Combining Aspect-Based Semantic Analysis with Multi-Criteria Decision Support Systems

UNIVERSITAT ROVIRA i VIRGILI

Author: Najlaa Maaroof

Supervisors: Prof. Dr. Antonio Moreno and Dr. Aida Valls

Universitat Rovira i Virgili

A thesis submitted in partial fulfilment of the requirements of the degree of MASTER IN COMPUTER SECURITY ENGINEERING AND ARTIFICIAL INTELLIGENCE

Master Degree

Tarragona 2019

This thesis is dedicated to
my daughter Lyan and my son Loay.
You are my dream come true, the greatest gift of my life I have ever received.

Acknowledgements

I would first like to thank my thesis advisors *Prof. Antonio Moreno* and *Dr. Aida Valls*. The doors to Prof. Moreno and Dr. Valls offices were always open whenever I ran into a trouble spot or had a question about my research or writing. They consistently allowed this paper to be my own work, but steered me in the right direction whenever they thought I needed it.

I would also like to thank the expert who was involved in the sentiment analysis part of this research project: my husband, Mr. Mohammed Jabreel. Without his passionate participation and input, this project could not have been successfully conducted.

Finally, I must express my very profound gratitude to my parents for providing me with unfailing support and continuous encouragement throughout my years of study. This accomplishment would not have been possible without them. Thank you.

Abstract

Nowadays, most decision processes rely on two sources of information: the decision maker's preferences and the reviews, about alternatives, published in online platforms. In the literature there are many systems, based on Multi-Criteria Decision Aid, that have been proposed to automate the decision-making processes taking into account the decision maker's preferences. On the other hand, Sentiment Analysis is the field of study that aims at developing systems to analyze consumers' reviews. However, to the best of our knowledge, there are not any works that utilize these two sources of information to support and improve the decision-making process.

In this thesis, we propose a novel outranking system, called SentiRank, which combines the decision maker's preferences with the users' reviews about the alternatives that the decision maker needs to rank. The contribution of this work has two sides: theoretical and practical. From the theoretical side, we propose an automatic method to integrate the users' feedback with the decision maker's preferences in the ranking process. We define a methodology to use the aspect categories, which can be extracted from the users' reviews, as additional attributes in the decision process. From the practical side, we have implemented the system using Java and applied it to a dataset of restaurants in Tarragona. The experimental results confirm the validity of our hypothesis on the interest of considering the customers' reviews in the decision-making process.

Resumen:

Hoy en día, la mayoría de los procesos de decisión se basan en dos fuentes de información: las preferencias de los tomadores de decisiones y las opiniones sobre las alternativas, que se publican en las plataformas en línea. En la literatura se han propuesto muchos sistemas, basados en la Ayuda a la Toma de Decisiones Multi-Criterio, para automatizar los procesos de toma de decisiones teniendo en cuenta las preferencias del tomador de decisiones. Por otro lado, el Análisis de Sentimientos es el campo de estudio en que se desarrollan sistemas para analizar las opiniones de los consumidores. Sin embargo, hasta donde sabemos, no hay ningún trabajo que utilice estas dos fuentes de información para apoyar y mejorar el proceso de toma de decisiones.

En esta tesis, proponemos un novedoso sistema de ordenación, llamado Senti-Rank, que combina las preferencias del tomador de decisiones con las opiniones de los usuarios sobre las alternativas que el tomador de decisiones necesita ordenar. La contribución de este trabajo tiene dos lados: teórico y práctico. Desde el punto de vista teórico, proponemos un método automático para integrar los comentarios de los usuarios con las preferencias de los tomadores de decisiones en el proceso de clasificación. Definimos una metodología para utilizar los tipos de aspectos que pueden extraerse de las revisiones de los usuarios como atributos adicionales en el proceso de decisión. Desde el punto de vista práctico, hemos implementado el sistema utilizando Java y lo hemos aplicado a un conjunto de datos de restaurantes en Tarragona. Los resultados experimentales confirman la validez de nuestra hipótesis sobre el interés de considerar las opiniones de los clientes en el proceso de toma de decisiones.

Resum:

Avui en dia, la majoria dels processos de decisió es basen en dues fonts d'informació: les preferències dels responsables de prendre la decisió i les opinions sobre les alternatives que es publiquen a les plataformes en línia. A la literatura s'han proposat molts sistemes, basats en el suport a la presa de decisions multi-criteri, que intenten automatitzar els processos de presa de decisions tenint en compte les preferències del responsable de la decisió. D'altra banda, l'anàlisi de sentiments és el camp d'estudi que té com a objectiu desenvolupar sistemes per analitzar les opinions dels consumidors. Tot i això, segons el que sabem, no hi ha treballs previs que utilitzin aquestes dues fonts d'informació per donar suport i millorar el procés de presa de decisions.

En aquesta tesi, proposem un nou sistema d'ordenació, anomenat SentiRank, que combina les preferències del usuari amb les opinions on-line sobre les alternatives. La contribució d'aquest treball té dues vessants: teòrica i pràctica. Des del vessant teòric, proposem un mètode automàtic per integrar el feedback dels usuaris amb les preferències del usuari en el procés de classificació. Definim una metodologia per utilitzar els tipus d'aspectes, que es poden extreure de les opinions dels usuaris, com a atributs addicionals en el procés de decisió. Des del vessant pràctic, hem implementat el sistema en Java i l'hem aplicat a un conjunt de dades de restaurants de Tarragona. Els resultats experimentals confirmen la validesa de la nostra hipòtesi sobre l'interès de considerar les opinions dels clients en el procés de presa de decisions.

Contents

List of Figures

List of Tables

"Public opinion is no more than this: what people think that other people think."

— Alfred Austin

1 Introduction

Contents

1.1 Motivation

The massive spread of online platforms, such as social networks and online shopping sites, which allow sharing reviews and opinions has changed different decision-making processes, especially the purchase process. "What other people think" has become an essential piece of information, as much as what we prefer, for most of our decision-making process. It is well known that price and other criteria have a significant influence on people' purchases; however, nowadays, online consumers' reviews have also become a common source of influence and an important input in people decision making [1–3]. Some studies have shown that 90% of consumers read online reviews before making a purchase decision [4, 5].

Commonly, service and products providers try to attract customers by publishing the best description of the features of their services. In such a case, it easy to find most of the alternatives that fit the preferences of decision-makers. However, it is also important to consider the quality of the advertised features and the aspects of alternatives. For instance, people usually prefer to visit restaurants with free WiFi which we can find in the description of most of the restaurants. Without customers' review, it is not possible to figure out the quality of this service.

Thus, to obtain better recommendations for each user we will consider two sources of information: the decision-maker preferences and online reviews. In this case, an important question is how to integrate these two sources to support and improve the decision-making process. We can find in the literature numerous systems proposed for modeling decision-makers preferences and ranking alternatives to support the decision process [6, 7]. On the other hand, there are different systems for analyzing people' reviews on products or services [8]. However, there are not any works on exploiting user reviews alongside decision-maker preferences to support the decision process.

Multiple Criteria Decision Aid (MCDA) [6] is the research field that deals with different conflicting criteria, establishing scientific bases to elaborate recommendations according to the needs of decision-makers. Artificial Intelligence techniques can be used to elicit and construct a knowledge model for each particular decision-maker.

Sentiment Analysis (SA), also known as Opinion Mining, is the problem of identifying people's opinions, sentiments, or attitudes expressed in a text. It usually involves the classification of a text into categories such as "positive," "negative," and "neutral." Due to the rapid growth of social networks and online platforms on the Internet, SA has been applied to analyze opinions on Twitter, Facebook, and other digital communities in real-time. SA has now a wide range of applications in fields like marketing, management, e-health, politics, and tourism [8–10]. For instance, it can enhance the capabilities of customer relationship management systems and recommenders by finding out which features customers are particularly interested in or avoiding the recommendation of items that have received unfavorable feedback.

We can use SA at different levels. Coarse-grained analysis attempt to extract the overall polarity on a document or sentence level, whereas, in a fine-grained level of analysis, the problem is to identify the sentiment polarity towards a specific aspect in a given text; this level of analysis is called Aspect-Based Sentiment Analysis (ABSA)[11]. An "aspect category" of an entity is a concept represents a unique aspect of the entity, while an "aspect expression" is an actual word or phrase that appears in the text indicating an aspect category. For example, picture, image, and photo are the same aspect for the entity camera. Thus, we need and group them in one aspect category. Each aspect category should also have a unique name in a specific application. Aspect expressions are usually nouns and noun phrases but can also be verbs, verb phrases, adjectives, and adverbs. For example, "picture quality" in "The picture quality of this camera is great" is an explicit aspect expression. Aspect expressions that are not nouns or noun phrases are called implicit aspect expressions. For instance, "expensive" is an implicit aspect expression in "This camera is expensive" and it implies the aspect category 'price'.

1.2 Objectives

The main aim of this work is to improve the current multi-criteria decision support methods by including the information contained in online reviews about a set of alternatives. Hence, our intention in this work is to understand how we can exploit SA to integrate the users' reviews in a multi-criteria ranking system to improve the quality of the decision-making process. Towards this objective we identify the following research questions:

- **Q1**: Which properties of user reviews can be used to improve the quality of the decision-making process?
- **Q2**: How can we extract those properties?
- **Q3**: How can we integrate the extracted properties in the decision-making problem formulation?
- **Q4**: How can we, automatically, solve the resulting decision-making problem?

1.3 Contributions

In response to the objectives described above, this master thesis proposes the SentiRank system. It is based on a multi-criteria decision aid system to give users the ability to exploit users' reviews alongside their preferences in the process of alternatives ranking. To achieve that, we primarily examine the task of ABSA. Given a collection of review documents about a set of alternatives, the goal is to automatically detect and analyze all expressions of sentiment towards a set of aspects in these documents. This problem setting involves mainly two subtasks.

- **Aspect Category Extraction**: Given a document of review about an alternative, we want to automatically extract all the aspects mentioned in the given document.

- **Aspect Category Polarity Identification**: Reviewers refer to product aspects in different contexts. They may use factual language and simply describe some aspects (e.g., "the camera has a 3x optical zoom") or they may express their opinion towards an aspect (e.g., "the 3x optical zoom works perfectly"). Our goal is to automatically detect expressions of sentiment in customer reviews towards the extracted aspects. We further aim at analyzing the polarity of these expressions. We want to know whether an utterance is predominantly positive (e.g., "works perfectly") or negative ("is totally crap").

After that, we exploit an MCDA method, named the ELECTRE-III, to develop a ranking system based on the decision-maker's preferences and the users' reviews.

To summarize, the main contributions of this thesis are as follows:

1. We have developed an ABSA system to extract the set of aspect categories mentioned in reviews and identify the expressed sentiments towards the extracted aspects and used it as unit in the developed outranking system.

2. The experimental results reveal that our ABSA system achieves the state-of-the-art systems' performance.

3. We have proposed a new multi-criteria ranking model by integrating decision-makers' preferences and users' reviews and used the ELECTRE-III method to solve the proposed problem.

4. We have evaluated the developed system in the Restaurant domain by collecting information on a set of restaurants in Tarragona and the reviews about them. After that we have used them to test the system.

1.4 Thesis Structure

The thesis is structured into five chapters and one appendix, namely Chapter 1 "Introduction", Chapter 2 "Background", Chapter 3 "Methodology", Chapter 4 "Experiments and Results" and Chapter 5 "Conclusion and Future Work". This chapter (i.e., Chapter 1) has introduced the work. Chapter 2 explains the background of the work. It provides an overview of sentiment analysis in general and aspect-based sentiment analysis in particular. We also overview, in chapter 2, the Multi-Criteria Decision Aids and the Outranking problem and cover the state of the art. In chapter 3, we describe the methodology. In particular, we describe the development of the aspect-based sentiment analysis system, the architecture of our ranking system and its implementation details. Chapter 4 presents the experiments and discusses the results. Chapter 5 summarizes the thesis and points out some ideas for future work. Appendix A describes the graphical user interface of the implemented system.

"All opinions are not equal.
Some are a very great deal more robust, sophisticated and well supported in logic and argument than others."

— Douglas Adams, *The Salmon of Doubt*

2
Background

Contents

2.1 Introduction

In this chapter, we place the contributions of this thesis into a broader context. We provide a brief overview of the relevant research areas and fields of application. We discuss the most important problem settings, introduce the related terminology and point out some basic approaches.

Section 2.2 starts with an introduction to sentiment analysis, and Section 2.3 provides an overview and outlines the main research topic of this thesis, namely, aspect-based sentiment analysis. In Section 2.4, we provide an overview of the multi-criteria ranking problem and the most common approaches to solving it. Subsequently, Section 2.5 summarizes this chapter.

2.2 Sentiment Analysis

The term sentiment analysis refers to the field of study in Natural Language Processing (NLP), which aims to develop theories, models, and approaches to treat the expression of opinions in natural language texts automatically. The goal is to determine the author's opinion about a specific target, or in a more abstract way, about a specific aspect. According to Liu [8], there are also many names and slightly different tasks for sentiment analysis, e.g., opinion mining, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining; they are now all under the umbrella of sentiment analysis or opinion mining. In this thesis, we prefer to use the term *sentiment analysis*.

2.2.1 Sentiment Analysis Tasks

Sentiment analysis is a heterogeneous field of study. Based on the application scenario and the type of textual data, different subproblems are defined. In the following, we present the most common tasks in sentiment analysis.

Sentiment polarity is the most well-studied subproblem of sentiment analysis, which we regard as a multi-class text classification problem. The goal is to determine whether the overall tone in a given text is positive, negative, or neutral. There are many other variants of this task. One is to classify a text according to a rating scale, e.g., 1="worst" to 5=" best." We may consider different levels of analysis, for example, document-level classification vs. sentence-level classification. It is also possible to formulate the problem as a regression problem. The goal, in this case, is to predict the intensity of sentiment that best represents the mental state of the author [12]. The intensity is a real-value in the range from 0 to 1. The 0 score means the most negative and 1 is the most positive.

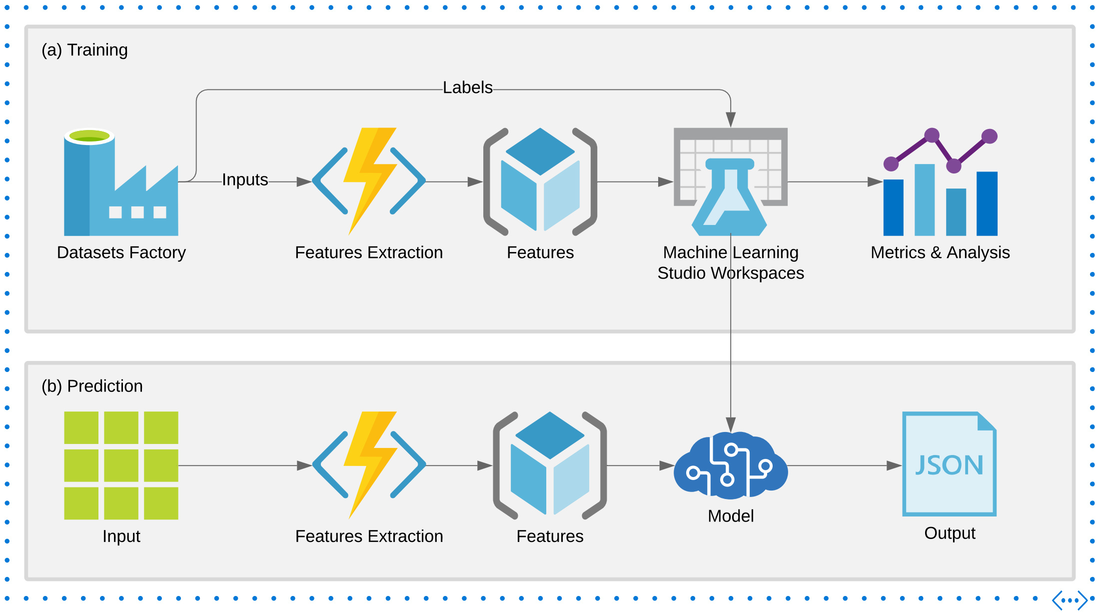


Figure 2.1: The text classification architecture.

Since it is a text classification problem, we can apply any existing supervised learning method, e.g., Naïve Bayes, and support vector machines (SVM) [13–15]. Figure 2.1 shows the pipeline of the general text classification problem. It is composed by two main phases: training phase and prediction phase. In the training phase, we learn a function or model to map an input into one of a set of predefined classes. The training phase requires labeled samples to train and validate the model. Each labeled sample is a tuple of the text and its corresponding label. Once we get the model, we use it in the prediction phase to classify unseen inputs. Both the training and the prediction phases share the same features extraction process. It aims to transform the input text into a numeric vector. In most cases, we precede the feature extraction process by some pre-processing steps to clean and normalize the text.

2.3 Aspect-Based Sentiment Analysis

Classifying text opinions at the document level or sentence level is usually insufficient for most of the applications. With these two levels, we can not identify the opinion targets or assign sentiments to them for two main reasons. First, people tend to mention multiple aspects of an entity in their reviews. Figure 2.2 shows the typical layout of online customer reviews. Second, a single opinion does not reflect the author's opinion about all aspects of an entity. Hence, we need to perform the analysis at high levels. To this end, we need two steps. First, we need to find all the aspects mentioned in the document. As soon as we get these aspects, we need to decide if the sentiment expressed towards each aspect is positive or negative. The first step is known as the aspect extraction task, whereas the second step is known as the aspect polarity detection task. The following sections describe these two tasks.

2.3.1 Aspect Extraction

Aspect extraction aims to extract the set of aspects mentioned in a given text. It is an essential step in opinion mining. The first work on aspect extraction was introduced by Hu and Liu, 2004 [16]. Although they distinguished between explicit and implicit aspects, they only dealt with explicit aspects. The main idea of their approach is to design a set of rules based on statistical observations. Popescu and Etzioni, 2007 [17] improved this method by assuming the product class is known prior. Based on that, they proposed an algorithm that detects whether a noun or noun phrase is a product feature by computing the point-wise mutual information between the noun and the product class. Scaffidi et al., 2007, [18] also tried to improve Hu and Liu's method by developing a language model to identify product features. They assumed that product features are more frequent in product reviews than in general text. Recent studies have shown that this syntactical approach performs quite well. This approach is beneficial in practice, as it is domain-independent and unsupervised. The idea is to employ rules on grammar dependency relations between the opinion words and the aspects. The main issue of this approach is that the rules need to be selected and tuned manually to avoid producing too many errors. Although it is easy to evaluate the accuracy of each rule automatically, it is not easy to select a set of rules that produces the best overall result due to the overlapping coverage of the rules. Liu et al., 2015 [19] proposed a novel method to select a useful set of rules to extract aspects by

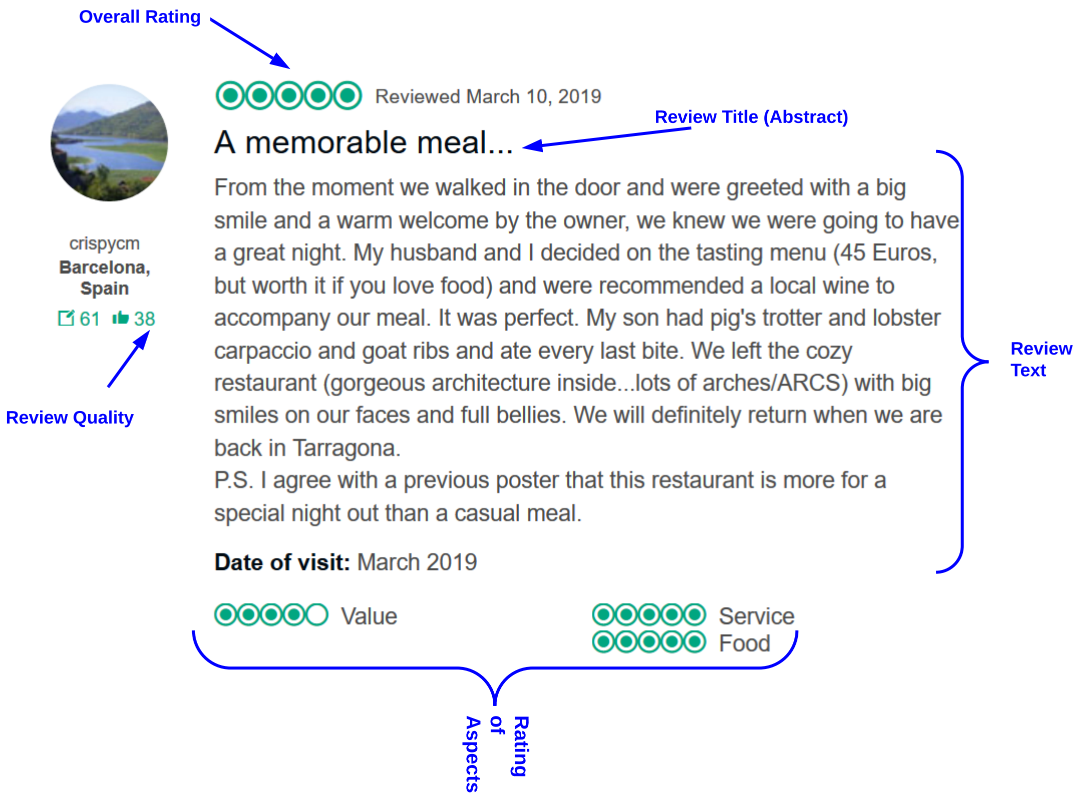


Figure 2.2: The layout of a typical online customer review.

exploiting the dependency relations between opinion words and aspects, as well as between opinion words and the aspects themselves.

We can regard the aspect extraction as a sequential labeling problem. In this case, the extracted aspects are terms rather than categories. The most popular methods in this context are Hidden Markov Models (HMM) and Conditional Random Fields (CRF). The authors in [20] used a lexicalized HMM to extract opinions along with their explicit aspects. Jakob and Gurevych, 2010 [21], Li et al., 2010 [22] and Choi and Cardie, 2010 [23] employed CRF to extract explicit aspects.

It is worth mentioning that most of the approaches mentioned above deal only with explicit aspects.

Some studies have employed the multi-label classification approach to extract aspect categories from a text. First, they define a set of aspect categories. Then, they develop a system to map sentences to none, one or more aspect categories from the defined set. For example, the authors in [24] developed five binary SVM classifiers. Each one specifies whether a sentence belongs to a specific category or not. The NLANGP team [25, 26] also modeled the aspect category extraction as a multiclass classification problem by developing a Logistic Regression for each category. The features used in their system are n-grams, parsing, and word clusters built from Amazon and Yelp data. The system of Sentiue [27] used a separate MaxEnt classifier with bag-of-word-like features (e.g., words, lemmas) for each entity and each attribute. Then, they applied heuristics on top of the output of the classifiers to determine which categories are assigned to each sentence. Following this approach, i.e., multi-label classification, we have developed a system to extract aspects from reviews as explained in chapters 3 and 4.

2.3.2 Aspect Polarity Detection

The goal of the aspect polarity detection task is to identify the polarity towards each aspect extracted from a given customer review about a target entity. The idea is to identify a set of $\{a, p\}$ tuples that summarize the opinions expressed in the review. The item a is extracted in the aspect extraction step while p can be positive, negative, neutral, or conflict. The task can be regarded as a classification problem where the input is a pair of a sentence, s, and an aspect, a, and the output is the polarity, p. Hence, a single sentence may have several opinions about different aspects, and each may have a different and independent polarity. Based on that, the key of solving this problem is to design a set of useful features. Such features need to be able to model the relations between the sentences and the aspects.

In this context, many studies have proposed efficient solutions. For example, the Sentiue team [27] developed a three class polarity classifier by extracting features from the text and the aspect category, including BoW, lemmas, bigrams after verbs, presence of polarized terms, and punctuation. The ECNU team [28] found that each aspect is related to specific fragments in the corresponding sentence. Based on that, they extracted from a given sentence what they called pending words by applying two steps. First, the segmentation step was used to split each sentence into several fragments. Then, the selection step selects out one or more fragments from the sentence for each aspect. The words in the selected fragments are the pending words. After that, they used four types of features extracted from the pending words to build logistic regression classifiers. The types of features used in their system were linguistic features, sentiment lexicon-based features, topic modeling-based features, and word embedding-based features. The NRC-Can-2014 team [24] trained one multi-class SVM classifier for all aspect categories. The feature set was designed to incorporate the information about a given aspect category using a domain adaptation technique. Following these approaches, we have designed our aspect category polarity detection, as explained in chapters 3 and 4.

2.4 Multi-Criteria Decision Aid

Decision problems [6], e.g., ranking, choice, and sorting problems, are usually complicated as they involve several criteria. People no longer consider only one criterion to make a decision. Moreover, in most of the cases, there is not any perfect alternative that matches all the criteria. An ideal option does not usually exist. Therefore, we must find a compromise. To this end, the decision-maker can make use of naive approaches such as a simple weighted sum. However, such a simple method can only be applied with the right constraints (e.g., correct normalization phase, independent criteria) to enable reasonable outputs. Moreover, it assumes the linearity of preferences which may not reflect the decision maker's preferences in many cases.

MCDA methods have been developed to support the decision-makers in their unique and personal decision process [29]. They are widely used in decision problems to find the "best possible" alternative solution. MCDA makes the process more explicit, rational, and efficient. It is a multidisciplinary field, deriving from Operations Research, and it uses mathematical approaches to deal with complex problems encountered in human activities. Nowadays, it also integrates artificial intelligence and economic welfare techniques.

A large number of methods have been developed to solve multi-criteria problems. In this section, we introduce the basic concepts required for modeling decision problems in MCDA and focus on outranking methods. We also present the ELECTRE-III method, one of the most well known outranking methods.

2.4.1 Multi-Criteria Decision Problem

The formalization of a multiple criteria decision aiding procedure is made from the point of view of the so-called decision-maker, who is the person that must make a decision. The data required in a decision problem is represented using two key elements:

1. Alternatives. They are the potential actions for the decision problem from which the decision-maker has to decide. Alternatives are represented as follows: $A = \{a_1, a_2, ..., a_n\}$ is the finite set of alternatives and n is the number of alternatives in A.

2. Criteria. A criterion is a tool constructed for evaluating and comparing potential actions according to the decision-maker subjective point of view about some reference indicators. The set of criteria is denoted $G = \{g_1, g_2, ..., g_m\}$

The type of MCDA problem depends on the goal of the decision-maker when approaching a certain problem [29]. Three main types are distinguished:

- Choice problem: it is a selection problem where the decision-maker has to choose the best option. MCDA helps the decision-maker to find among all the possibilities the best alternative(s).

- Ranking problem: in this type of problem, the decision-maker wants to get a ranking of the alternatives from the best to the worst. Two (or more) alternatives may be equivalent. The outcome of an MCDA method can be either a partial or a complete ranking of the set of alternatives.

- Sorting problem: the problem of sorting consists of assigning the alternatives to predefined categories. These categories are ordered from the worst to the best. The outcome of an MCDA sorting method is an assignment of the alternatives among the different classes.

2.4.2 Outranking

The outranking methods are based on pairwise comparisons of the options. This means that every option is compared to all the other options. Based on these pairwise comparisons, final recommendations can be drawn. The next subsection explains one of the most common outranking methods in MCDA and especially in ranking problems.

2.4.3 ELECTRE Methods

The first ELECTRE method was presented by Benayoun, Roy, and Sussman (1966) who reported on the works of the European consultancy company SEMA for a specific real-world problem. However, the first journal article did not appear until 1968, when Roy described the method in detail [30]. Later, it was renamed to ELECTRE I. The name ELECTRE Iv (v for veto) is sometimes used when veto thresholds are considered, but it did not become an official name [29]. Several other ELECTRE methods were developed during the following two decades: ELECTRE II [31], ELECTRE III [32], ELECTRE IV [33], ELECTRE TRI [34] and ELECTRE IS [35]. ELECTRE TRI was later renamed to ELECTRE TRI-B [36], when a new version, ELECTRE TRI-C [37] was developed, but most of the literature still use the name ELECTRE TRI for the original version. Recently, ELECTRE TRI-nC, was presented by Almeida-Dias, Figueira, and Roy (2012) [38] as an extension of ELECTRE TRI-C. All methods, except ELECTRE I, Iv and II, consider the concept of pseudo-criteria. Thanks to indifference and preference thresholds, this concept allows to model imperfect knowledge, which may be a result of uncertainty, imprecision, and ill-determination of specific data. The main advantage of the ELECTRE methods is that they avoid compensation between criteria and any normalization process, which distorts the original data. On the other hand, their main drawback is that they require various technical parameters, which means that it is not always easy to fully understand them. A more detailed history and overview of the ELECTRE methods can be found in [29]. In the following, we discuss in details ELECTRE III as we employ it in our system.

ELECTRE-III

ELECTRE-III is composed by two phases. First, we construct the outranking relationships among the alternatives, and then we exploit them. The decision-maker is required to provide most of the information, including the weight of the criteria, the indifference, the preference, and the veto thresholds in the first phase.

In the construction phase, ELECTRE-III makes use of outranking relations. An outranking relation, where an alternative a outranks another alternative b, denoted by aSb, expresses the fact that there are sufficient arguments to decide whether a is at least as good as b. Also, there are no essential reasons to refute this fact (Roy 1974). The exploitation phase uses one of two methods: Net Flow Score (NFS) or Distillation. In this work, we use NFS as explained below.

We first explain the construction phase steps. We compute the outranking degree between each pair a and b of alternatives to measure or evaluate the outranking relation assertion as the following.

The credibility or outranking degree, $S(a, b)$, is a score between 0 and 1, where the closer $S(a, b)$ is to 1, the stronger the assertion. It considers two aspects: the *concordance* and the *discordance* of the statement that a outranks b. The concordance and discordance are measured respectively while incorporating the decision maker's preference on multiple, usually conflicting, criteria. The user is required to provide the indifference and preference thresholds to compute the concordance degree. The veto threshold is needed to calculate the discordance degree.

Concordance A partial concordance degree $c_i(a, b)$ measures the assertion "a outranks b" or "a is at least as good as b" on the specific criterion g_i. The uncertainty of the decision-maker preference can be represented, as shown in Figure 2.3, by the indifference, q_i, and preference, p_i, thresholds.

The *indifference* threshold indicates the largest difference between the performances of the alternatives on the criterion considered such that they remain indifferent for the decision maker. The *preference* threshold indicates the largest difference between the performances of the alternatives such that one is preferred

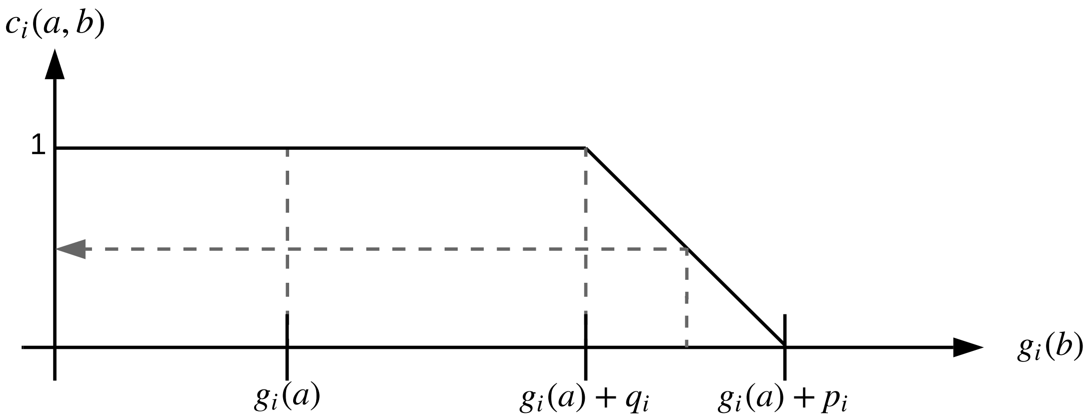


Figure 2.3: The partial concordance index $c_i(a, b)$.

over the other on the considered criterion. Between these two thresholds, the partial concordance degree is computed based on a linear interpolation as shown in Figure 2.3. Given that, we can define the partial concordance index as follows:

$$c_i(a, b) = \begin{cases} 1 & \text{if, } g_i(a) \geqslant g_i(b) - q_i \\ 0 & \text{if, } g_i(a) \leqslant g_i(b) - p_i \\ \frac{p_i - g_i(a) - g_i(b)}{p_i - q_i} & otherwise \end{cases} \tag{2.1}$$

Once we get the values of all the partial concordances, the global concordance index can be computed for each pair of alternatives (a, b) as follows:

$$C(a, b) = \frac{1}{W} \sum_{i=1}^{m} w_i c_i(a, b) \tag{2.2}$$

In this expression w_i is the weight of the criterion g_i, W is the sum of the weights and m is the number of criteria.

Discordance On the other hand, the partial discordance degree $d_i(a, b)$ measures the decision-maker discordance with the assertion "a is at least as good as b" on criterion g_i. If the decision-maker, when considering criterion g_i, strongly disagrees with the assertion, the discordance degree reaches its maximum value, i.e., 1, and reflects the fact that g_i sets its veto. This is the case if the difference in performances (i.e., $g_i(b) - g_i(a)$) is higher than a so-called veto threshold, denoted by v_i. The discordance degree has minimum value 0 when there is no reason to refute the

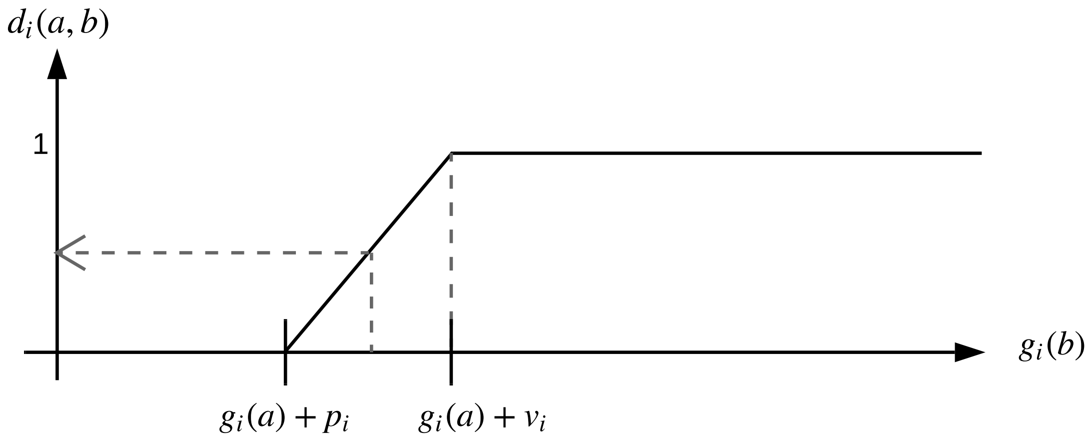


Figure 2.4: The partial discordance index $c_i(a,b)$.

assertion. As in the case of the partial concordance degree, between these two extremes, $d_i(a,b)$ changes linearly between the preference and veto thresholds as a function of the difference $g_i(b) - g_i(a)$, as shown in Figure 2.4. We define the partial discordance index as the following.

$$d_i(a,b) = \begin{cases} 1 & \text{if, } g_i(a) \leqslant g_i(b) - v_i \\ 0 & \text{if, } g_i(a) \geqslant g_i(b) - p_i \\ \frac{g_i(b)-g_i(a)-p_i}{v_i-p_i} & \textit{otherwise} \end{cases} \tag{2.3}$$

Outranking Degree Finally, a global outranking degree $S(a,b)$ summarizes the concordance and discordance degrees into one measure of the assertion "a outranks b" as shown below.

$$p(a,b) = \begin{cases} C(a,b) & \text{if, } d_i(a,b) \leqslant C(a,b), \forall i \\ \prod_{i \in I(a,b)} \frac{1-d_i(a,b)}{1-C(a,b)} & \textit{otherwise} \end{cases} \tag{2.4}$$

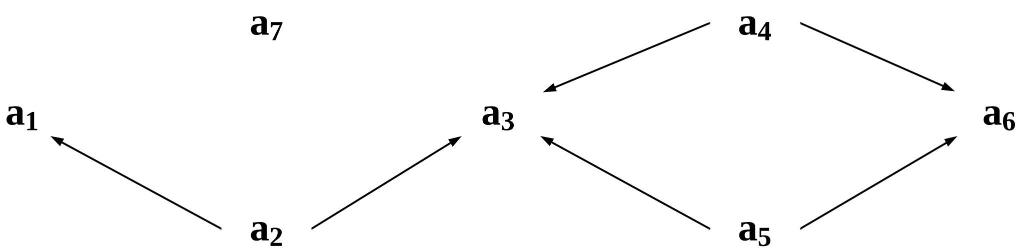
Where $I(a,b)$ is the set of criteria in which $d_i(a,b) > C(a,b)$.

Net Flow Score

The matrix that contains the credibility values for each pair of alternatives may be interpreted as the adjacency matrix of a labelled directed graph, in which the nodes are the alternative and the label of the edge between two alternatives a and b is the value $p(a,b)$. The NFS procedure analyzes this graph to calculate two evidences:

strength and weakness. The strength of an alternative a is defined as the sum of the credibility values of the edges that leave from the node a. The weakness of an alternative a is defined as the sum of the credibility values of the edges that reach the node a. The NFS of an alternative a is the difference between its strength and its weakness. This value allows ranking the alternatives in a descending partial order. Figure 2.5 shows an example of NFS.

$$NFS(a) = \sum_{b \in A} [S(a,b) - S(b,a)] \tag{2.5}$$



S	a_1	a_2	a_3	a_4	a_5	a_6	a_7	Streangth
a_1	0	0	0	0	0	0	0	0
a_2	1	0	1	0	0	0	0	2
a_3	0	0	0	0	0	0	0	0
a_4	0	0	1	0	0	1	0	2
a_5	0	0	1	0	0	1	0	2
a_6	0	0	0	0	0	0	0	0
a_7	0	0	0	0	0	0	0	0
Weakness	1	0	3	0	0	2	0	

NFS(a_1) = 0 - 1 = -1 Rank = 3

NFS(a_2) = 2 - 0 = 2 Rank = 1

NFS(a_3) = 0 - 3 = -3 Rank = 5

NFS(a_4) = 2 - 0 = 2 Rank = 1

NFS(a_5) = 2 - 0 = 2 Rank = 1

NFS(a_6) = 0 - 2 = -2 Rank = 4

NFS(a_7) = 0 - 0 = 0 Rank = 2

Figure 2.5: An example of Net Flow Score.

2.5 Summary

This chapter provided a brief overview of the relevant research areas and fields of application. We discussed the most important problem settings, introduced the related terminology and pointed out some basic approaches. The aim of this overview was to make the understanding of the methodology of this work simpler and easier. We explained briefly the following concepts: sentiment analysis, aspect-based sentiment analysis, the multi-criteria decision problem and the outranking methods. We also presented a review and summary of the state-of-the-art methods that have been used to extract the aspects and analyze the polarities towards them. The next chapters (Chapter 3 and Chapter 4) explain how we utilised the aspect-based sentiment analysis approach to answer the research questions pointed in Chapter 1 and discuss the results that were obtained.

*"You are not entitled to your opinion.
You are entitled to your informed opinion. No one is
entitled to be ignorant"*

— Harlan Ellison

3

Methodology

Contents

3.1 Introduction

This chapter explains the proposed system, SentiRank, the tools, and the resources used to implement it. SentiRank receives as inputs a set of alternatives, e.g., restaurants, hotels, or laptops, to be ranked and their corresponding consumers' reviews. Figure 3.1 shows the overall architecture of SentiRank. It is composed by three primary units. The first one is the customers' reviews unit, which transforms the reviews about alternatives into a matrix of real numbers which we call the social performance table. The second one is the domain analysis unit; it takes as an input the description of the alternatives and the goals of the decision-makers

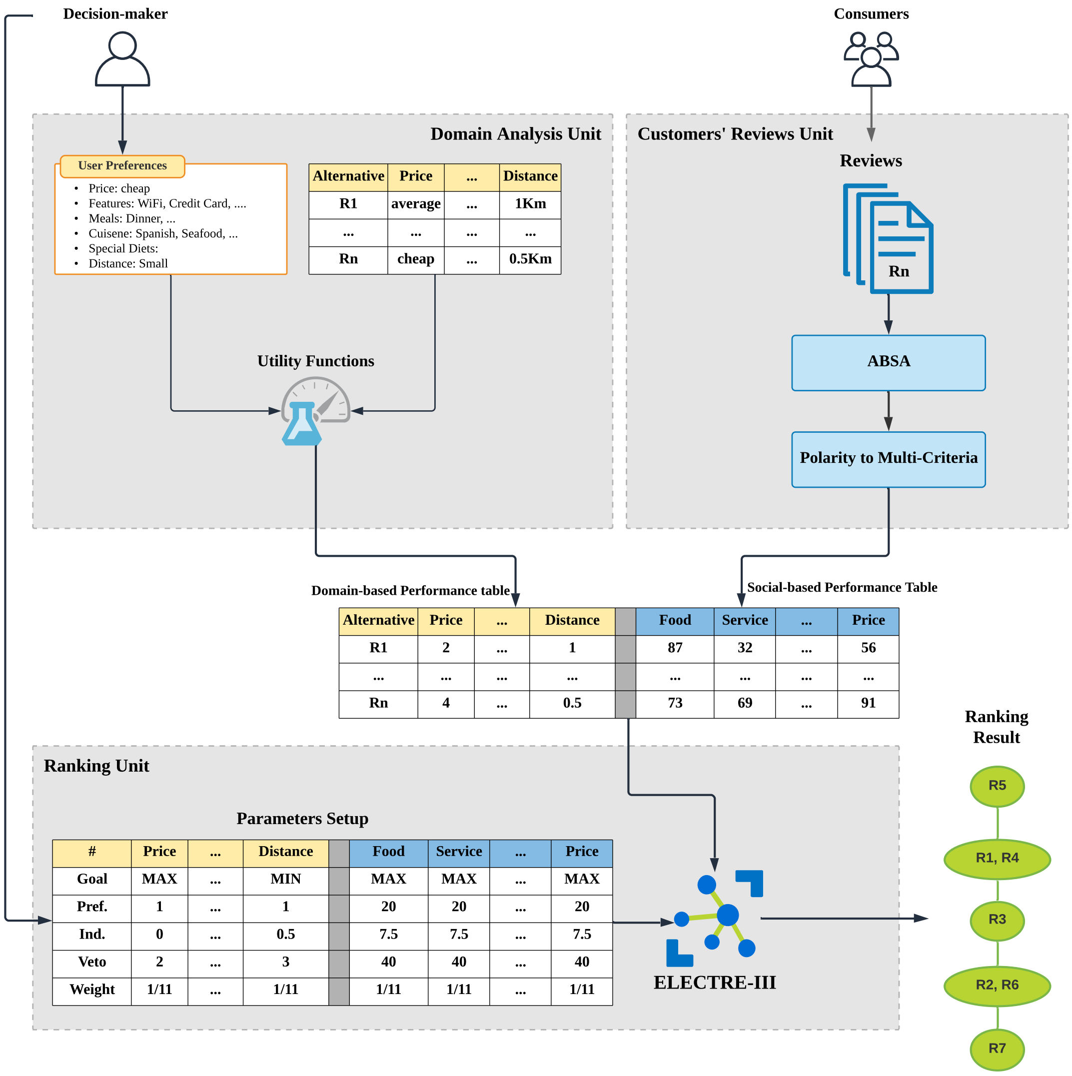


Figure 3.1: The architecture of SentiRank.

and converts them to another matrix of real numbers called domain performance table. As soon as we get the output of the two units, we merge them to get the final performance table. Finally, we feed the performance table to the last unit, i.e., the ranking unit, to get the alternatives ranked.

3.2 Customers' Reviews Unit

The input to this unit is a collection of customers' reviews about given alternatives, and the output is a numerical matrix called social-based performance table. Each

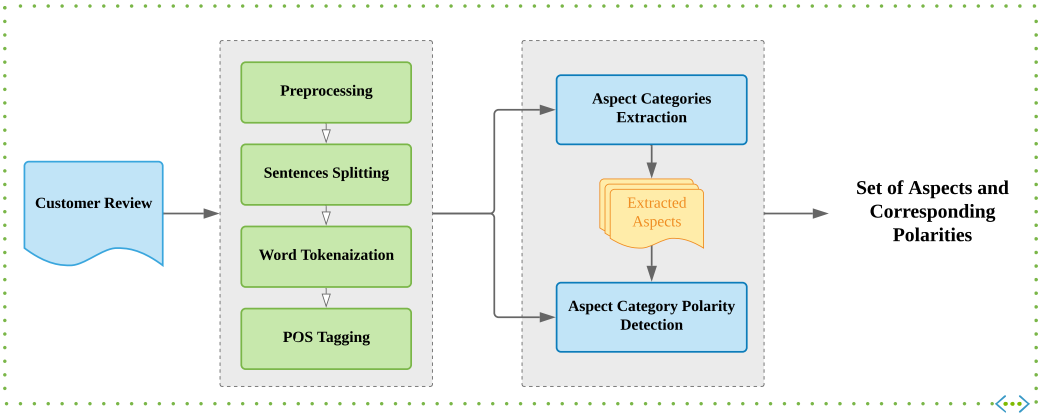


Figure 3.2: The architecture of ABSA model.

column in the matrix represents a specific criterion. The main component of the customers' review unit is the aspect-based sentiment analysis (ABSA) model. Thus, we start by explaining how we build it below.

The ABSA model contains two sub-models. The first one aims to extract the set of aspects commented in a given review. We use the second one to identify the sentiment expressed in the review towards each extracted aspect. As shown in Figure 3.2, the first step is to preprocess the text in the review. After that, we split it into a list of sentences and tokenize each one. The next step is to use the aspect extraction sub-model to extract the aspects from each sentence. Finally, we feed the sentence and the extracted aspects to the sentiment classifier. We explain in the following sections these steps and how we build the models.

3.2.1 Text Preprocessing

Text preprocessing is an essential step for most text analytics problems in general and especially for the sentiment analysis problems. The text in reviews and social posts (like tweets) is usually noisier than the one in articles and blogs. People tend to write their posts, messages, and reviews in a less formal way. Thus, in our system, text cleaning and processing is essential. We developed the text processing steps to utilize most of the information in the text. We perform the following text normalization steps in SentiRank.

- Text lowercasing.
- Removal of HTML tags.
- The URLs are replaced by the <URL> symbol.
- The user mentions are replaced by the <USER> symbol.
- The emails are replaced by the <EMAIL> symbol.
- The dates are replaced by the <DATE> symbol.
- The times are replaced by the <TIME> symbol.
- The numbers are replaced by the <NUMBER> symbol.
- The money is replaced by the <MONEY> symbol.

We use Stanford's tokenizer and Part-of-Speech (POS) tagger to split the text into a set of sentences and perform the word tokenization and tagging steps.

3.2.2 Aspect Category Extraction

Unlike the traditional single-label classification problem (i.e., multi-class or binary), where an instance is associated with only one label from a finite set of labels, in the multi-label classification problem [39], an instance is associated with a subset of labels. If we look to the problem of the aspect category extraction, we can notice that it is a typical multi-label classification problem, where a sentence may belong to none, one or more aspects. Thus, in this work, we regard the problem of *aspect category extraction* as a multi-label classification problem. As a result, we first present the definition of the multi-label classification problem and next explain our model.

Let $X = \{x_1, x_2, \ldots, x_n\}$ be the set of all instances and $A = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ be the set of all aspects. Where n is the number of samples and m is the number of aspects. We can define the set of data:

$$D = \{(x_i, \hat{A}_i) | x_i \in X \text{ and } \hat{A}_i \subseteq A \text{ is the set of aspects associated with } x_i\} \quad (3.1)$$

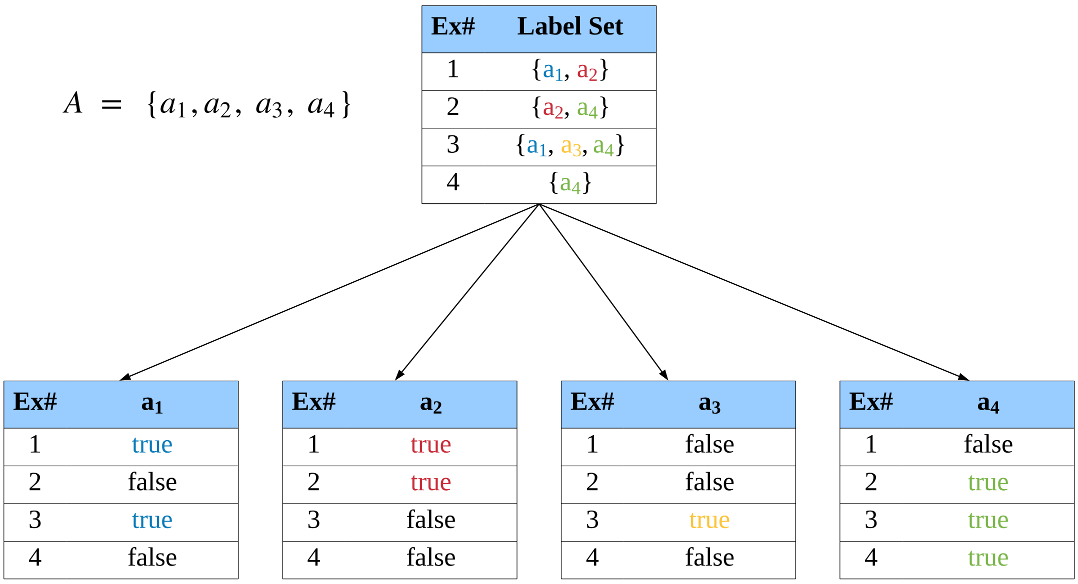


Figure 3.3: Illustration of the decomposition of a multi-label dataset into multiple binary datasets.

In this case, D is called a supervised multi-label dataset. One common strategy to solve this problem is to transform the problem into a traditional classification problem. To this end, we use the most common method, binary relevance [39–41]. The idea of the binary relevance method is simple and intuitive. The aspect extraction problem is decomposed into multiple binary problems, one problem for each aspect. Then, an independent binary classifier is trained to predict the relevance of each one of the aspects. In order to train the binary classifier for each corresponding aspect α_j, $1 \leqslant j \leqslant m$, we first construct the following binary dataset:

$$D_j = \{(x_i, \phi(\hat{A}_i, \alpha_j))|1 \leqslant i \leqslant n\} \tag{3.2}$$

In this expression, $\phi(\hat{A}_i, \alpha_j)$ is a function that returns 1 if $\alpha_j \in \hat{A}_i$ and 0 otherwise.

Figure 3.3 illustrates the decomposition process.

After that, we build a SVM binary classifier $h_j : X \rightarrow \{0, 1\}$ from the training set D_j.

During the inference, and when we want to extract the aspects of a new and unseen sentence x, the system predicts as relevant labels the ones that are predicted to be 1 by the binary classifier h_j for the instance x. Hence, the function $h : X \rightarrow 2^A$ is defined as follows:

$$h(x) = \{\alpha_j | h_j(x) = 1, 1 \leqslant j \leqslant m\} \tag{3.3}$$

3.2.3 Aspect Polarity Detection

The goal of this sub-task is to detect the sentiment expressed towards a given aspect in a given sentence. For each input pair (sentence, aspect), the output is a single sentiment label: positive, negative, neutral, or conflict. We train one multi-class SVM classifier for each aspect category. The feature set is extended to incorporate the information about a given aspect category as explained in the next chapter. We define the aspect polarity detection function $f : X \times A \rightarrow \{positive, negative, neutral, conflict\}$ as the following:

Algorithm 1: Aspect extraction and polarity detection algorithm.

```
input  : A review rev.
output : A list of pairs L = {(α_1, p_1), ..., (α_l, p_l)}, where α_k is an aspect and
         p_k its corresponding polarity.
L ⟵ Empty;
Split rev to a set of normalized sentences and store them in Sents;
for x ∈ Sents do
    Let Â be the set of aspects extracted from x using 3.3;
    for α ∈ Â do
        Let p be the polarity obtained by 3.4;
        Add the pair (α, p) to L;
    end
end
```

$$f(x, \alpha) = \begin{cases} \text{positive} & \text{If, } \arg\max \sigma(x, \alpha) = 1 \\ \text{negative} & \text{If, } \arg\max \sigma(x, \alpha) = 2 \\ \text{neutral} & \text{If, } \arg\max \sigma(x, \alpha) = 3 \\ \text{conflict} & \text{If, } \arg\max \sigma(x, \alpha) = 4 \end{cases} \tag{3.4}$$

Where, $\sigma : X \times A \rightarrow \mathbf{R}^4$ is the classification support function. We summarize the aspect extraction and polarity detection steps in the algorithm 1.

3.2.4 From Polarity to Multi-Criteria

Recall that our main goal is to integrate consumers' reviews about a set of alternatives with the decision-maker's preferences in the decision making process. Hence, we consider each aspect in the social domain as a criterion. We call these set of aspects as the social-criteria. The algorithm 2 shows the transformation steps. The inputs to the algorithm are a set of alternatives O, a set of aspects A (from a specific domain, e.g., restaurants) and a collection of reviews, RV, about each alternative. The output is a matrix with r rows and m columns, where r denotes the number of alternatives and m is the number of aspects. Each value in the cell i, j is the polarity index which represents the degree of consumers' satisfaction about the aspect $\alpha_j \in A$ of the alternative $o_i \in O$. We use the polarity index defined in [9] as the following:

$$P_{index}(P, N) = \begin{cases} 1 - \frac{N}{P} & \text{if, } P > N \\ \frac{P}{N} - 1 & \text{if, } P < N \\ 0 & otherwise \end{cases} \tag{3.5}$$

Here, P is the number of positive opinions and N is the number of negative opinions. The range of P_{index} is $[-1, 1]$, so w transform it to $[0, 100]$ by adding 1 and multiply the result by 50. Where 0 means the lowest satisfaction and 100 refers to the highest degree of satisfaction. The algorithm 2 illustrates the transformation steps.

3.3 Domain Analysis Unit

As shown in Figure 3.1, this unit receives as an input the decision-maker preferences and the description of the alternatives and feeds them to the utility functions to transform the preferences into numerical values. Both the description of the alternatives and the user preferences values are given for a specific criteria which we call the domain criteria in this work.

Decision-makers can express their preferences in different ways depending on the type of each criterion. For example, users can specify their preferences on the price criterion as a range of amounts or as a linguistic value (e.g., very cheap,

Algorithm 2: From polarity to performance table algorithm.

```
input  : A set of alternatives O, a set of aspects A and reviews RV.
output : The social performance table, M.
initialization;
r ⟵ |O|;
m ⟵ |A|;
POS ⟵ Zeros[r, m];
NEG ⟵ Zeros[r, m];
M ⟵ Zeros[r, m];
i ⟵ 1;
while i ⩽ r do
    RV_i ⟵ RV[i];
    for rev ∈ RV_i do
        let L be the list of extracted aspects from rev using Algorithm 1;
        for (α, p) ∈ L do
            let j be the index of aspect α;
            if p == "positive" then
                POS[i, j] = POS[i, j] + 1;
            else
                if p == "negative" then
                    NEG[i, j] = NEG[i, j] + 1;
                end
            end
        end
    end
    i = i + 1;
end
i ⟵ 1;
while i ⩽ r do
    j ⟵ 1;
    while j ⩽ m do
        P ⟵ POS[i, j];
        N ⟵ NEG[i, j];
        M[i, j] = P_index(P, N);
        j = j + 1;
    end
    i = i + 1;
end
```

cheap, average or expensive). Recall that the goal of this unit is to convert the expression of user's preferences on the domain criteria into a numerical value that reflects the satisfaction degree of the user. Hence, we define in this work three different utility functions that measure the performance of a feature of the alternative. These functions can be used in different domains, such as the price criteria in restaurants, hotels or laptops domain. However, it is possible to extend the functions by defining new ones.

Identity

This utility function does not need any processing, the preferred value of this criterion is given as a real value x in its range. We simply express this criterion as follows.

$$g(x) = x \tag{3.6}$$

Categorical

The range of this kind of utility function is always $[0, 1]$ and the goal is maximize. The value 0 means the lowest degree of satisfaction and 1 means the highest. The utility function is defined as the following:

$$g(U; F) = \frac{|U \cap F|}{|U|} \tag{3.7}$$

In this expression, U refers to the items the user prefers and F is the set of available items in this criteria; it is also called the domain of the criterion.

Example : let us have a categorical criterion and the user preference is $U = \{a, b, c, d, e\}$ and what an alternative provides is $F = \{a, d, f, g, h\}$, then the utility value of this alternative given the preferred value is $g(U; F) = \frac{|\{a,d\}|}{|\{a,b,c,d,e\}|} = \frac{2}{5} = 0.4$.

Linguistic

The range of this kind of function is $[1, |L|]$, where L is the set of linguistic terms in the domain of the function and the goal is maximum. The value 1 means the lowest degree of satisfaction and $|L|$ means the highest. The utility function is defined as the following:

$$g(x;p) = |L| - abs\,(pos(x) - pos(p)) \tag{3.8}$$

Where $x, p \in L$ are the alternative value and the preferred value respectively, pos is the position of the linguistic value in L and abs is the absolute value.

Example : let us have the price feature and it is represented as a linguistic variable. Let $L = \{very\ cheap, cheap, average, expensive\}$ be the linguistic values. Here, $pos(cheap)$ is 2. Assuming that the user prefers $p = cheap$ and the alternative value is $x = expensive$, then the utility value of this alternative given the user preference is $g(expensive; cheap) = 4 - abs(4 - 2) = 4 - 2 = 2$.

3.4 Ranking Unit

The inputs to this unit are the performance table that contains both the domain and the social-based matrices and the criteria properties setup. Each criterion, either domain-based or social-based, has the following properties: minimum value, maximum value, the goal (MAX or MIN), the preference threshold, the indifference threshold, the veto threshold and the weight. The decision-maker is responsible of setting the values of these properties.

The outranking procedure used in SentiRank is as explained in chapter 2. Given the performance table and the criteria' properties, we first apply the construction procedure to get the outranking relation. Afterwards, we apply the NFS procedure to perform the exploitation step to get the final ranking of the alternatives.

3.5 Summary

In this chapter, we explained the architecture of the proposed system and its implementation details in more general view. SentiRank is composed by three main units: Customers' review unit, domain analysis unit and ranking unit. The first two units can be customized to any domain such as restaurant, hotels, laptops, etc. Next chapter shows as a case study how we can implement SentiRank for a specific domain, presents the obtained results and outline it.

"Don't judge a man by his opinions, but what his opinions have made of him."

— Georg Christoph Lichtenberg

4

Experiments and Results

Contents

4.1 Introduction

We have evaluated and tested the proposed system in the restaurant domain. To this end, we built an aspect extraction model and aspect polarity detection for this domain. After that, we collected the details and reviews of a set of restaurants from Tarragona as alternatives. Finally, we tested the system with different users' profiles (i.e., user preferences).

This chapter is structured as follows. First, we describe in section 4.2 the data and the criteria used in this study. Section 4.3 explains how we built the sentiment analysis models. We present the features setups of aspect extraction models and aspect polarity detection and the parameters' values of each classifier. We also show and discuss the performance of the developed models with a publicly available benchmark on the restaurant domain presented in SemEval-2014 [42]. In Section 4.4, we show the performance of our system, i.e., SentiRank, against different users' preferences and analyse the obtained results. Finally, in section 4.5, we summarize this chapter.

4.2 Data

Table 4.1 shows the description of the domain criteria and the social-based criteria. We chose the domain criteria from the details of the restaurants on TripAdvisor. The social criteria are the aspects defined in SemEval-2014 Task 4: Aspect Based Sentiment Analysis [42].

Criterion	Utility Function	Domain	Goal
A- Domain			
Price	Linguistic	{Very Cheap, Cheap, Average, Expensive}	MAX
Cuisines	Categorical	Spanish, Seafood, Bar, ...	MAX
Special Diets	Categorical	Vegetarian Friendly, ...	MAX
Meals	Categorical	Lunch, Drinks, ...	MAX
Features	Categorical	Private Dining, Seating, WiFi, ...	MAX
Distance	Identity	$[0, \infty]$	MIN
B-Social			
Food	Identity	[0, 100]	MAX
Price	Identity	[0, 100]	MAX
Service	Identity	[0, 100]	MAX
Ambience	Identity	[0, 100]	MAX
Anecdotes	Identity	[0, 100]	MAX

Table 4.1: The description of the criteria.

We trained our aspect extraction and aspect polarity detection models on the publicly available dataset provided by the organizer of SemEval-2014 Task

Sentence	Aspect/Polarity
But the staff was so horrible to us.	(service, negative)
All the money went into the interior decoration, none of it went to the chefs.	(ambience, positive), (food, negative)
The food was delicious but do not come here on a empty stomach.	(food, conflict)
I grew up eating Dosa and have yet to find a place in NY to satisfy my taste buds.	(anecdotes, neutral)

Table 4.2: Examples of sentences from the training dataset.

Aspect	Positive		Negative		Conflict		Neutral		Total	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
FOOD	867	302	209	69	66	16	90	31	1232	418
PRICE	179	51	115	28	17	3	10	1	321	83
SERVICE	324	101	218	63	35	5	20	3	597	172
AMBIENCE	263	76	98	21	47	13	23	8	431	118
ANECD	546	127	199	41	30	15	357	51	1132	234
Total	2179	657	839	159	163	52	500	94	3713	1025

Table 4.3: Aspect categories distribution per sentiment class.

4: Aspect Based Sentiment Analysis [42]. The training data was composed by 3041 English sentences, whereas the test data size is 800 English sentences. Each sentence is labeled with none, one or more aspects with their corresponding polarities as shown in Table 4.2. We present in Table 4.3 the distribution of the polarity classes per category. We can notice that "positive" is the majority polarity class while the dominant aspect category is "food" in both the training and test datasets. Additionally, we collected the details of 23 restaurants from Tarragona as explained in section 4.4.

4.3 Sentiment Analysis Model: Training Setup

We describe in this section the training setup of the aspect extraction and the aspect polarity detection models. We used LIBLINEAR[1] to build all the linear classifiers. First, we define all the features used in our system and then we show the features configuration of each model. After that, we present the evaluation metrics, the performance of the models and discuss the results.

Features

The set of features used in our system are the following:f

- **Word ngrams**: presence or absence of contiguous sequences of $1, 2, ..., n$ tokens.

[1] https://www.csie.ntu.edu.tw/ cjlin/liblinear/

- **Character ngrams**: presence or absence of contiguous sequences of $3, 4, ..., n$ characters.
- **Part-Of-Speach tags**: the number of occurrences of each part-of-speech tag.
- **Lexicon-Based features**: Sentiment lexicons are lists of words with associations to positive and negative sentiments. For each token w and emotion or polarity score $score(w; l)$ in a given lexicon l, we use the sentiment/emotion score to obtain the following features: (1) the total score $\sum_w score(w; l)$; (2) the maximal score $\max_w score(w; l)$; (3) the number of tokens in the input sentence with score(w; l) $>$ 0; (4) the minimal score $\min_w score(w; l)$ (5) the number of tokens in the input sentence with score(w; l) $<$ 0; (6) the score of the last token in the sentence. The set of lexicons used in this work are the following:
  - Yelp Restaurant Sentiment Lexicon (Yelp-Res): this lexicon was automatically generated from the customer reviews for food-related businesses from the Yelp Dataset [2]. It has 39,274 unigram entries [43].
  - NRC Hashtag Sentiment Lexicon (NRC-Hashtag): this lexicon has entries for 39,413 unigrams and 178,851 bigrams. This lexicon was constructed from a pseudo-labeled corpus of tweets [43].
  - NRC Yelp Word-Aspect Association (WA-Lexicon): 183,935 reviews from the Yelp Dataset were used to generate lexicons of terms associated with the aspect categories of food, price, service, ambiance, and anecdotes. For each term w and each category c an association score was calculated using pointwise mutual information [24]. The difference between this lexicon and the others is that these lexicon is aspect associated whereas the others are general.

[2]http://www.yelp.com/dataset_challenge

- **Word Clusters**: Word clusters can provide an alternative representation of text, significantly reducing the sparsity of the token space. The CMU pos-tagging[3] tool provides 1000 clusters produced with the Brown clustering algorithm on 56 million English language tweets. We use number of occurrences of tokens from each of the 1000 clusters as clusters-based features.

Configurations

Table 4.4 shows the set of features used to build the aspect category detection models. All these configurations were obtained using K-Fold cross-validation, where $k = 5$.

	Word-Ngram	Char-Ngram	POS	Clusters	WA-Lexicon
Food	✓, N=3		✓	✓	✓
Price	✓, N=4		✓	✓	
Service	✓, N=1	✓, N=6		✓	
Ambience	✓, N=3		✓	✓	✓
Anecdotes	✓, N=3		✓	✓	✓

Table 4.4: The set of features used with each aspect category.

Following the approach used in [24], the feature set used to train the aspect-category polarity detection model was extended to incorporate the information about a given aspect category c using a domain adaptation technique as the following: each word w has two copies, *w#general* (for general domain) and *w#c* (for the specific category of the instance). For example, for the input pair ("The bread is top notch as well.", 'food') two copies of the word "top" would be used: "top#general" and "top#food". This technique allows the classifier to learn common sentiment features (e.g., the word "good" is associated with positive sentiment for all aspect categories). At the same time, aspect-specific sentiment features can be learned from the training instances pertaining to a specific aspect category (e.g., the word "delicious" is associated with a positive sentiment for the category 'food').

After sentences were tokenized and tagged as explained in chapter 3, the following features were extracted: General Word-Ngrams with $N = 3$, Category-Based Word-Ngrams with $N = 3$, POS tags, Clusters and lexicon-based features. Again,

[3] http://www.cs.cmu.edu/ ark/TweetNLP/clusters/50mpaths2

we used two types of lexicons (global lexicons and target-specific lexicon). The global lexicons used are Yelp-Res and NRC-Hashtag whereas W–A Lexicon is used as a target-specific lexicon.

Evaluation Metrics

To evaluate the aspect category detection model, we use the precision (P), recall (R) and F1 measures.

$$P = \frac{|S \cap G|}{|S|} \tag{4.1}$$

$$R = \frac{|S \cap G|}{|G|} \tag{4.2}$$

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \tag{4.3}$$

Where S is the set of aspect categories returned by the system for all the test sentences, and G is the set of the gold (correct) aspect categories.

To evaluate the aspect-category polarity detection, we use the accuracy which is the number of correctly predicted aspect-category polarity labels divided by the total number of the actual aspect-category polarity labels.

Results

Table 4.5 shows the performance of the proposed aspect detection model for each category. In terms of $F1$, in general, the model shows a very good performance in all the categories. The model gives the best performance with the food category followed by the price category. The worst performance is registered with the anecdotes category.

We compare the performance of the proposed system with the top three systems in the SemEval-2014 Task 4: Aspect Based Sentiment Analysis [42] as shown in tables 4.6 and 4.7. The best values are shown in bold, where the underlined values indicate that our system gives the second best performance. Our system almost achieves the performance of the best system, NRC-Can-2014 [24] in both tasks.

Aspect	P	R	F1	Accuracy
Food	94.43	89.23	91.76	91.63
Service	93.75	87.21	90.36	96
Price	98.61	85.54	91.61	98.25
Ambience	86.73	83.05	84.85	95.63
Anecdotes	80	75.21	77.53	87.25
Average	90.704	84.048	87.222	93.752

Table 4.5: The performance of the proposed aspect detection model.

System	P	R	F1
NRC-Can-2014 [24]	**91.04**	**86.24**	**88.58**
UNITOR [44]	84.98	85.56	85.26
XRCE [45]	83.23	81.36	82.28
Our System	<u>90.41</u>	84.68	<u>87.45</u>

Table 4.6: The comparison of the proposed aspect detection model with the state-of-the-art.

The next section presents a case study that confirms the robustness of the proposed models.

4.4 Case Study

As a case study we collected the information and the users' feedback of 23 restaurants in Tarragona from TripAdvisor. As shown in Figure 4.1, the details of each alternative (i.e., restaurant) are the domain aspects (table 4.1) used in this study. In addition to the users' reviews, TripAdvisor provides the overall rating and the rating of each aspect, see the left side of Figure 4.1.

Table 4.8 shows the ratings of the collected alternatives. The rest of details are suppressed because of the layout issue. However, readers can find this information

System	Accuracy
NRC-Can-2014 [28]	**82.92**
UNITOR [44]	76.29
XRCE [45]	78.14
Our System	<u>82.44</u>

Table 4.7: The comparison of the proposed aspect-category polarity detection model with the state-of-the-art.

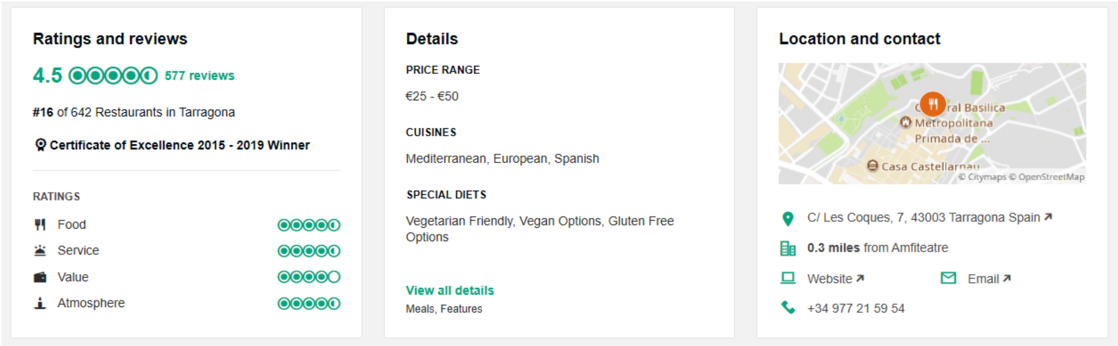


Figure 4.1: Restaurant details in TripAdvisor.

Id	Restaurant Name	Overall	Food	Service	Price	Amb.
1	AQ	4.5	4.5	4.5	4	4.5
2	ELIAN Cafe Restaurant	4.5	4.5	4.5	4.5	4.5
3	La Caleta	4.5	4.5	4.5	4	4.5
4	Tarakon	4.5	4.5	4.5	4.5	4
5	Sadoll Restaurant	4.5	4.5	4.5	4.5	4.5
6	El Taller	4.5	4.5	4.5	4	4
7	El Trull	4.5	4.5	4.5	4.5	4.5
8	Les Coques	4	4	4	4	4
9	Mas Rosello	4	4.5	4.5	4	4.5
10	Les Fonts de Can Sala	4.5	4.5	4.5	4	4.5
11	La Xarxa	4.5	4.5	4.5	4.5	4
12	El Encuentro	4.5	4	4	4.5	3.5
13	La Capital	4	4	4.5	4	3.5
14	Barquet	4.5	4.5	4	4	4
15	Arcs Restaurant	4.5	4.5	4.5	4	4.5
16	Restaurante Ca L Eulalia	4.5	4.5	4	3.5	3.5
17	Restaurante Club Nautico Salou	4.5	4.5	4.5	4	4.5
18	Octopussy	3.5	3.5	3.5	3.5	3.5
19	Palermo 1962 S.c.p.	4	4	4	4	3
20	The Cotton Club Restaurant & Cocktails	4.5	4.5	4.5	4.5	3.5
21	Lizarran Parc Central	3	3.5	3.5	3	3
22	Buffalo Grill	3	3.5	3.5	3	4
23	Indian Restaurant Mirchi Tarragona	3	2	2	3	3

Table 4.8: The alternatives and the rating based on TripAdvisor.

in TripAdvisor.

We analyse the correlation between the output of our sentiment analysis system and the ratings provided by TripAdvisor as follows: first, we calculate the polarity index of each pair of alternative and aspect given a set of reviews as described in

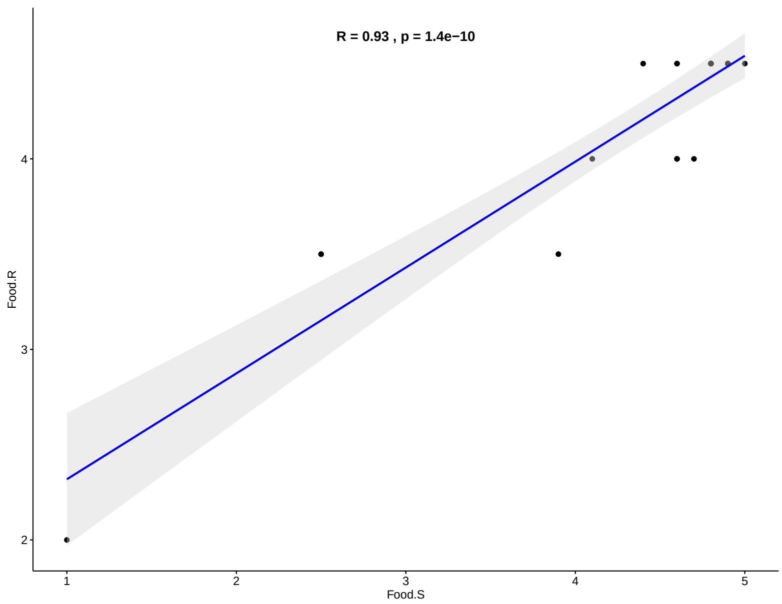


Figure 4.2: The correlation analysis between food-review and food-rating. Food.S refers to the output of our system whereas food.R refers to the rating on the food aspect provided by TripAdvisor.

algorithm 2; then, the overall polarity index of an alternative is obtained by taking the average of all the aspects; finally, we convert the polarity index into a rating value range from 0 to 5 by dividing the polarity index value by 20. Once we get the rating extracted from the reviews we perform the correlation analysis.

Figures 4.2-4.6 show the results of the analysis. In each case, we assumed that the relation between the variables is linear; the slope of the line indicates the strength of the relation. In general, the correlation between the overall rating provided by our system and the overall rating provided by TripAdvisor is strong. This shows the robustness of our sentiment analysis system and indicates that it is applicable to the restaurant domain. If we take a deeper the analysis, we can find that the "food" aspect has the strongest correlation with 0.93 whereas the "price" shows the weakest correlation with 0.55. We attribute that to the fact that people tend to give their opinion towards the quality of the service rather than the price as we found that the "price" aspects has the lowest number of reviews. The second best correlation is shown with the "service" aspect. The "ambience" aspect shows an acceptable correlation with value of 0.73, however it is not quite strong. This can be attributed

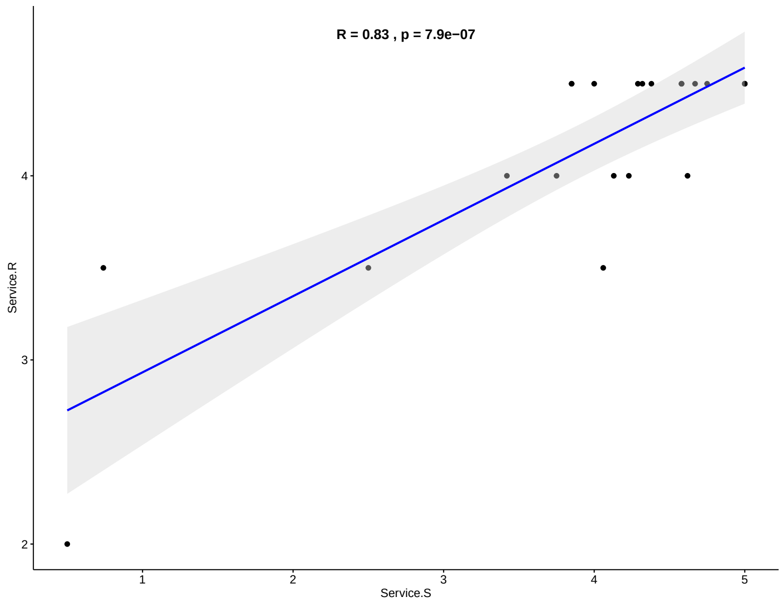


Figure 4.3: The correlation analysis between service-review and service-rating. Service.S refers to the output of our system whereas service.R refers to the rating on the service aspect provided by TripAdvisor.

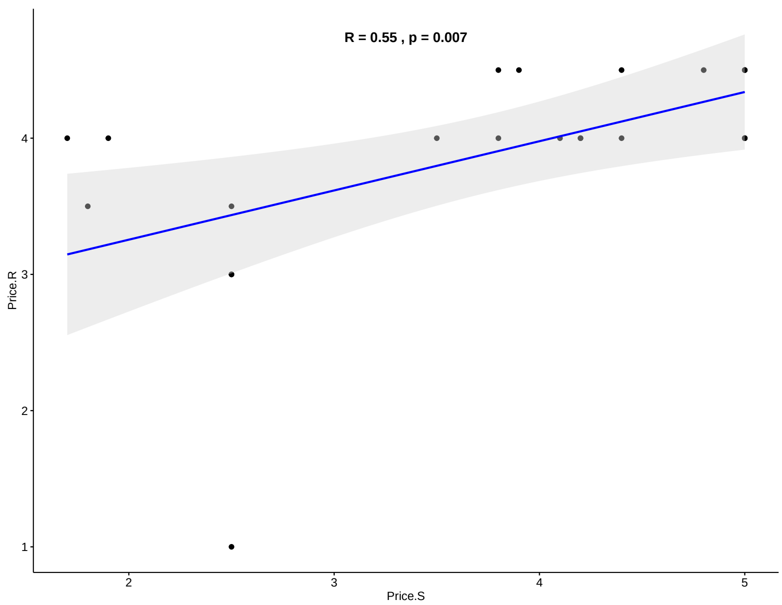


Figure 4.4: The correlation analysis between price-review and price-rating. Price.S refers to the output of our system whereas price.R refers to the rating on the price aspect provided by TripAdvisor.

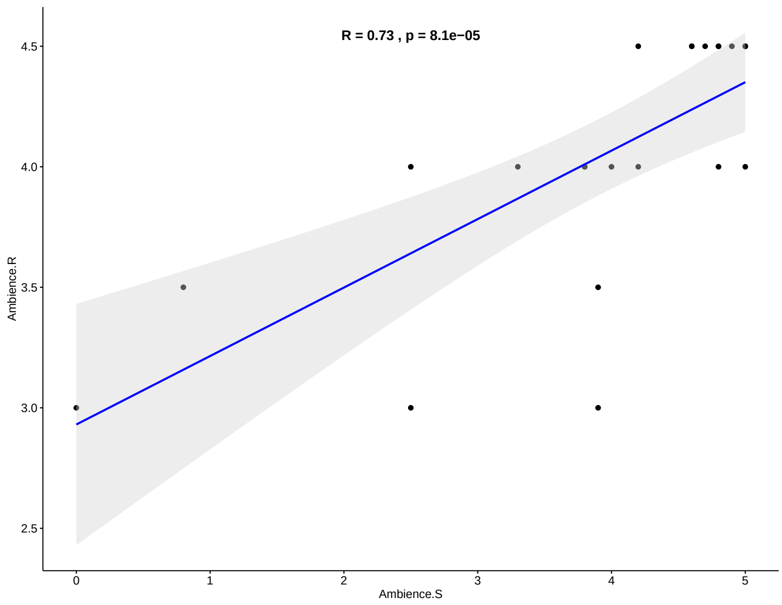


Figure 4.5: The correlation analysis between ambience-review and ambience-rating. Ambience.S refers to the output of our system whereas Ambience.R refers to the rating on the ambience aspect provided by TripAdvisor.

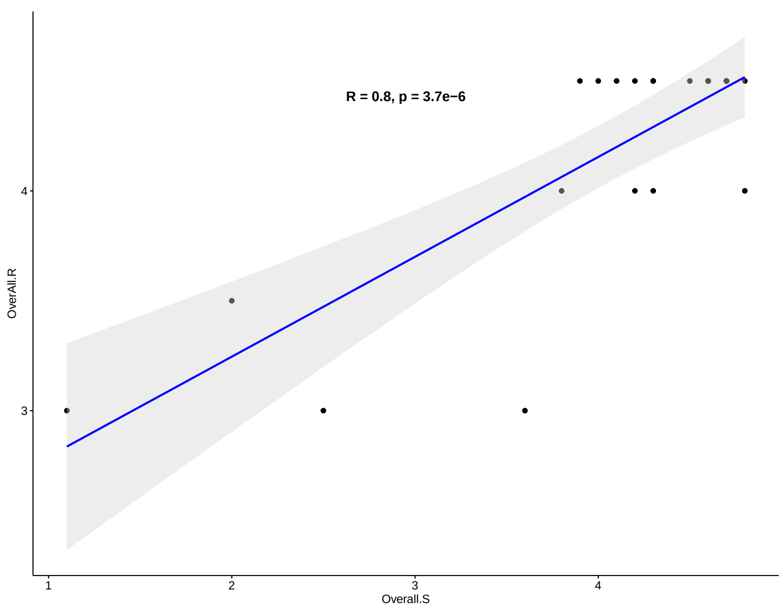


Figure 4.6: The overall correlation analysis of all aspects. Overall.S refers to the output of our system whereas Overall.R refers to the overall rating provided by TripAdvisor.

to the low performance of our system with this aspect as shown in table 4.5.

As the results discussed above show the effectiveness of our sentiment analysis system, now we can use it to combine the users' preferences with the consumers' feedback to rank the alternatives. The next subsection shows the output of SentiRank on two cases with two user profiles. In the first case, we only use the user's preferences (i.e., the domain criteria) as inputs to the ranking procedure. The second case combines the user's preferences with the consumers' feedback.

Profile 1

In this profile, we assume the decision-maker has the following preferences:

- **Price**: Cheap.
- **Cuisines**: Spanish, Seafood, Bar.
- **Special Diets**: Vegetarian Friendly, Gluten Free Options,
- **Meals**: Lunch, Drinks, Branch.
- **Features**: Private Dining, Seating, Wheelchair Accessible, Free Wifi, Accepts Credit Cards, Table Service, Parking Available, Highchairs Available, Full Bar.
- **Distance**: the distance from the town's centre must be as small as possible.

We also assume that the decision-maker is strict with the Special Diets and Meals criteria and flexible or tolerant with the Cuisine and Features criteria. Based on that the decision-maker provides the values of the indifference, preference and veto thresholds as shown in Table 4.9. All the social-based criteria have the same values of the thresholds.

Figures 4.7 and 4.8 show the domain and social performance tables respectively. Figure 4.9 shows the ranking results. The left side graph is the ranking when we only consider the domain criteria. The graph in the right side shows the ranking of the same alternative by adding the social-based criteria to the process. The alternatives

Criterion	Indifference	Preference	Veto
A - Domain			
Price	0	1	2
Cuisine	0	0.4	No Veto
Special Diets	0	0	0.5
Meals	0	0	0.5
Features	0	0.4	No Veto
Distance	0.5	1	3
B - Social			
Food	7.5	20	40
Price	7.5	20	40
Service	7.5	20	40
Ambience	7.5	20	40
Anecdotes	7.5	20	40

Table 4.9: The values of the criteria thresholds.

are ranked based on the NFS value. The first indication from the results is the effect of the social-based criteria. There are clearly differences in the ranking results.

The disagreement between the two cases can be attributed to the fact that most of the alternatives try to attract customers as much as possible by advertising, in the online platform, features that match the customers' preferences. On the contrary, consumers give their feedback and experiences with each alternative. This can be confirmed by checking the utility values of the domain criteria in Figure 4.7. We found most of the alternatives have similar values in most of the criteria especially with the special diets and meals criteria which we know that the user is strict on them. Moreover, we found that the most discriminative criteria in the domain criteria are the distance followed by the features.

Hence, by considering the social-based criteria we can comment the following observations:

- Although the alternatives 'El Taller' and 'Barquet' have similar utility values in terms of the domain criteria (highlighted in Figure 4.7), the former is promoted and it outranks the later. 'El Taller' is in position 2 with the first case (a) and in position 1 with the second one (b). Inspecting the utility

Seq	Restaurant	Price	Cuisines	Special Diets	Meals	Features	Distance
1	AQ	2	0.33	1	0.33	0.67	2.6
2	ELIAN Cafe	3.5	0.33	1	0.33	0.33	0.5
3	La Caleta	2	0.33	1	0.33	0.89	3.5
4	Tarakon	3	0.33	0.5	0.33	0.33	1.9
5	Sadoll	3.5	0.33	1	0.33	0.67	2.1
6	**El Taller**	**3.5**	**0.33**	**1**	**0.33**	**0.78**	**1.4**
7	El Trull	3.5	0.33	1	0.33	0.67	7.6
8	Les Coques	2	0.33	1	0.33	0.56	2.3
9	Mas Rosello	2	0.33	1	0.33	0.67	6.1
10	Les Fonts de Can Sala	2	0.00	1	0.33	1.00	3
11	La Xarxa	3.5	0.67	0.5	0.33	0.78	1.6
12	El Encuentro	3	0.33	0	0.33	0.67	0.9
13	La Capital	3	0.67	0.5	0.00	0.78	1
14	**Barquet**	**3.5**	**0.67**	**1**	**0.33**	**0.56**	**1.1**
15	Arcs	2	0.33	1	0.33	0.67	2.5
16	Ca L Eulalia	2	0.67	0.5	0.33	0.56	2.3
17	Club Nautico Salou	3.5	0.67	0	0.33	0.89	14.3
18	Octopussy	3.5	0.67	0.5	0.67	0.89	2
19	Palermo 1962 S.c.p.	3.5	0.67	0.5	0.33	0.44	11.2
20	The Cotton Club & Cocktails	3.5	0.33	1	0.67	0.78	11.3
21	Lizarran Parc Central	3.5	0.33	0	0.33	0.56	1.3
22	Buffalo Grill	3.5	0.00	0	0.33	0.33	5.7
23	Indian Mirchi	3.5	0.00	0	0.00	0.11	0.75

Figure 4.7: Profile 1: the domain performance-table. The colors indicate the degree of satisfaction and range from "green", the worst, to the "red", the best.

values of these two alternatives in the social performance table in Figure 4.8 reveals that the alternative 'El Taller' shows btter performance than 'Barquet'.

- The alternatives 'ELIAN Cafe' and 'Sadoll' are promoted to position 2.
- The position of the alternatives 'La Caleta' and 'Les Fonts de Can Sala' are changed from 7 and 13 to 3 respectively. If we look at the utility values of the domain criteria (shown in Figure 4.7) of these two alternatives, we can see that they show very similar performance to the best two (i.e., Barquet and El Taller) except with the distance and the price criteria. Looking at

Seq	Restaurant	Food	Service	Price	Ambience	Anecdotes
1	AQ	95.51	91.67	38.89	91.30	93.75
2	ELIAN Cafe	98.75	93.33	100.00	83.33	96.88
3	La Caleta	97.22	76.92	100.00	100.00	96.15
4	Tarakon	96.43	100.00	100.00	75.00	90.38
5	Sadoll	95.28	95.00	95.00	96.15	96.30
6	El Taller	97.83	91.67	83.33	100.00	90.91
7	El Trull	97.22	76.92	100.00	100.00	96.15
8	Les Coques	92.47	68.42	33.33	96.15	91.67
9	Mas Rosello	96.15	100.00	87.50	97.22	100.00
10	Les Fonts de Can Sala	97.22	100.00	100.00	100.00	75.00
11	La Xarxa	100.00	80.00	75.00	75.00	97.37
12	El Encuentro	92.00	75.00	87.50	66.67	83.33
13	La Capital	81.25	87.50	100.00	83.33	73.08
14	Barquet	92.50	82.50	81.25	80.77	82.00
15	Arcs	95.08	91.59	75.00	94.26	95.08
16	Ca L Eulalia	96.88	84.62	35.71	78.57	92.31
17	Club Nautico Salou	91.67	85.71	83.33	96.15	100.00
18	Octopussy	50.00	14.71	50.00	16.67	70.00
19	Palermo 1962 S.c.p.	93.65	92.31	70.00	77.27	90.48
20	The Cotton Club & Cocktails	87.01	86.50	78.89	92.22	87.59
21	Lizarran Parc Central	50.00	50.00	50.00	50.00	50.00
22	Buffalo Grill	78.95	81.25	50.00	50.00	100.00
23	Indian Mirchi	20.00	10.00	50.00	0.00	25.00

Figure 4.8: Profile 1: the social performance-table. The colors indicate the degree of satisfaction and range from "green", the worst, to the "red", the best.

the values of the social-based criteria (i.e., the polarity index) of these two alternatives, we can notice that they have very high scores, which cause the large jump. This can indicate that price and distance are no longer critical criteria in people's decisions. This can also be observed with the alternative 'Trakon'.

- The alternatives in the worst positions have small differences.

Thus, in general, the social-based criteria impact the ranking process.

Profile 2

In this profile, we assume the decision maker has the following preferences:

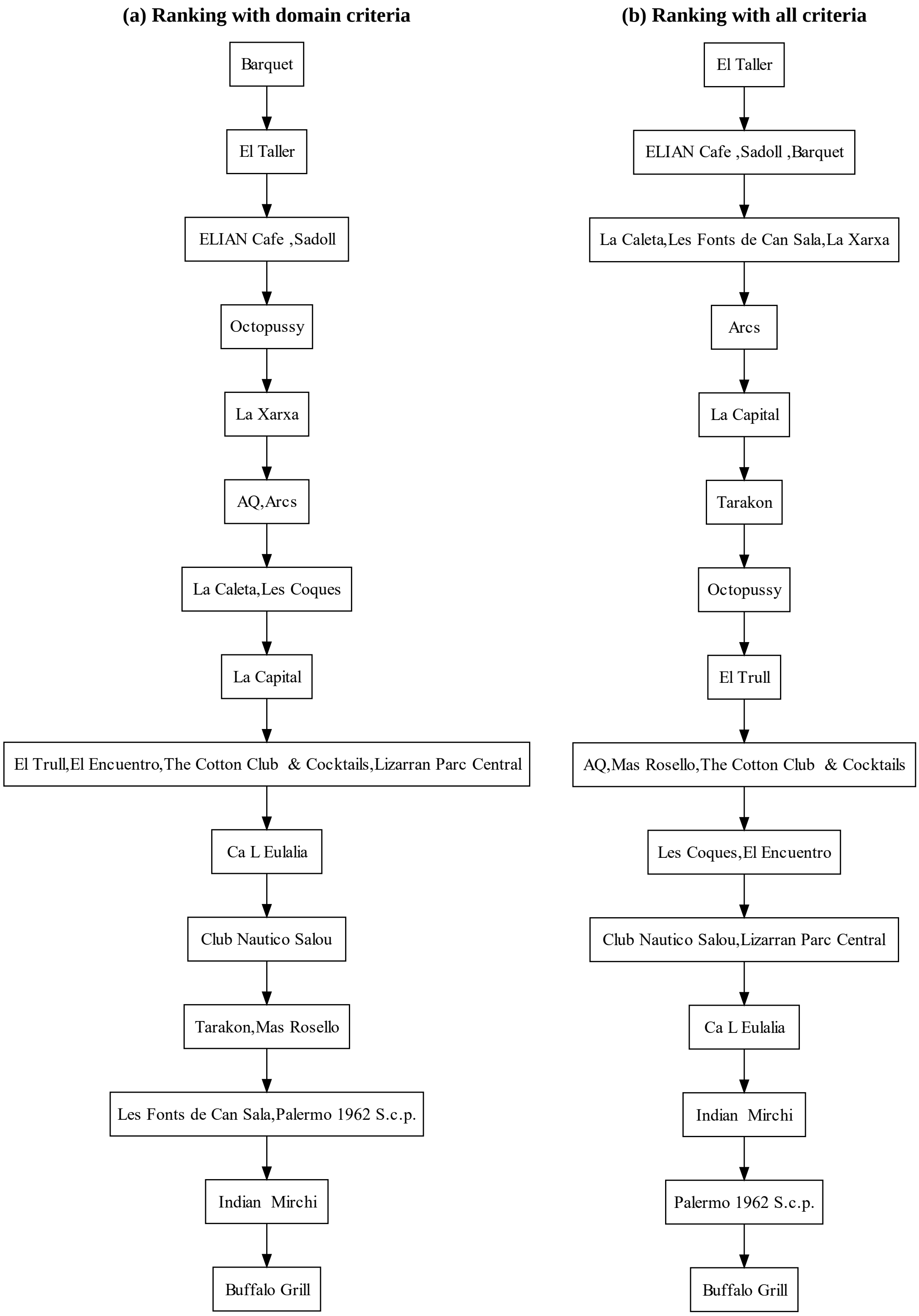


Figure 4.9: Profile 1: ranking result.

Seq	Restaurant	Price	Cuisine	Special Diets	Meals	Features	Distance
1	AQ	4.00	0.00	1.00	0.33	0.63	2.60
2	ELIAN Cafe	2.50	0.00	1.00	0.33	0.50	0.50
3	La Caleta	4.00	0.00	1.00	0.33	1.00	3.50
4	Tarakon	1.00	0.00	0.50	0.33	0.25	1.90
5	Sadoll	2.50	0.00	1.00	0.33	0.50	2.10
6	El Taller	2.50	0.00	1.00	0.33	0.63	1.40
7	El Trull	2.50	0.67	1.00	0.33	0.75	7.60
8	Les Coques	4.00	0.00	1.00	0.67	0.50	2.30
9	Mas Rosello	4.00	0.00	1.00	0.67	0.63	6.10
10	Les Fonts de Can Sala	4.00	0.00	1.00	0.00	1.00	3.00
11	La Xarxa	2.50	0.00	0.50	0.33	0.63	1.60
12	El Encuentro	1.00	0.00	0.00	0.33	0.63	0.90
13	La Capital	1.00	0.33	0.50	0.00	0.75	1.00
14	Barquet	2.50	0.00	1.00	0.67	0.38	1.10
15	Arcs	4.00	0.00	1.00	0.33	0.50	2.50
16	Ca L Eulalia	4.00	0.00	0.50	0.33	0.50	2.30
17	Club Nautico Salou	2.50	0.00	0.00	0.33	0.88	14.30
18	Octopussy	2.50	0.00	0.50	0.67	0.88	2.00
19	Palermo 1962 S.c.p.	2.50	0.00	0.50	0.33	0.38	11.20
20	The Cotton Club & Cocktails	2.50	0.00	1.00	1.00	0.75	11.30
21	Lizarran Parc Central	2.50	0.00	0.00	0.33	0.38	1.30
22	Buffalo Grill	2.50	0.33	0.00	0.33	0.13	5.70
23	Indian Mirchi	2.50	0.00	0.00	0.00	0.13	0.75

Figure 4.10: Profile 2: the domain performance-table. The colors indicate the degree of satisfaction and range from "green", the worst, to the "red", the best.

- **Price**: Expensive.
- **Cuisines**: Barbecue, Bar, Grill.
- **Special Diets**: Vegetarian Friendly, Gluten Free Options.
- **Meals**: Dinner, Drinks, Late Night.
- **Features**: Private Dining, Serves Alcohol, Table Service, Free Wifi, Accepts Credit Cards, Outdoor Seating, Parking Available, Full Bar.
- **Distance**: the distance from the town's centre must be as small as possible.

We used the same thresholds' values used with Profile 1. Similarly, all the criteria are given the same importance (i.e., the same weight). The utility values of the domain criteria are shown in Figure 4.10 and the ranking results of the two

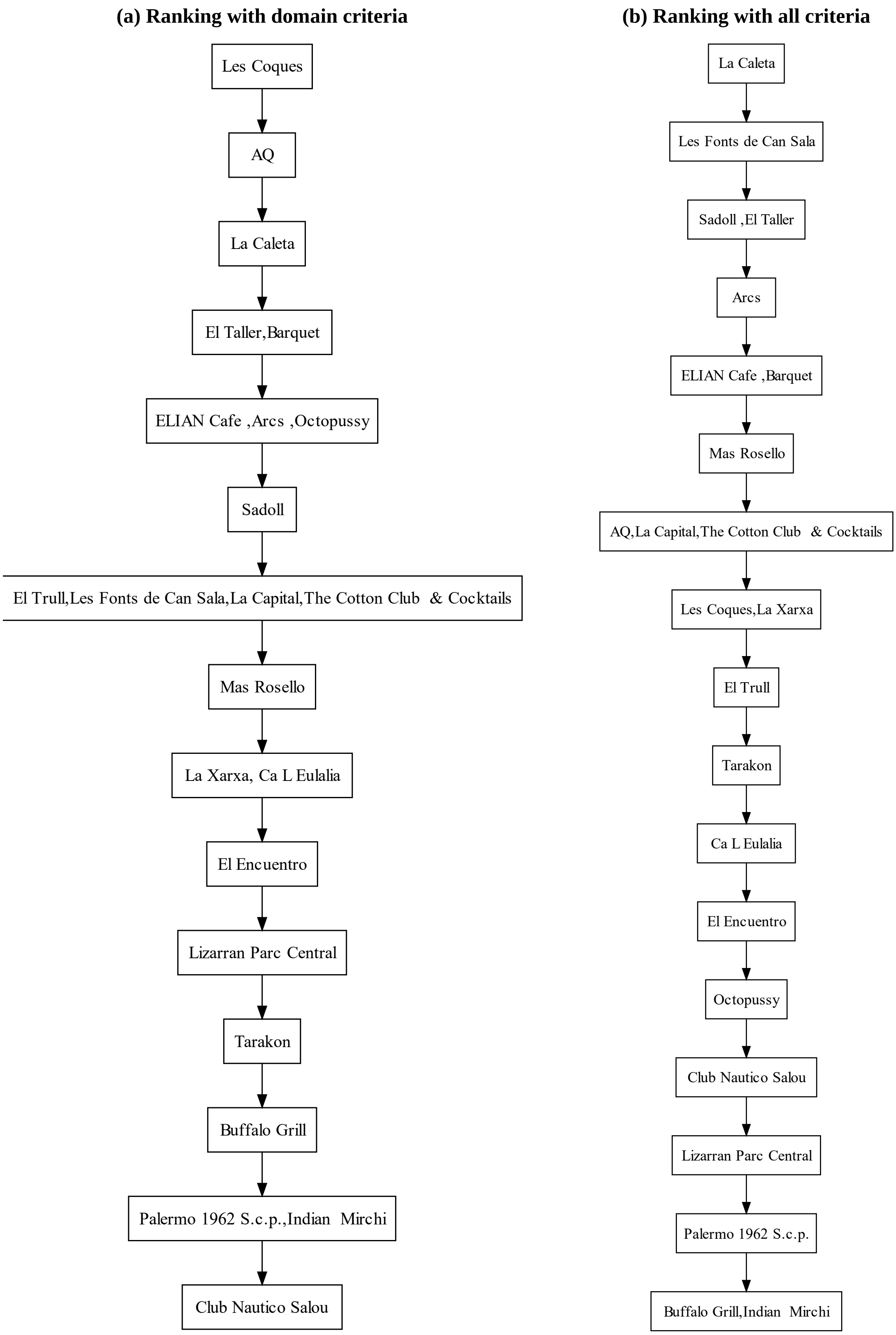


Figure 4.11: Profile 2: ranking result.

cases (without and with the social-based criteria) are shown in Figure 4.11. As we can see, like the results of the profile 1, there are clear differences in the two ranking graphs. For example, the alternative 'La Caleta' is changed from position 3 to the first position. On the other hand, the alternative 'Les Coques' is downgraded from position 1 to position 8 and 'AQ' is downgraded from position 2 to position 7. We analysed the domain performance table, Figure 4.10, of this profile and found the price criterion plays a key role in the ranking process followed by the distance and features. The alternative 'Les Coques' shows good performance in terms of the social criteria (the price criterion is an exception). However, all the alternatives that outperformed the alternative 'Les Coques' by adding the social-criteria show better performance than it in most of the social criteria.

Given all that, we can conclude that adding the social-based criteria adds more constraints on the ranking process. However, we believe that the integration of the users' reviews with the decision-maker preference needs more investigation to improve the results. The next chapter discusses this point in more details.

4.5 Summary

In this chapter, we showed as a case study how we can implement SentiRank for the restaurant domain. First, we explained the training setups of the aspect-based sentiment analysis model. After that, we presented the information collected about 23 restaurants in Tarragona. The collected dataset was used as a case study with two different profiles. We reported and analysed the obtained ranking results in two cases. In the first case, we only used the user's preferences (i.e., the domain criteria) as inputs to the ranking procedure. The second case combined the user's preferences with the consumers' feedback. The next chapter concludes the work and points to some ideas for future work.

"Don't change your mind just because people are offended; change your mind if you're wrong."

— Criss Jami, *The Salmon of Doubt*

5 Conclusion and Future work

Contents

5.1 Conclusion

Decision making is a very hard task as it, often, requires the analysis of hundreds of potential alternatives defined on multiple and conflict criteria. Methods based on decision rules, utility functions or outranking methods have been proposed in the multi-criteria decision aid field. All of these methods use only the utility values of, what we call in this dissertation, the domain criteria based on the decision maker's preferences. With the rapid growth of the web, online platforms such as online shopping websites and social networks give people a great space to share their feedback and experiences about products, services, etc. Thus, in this thesis, we have proposed an outranking system, called SentiRank, that combines the decision-maker preferences with the users' reviews about alternatives that the decision-maker needs to rank. The contribution of this work has two sides: theoretical and practical. From the theoretical part, we have given answers to the main research questions about integrating the users' feedback with the decision-maker preferences in the ranking process. We have defined a methodology to use the aspect categories, which can be extracted from the users' reviews, as additional attributes in the decision process. From the practical side, we have implemented the system using Java and applied it it to a dataset of restaurants in Tarragona. Appendix A shows the graphical user interface of the implemented system.

5.2 Future Work

The work done in this thesis opens an interesting and promising research line that enables the use of people' feedback and reviews for decision support.

For the future work, and from a practical perspective, we plan to extend the implemented system by adding more domains such as hotels, laptops, and cameras. We also plan to make the system a web-based application and implement RESTfull APIs to allow developers and users to use the system remotely. From the theoretical part, we would like to work on different levels of sentiment analysis. For example, in this thesis we have performed the analysis on the aspect category level, hence, in the future, we plan to perform the analysis on the aspect term level [11]. Moreover, in this thesis, we have only considered the reviews written in English language. Thus, we would like to work in a multi-lingual system. Recently, deep learning has been used in sentiment analysis and recommender systems [46], so, we plan to investigate how we can use deep-learning in future work.

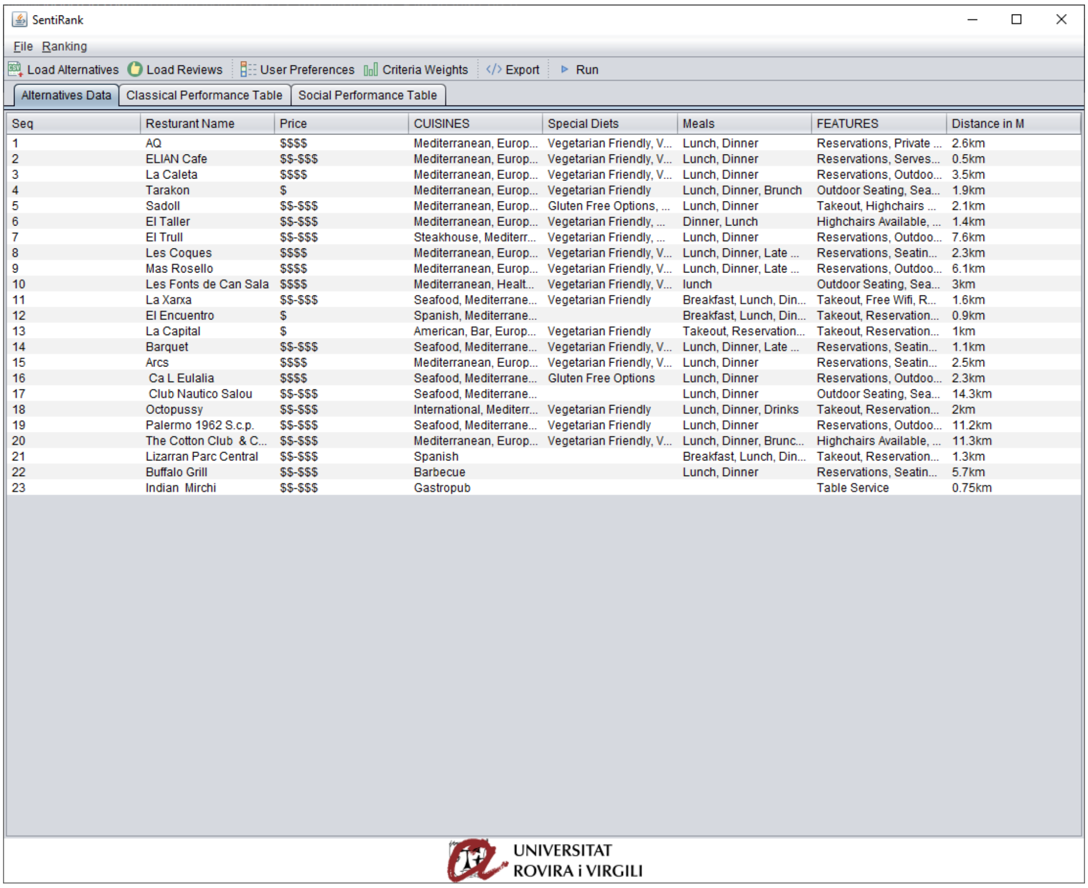
Appendices

A
SentiRank: GUI

In this appendix, we present the GUI of the implemented system with some guidance. The main interface of the system is shown in Figure A.1. First, the user needs to load the set of alternatives by clicking on the button "Load Alternatives". The supported format is Comma Separated Values (CSV) and the file must contain the alternative names and the criteria' values.

Once the alternatives data is loaded, the end user needs to setup his/her preferences as shown in Figure A.2. We omitted the distance criterion from this window as the default preference on this criteria is always the minimum distance. The user can open the preferences window by clicking on the button "User Preferences" in the main window. When the user clicks on the "Ok" button in the user preferences window, the utility function of each criterion will be applied based on the user preference on this criteria and what each alternative provides. The result of this process will be shown in the "Domain Performance Table" tab page in the main window.

Next, the user can load and analyse people' reviews by clicking on the button "Load Reviews". As a result the window depicted in Figure A.3 will be shown up. The user through this window can browse to the directory where the reviews exist. Then, the system starts to read the reviews and analyse them. The result of this step will be shown in the "Social Performance Table" tab page in the main window.



Seq	Resturant Name	Price	CUISINES	Special Diets	Meals	FEATURES	Distance in M
1	AQ	$$$$	Mediterranean, Europ...	Vegetarian Friendly, V...	Lunch, Dinner	Reservations, Private ...	2.6km
2	ELIAN Cafe	$$-$$$	Mediterranean, Europ...	Vegetarian Friendly, V...	Lunch, Dinner	Reservations, Serves...	0.5km
3	La Caleta	$$$$	Mediterranean, Europ...	Vegetarian Friendly, V...	Lunch, Dinner	Reservations, Outdoo...	3.5km
4	Tarakon	$	Mediterranean, Europ...	Vegetarian Friendly	Lunch, Dinner, Brunch	Outdoor Seating, Sea...	1.9km
5	Sadoll	$$-$$$	Mediterranean, Europ...	Gluten Free Options, ...	Lunch, Dinner	Takeout, Highchairs ...	2.1km
6	El Taller	$$-$$$	Mediterranean, Europ...	Vegetarian Friendly, ...	Dinner, Lunch	Highchairs Available, ...	1.4km
7	El Trull	$$-$$$	Steakhouse, Mediterr...	Vegetarian Friendly, ...	Lunch, Dinner	Reservations, Outdoo...	7.6km
8	Les Coques	$$$$	Mediterranean, Europ...	Vegetarian Friendly, V...	Lunch, Dinner, Late ...	Reservations, Seatin...	2.3km
9	Mas Rosello	$$$$	Mediterranean, Europ...	Vegetarian Friendly, V...	Lunch, Dinner, Late ...	Reservations, Outdoo...	6.1km
10	Les Fonts de Can Sala	$$$$	Mediterranean, Healt...	Vegetarian Friendly, V...	lunch	Outdoor Seating, Sea...	3km
11	La Xarxa	$$-$$$	Seafood, Mediterrane...	Vegetarian Friendly	Breakfast, Lunch, Din...	Takeout, Free Wifi, R...	1.6km
12	El Encuentro	$	Spanish, Mediterrane...		Breakfast, Lunch, Din...	Takeout, Reservation...	0.9km
13	La Capital	$	American, Bar, Europ...	Vegetarian Friendly	Takeout, Reservation...	Takeout, Reservation...	1km
14	Barquet	$$-$$$	Seafood, Mediterrane...	Vegetarian Friendly, V...	Lunch, Dinner, Late ...	Reservations, Seatin...	1.1km
15	Arcs	$$$$	Mediterranean, Europ...	Vegetarian Friendly, V...	Lunch, Dinner	Reservations, Seatin...	2.5km
16	Ca L Eulalia	$$$$	Seafood, Mediterrane...	Gluten Free Options	Lunch, Dinner	Reservations, Outdoo...	2.3km
17	Club Nautico Salou	$$-$$$	Seafood, Mediterrane...		Lunch, Dinner	Outdoor Seating, Sea...	14.3km
18	Octopussy	$$-$$$	International, Mediterr...	Vegetarian Friendly	Lunch, Dinner, Drinks	Takeout, Reservation...	2km
19	Palermo 1962 S.c.p.	$$-$$$	Seafood, Mediterrane...	Vegetarian Friendly	Lunch, Dinner	Reservations, Outdoo...	11.2km
20	The Cotton Club & C...	$$-$$$	Mediterranean, Europ...	Vegetarian Friendly, V...	Lunch, Dinner, Brunc...	Highchairs Available, ...	11.3km
21	Lizarran Parc Central	$$-$$$	Spanish		Breakfast, Lunch, Din...	Takeout, Reservation...	1.3km
22	Buffalo Grill	$$-$$$	Barbecue		Lunch, Dinner	Reservations, Seatin...	5.7km
23	Indian Mirchi	$$-$$$	Gastropub			Table Service	0.75km

Figure A.1: SentiRank: main window.

Now, the user must setup the weights and the thresholds of each criterion by pressing on the button "Criteria Weights" to open the window shown in Figure A.4

Finally, the user either exports, by clicking on the export button, the data into XMCDA format to perform the ranking externally using the Diviz software or perform the analysis internally using our system by pressing on the button "Run". The ranking results window shown in Figure A.5 will be invoked. In this window, the user can export the analysis results to CSV files. He/she can also try different thresholds through the slider shown in the top of the window and observe the results in interactive mode.

Figure A.2: User preferences window.

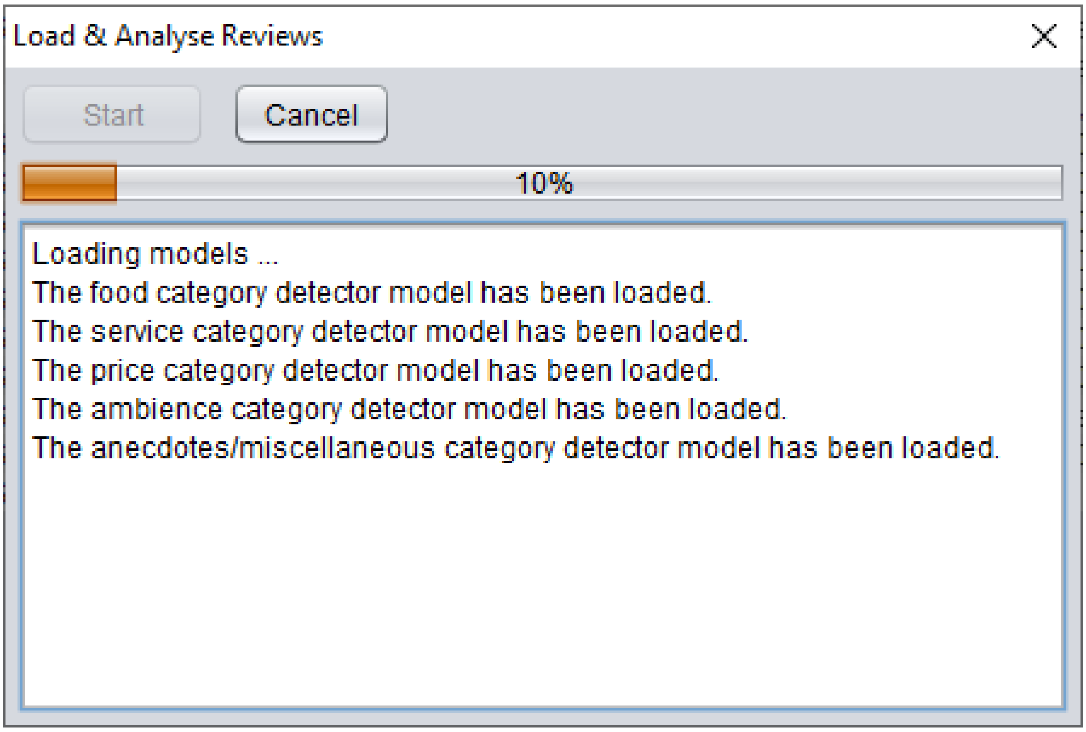


Figure A.3: Loading and analysing users' reviews.

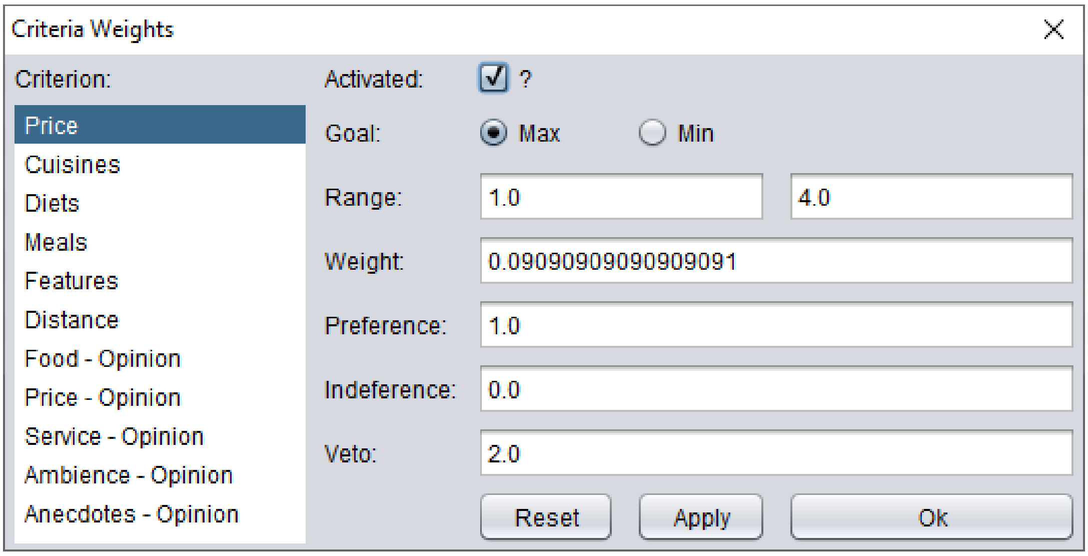


Figure A.4: Setup the weights and the thresholds of criteria.

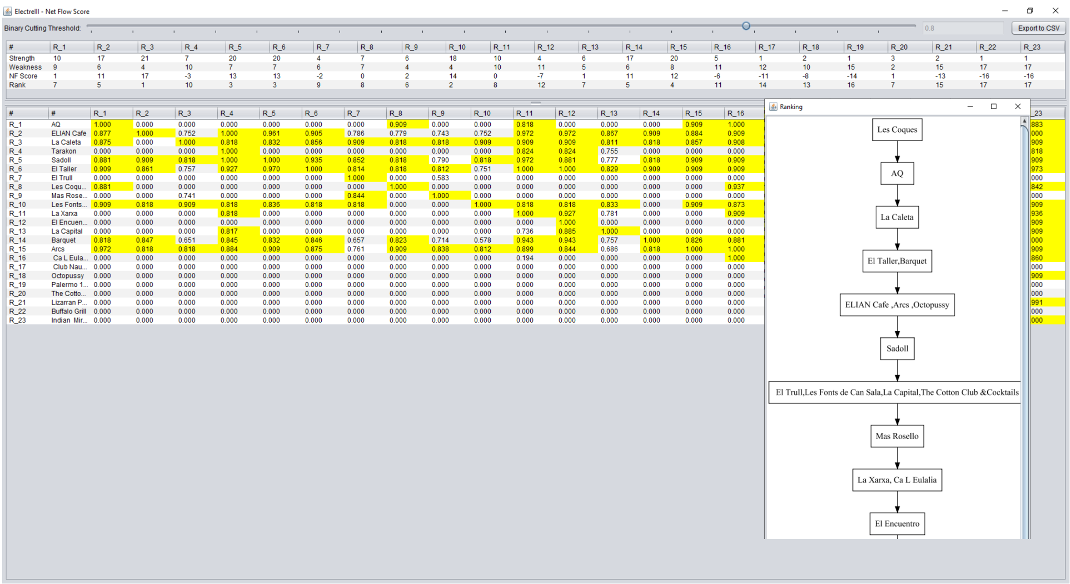


Figure A.5: SentiRank: ranking results.

References

[1] Kristopher Floyd, Ryan Freling, Saad Alhoqail, Hyun Young Cho, and Traci Freling. “How online product reviews affect retail sales: A meta-analysis”. In: *Journal of Retailing* 90.2 (2014), pp. 217–232.

[2] Itamar Simonson. “Imperfect progress: An objective quality assessment of the role of user reviews in consumer decision making, a commentary on de Langhe, Fernbach, and Lichtenstein”. In: *Journal of Consumer Research* 42.6 (2016), pp. 840–845.

[3] Elizabeth C Webb and Itamar Simonson. “Using Reviews to Determine Preferences: How Variance in Customer-Generated Reviews Affects Choice”. In: *ACR North American Advances* (2017).

[4] Li Shengli and Li Fan. “The interaction effects of online reviews and free samples on consumers’ downloads: An empirical analysis”. In: *Information Processing & Management* 56.6 (2019), p. 102071.

[5] K Saleh. *The importance of online customer reviews [Infographic](2017)*. 2017.

[6] Alessio Ishizaka and Philippe Nemery. *Multi-criteria decision analysis: methods and software*. John Wiley & Sons, 2013.

[7] Denis Bouyssou, Thierry Marchant, Marc Pirlot, Alexis Tsoukias, and Philippe Vincke. *Evaluation and decision models with multiple criteria: Stepping stones for the analyst*. Vol. 86. Springer Science & Business Media, 2006.

[8] Bing Liu. “Sentiment analysis and opinion mining”. In: *Synthesis lectures on human language technologies* 5.1 (2012), pp. 1–167.

[9] Mohammed Jabreel and Antonio Moreno. “SentiRich: Sentiment Analysis of Tweets Based on a Rich Set of Features.” In: *CCIA*. 2016, pp. 137–146.

[10] Mohammed Jabreel, Antonio Moreno, and Assumpció Huertas. “Do Local Residents and Visitors Express the Same Sentiments on Destinations Through Social Media?” In: *Information and Communication Technologies in Tourism 2017*. Ed. by Roland Schegg and Brigitte Stangl. Cham: Springer International Publishing, 2017, pp. 655–668.

[11] Mohammed Jabreel, Fadi Hassan, and Antonio Moreno. “Target-dependent sentiment analysis of tweets using bidirectional gated recurrent neural networks”. In: *Advances in Hybridization of Intelligent Methods*. Springer, 2018, pp. 39–55.

[12] Saif Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. “Semeval-2018 task 1: Affect in tweets”. In: *Proceedings of The 12th International Workshop on Semantic Evaluation*. 2018, pp. 1–17.

[13] Simon Tong and Daphne Koller. “Support vector machine active learning with applications to text classification”. In: *Journal of machine learning research* 2.Nov (2001), pp. 45–66.

[14] Wen Zhang, Taketoshi Yoshida, and Xijin Tang. "Text classification based on multi-word with support vector machine". In: *Knowledge-Based Systems* 21.8 (2008), pp. 879–886.

[15] Sang-Bum Kim, Kyoung-Soo Han, Hae-Chang Rim, and Sung Hyon Myaeng. "Some effective techniques for naive bayes text classification". In: *IEEE transactions on knowledge and data engineering* 18.11 (2006), pp. 1457–1466.

[16] Minqing Hu and Bing Liu. "Mining and summarizing customer reviews". In: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2004, pp. 168–177.

[17] Ana-Maria Popescu and Orena Etzioni. "Extracting product features and opinions from reviews". In: *Natural language processing and text mining*. Springer, 2007, pp. 9–28.

[18] Christopher Scaffidi, Kevin Bierhoff, Eric Chang, Mikhael Felker, Herman Ng, and Chun Jin. "Red Opal: product-feature scoring from reviews". In: *Proceedings of the 8th ACM conference on Electronic commerce*. ACM. 2007, pp. 182–191.

[19] Qian Liu, Zhiqiang Gao, Bing Liu, and Yuanlin Zhang. "Automated rule selection for aspect extraction in opinion mining". In: *Twenty-Fourth International Joint Conference on Artificial Intelligence*. 2015.

[20] Wei Jin and Hung Hay Ho. "A novel lexicalized HMM-based learning framework for web opinion mining". In: *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM. 2009, pp. 465–472.

[21] Niklas Jakob and Iryna Gurevych. "Extracting opinion targets in a single-and cross-domain setting with conditional random fields". In: *Proceedings of the 2010 conference on empirical methods in natural language processing*. Association for Computational Linguistics. 2010, pp. 1035–1045.

[22] Fangtao Li, Chao Han, Minlie Huang, Xiaoyan Zhu, Ying-Ju Xia, Shu Zhang, and Hao Yu. "Structure-aware review mining and summarization". In: *Proceedings of the 23rd international conference on computational linguistics*. Association for Computational Linguistics. 2010, pp. 653–661.

[23] Yejin Choi and Claire Cardie. "Hierarchical sequential learning for extracting opinions and their attributes". In: *Proceedings of the ACL 2010 conference short papers*. Association for Computational Linguistics. 2010, pp. 269–274.

[24] Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif Mohammad. "NRC-Canada-2014: Detecting aspects and sentiment in customer reviews". In: *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. 2014, pp. 437–442.

[25] Zhiqiang Toh and Jian Su. "NLANGP: Supervised Machine Learning System for Aspect Category Classification and Opinion Target Extraction". In: *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. Denver, Colorado: Association for Computational Linguistics, June 2015, pp. 496–501. URL: https://www.aclweb.org/anthology/S15-2083.

[26] Zhiqiang Toh and Jian Su. “NLANGP at SemEval-2016 Task 5: Improving Aspect Based Sentiment Analysis using Neural Network Features”. In: *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016).* San Diego, California: Association for Computational Linguistics, June 2016, pp. 282–288. URL: `https://www.aclweb.org/anthology/S16-1045`.

[27] José Saias. “Sentiue: Target and Aspect based Sentiment Analysis in SemEval-2015 Task 12”. In: *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015).* 2015, pp. 767–771.

[28] Fangxi Zhang, Zhihua Zhang, and Man Lan. “Ecnu: A combination method and multiple features for aspect extraction and sentiment polarity classification”. In: *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014).* 2014, pp. 252–258.

[29] José Figueira, Salvatore Greco, and Matthias Ehrgott. *Multiple criteria decision analysis: state of the art surveys.* Vol. 78. Springer Science & Business Media, 2005.

[30] Bernard Roy. “Critères multiples et modélisation des préférences (l’apport des relations de surclassement)”. In: *Revue d’Economie Politique* 84.1 (1974), pp. 1–44.

[31] Berthier Roy and Patrice Bertier. *La méthode ELECTRE II: une méthode de classement en prédence de critères multiples.* 1971.

[32] Bernard Roy. “ELECTRE III: Un algorithme de classements fondé sur une représentation floue des préférences en présence de critères multiples”. In: *Cahiers du CERO* 20.1 (1978), pp. 3–24.

[33] Bernard Roy and J Hugonnard. “Classement des prolongements de lignes de métro en banlieue parisienne (présentation d’une méthode multicritère originale)”. In: *Cahiers du CERO* 24.2-4 (1982), pp. 153–171.

[34] Wei Yu. “ELECTRE TRI(aspects méthodologiques et manuel d’utilisation)”. In: *Document- Université de Paris-Dauphine, LAMSADE* (1992).

[35] Bernard Roy and Jean Michel Skalka. *ELECTRE IS: Aspects méthodologiques et guide d’utilisation.* LAMSADE, Unité associée au CNRS no 825, Université de Paris Dauphine, 1987.

[36] Jose Rui Figueira, Salvatore Greco, Bernard Roy, and Roman Słowiński. “ELECTRE methods: main features and recent developments”. In: *Handbook of multicriteria analysis.* Springer, 2010, pp. 51–89.

[37] Juscelino Almeida-Dias, José Rui Figueira, and Bernard Roy. “Electre Tri-C: A multiple criteria sorting method based on characteristic reference actions”. In: *European Journal of Operational Research* 204.3 (2010), pp. 565–580.

[38] Juscelino Almeida-Dias, José Rui Figueira, and Bernard Roy. “A multiple criteria sorting method where each category is characterized by several reference actions: The Electre Tri-nC method”. In: *European Journal of Operational Research* 217.3 (2012), pp. 567–579.

[39] Min-Ling Zhang, Yu-Kun Li, Xu-Ying Liu, and Xin Geng. “Binary relevance for multi-label learning: an overview”. In: *Frontiers of Computer Science* 12.2 (Apr. 2018), pp. 191–202. URL: `https://doi.org/10.1007/s11704-017-7031-7`.

[40] Grigorios Tsoumakas and Ioannis Katakis. "Multi-label classification: An overview". In: *International Journal of Data Warehousing and Mining (IJDWM)* 3.3 (2007), pp. 1–13.

[41] Jesse Read. "Scalable multi-label classification". PhD thesis. University of Waikato, 2010.

[42] Maria Pontiki, Haris Papageorgiou, Dimitrios Galanis, Ion Androutsopoulos, John Pavlopoulos, and Suresh Manandhar. "SemEval-2014 Task 4: Aspect Based Sentiment Analysis". In: *SemEval 2014* (2014), p. 27.

[43] Svetlana Kiritchenko, Xiaodan Zhu, and Saif M Mohammad. "Sentiment analysis of short informal texts". In: *Journal of Artificial Intelligence Research* 50 (2014), pp. 723–762.

[44] Giuseppe Castellucci, Simone Filice, Danilo Croce, and Roberto Basili. "Unitor: Aspect based sentiment analysis with structured learning". In: *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)*. 2014, pp. 761–767.

[45] Caroline Brun, Diana Nicoleta Popa, and Claude Roux. "Xrce: Hybrid classification for aspect-based sentiment analysis". In: *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. 2014, pp. 838–842.

[46] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, et al. "Wide & deep learning for recommender systems". In: *Proceedings of the 1st workshop on deep learning for recommender systems*. ACM. 2016, pp. 7–10.