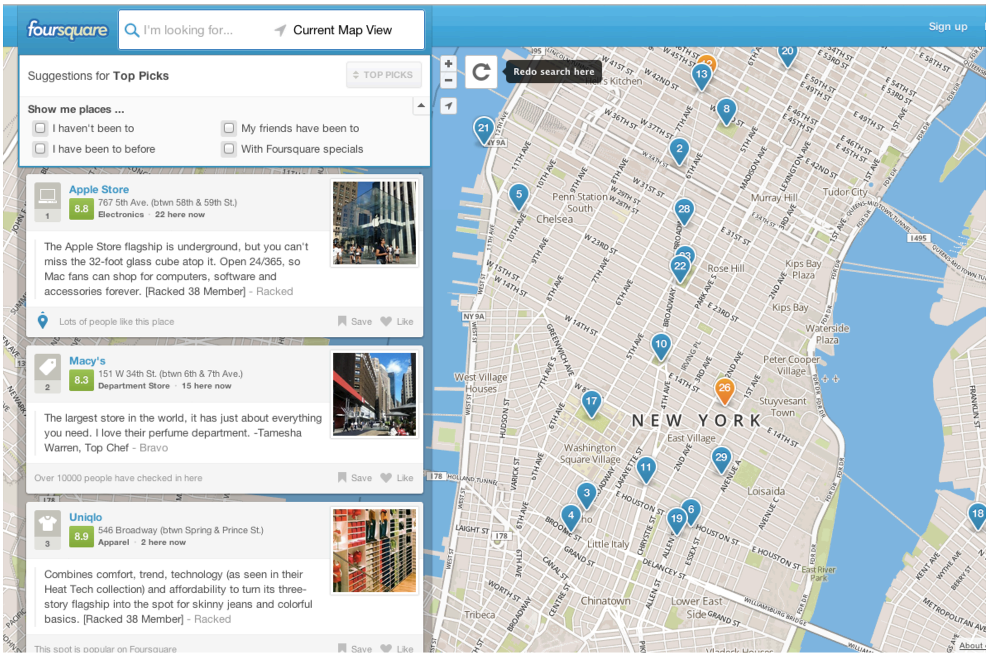


Figure 2.1. Foursquare search interface: characteristics of the use

friends recommendations, trusted reviews and other signals in order to reinvent maps as a more social experience (e.g Figure 2.2). Initially available as a web and mobile application, Spotsetter uses an algorithm to pull in users content from social networks like Facebook, Twitter, Instagram and Foursquare, as well as venue content from over 30 review sites and lists from trusted sources like Yelp and TripAdvisor. Spotsetter is similar in a way to Foursquare, which also plots social data on a map, but instead of being limited to one source, the app pulls from multiple platforms.

2.2 Recommendation Systems

Recommendation systems combine several computational techniques to select custom items based on the interests of users and the context in which they are inserted [Macedo et al., 2015]. Such items may take varied forms, including reviews, places, books, movies, news, music, videos, ads, people and products from a virtual store [Bobadilla et al., 2013]. One common form of recommendation is to order these items in a rank according to their relevance to the target user [Liu and Aberer, 2014; Raghavan et al., 2012; Vargas et al., 2014].

A recommender system is traditionally divided into two types:

- Collaborative filtering is a domain-independent prediction technique that cannot easily and adequately be described by metadata, e.g. movies and music. Col-

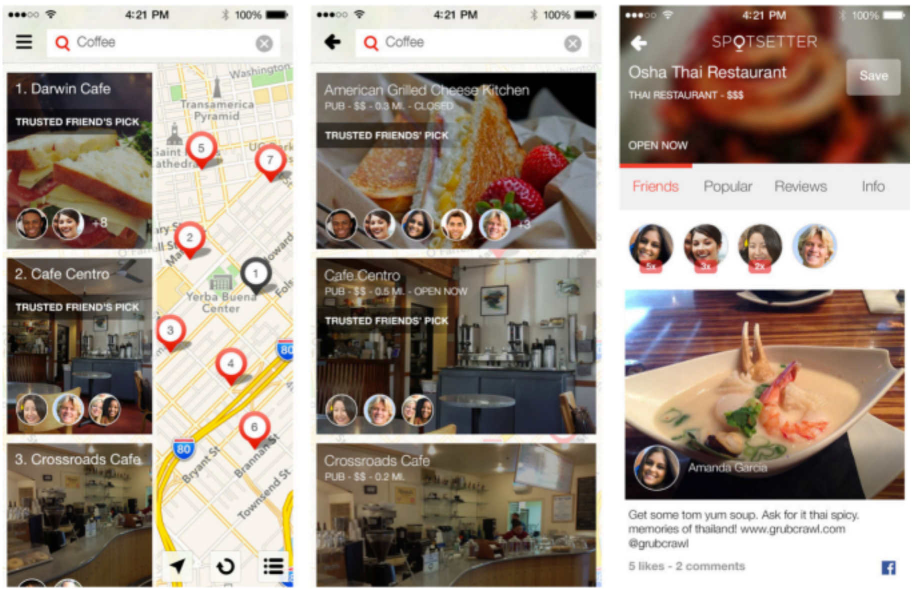


Figure 2.2. Spotsetter recommendation interface by pulling data from multiple sources such as social networks and review websites

laborative filtering techniques work by building a database (user-item matrix) of preferences for items by users, and then matching users with relevant interest and preferences by calculating similarities between their profiles to make recommendations [Herlocker et al., 2004]. Such users build a group called neighborhood. A user gets recommendations to those items that he/she has not rated but were already positively rated by users in his/her neighborhood. Figure 2.3 shows an example in which users are considered similar by reading some articles. Since they share preferences, an article read by one must interest the other.

- Content-based is a domain-dependent method and emphasizes the analysis of the attributes of items in order to generate predictions. When documents (such as web pages, publications and news) are to be recommended, content-based filtering technique is the most successful. In content-based filtering techniques, recommendation is made based on the user profiles by using features extracted from the content of the items the user has evaluated in the past [Bobadilla et al., 2013]. Items that are mostly related to the positively ones are recommended to the user. Figure 2.4 shows an example where a user reads a few articles. Other articles are considered similar by sharing some characteristics like subject (e.g photograph, computer science), writer, publishing company, etc. Since these articles are similar to previous articles read by the user, it is more likely that the user will like them as well.

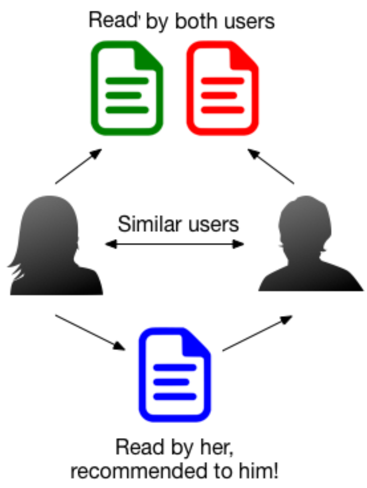


Figure 2.3. Collaborative filtering example

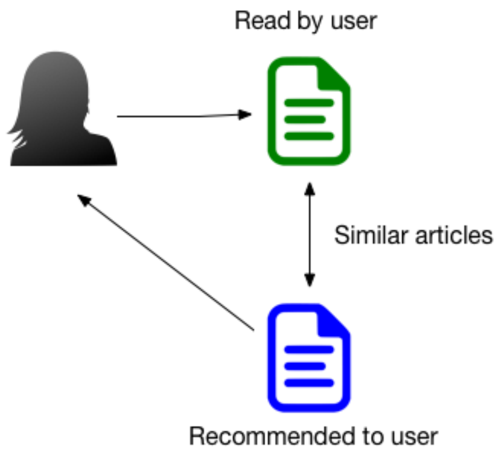


Figure 2.4. Content-based filtering example

2.3 Online Review Systems

Evaluating the helpfulness of reviews requires a dataset built from a social network that allows (*i*) users to write reviews for POIs, and (*ii*) POIs' owners to answer such reviews. The first restriction is easily satisfied by most POI-related social network (e.g FourSquare, TripAdvisor), but the second is not a common feature. TripAdvisor satisfies both and is the world's largest travel site, offering advice from millions of travelers: with 500 million reviews and opinions covering 7 million accommodations, restaurants and attractions. Although it allows POIs owners to answer their reviews,

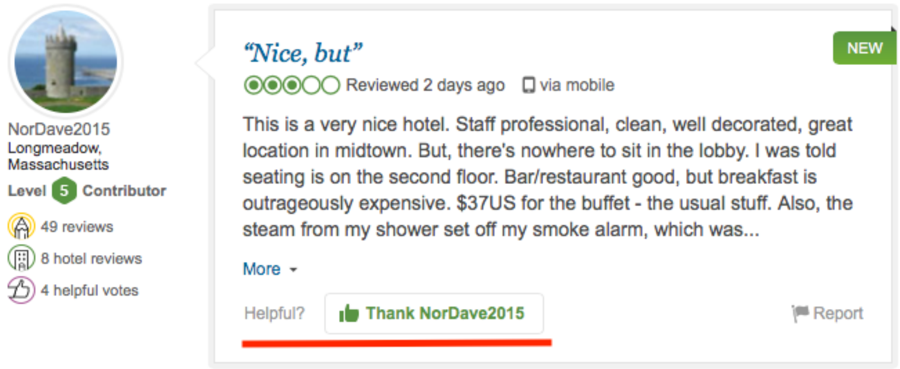


Figure 2.5. Tripadvisor review vote

such feature is not present for all POI categories (e.g Concerts and Shows). Indeed, we have empirically evaluated Triapadvisor POIs categories and discovered that *hotel* is the category with largest answer rate. At the end, we use TripAdvisor hotel category to construct our evaluation datasets.

A review refers to an evaluation written by a user or consumer for a product or a service based on an opinion and/or experience as a user of the reviewed item. Reviews are in the form of several lines of text accompanied by a numerical rating. This text aims to help in shopping decision of a possible buyer, for example. A consumer review of a product usually evaluates how well the product measures up to expectations based on the specifications provided by the manufacturer or seller. It focuses on performance, reliability, quality, defects if any, and value for money. Often it includes comparative evaluations against competing products. Finally, observations are factual as well as subjective in nature.

For example, in Tripadvisor, travelers may vote for a review that they find particularly helpful to them, by clicking the button at the bottom of each review (Figure 2.5). Helpful vote counts are displayed on each review and in member profiles, but these votes are not a factor in Tripadvisor ranking algorithms. TripAdvisor removes votes that fall under any of the following scenarios: member votes for their own review; duplicate votes; owner votes for a review about their own business and owner votes for a review about a competitors business.

Considering a website in which users may give feedback about the consumed services or products, the user can act in three different ways: (*i*) writer - user writes a review about the consumed item; (*ii*) reader - user reads reviews looking for important information about a target item; and (*iii*) voter - user grades an existing review indicating how useful it is.

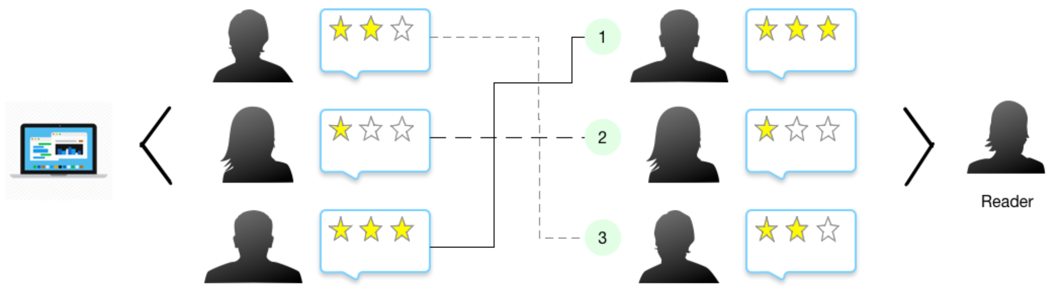


Figure 2.6. Review Recommender System: given a reader and an item (e.g., a notebook) with reviews written by authors, the system retrieves a personalized ranking of reviews.

2.4 Review Recommendation

One specific scenario is when the items being reviewed are locations. Specially, location based social networks, such as TripAdvisor and FourSquare, are now important tools for users to choose hotels, restaurants and tourist attractions – locations known as *points of interest* or POI. The contents of these social networks are generated by users, thus providing access to the opinions of many individuals. A user may: contribute with an opinion, evaluate a POI by a rating or indicating whether liked or not, and write a review. Then, the problem with big volumes of data persists: TripAdvisor handles more than 500 million reviews and opinions from 7 million places in more than 135,000 destinations; whereas FourSquare handles data of more than 93 million places. With so many reviews, how can a user finds a proper one?

Ideally, a review recommender system provides a review ranking in descending order of helpfulness for a given user and item, as illustrated in 2.6. Thereafter, whenever a reader accesses a product, the most helpful reviews are on top, eliminating the problem of manually looking for a needle in a haystack. The main questions to solve the problem of recommending a review and its importance are discussed in Chapter 3.

2.5 Our Problem Definition

To the best of our knowledge, none of the existing methods estimates the helpfulness of a review for a business administrator or POI owner. A recent survey shows that all current works on review quality prediction identify the relevance of a review for clients (users) [Maroun et al., 2016]. However, the relevance of a review to an establishment differs from the relevance to the customer because now the goal is not to help decide which product to buy or which place to visit. The focus is to identify

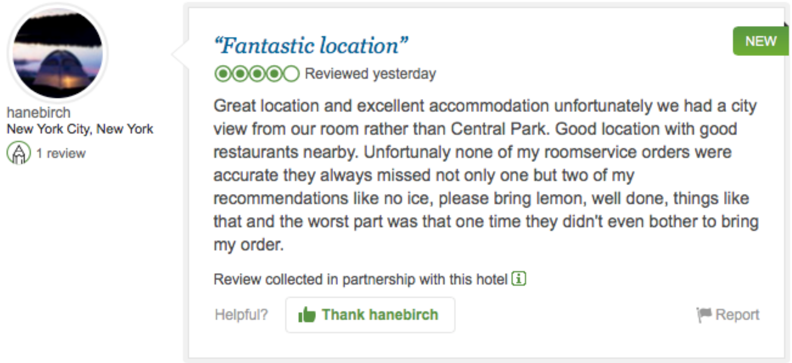


Figure 2.7. Review: Fantastic location and terrible room service.

comments that address important issues, especially those related to the establishment that can be improved to increase customer satisfaction and help in making strategic and administrative decisions. It is important to note that an unsatisfied user can cause considerable damage, since comments on the Internet can spread quickly and influence other potential customers.

Motivating Example Assume that three travelers, who have joined an online reviewing website, have written a review about the hotel they booked in different dates as ilustrated in Figures 2.7, 2.8 and 2.9. The first one covered the aspects about location, the second detailed the quality of breakfast, and the third enjoyed bedroom size. The clients made important and different compliments about hotel, but all off them also complained about room service. Service quality is one key point to make customers happy. If the hotel manager had an easy and efficient way to detect the problem with room service, such manager would be able to propose solutions and avoid more clients dissatisfaction after the first client review. Once again, providing the owner an averaged rate or summary of all complains is also not as useful as the individual and textual ones, as the owner may want to not only address and solve such issues but also properly give answers to the three users who complained about room service.

In the previous example, helpfulness of users review can be evaluated by focusing on the problems they mention. We argue that a personalized review quality for POIs owner may be more accurate than a user review recommendation approach and introduce the problem of *owner-oriented review ranking*. As exemplified in the introduction, a business owner cannot change the distance to the nearest train station, but can surely improve its staff.

Problem Statement Given a set of reviews $R = r_1, r_2, ..., r_x$ about a POI, a set of reviewers $V = v_1, v_2, ..., v_y$ who have written R and a set of owners of POIs $O =$

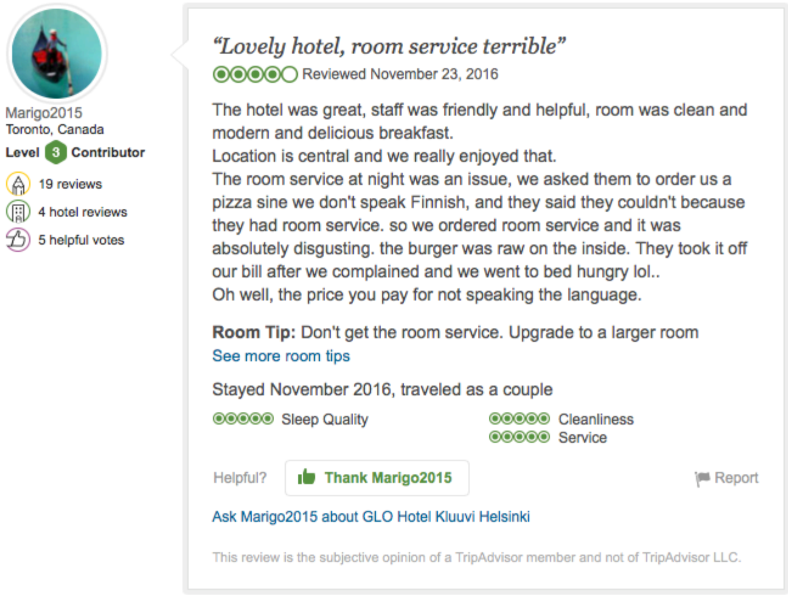


Figure 2.8. Review: Delicious breakfast and terrible room service

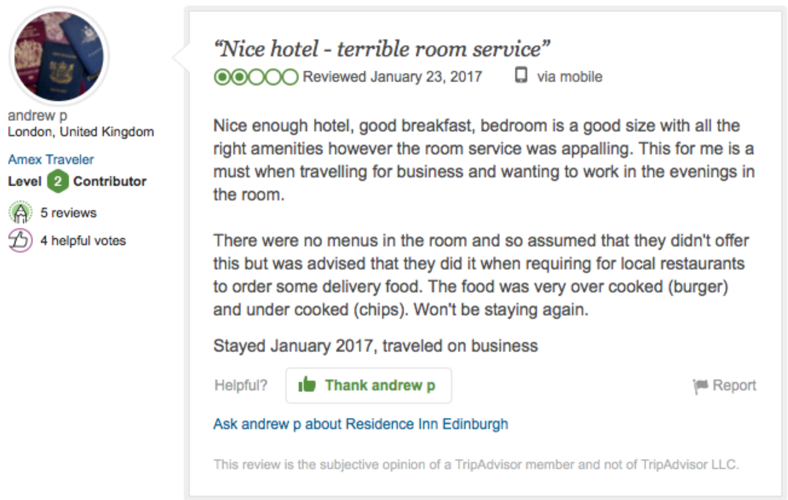


Figure 2.9. Good bedroom size and terrible room service

$o_1, o_2, ..., o_z$ who manage the POIs and are responsible to get clients feedback. The task is to rank R according to the quality rating (helpfulness) of each $r \in R$ for a given owner o_j.

Chapter 3

Background and Related Work

There is no work that uses the explicit sentiments and topics extracted from reviews for owners recommendation purposes. Nonetheless, we now discuss related work on specific topics of our problem and solutions. Section 3.1 explains the importance of opinions and the main techniques to identify them. Section 3.2 points the danger of fake reviews and how to detect them. Then, Section 3.3 shows the importance of reviews and how they can influence sales, and Section 3.4 goes over techniques for predicting a review helpfulness.

3.1 Sentiment Analysis

Opinions are central to almost all human activities because they are key influencers of our behaviors. Whenever we need to make a decision, we may want to know others opinions first. In the real world, businesses and organizations always want to find consumer or public opinions about their products and services. Individual consumers also want to know the opinions of existing users of a product before purchasing it, and opinions about political candidates before making a voting decision in a political election. In the past, when an individual needed opinions, he/she asked friends and family. When an organization or a business needed public or consumer opinions, it conducted surveys, opinion polls and focus groups. Acquiring public and consumer opinions has long been a huge business itself for marketing, public relations and political campaign companies. Today, such opinion mining tasks have been automatized.

Previous works studying aspects and sentiments of reviews can be defined as *opinion mining*, which operates on text portions of any size and shape, such as web pages, comments, tweets, etc. Every opinion is composed of at least two elements: a target (topic, product, person, etc.) and a feeling (attitude, opinion, emotion) about this target [Liu and Zhang, 2012].

An opinion without its target being identified is of limited use. Realizing the importance of opinion targets also helps to understand the sentiment analysis problem better. For example, the sentence "although the service is not that great, I still love this hotel" clearly has a positive tone; however, one cannot say that this sentence is entirely positive. In fact, the sentence is positive about the hotel, but negative about its service (not emphasized). In many applications, opinion targets are described by entities and their different aspects. Thus, the goal of this level of analysis is to discover sentiments on entities and their aspects. For example, the sentence "The iPhone's call quality is good, but its battery life is short" evaluates two aspects (call quality and battery life) of iPhone (entity). The sentiment on iPhone's call quality is positive, but the one on its battery life is negative. The call quality and battery life of iPhone are the opinion targets. Based on this level of analysis, a structured summary of opinions about entities and their aspects can be produced, which turns unstructured text to structured data and can be used for all kinds of qualitative and quantitative analyses.

To make things even more interesting and challenging, there are two types of opinions, i.e., regular opinions and comparative opinions [Jindal and Liu, 2007]. A regular opinion expresses a sentiment only on a particular entity or an aspect of the entity, e.g., Coke tastes very good, which expresses a positive sentiment on the aspect taste of Coke. A comparative opinion compares multiple entities based on some of their shared aspects, e.g., Coke tastes better than Pepsi, which compares Coke and Pepsi based on their tastes (an aspect) and expresses a preference for Coke.

Another perspective is to consider the time of the opinion as well. The process of mining temporal opinions involves defining the average opinion on a particular topic in two or more different points in time. Changes in opinion can then be identified and used to find patterns or summarize the opinion regarding a specific aspect [Cheng et al., 2011]. For example, Lourenço Jr. et al. [2014] have studied an efficient way to analyze people opinions about topics and entities on social networks like Twitter. However, changes in opinions are not necessarily useful by themselves, as they need some factor to compare them. Therefore, the utility of opinion change detection becomes more evident when combined with the understanding of why this change occurred.

Yet a different perspective is how to visualize sentiments and opinions in an intuitive way. Considering the customer's perspective, we have developed *POIView*[1] [Prado and Moro, 2015], a tool that uses sentiment analysis techniques to visualize

[1]POIView: `http://www.dcc.ufmg.br/~mirella/Tools/POIView/`

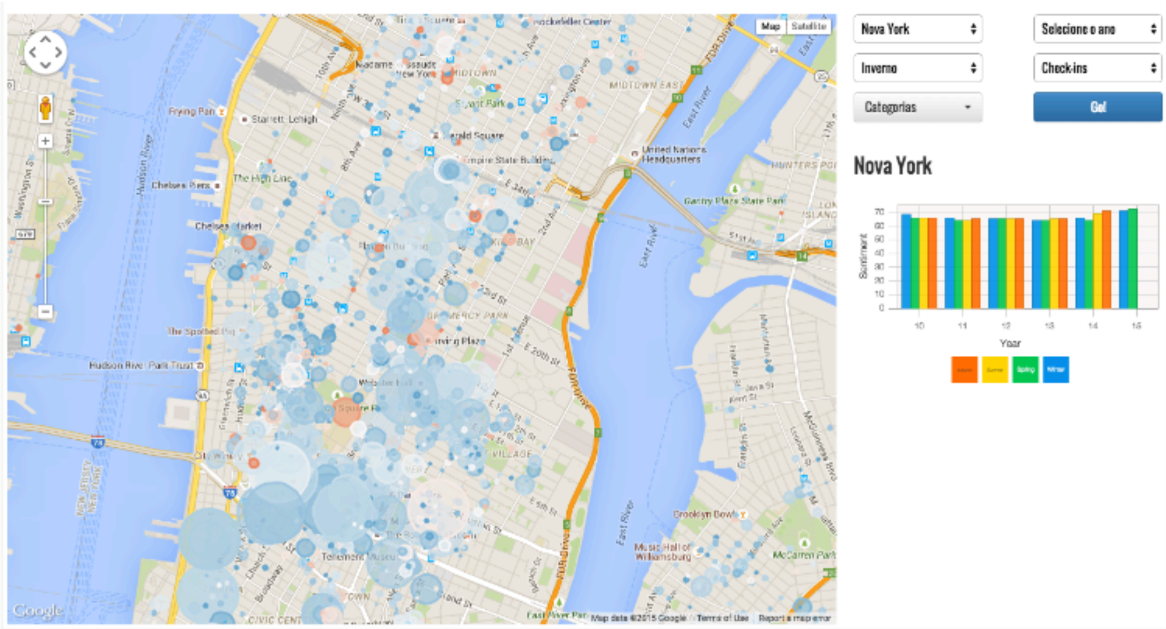


Figure 3.1. POIView - POIs opinion visualization, allowing to detect main areas according to location, seasons, checkins or categories

opinion of POIs (Figure 3.1). The tool is based on users reviews about FourSquare's locations. POIView provides a way to detect areas with well and poorly evaluated places on a map based on their ratings. It also includes several filters where users can choose city, year, season and categories to display only the desired information. For commercial purposes, POIView allows a constant analysis of how views on a property vary over time in order to identify if current business strategies provide good results.

It is possible to perform sentiment classification at the document level in order to get an overall opinion on an entity, topic or event. However, this level of classification has some shortcomings for applications:

- In many applications, the user needs to know additional details, e.g., what aspects of entities are liked and disliked by consumers. In typical opinion documents, such details are provided, but document sentiment classification does not extract them for the user.

- Document sentiment classification is not easily applicable to non-reviews such as discussion forum , blogs and news articles, because many such post can evaluate multiple entities and compare them. In many cases, it is hard to determine whether a post actually evaluates the entities that the user is interested in, and whether the post expresses any opinion at all, let alone to determine the sentiment about them.

Furthermore, sentiment (and opinions) analysis requires to differentiate between statements of fact and opinions, and to detect the polarity of the sentiment expressed. Recent studies have tackled sentiment analysis from different perspectives. Turney [2002] ranked the polarity of feeling reviews on document level. Wiebe et al. [1999] classified subjectivity level of sentences using classes as adjectives, verbs and attributes. Riloff and Wiebe [2003] drew subjective expressions of sentences using the learning process *bootstrapping pattern*. Yu and Hatzivassiloglou [2003] identified the polarity of sentences opinions using semantically oriented words. Recently, Liu et al. [2017] proposed a method based on sentiment analysis and fuzzy set theory to rank the products through online reviews. Salehan and Kim [2016] investigated the predictors of readership and helpfulness of OCR using a sentiment mining approach for big data analytics and showed that reviews with higher levels of positive sentiment in the title receive more readerships.

The aforementioned techniques have been applied and examined in different fields, such as user reviews and news articles. In our proposed solution, the sentiment analysis serves to identify opinions about topics and the polarity of the sentiment expressed in the reviews. Knowing what customers are talking about and how they like a product (or not) allows to identify the most relevant aspects of service or product, and to determine if there is anything the POIs owners should do to improve their establishment.

3.2 Spam, Fake Review Detection

Online reviews have become a valuable resource for decision making. Unfortunately, this importance of reviews also gives good incentive for spam. The key challenge of opinion spam detection is that unlike other forms of spam, it is very hard, if not impossible, to recognize fake opinions by manually reading them. Then, finding opinion spam data to help design and evaluate detection algorithms is equally hard. For other forms of spam, one can recognize them fairly easily. Nonetheless, three types of spam reviews were identified in Jindal and Liu [2007].

- **Fake reviews.** These are untruthful reviews written not based on the reviewers genuine experience of using the products or services, but written with hidden motives. They often contain undeserving positive opinions (products or services) in order to promote the entities and/or unjust or false negative opinions about others in order to damage their reputations.

- **Reviews about brands only.** These reviews do not comment on the specific products or services that they are supposed to review, but only comment on the brands or the manufacturers of the products. Although they may be genuine, they are considered as spam as they are not targeted at the specific products and are often biased. For example, a review for a specific HP printer says I hate HP. I never buy any of their products.

- **Non-reviews.** These are not reviews. There are two main sub-types: advertisements and other irrelevant texts containing no opinions (e.g., questions, answers and random texts). Strictly speaking, they are not opinion spam as they do not give user opinions.

To solve such issues, Jindal and Liu [2007] propose to perform spam detection based on duplicate finding and classification, then applying Logistic regression to learn a predictive model. Mukherjee et al. [2013] study how well existing research methods work in detecting real-life fake reviews in a commercial website, and compare them to Yelps filtered and unfiltered reviews. Then, Lappas et al. [2016] found that even limited injections of fake reviews can have a significant effect and explore the factors that contribute to this vulnerable state and in certain markets. Specifically, 50 fake reviews are sufficient for an attacker to surpass any of its competitors in terms of visibility. Overall, so much has been done to filter actual opinions aiming at cleaning the data. Here, filtering ads and detecting spam reviews are important steps before recommending reviews for owners: ads have no value for a manager, and spam reviews could lead to wrong decisions and analysis about services and products.

3.3 Influence of Reviews on Sales

The creation of online consumer communities to provide product reviews and advice has been touted as an important (although somewhat expensive) component of Internet retail strategies. For example, Chevalier and Mayzlin [2003] examine the effect of consumer reviews on relative sales of books at Amazon[2] and BarnesAndNoble[3] and find that positive reviews and book sales positively correlate. Dellarocas et al. [2007] propose concrete models for decision support in a specific business context and provide

[2]Amazon: `http://www.amazon.com`

[3]Barnesandnoble: `http://www.barnesandnoble.com`

quantitative assessments of the information value of online review metrics relative to more traditional metrics. Hence, it is crucial for a POI owner to have an efficient way to know the most important problems that clients are complaining about, making it possible to propose solutions and fix them shortly.

According to Reevoo[4], 50 or more reviews per product can mean a 4.6% increase in conversion rates and 63% of customers are more likely to make a purchase from a site which has user reviews. All reviews are valuable, and a mix of positive and negative reviews helps to improve consumer trust in the opinions they read. Indeed, recent stats from Reevoo suggest that the presence of bad reviews actually improves conversions by 67%. Reevoo found that people that seek out and read bad reviews convert better, because their paying such close attention means they are more likely to be in purchase mode. Then, 68% of consumers trust reviews more when they see both good and bad scores, while 30% suspect faked reviews when they do not see any negative opinions on the page. The benefits of bad reviews depends on the proportion of good to bad. The negative reviews make the positive ones more believable, but there is a point at which they ring alarm bells for consumers. If, for instance, a product page contains 15 reviews, and two are negative, then the other 13 look trustworthy. If that proportion changes, it is a different matter.

3.4 Predicting Review Helpfulness/Quality

The problem of automatically determining the quality of content generated by online social networking users has attracted much attention [Hu and Liu, 2004; Ngo-Ye et al., 2017; Pang and Lee, 2005; Qazi et al., 2016; Turney, 2002; Zhou and Guo, 2017]. Mudambi et al. [2011] studies what makes a helpful consumer review. By conducting a controlled experiment that manipulated both the review content and the description of the reviewer, they found that reviews written by a self-described expert are more helpful than those that are not. For instance, Bigonha et al. [2012] defined quality metrics for influential users on Twitter. Pang and Lee [2005] studied products reviews prediction, which can be significant due to the correlation between a product evaluation and the usefulness of a review. However, the overall assessment of a product is known.

Other previous works [Ghose and Ipeirotis, 2011; Kim et al., 2006; Liu et al., 2007; Tsur and Rappoport, 2009; Zhang and Varadarajan, 2006] have focused on automatically determining the quality of reviews through textual attributes and social aspects [McAuley and Leskovec, 2013; Tang et al., 2013]. Textual attributes include

[4]Reevoo: `http://www.reevoo.com`

text statistics, such as the text size, average size of the sentences and percentage of adjectives. Social attributes relate to the reviews authors and are extracted from their social context, as the number of evaluations made by the author, author connection degree in the social network, average grades given, among others. Korfiatis et al. [2012] explore the interplay between online review helpfulness, rating score and the qualitative characteristics of the review text as measured by readability tests. Zhou and Guo [2017] examine social influence effects on review helpfulness through effect of review order, finding that the order of a review negatively relates to review helpfulness.

Another direction is to formulate the problem as classification or regression with users votes serving as ground truth. These approaches consider that all users have a common perception of helpfulness and are limited to discovering helpfulness for an average person. For example, Zhang and Varadarajan [2006] discovered that syntactical features of review text are most useful, whereas review size seems poorly related with its quality. Suleman and Vechtomova [2016] applied a two-step hierarchical clustering process that first clusters words representing aspects based on the semantic similarity of their contexts and then on the similarity of the hypernyms of the cluster members for product aspect discovery. Only a few works [Moghaddam et al., 2012; Tang et al., 2013] dealt with the problem in a personalized fashion, considering users idiosyncrasy, then exploring the hypothesis that users do not perceive helpfulness in the same way (what is useful for one, is not for another) by incorporating relevant personalized information for review recommendation.

In addition to textual attributes, Kim et al. [2006] included metadata information (such as notes given to an item under evaluation) and concluded that review size and the number of stars in product ratings are more useful to their regression model. Ghose and Ipeirotis [2011], based on subjective analysis of the text, showed evidence that reviews addressing extreme aspects are considered more useful. Liu et al. [2008] considered user experience as reviewer and the frequency they create reviews in addition to writing style in a non-linear regression model. Recently, Qazi et al. [2016] focused on uninvestigated factors by looking at not just the quantitative factors (such as the number of concepts), but also qualitative aspects of reviewers (including review types such as the regular, comparative, suggestive and reviewer helpfulness) and built a conceptual model for helpfulness prediction. Ngo-Ye et al. [2017] identified the most helpful reviews quickly and enhanced the review website's usefulness by examining the utility of script (script includes the salient elements that readers look for before determining whether a review is helpful) analysis for predicting the helpfulness of online customer reviews.

Despite their differences, all these previous works consider the problem of evaluating the quality of reviews and classify them as best as possible from the user's point of view (e.g., product buyer in an e-commerce system). Here, we introduce a new problem: identifying the usefulness of a review for the *business owner*. To address it, we propose using aspects and sentiments of reviews, and creating a rank with reviews that can be most useful for the management and development of establishment at the top. For the owner of an establishment, the quality of a review can help identify negative aspects and problems of products and services. Moreover, it is possible to determine the most successful products to expand their production as well.

Chapter 4

Identifying Helpful Reviews for POI's Owners

We treat the task of finding helpful reviews as a ranking problem in which the goal is to obtain a scoring function that gives higher scores to helpful reviews in a given set of reviews. Our strategy is to rank the reviews based on their aspects and sentiments. Knowing what clients are talking about and how they feel allows to detect important reviews, but we must consider the opinion of as many clients as possible to identify main problematic topics.

Overall, our solution (*OwnerView*) to rank reviews according to the helpfulness for the business owner is composed of three steps. First, we define a strategy to extract aspects and their sentiment from a set of reviews of a given POI (Section 4.1). We then measure the weight of each aspect by considering its relevance for the whole set (Section 4.2). Finally, we score the helpfulness of a target review by using its aspects relevance (Section 4.3).

4.1 Aspects Extraction and Sentiment Analysis

We now turn to aspect extraction, which can also be considered as an information extraction task. The goal is: given an opinion, return the aspects and the user's sentiment about them. The key characteristic is that an opinion always has a target. The target is often the aspect or topic to be extracted from a sentence. Thus, it is important to recognize each opinion expression and its target from a sentence. Also, we should note that some opinion expressions can play two roles, i.e., indicating a positive or negative sentiment, and implying an (implicit) aspect (target). For example, in "this car is expensive", the word *expensive* indicates both a sentiment and the aspect price.

Review	**Sentiment**	**Aspect**	**Score**
"The air conditioning unit was very noisy and there was also a lot of noise from the lift at night, not only noise but vibrations in the room. Saying that the room although small was clean as was the rest of the hotel. Staff were friendly"	clean	room	0.62
	friendly	staff	0.79
	very noisy	air conditioning unit	-0.778

Figure 4.1. Review with positive and negative aspects (total score: 0.21)

Review	**Sentiment**	**Aspect**	**Score**
"Staff was very friendly upon checkin and at checkout. The hotel was very clean and quiet. Breakfast was good. Would consider staying here again if in town. I would highly recommend this hotel to anyone in the St. Louis area, especially anyone on a business trip."	clean and quiet	hotel	0.884
	friendly	staff	0.929
	would highly recommend	hotel	0.798
	good	breakfast	0.79

Figure 4.2. Review with only positive aspects (total score: 0.85)

Identifying aspects and analyzing sentiment are complex data processing tasks. Luckily, there are many solutions available such as the whole framework described in [Jo and Oh, 2011]. There is also publicly available tools, as the services provided by *HPE HEAVEN OnDemand*[1]. For practical reasons and good overall performance, we use the latter as explained next.

HPE HEAVEN OnDemand is a platform for building cognitive computing solutions using text analysis, speech recognition, image analysis, indexing and search APIs. The sentiment analysis API processes text and returns a list of aspects mentioned from the text along with the sentiment and the score associated to them. It returns the sentiment as positive, negative, neutral or mixed. It uses a dictionary of positive and negative words of different types, and defines patterns that describe how to combine these words to form positive and negative phrases. Automatically classifying text by sentiment allows to easily find out the general opinions of people in an area of interest. For example, a manager might want to analyze reviews from a restaurant to help improving services provided, menu choices or to enhance customer experience.

[1]HPE HEAVEN OnDemand: `http://www.havenondemand.com`

Review	Sentiment	Aspect	Score
"The worst wifi I have experienced in a hotel for some time. My shower was so poorly designed that I flooded the bathroom twice! Someone had left their tee shirt in my colleagues shower on her second night which made you wonder what was going on in her room. Her ac was broken requiring a man to fix it for 1 HR despite arriving very late. Altogether not what you would expect for $400+ a night. Won't be coming back."	the worst	with	-0.798
	poorly designed	shower	-0.696
	broken	ac	-0.535

Figure 4.3. Review with only negative aspects (total score: -0.645)

The API splits the input text into entities, which describe a part of the text with a particular sentiment. The API returns details of the extracted entities, including the length and the detected sentiment. Each sentiment extracted contains valuable information. For example, Figures 4.1, 4.2 and 4.3 show the text of three real reviews and their aspects as extracted by HPE tool. Specifically:

- **their sentiment** is the qualifier for the sentiment or opinion (e.g. the adjectives *clean* and *friendly*);
- **aspect** indicates what the positive or negative sentiment is about (e.g *room* and the *staff*); and
- **score** is a value between 0 and 1 (0 and -1 in the negative case), which indicates the strength and confidence of the sentiment.

4.2 Calculating Aspect Weight

Using the sentiment and aspect identification API (described in Section 4.1) enables to extract all necessary knowledge to calculate the weight of each aspect considering the set of reviews of a POI. The sentiment of an aspect may be positive or negative, and both are useful:

- **positive** aspects help identify reviews with compliments about features or services. This kind of review allows to identify for example what clients like and if a given promotion or service had the expected impact; and

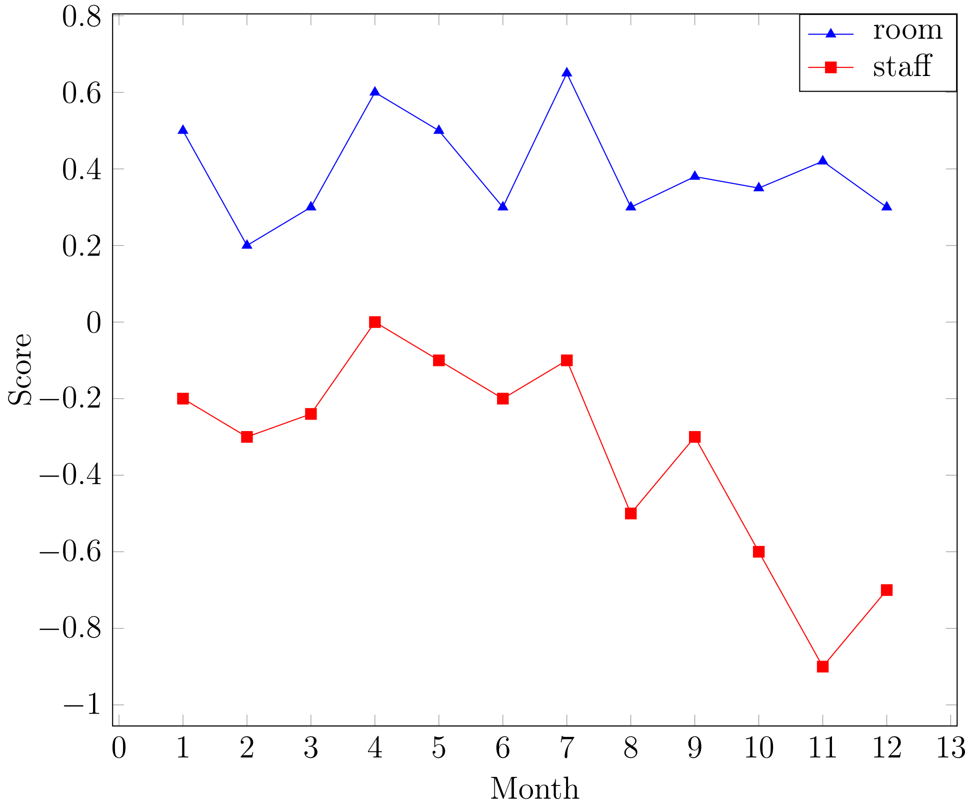


Figure 4.4. Aspect scores along a year for the same hotel

- **negative** aspects help to find reviews pointing out possible problems or features that clients do not like. It is probably more important than a positive aspect since it may lead to losses to the POI in a short amount of time, as it may spread quickly on social networks and specialized media, for example.

Analyzing POI reviews in the owners perspective is important to discover how the place is being evaluated by clients. Not identifying how the opinion of clients is varying, mainly when it is negative, may damage the POI. This kind of problem could be avoided if POIs owners had an efficient way to identify what aspects are not pleasing their clients. However, having just the rates given by clients or their average is not enough, as argued in Section 1. By knowing the specific problems, it should be simpler to take faster decisions to address them.

Considering a hotel as POI, an real example of how the sentiment about *room* and *staff* (from a real hotel) varies through 2016 is shown in Figure 4.4. It presents the aggregated sentiment for each aspect considering all reviews from each month in that year. In this case, room sentiment varies, but is always positive. On the other hand, the opinion about the hotel staff is bad over the whole year, with some oscillation in the

Algorithm 1 Aspect weight algorithm

Input R: a set of reviews from a POI processed by sentiment analysis API (Section 4.1).

Output: a list of aspects and their processed weights.

1: **procedure** WEIGHT(R)
2: $lastDate \leftarrow S.getLastReviewDate()$
3: $aspects \leftarrow \{\}$ ▷ Each aspect has two counters: a negative (n) and positive (p)
4: **for** each review r in R **do**
5: $d \leftarrow 1/\log(lastDate - review.getDate())$ ▷ date score of the review **r**
6: **for** a in $r.getAspects()$ **do** ▷ **a** is an aspect of r
7: **if** a in $aspects$ **then**
8: $v \leftarrow a.getSentimentVal()$
9: **if** $v < 0$ **then** ▷ negative sentiment
10: $aspects[a].n \leftarrow aspects[a].n + v * d.$
11: **else** ▷ positive sentiment
12: $aspects[a].p \leftarrow aspects[a].p + v * d.$
13: **else** ▷ initialize aspect values neg. and pos.
14: **if** $v < 0$ **then**
15: $aspects[a] \leftarrow \{n = v * d, p = 0\}$
16: **else**
17: $aspects[a] \leftarrow \{n = 0, p = v * d\}$
18: **return** $aspects$

first semester and getting sour in the second one. Had the manager properly answered and solved such staff problems over the first months, its score curve could have been in the positive numbers by the end of the year. That is exactly the expected outcome of our work.

There are two options when considering the relevance of an aspect: (i) ignore review date - useful to identify the main reviews and discover general aspects of POI over a set of reviews; and (ii) consider review date - useful to monitor the variation of customer opinion over time. Our assumption is that when considering the relevance of an aspect, we must also use the date of the review to weight its importance. In a real case scenario, a hotel manager (for example) is most worried about recent reviews than from last year, having this in mind, the process of assigning a weight to an aspect is described in Algorithm 1. When evaluating review aspects, our solution treats negative and positive sentiments separately (lines 10, 15 and 12, 17) to emphasize that there are two perspectives of the same aspect. We also apply a log function (line 5) to give more relevance for recent reviews.

4.3 Review Score

Identifying the helpfulness of a review for a POI owner (administrator, manager, etc) requires analyzing the topics and opinions expressed by a client in a review. Positive and negative comments are important feedback to POI owner and help to identify main problems and quality good services. Evaluating such reviews by considering the perspective of all clients is also important to avoid biased decisions. Automatically identifying helpful reviews is a complex task, since a POI usually has thousands of comments. To solve such a problem, we introduce Algorithm 2 for measuring the helpfulness of a review by aggregating the importance of each aspect computed by Algorithm 1.

One challenge is how to aggregate all aspect scores to maximize the level of helpfulness agreement of a review for a POI owner. There are two main challenges that lead to different scores: (*i*) how to consider positive and negative aspects; and (*ii*) how to weight the review by the writer's reputation. The first one is solved by procedure *ScoreHelper()*: it allows the algorithm to use only negative, only positive or a combination of both sentiments about an aspect (the formula was empirically obtained by testing values from 1 to 20 with step 0.05; and the value 9.75 produced better results). The second uses valuable information about the writer: the number of reviews he/she has written and the number of positive votes received. It is calculated by procedure *User()*, and a user who always receives positive votes should be more reliable. This step helps to avoid fake users or spammers, since the reputation was evaluated by other users.

Having both challenges solved allows to calculate review score by procedure *Score()*: it loops through its aspects and invokes *ScoreHelper()* and *User()* procedures passing the desired tuning options. After calculating a score for each POI review, the algorithm sorts them in descending order and returns a rank. Such a rank is oriented towards the owner and not the client of a POI. Therefore, the most helpful reviews for the owners are at the top.

Algorithm 2 Review score algorithm

Input A, R, O, V: A - a set of aspects (each with two counters a negative (n) and positive (p) counter) from a POI and their relevance obtained from Algorithm 1; R - a set of reviews from a POI processed by the sentiment analysis API (section 4.1); O - option to consider negative, positive, or negative and positive aspects; V - aspect sentiment value;

Output: a list of reviews ordered by the most helpful to the owner.

```
procedure ScoreHelper(A, O)
                                ▷ A.n and A.p represent neg. and pos. values of aspect A
    if O = "negative" then
        return A.n
    if O = "positive" then
        return A.p
    if O = "both" then
        return 9.75 * A.n + A.p

procedure User(U)
    return 1 + (U.helpfulVotes/U.numberOfReviews)

procedure Score(A, R, O)
    rank ← {}
    for each review r in R do
        for each aspect a in r.getAspects() do
            if r in rank then
                                ▷ increment review score with aspect score
                rank[r] ← rank[r] + ScoreHelper(a, O)
            else
                                ▷ initialize review score with aspect score
                rank[r] ← ScoreHelper(a, O)
        rank[r] ← rank[r] * User(r.user)            ▷ writer's reputation
    return rank.sort("descending")
```

Chapter 5

Experimental Evaluation

Overall, Figure 5.1 illustrates the process for our whole experimental evaluation. It also serves as a guide for this section as follows. OwnerView is evaluated over real TripAdvisor review data, whose collecting and pre-processing are detailed in Section 5.1. We evaluate our solution against ground truth dataset built from scratch, as presented in Section 5.2. We note that our method may be applied over any review dataset, not being limited by TripAdvisor or a POI category. The baseline for experimental comparisons is described in Section 5.3. Then, Section 5.4 presents the evaluation metrics, whereas Section 5.5 goes over the results and discussions.

5.1 Data and Pre-processing

Our experimental evaluation considers a dataset consisting of 72,876 reviews from 9,676 hotels, which were randomly selected from 55,238[1] in the United States. The data were collected from TripAdvisor in July 2016, representing all reviews from the selected hotels on the collecting date. A pre-processing is also necessary to eliminate incomplete and noisy data. We do so by filtering out: repeated reviews from a hotel, reviews with less than ten words, reviews with non English words (more than 10%), and meaningless words such as URL. The statistics of such filtering is presented in Table 5.1. Such step reduced the number of reviews by about 16%, thus obtaining a final set of 61,815 reviews.[2] Finally, the API (Section 4.1) was applied to the final set of reviews and POIs answers.

Each review has the following data: writer's username, number of reviews made

[1] Total of US hotels on TripAdvisor
[2] The final dataset is available at `http://www.dcc.ufmg.br/~mirella/projs/apoena`

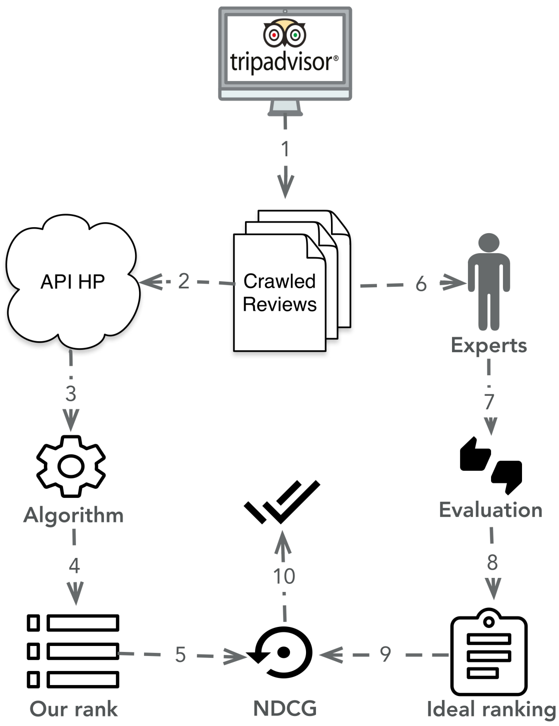


Figure 5.1. Ranking evaluation steps

Table 5.1. Filtered reviews in pre-processing step

	# reviews	%
Repeated reviews	728	0.99
Reviews with less than ten words	2,796	3.83
Reviews with non English words	7,537	10.34

by the writer, number of positive votes[3] received by the writer, review date, and POI's owner answer. Optionally, a POI administrator may answer a review, as the example in Figure 5.2. We have also analyzed the dataset by focusing on the data necessary to identify helpful reviews (e.g., review text and answer). Table 5.2 presents the number of positive and negative aspects obtained after processing it with the API described in Section 4.1.

We notice that there are more positive than negative reviews, and negative reviews usually have between zero and three negative aspects. The distribution of the number of positive and negative aspects and answers has a clear trend: the concentration of positive aspects are between three and five, and negative aspects between zero and two. Such analysis points to a possible relation in the way a POI owner answers a user, as discussed in Section 5.3.

[3]Positive vote: a writer receives a positive vote when another user marks one of his/her reviews as useful

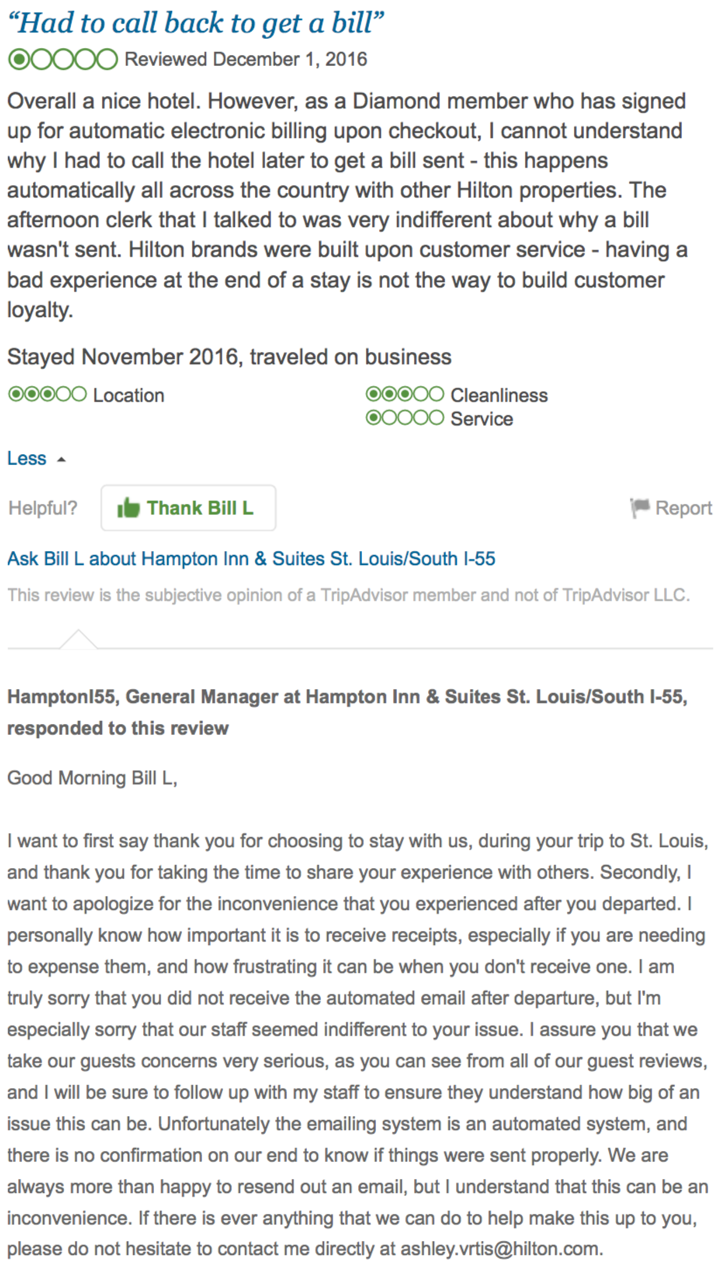


Figure 5.2. A real example of a POI owner review answer

Another interesting fact is that only 1,673 from 9,676 POI owners answer user reviews, representing just 17% of the dataset hotels. Such POIs owners answered 7,271 reviews, representing almost 12% of answered reviews. Despite such a small fraction of answers, POI owners are consistent, as when answering one, they usually answer almost all received reviews.

Table 5.2. Dataset statistics with number of aspects in one review, number of positive/negative reviews and answers.

#Aspects	# pos.rev.	# neg.rev.	# pos.ans.	# neg.ans.
0	490	3,900	96	5,468
1	685	1,508	435	945
2	953	783	840	537
3	1,165	424	1,222	181
4	1,147	244	1,374	80
5	859	147	1,123	35
6	663	90	892	12
7	416	57	530	5
8	273	38	344	4
9	195	23	201	2
10	143	16	86	-
11	82	9	57	1
12	47	6	29	-
13	39	8	16	-
14	18	3	11	-
15	26	5	4	1
16	19	2	4	-
17	13	3	5	-
18	10	2	1	-
19	8	1	-	-
20	5	1	-	-
21	5	-	-	-
22	4	1	1	-

5.2 Ground Truth

Evaluating a recommendation (ranking) algorithm requires to compare the generated output with a ranking from a ground truth dataset (i.e., the ground truth represents the ideal order for a rank). Then, the goal of the experimental evaluation is to check whether the proposed algorithm ranks a given collection as close as the ideal one.

Being one of our big motivations, practically none of the existing methods estimates the helpfulness of a review for a POI owner. Therefore, evaluating our solution requires to build (from scratch) a ground truth dataset from experts' evaluations. This dataset represents the *ideal* ranking of reviews for a POI owner.

Therefore, for the ground truth dataset, we randomly selected 25 hotels (from the United States at TripAdvisor, with more than 1,000 reviews in the last 5 years) and the last 200 reviews from each hotel, giving a total of 5,000 reviews. To rate how helpful those reviews are, we defined a simple web interface to allow experts to evaluate them,

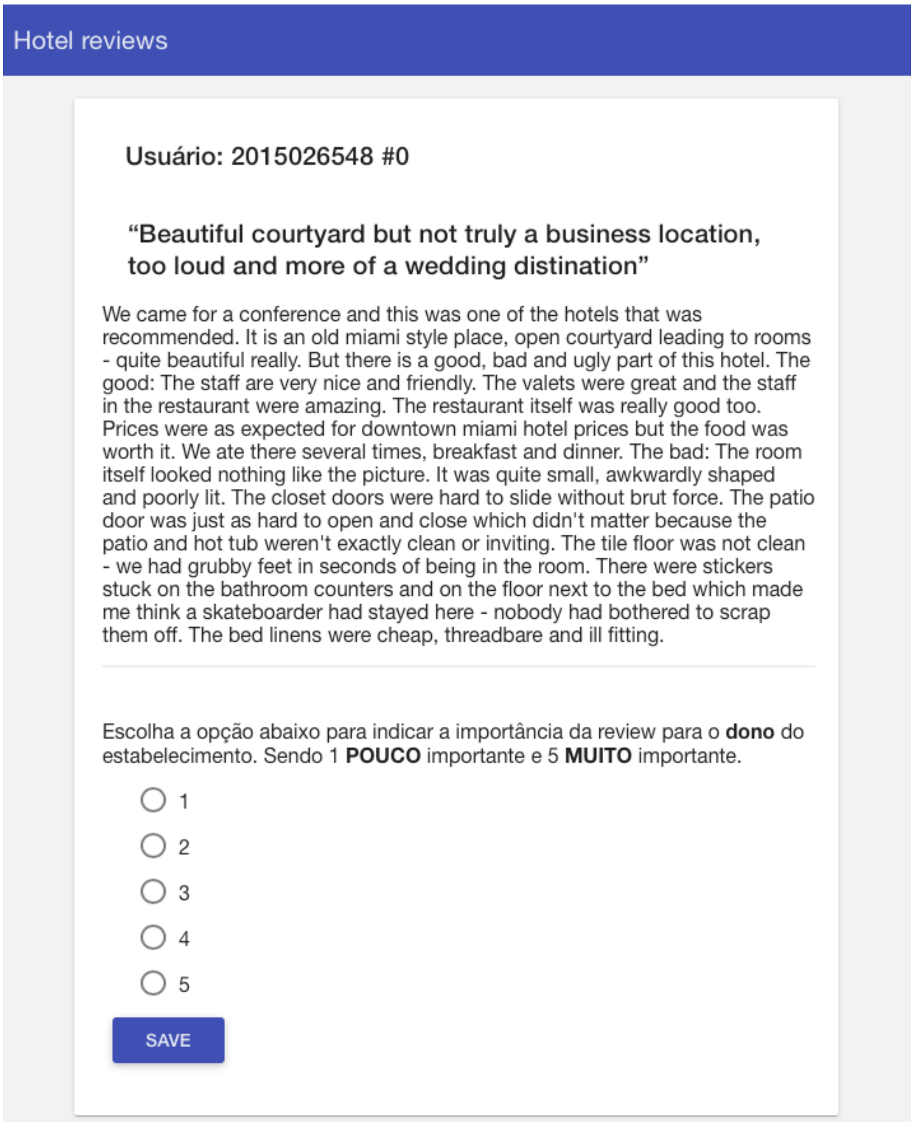


Figure 5.3. Web interface used by experts to evaluate reviews

as illustrated in 5.3. Overall, we asked more than 100 experts (whose profiles are inn Figure 5.4) to score the reviews from 1 to 5 – where 1 means the review is useless for a POI owner and 5 otherwise. Each review was evaluated five times, and then computed the average. The experts were oriented to play as they were the hotel owner and give their rate for the reviews helpfulness.

With such experts evaluations, we built 25 test collections to use as ground truth, one per each hotel. Despite being a small collection, we believe it to be highly reliable, once it was constructed by considering experts opinion for randomly selected hotels.

Table 5.3 shows the average experts scores of each hotel, and Figure 5.5 shows the distribution of reviews score (average) given by experts. One may notice that the

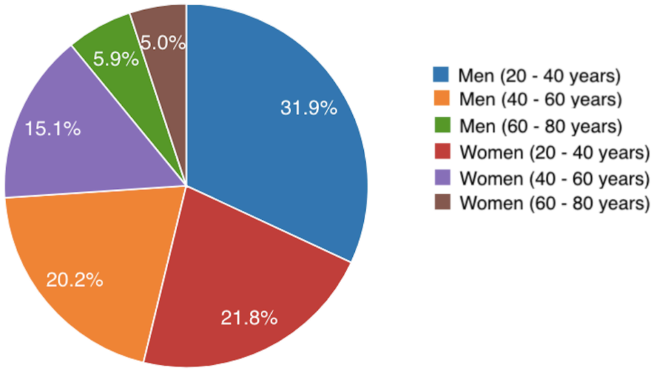


Figure 5.4. Experts age group and gender distribution

Table 5.3. Ground Truth statistics with score, number of evaluations and percentage representation.

Score	# of evaluations	%
5/5	2,041	10.12%
4/5	3,917	19.6%
3/5	5,130	25.6%
2/5	4,099	20.5%
1/5	4,813	24.0%

average evaluation for each hotel is close or less than 3. Considering all reviews, about 70% evaluations were indeed 3 or less. This means that the majority of reviews was not rated as helpful to the owner. Considering this scenario, a hotel has hundreds of reviews, and only a few give relevant information, making it even harder for a POI owner to identify the important ones. Such empirical evaluation shows that our motivation for this work holds: most reviews are indeed not useful for the POI owners.

5.3 Methods for Comparison

In our POI owner oriented context, there is also no baseline to compare our solution against. After analyzing the reviews in Section 5.1, we define a hypothesis that there could be a relation between review quality (helpfulness) and the answer given by the POI owner. If such a relation exists, the POI owners may *explicitly* mention the important aspects in the reviews. Therefore, we propose to analyze such a relation to build (again from scratch) a ranking baseline.

Specifically, a user writes a review mentioning aspects with positive and negative opinions. The negative reviews usually point out problems and dissatisfaction on the establishment and its surroundings. The POI owner may answer such a review by

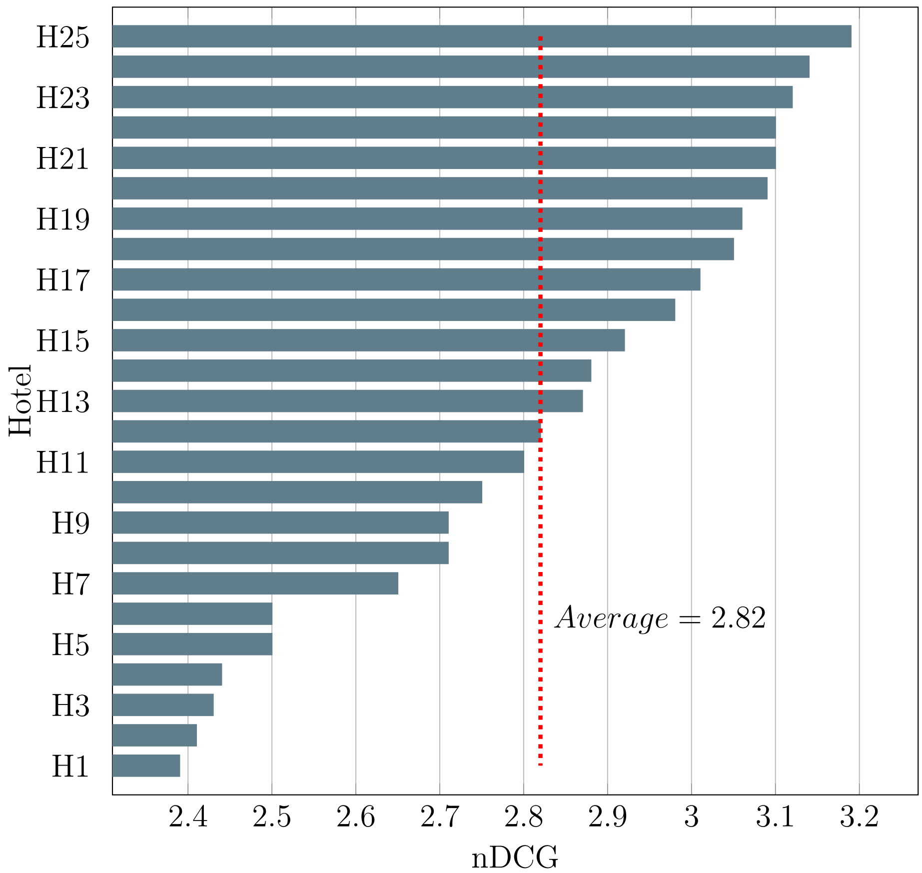


Figure 5.5. Hotels average evaluation score by experts in a 5.0 scale

(usually) thanking the positive compliments and making excuses, explaining or even informing that a solution is on the way for the negative aspects. Either positive or negative, the owner usually mentions again the aspect being answered, e.g., "the *room size* will be improved in the upcoming restoration" and "we are planing regular meetings with the *staff* regarding how to properly treat our guests". Then, such feedback allows to determine the relevance of a review based on the cosine similarity between its aspects and the response given by the establishment. High similarity means the establishment took time in answering the points addressed in the review, i.e., this is a relevant review for the establishment and deserves a proper answer. On the other hand, low similarity may indicate generic responses of establishments (e.g., "thanks for point it out, we will improve it soon").

We note that it is in the interest of an owner to answer to a customer's complaints as informative as possible to prevent losing clients and reputation. Thus, based on the similarity between review and response, our baseline considers a review ranking for each hotel (only reviews answered), in descendant order of similarity.

5.4 Evaluation Metrics

We evaluate our ranking method by computing the Normalized Discounted Cumulative Gain [Balakrishnan and Chopra, 2012] of the top-k reviews in the ranking produced by it (i.e, NDCG@k). NDCG is one of the existing metrics to compare rankings, and is our choice because it is simple, easy and produces good results. NDCG ranges from 0 to 1 indicating greater agreement between the ranking produced by the method and an ideal ranking determined by experts evaluations of hotels reviews. This metric is based on two rules:

- Extremely relevant documents are more important (valuable) than documents with marginal relevance; and
- The lower the position of the document in the ranking, the lower the value of this document for the user.

NDCG@k is built from DCG@k, the discounted cumulative gain in the top-k reviews, which is computed as:

$$DCG_k = \sum_{i=1}^{K} \frac{2^{rel_i} - 1}{\log{(i+1)}} \tag{5.1}$$

where rel_i is the score given by ranking algorithm for the review at position i. Then, the $IDCG_k$ is the ideal value of DCG@k obtained when the reviews are sorted in decreasing order of their actual helpfulness. Overall:

$$nDCG_k = \sum_{i=1}^{K} \frac{DCG_k}{IDCG_k} \tag{5.2}$$

5.5 Experimental results

We start by validating our method against the ground truth dataset to evaluate the quality of the rank produced by our algorithm. Then we compare the baseline versus the ground truth to verify our hypothesis that there is a relation between review and POI answer. When evaluating against the ground truth, it is important to show each hotel evaluation to avoid bias data. Aggregating all data could lead to a satisfactory result on average, but could hide bad evaluations. We also evaluate *OwnerView* against the baseline. We do not validate *OwnerView* explicitly considering temporal data because we would need a ground truth incrementally constructed or a baseline that considers temporal aspects too.

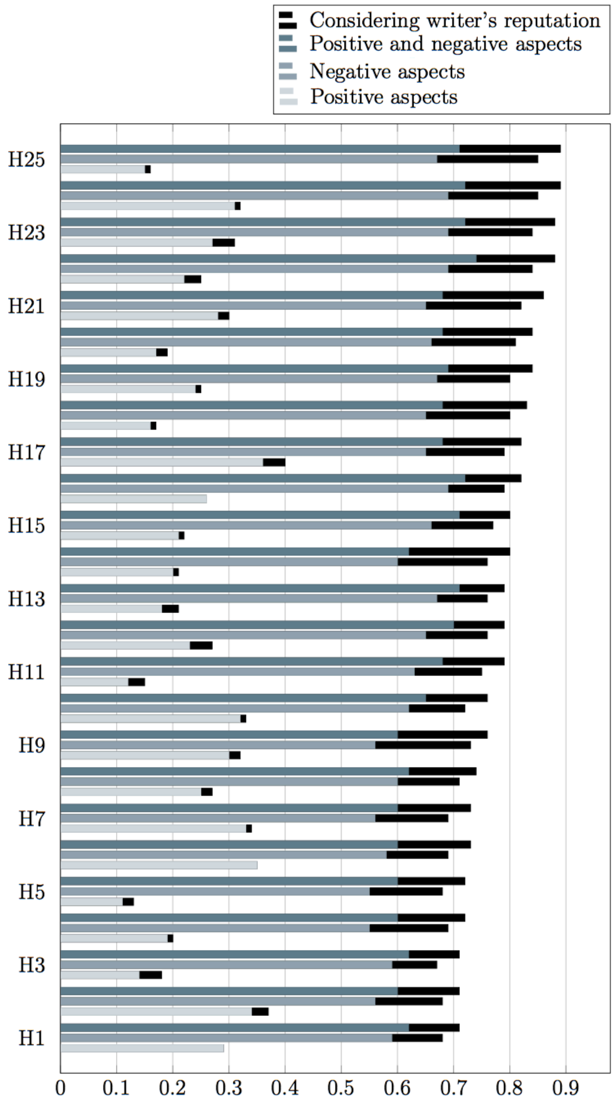


Figure 5.6. Algorithm evaluation scenarios

***OwnerView* against the 25 hotels of the Ground Truth.** We start by evaluating which set of aspects and tuning options presents better results for our algorithm. There are three configurations for which aspects to consider: (i) only positive aspects, (ii) only negative aspects, and (iii) both positive and negative aspects combined (Score-Helper() from Section 5.3). Besides the aspects, we also need to evaluate if considering the review writers reputation score improves the ranking. All cases are evaluated by the nDCG metric described in Section 5.4.

Figure 5.6 shows the results of all evaluated configurations. The worst results are by considering *only* positive aspects, with nDCG lower than 0.4. In contrast, when considering only negative aspects, the nDCG is close or higher than 0.6. The combined option (empirically obtained by tuning *ScoreHelper()*) improves a little the ranking quality, by increasing nDCG in 0.035 on average and keeping all values greater than 0.6. Finally, we evaluate the review writers' reputation allied to both positive and negative aspects. Such combination provides considerable improvement on the ranking, by increasing nDCG in 0.1 on average.

This last configuration is also the best one, giving average nDCG around 0.8, implying that it produces ranking close to the ideal one. Overall, such results confirm our hypothesis that negative aspects are indeed more relevant to POI owners, and giving higher weight to them on our ranking algorithm is a distinct strategy.

Baseline Validation against the Ground Truth. Our baseline was built by considering the hypothesis that there is a relation between the review and its answer from the POI owner. Before considering it as a valid baseline ranking, we must evaluate if such hypothesis holds.

Figure 5.7 shows the nDCG results for all hotels in the ground truth dataset. These results are very promising, as the lowest nDCG value is 0.76, and the greatest is 0.9. The average nDCG is 0.82, indicating that the baseline produces ranking close to the ideal one. In other words, we may use it as baseline to evaluate big data volumes without asking for expert evaluations, which may be expensive or not available. Then, we solve the problem of automatically obtaining a large dataset to use as a comparison ranking, being one of the contributions of this dissertation.

***OwnerView* against the Baseline.** We now present the comparison between our proposed method and the baseline on a large dataset (Section 5.1). We rank the reviews with both methods and then compare the results by using nDCG to show how close our algorithm is to the baseline. Figure 5.8 shows the Cumulative Distribution Function on nDCG value obtained for all 9,676 hotels. Both rankings are tied, and the set of reviews from a hotel has a chance larger than 20% of having a nDCG greater than 0.6. It is important to mention that the set of reviews used is different for each test case, since each is built based on the responses of one establishment at a time.

Overall, OwnerView obtains great results for most test cases and medium results for just a few. Investigating the reason for such a discrepancy, we found that the hotels with worst results were those that provided generic or vague answers to the reviews, without focusing on the complaints of the users (e.g Figure 5.9).

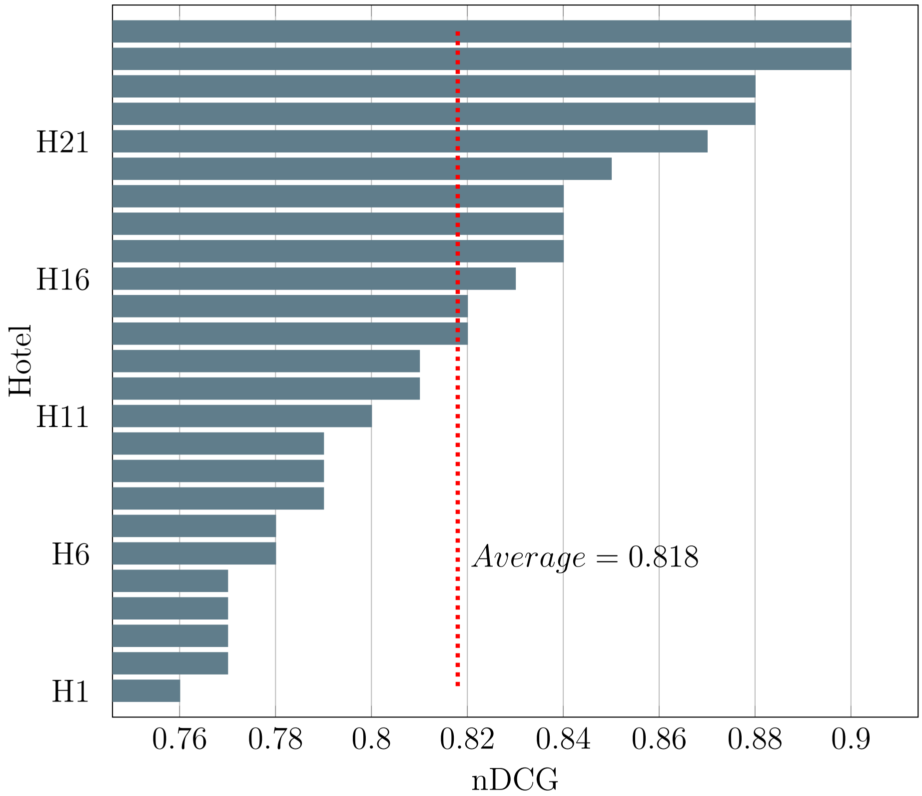


Figure 5.7. Baseline validation against the ground truth

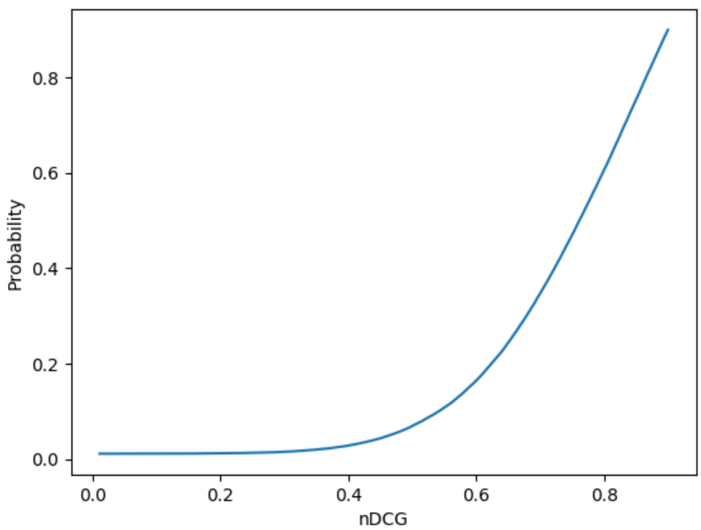
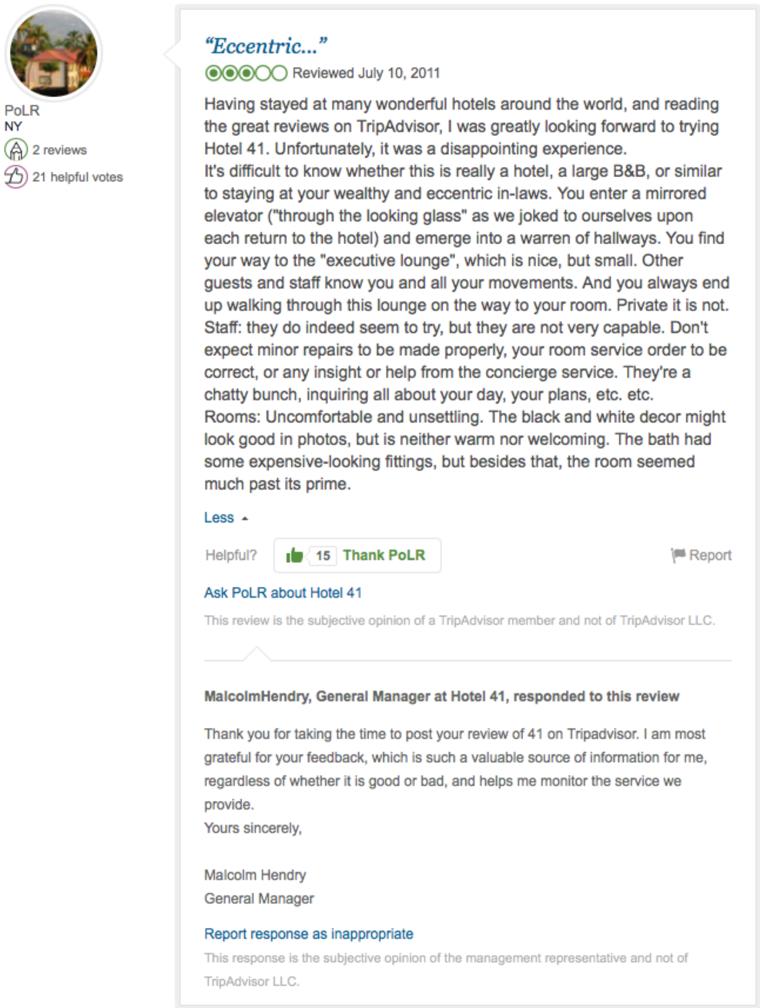


Figure 5.8. Cumulative Distribution Function on nDCG for OwnerView against the baseline


PoLR
NY
2 reviews
21 helpful votes

"Eccentric..."

Reviewed July 10, 2011

Having stayed at many wonderful hotels around the world, and reading the great reviews on TripAdvisor, I was greatly looking forward to trying Hotel 41. Unfortunately, it was a disappointing experience.
It's difficult to know whether this is really a hotel, a large B&B, or similar to staying at your wealthy and eccentric in-laws. You enter a mirrored elevator ("through the looking glass" as we joked to ourselves upon each return to the hotel) and emerge into a warren of hallways. You find your way to the "executive lounge", which is nice, but small. Other guests and staff know you and all your movements. And you always end up walking through this lounge on the way to your room. Private it is not.
Staff: they do indeed seem to try, but they are not very capable. Don't expect minor repairs to be made properly, your room service order to be correct, or any insight or help from the concierge service. They're a chatty bunch, inquiring all about your day, your plans, etc. etc.
Rooms: Uncomfortable and unsettling. The black and white decor might look good in photos, but is neither warm nor welcoming. The bath had some expensive-looking fittings, but besides that, the room seemed much past its prime.

Less

Helpful? 15 Thank PoLR Report

Ask PoLR about Hotel 41

This review is the subjective opinion of a TripAdvisor member and not of TripAdvisor LLC.

MalcolmHendry, General Manager at Hotel 41, responded to this review

Thank you for taking the time to post your review of 41 on Tripadvisor. I am most grateful for your feedback, which is such a valuable source of information for me, regardless of whether it is good or bad, and helps me monitor the service we provide.
Yours sincerely,

Malcolm Hendry
General Manager

Report response as inappropriate

This response is the subjective opinion of the management representative and not of TripAdvisor LLC.

Figure 5.9. Review: generic answer from hotel

Chapter 6

Conclusion and Future Work

In this dissertation, we defined a new twist to a known problem: how to rank reviews according to their helpfulness to the business owner, instead of to the client (which has been the only way, so far). Indeed, we have empirically showed that most reviews are not helpful at all to the owners, as evaluated by experts.

Regarding our solution, we proposed that positive and negative aspects of the review be considered, with more weight to the negative ones. Then, we have empirically showed that reviews on the negative aspects of a POI are more relevant to its owner. Hence, our solution of giving more weight to negative aspects holds. Furthermore, we proposed that not only positive and negative aspects of the review be considered, but also the reputation of its writer. Our experimental evaluation has indeed showed that considering such reputation is paramount for obtaining better ranking results.

As ranking reviews useful for owners (and not clients) is a new problem, there is no current state-of-the-art to compare our solutions against. Therefore, our contributions also include creating two datasets: one ground truth based on experts' evaluation, and one baseline considering the similarity between reviews and their answers. To build the ground-truth dataset, we asked more than one hundred experts to evaluate a subset of 25 hotels with 200 reviews each. Each review was evaluated at least five times, receiving a score between 1 and 5. These evaluations allowed to define the ideal ranking of reviews from the 25 evaluated hotels. The result of the baseline also allowed to solve the problem of automatically obtaining a large dataset to evaluate a review ranking focused on POIs owners. This evaluation also provides great results demonstrating that our method could be experimented in a real scenario to help POI owners making decisions. Both are publicly available for download and further exploration.[1]

Given its complexity, there are different directions to keep pursuing a solution

[1] The final dataset is available at `http://www.dcc.ufmg.br/~mirella/projs/apoena`

for proposing owner-oriented review ranking. For example, we plan to investigate the benefits of adding new features to *OwnerView* such as geographic ones. We also plan to further analyze the temporal dynamics of the reviewing process and its correlation to POI popularity.

Finally, a paper presenting our problem and its solution has been accepted for publication at ACM Hypertext 2017 [Prado and Moro, 2017].

Bibliography

Askalidis, G. and Malthouse, E. C. (2016). The value of online customer reviews. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 155--158, Boston, Massachusetts, USA.

Balakrishnan, S. and Chopra, S. (2012). Collaborative ranking. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, pages 143--152, Washington, USA.

Bigonha, C., Cardoso, T. N. C., Moro, M. M., Gonçalves, M. A., and Almeida, V. A. F. (2012). Sentiment-based influence detection on twitter. *Journal of the Brazilian Computer Society*, 18(3):169–183.

Bobadilla, J., Ortega, F., Hernando, A., and GutiéRrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, 46:109--132.

Chatterjee, P. (2001). Online reviews: Do consumers use them? *Association for Consumer Research*, 28(1):129–134.

Cheng, L., Ke, Z., and Shiue, B. (2011). Detecting changes of opinion from customer reviews. In *Eighth International Conference on Fuzzy Systems and Knowledge Discovery*, pages 1798--1802, Shanghai, China.

Chevalier, J. A. and Mayzlin, D. (2003). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 10(1):345–354.

Dellarocas, C., Zhang, X. M., and Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21(4):23 – 45.

Duan, W., Gu, B., and Whinston, A. B. (2008). Do online reviews matter? - an empirical investigation of panel data. *Decision Support Systems*, 45(4):1007--1016.

Ghose, A. and Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23(10):1498--1512.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., and Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1):5--53.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 168--177, Seattle, Washington.

Hu, Y., Chen, Y., and Chou, H. (2017). Opinion mining from online hotel reviews - A text summarization approach. *Information Processing and Management*, 4(1):436--449.

Jindal, N. and Liu, B. (2007). Review spam detection. In *Proceedings of the 16th International Conference on World Wide Web*, pages 1189--1190, Calgary, Canada.

Jo, Y. and Oh, A. H. (2011). Aspect and sentiment unification model for online review analysis. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, pages 815--824, Hong Kong, China.

Kim, S.-M., Pantel, P., Chklovski, T., and Pennacchiotti, M. (2006). Automatically assessing review helpfulness. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 423--430, Sydney, Australia.

Korfiatis, N., García-Bariocanal, E., and Sánchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications*, 11(3):205--217.

Krawczyk, M. and Xiang, Z. (2016). Perceptual mapping of hotel brands using online reviews: a ext analytics approach. *Journal of Information Technology & Tourism*, 16(1):23--43.

Lacic, E., Kowald, D., and Lex, E. (2016). High enough?: Explaining and predicting traveler satisfaction using airline reviews. In *ACM Hypertext and Social Media*, pages 249--254, Halifax, Canada.

Lappas, T., Sabnis, G., and Valkanas, G. (2016). The impact of fake reviews on online visibility: A vulnerability assessment of the hotel industry. *Information Systems Research*, 9(1):940--961.

Law, D., Gruss, R., and Abrahams, A. S. (2017). Automated defect discovery for dishwasher appliances from online consumer reviews. *Expert System Applications*, 67(c):84--94.

Liu, B. and Zhang, L. (2012). A Survey of Opinion Mining and Sentiment Analysis. In *Mining Text Data*, pages 415–463. Amsterdam, The Netherlands.

Liu, J. et al. (2007). Low-quality product review detection in opinion summarization. In *In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 334--342, Prague, Czech Republic.

Liu, X. and Aberer, K. (2014). Towards a dynamic top-n recommendation framework. In *Proceedings of the 8th ACM Conference on Recommender Systems*, pages 217--224, San Francisco, California, USA.

Liu, Y., Bi, J., and Fan, Z. (2017). Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory. *Information Fusion*, 29(2):149--161.

Liu, Y., Huang, X., An, A., and Yu, X. (2008). Modeling and predicting the helpfulness of online reviews. In *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, pages 443--452, Pisa, Italy.

Lourenço Jr., R. et al. (2014). Economically-efficient sentiment stream analysis. In *Proceedings of the 37th International ACM SIGIR Conference on Research; Development in Information Retrieval*, pages 637--646, Gold Coast, Australia.

Macedo, A. Q., Marinho, L. B., and Santos, R. L. (2015). Context-aware event recommendation in event-based social networks. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 123--130, Vienna, Autria.

Maroun, L. B., Moro, M. M., Almeida, J. M., and Silva, A. P. C. (2016). Assessing review recommendation techniques under a ranking perspective. In *Proceedings of the 27th ACM Conference on Hypertext and Social Media*, pages 113--123, Halifax, Canada.

McAuley, J. and Leskovec, J. (2013). Hidden factors and hidden topics: Understanding rating dimensions with review text. In *Proceedings of the 7th ACM Conference on Recommender Systems*, pages 165--172, Hong Kong, China.

Moghaddam, S., Jamali, M., and Ester, M. (2012). Etf: Extended tensor factorization model for personalizing prediction of review helpfulness. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, pages 163--172, Seattle, Washington.

Mudambi, S. M., Connors, L., and Schuff, D. (2011). Is it the review or the reviewer? a multi-method approach to determine the antecedents of online review helpfulness. pages 1–10, Manoa, Hawaii.

Mukherjee, A., Venkataraman, V., Liu, B., and Glance, N. S. (2013). What yelp fake review filter might be doing? In *Seventh International AAAI Conference on Weblogs and Social Media*, pages 409–418, Michigan, USA.

Ngo-Ye, T. L., Sinha, A. P., and Sen, A. (2017). Predicting the helpfulness of online reviews using a scripts-enriched text regression model. *Expert Systems with Applications*, 1(1):98--110.

O'Mahony, M. P. and Smyth, B. (2009). Learning to recommend helpful hotel reviews. In *Proceedings of the Third ACM Conference on Recommender Systems*, pages 305--308, New York, USA.

Pang, B. and Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 115--124, Michigan, USA.

Phillips, P., Barnes, S., Zigan, K., and Schegg, R. (2009). Tried and tested: the impact of online hotel reviews on consumer consideration. *Tourism Management*, 30(1):235–249.

Prado, T. R. P. and Moro, M. M. (2015). POIView: Análise de Tendências a partir de Revisões Online. In *Brazilian Symposium on Databases, Demo Session*, pages 167–172, Rio de Janeiro, Brazil.

Prado, T. R. P. and Moro, M. M. (2017). Review Recommendation for Points of Interest's Owners. In *Proceedings of the ACM Conference on Hypertext and Social Media*, Prague, Czech Republic - accepted for publication.

Qazi, A., Shah Syed, K. B., Raj, R. G., Cambria, E., Tahir, M., and Alghazzawi, D. (2016). A concept-level approach to the analysis of online review helpfulness. *Computers in Human Behavior*, 58(1):75--81.

Raghavan, S., Gunasekar, S., and Ghosh, J. (2012). Review quality aware collaborative filtering. In *Proceedings of the Sixth ACM Conference on Recommender Systems*, pages 123--130, Dublin, Ireland.

Riloff, E. and Wiebe, J. (2003). Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, pages 105--112, Seattle, USA.

Salehan, M. and Kim, D. J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, 14(2):30–40.

Suleman, K. and Vechtomova, O. (2016). Discovering aspects of online consumer reviews. *Journal of Information Science*, 42(4):492–506.

Tang, J. et al. (2013). Context-aware review helpfulness rating prediction. In *Proceedings of the 7th ACM Conference on Recommender Systems*, pages 1--8, Hong Kong, China.

Tsur, O. and Rappoport, A. (2009). Revrank: A fully unsupervised algorithm for selecting the most helpful book reviews. In *Third International AAAI Conference on Weblogs and Social Media*, pages 154--161, San Jose, California.

Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 417 -- 424, Philadelphia, USA.

Vargas, S., Baltrunas, L., Karatzoglou, A., and Castells, P. (2014). Coverage, redundancy and size-awareness in genre diversity for recommender systems. In *Proceedings of the 8th ACM Conference on Recommender Systems*, pages 209--216, San Francisco, USA.

Wiebe, J. M., Bruce, R. F., and O'Hara, T. P. (1999). Development and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 246--253, Maryland, USA.

Ye, Q., Law, R., and Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, 28(1):180--182.

Yu, H. and Hatzivassiloglou, V. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In

Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, pages 129--136, Stroudsburg, PA, USA.

Zhang, Z. and Varadarajan, B. (2006). Utility scoring of product reviews. In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management*, pages 51--57, Arlington, Virginia, USA.

Zheng, Y. (2012). Tutorial on location-based social networks. In *In Proceedings of 21st World Wide Web Conference*, pages 679--688, Lyon, France.

Zheng, Y., Zhang, L., Ma, Z., Xie, X., and Ma, W.-Y. (2011). Recommending friends and locations based on individual location history. *ACM Transactions on the Web*, 5(1):5:1--5:44.

Zhou, S. and Guo, B. (2017). The order effect on online review helpfulness: A social influence perspective. *Decision Support Systems*, 93(1):77--87.