A Real-Time Aspect-Based Sentiment Analysis System of YouTube Cooking Recipes

Randa Benkhelifa, Nasria Bouhyaoui and Fatima Zohra Laallam

Abstract Nowadays, uploading, searching and downloading cooking recipes, as well as their rating and reviewing have become a daily habit. Millions of reviews seek for exchanging recipes over YouTube. A user spends a lot of time searching for the best cooking recipe through users' comments. Opinion Mining and Sentiment Analysis are critical tools for information-gathering to find out what people are thinking. In this chapter, we introduce a sentient based real-time system which mines YouTube meta-data (Likes, Dislikes, views, and comments) in order to extract important cooking recipes features and identify opinions polarity according to these extracted features. To improve the performance of our system, we construct a cooking recipe lexicon and propose some algorithms that constructed on sentiment bags, based on particular words related to food emoticons and injections.

Keywords YouTube comments · Opinion mining · Feature extraction
Opinion extraction · Subjectivity · Emoticon · Injection

1 Introduction

People in Web-based social applications such as Facebook, YouTube, and Twitter generate a considerable mass of information. That includes a wealth of opinions which requires analyzing of their aspects, reactions, emotions, etc. towards various entities like services, products, issues, events and their attributes are called opinion mining [18]. This latter, interests on methods of opinion detection and extraction

R. Benkhelifa (✉) · N. Bouhyaoui · F. Z. Laallam
Department of Computer Science and Information Technologies, Université Kasdi Merbah
Ouargla, BP 511 Route de Ghardaia, 30 000 Ouargla, Algeria
e-mail: randa.benkhelifa@univ-ouargla.dz

N. Bouhyaoui
e-mail: nasria.bouhyaoui@univ-ouargla.dz

F. Z. Laallam
e-mail: laallam.fatima_zohra@univ-ouargla.dz

A. E. Hassanien (ed.), *Machine Learning Paradigms: Theory and Application*, Studies in Computational Intelligence 801,
https://doi.org/10.1007/978-3-030-02357-7_11

of sentiments existed in a text [23]. In recent years opinion mining problems have attracted a great deal of attention [26], where several researchers have proposed solutions in order to help individuals or organizations in making decisions. The essential task of opinion mining is analyzing and classifying ‘polarity’ of given text, whether the expressed opinion is positive, or negative. Implementing opinion mining methods facilitate decision-making task in multiple fields such as efficient recommendation systems, financial study, market research and product growth.

Nowadays a large number of cooking recipes are available on the Web; YouTube is considered one of the most popular websites that contain a wealth of opinions about cooking recipes. People can share their experiences about various recipes through reviews, comments, like dislike or ratings. Cooking recipe reputation is based on what people think about it. As the numbers of reviews are very high, the user ends up spending a lot of time for searching the best cooking recipe based on the experiences shared by review writers. Therefore, automatic analysis of online user reviews has become a needed research area due to the rapid growth of user-generated reviews [3, 8, 13, 19, 32, 36].

Food recipe reviews published by users can help others make their decisions and select the best recipe for recipes for one cooking. In opinion mining, different levels of granularity analysis have been proposed which are Document level, Sentence level, and Aspect level. Document-level concerns the classification of the entire document as a positive or negative sentiment for a recipe. The Sentence level focuses on the categorization of each sentence whether the expressed opinion is positive, negative or neutral opinion for a recipe. The last one is Aspect level (Aspect-based opinion mining) [14, 18, 20], an aspect is an attribute or feature of a recipe mentioned in reviews.

This level focuses on the extraction of aspects associated with its opinion polarity from the reviews. For example, the following sentence “the recipe is so delicious but the preparation takes a lot of time” contains two aspects or opinion targets, namely Taste and Duration. In this case, applying a sentence level polarity detection technique would mistakenly result in a polarity value close to neutral, since the two opinions expressed by the users are opposite. Hence, aspect extraction is necessary first to deconstruct sentences into recipe features and then assign a separate polarity value to each of these features.

Most of the previous researchers focused on the extraction of the explicit aspects like in the researches [7] and [13], where others work on the extraction of implicit features [27, 43] and [44] in these works the authors propose A Rule-Based Approach to Aspect Extraction from Product Reviews.

In this chapter, we have performed opinion mining and sentiment analysis on YouTube cooking recipes by developing a system named “We Know What is the Best Recipe for You”. The objective of this system is to rank various cooking recipes in order to select the best one through the reviews and the meta-data (Likes, Dislikes, and the views) associated with each one. This saves the time of the users searching for the best recipe for a particular food recipe. For the aim of achieving that we propose an approach to extract online review cooking recipes aspects (features), then obtain the polarity of each opinion through the classification of the extracted aspects in the following features classes: Duration, Decoration, Difficulty, Healthy, Cost, and Taste. To improve the performance of our system we proposed some algorithms that constructed on sentiment bags, based on special words related to food emoticons and injections.

To the best of our knowledge, the proposed system is the first one that can find out the best YouTube cooking recipe for a given cooking. Also, the first system which exploits the meta-data (comments, likes, dislikes, and views) associated with the YouTube video in order to enhance its performance.

The structure of this chapter is as follows: Sect. 2 reviews relevant related work in the fields of opinion mining and sentiment analysis, Sect. 3 overviews the proposed system and approaches. Section 4 describes the data set, which forms the basis of our analyses and experiments. Section 5 presents results and discussion of the findings. Finally, Sect. 6 presents conclusion and potential future research directions.

2 Related Works

Relevant related research is collated in three main sub-sections: First, we review work evaluating the subjectivity detection and sentiment analysis. We continue to review research on aspect-based sentiment analysis then we highlight some studies on cooking and food recipes. All these points contribute to the formulation of our system which the details are shown in the next section.

The text mining plays an important role in many real-world applications such as machine translation, information extraction, sentiment detection, summarization, etc. Opinion mining and sentiment analysis are important tasks of text mining. Recently, several studies on opinion mining and sentiment analysis are crowned about comments in social media. Some research focuses on Subjectivity detection which can be defined as a process of selecting sentences containing opinion [10] and [37]. The purpose of subjectivity/objectivity classification in opinion mining research is to distinguish between factual and subjective (expressing an opinion or emotion) texts present in online comments [6, 9, 13, 25], and [39]. The authors in [15] have aimed to propose methods for identifying subjective sentences from customer reviews for mining product features and user opinions. Others works are interested in sentiment analysis which is a process of finding users opinion about the particular topic [38]. It performed on different domain data such as Movie [26], Books and Products [8, 13, 36], Restaurants [42], and cooking recipes [19, 32], etc.

In [19] the authors have explored various strategies for predicting recipe ratings based on user reviews. In [32] the authors present a sentiment based rating approach for food recipes which sorts food recipes present on various websites from sentiments of review writers. The authors in [33] investigate the influence of different types of biases on ratings, views and comment sentiment on online recipes. A big-data analytics-based approach to identify the supply chain management issues in food industries is proposed in [35]. In [17] the authors have used data from the cooking platform Epicurious and have attempted to predict ratings for recipes based on user reviews. The work in [4] has developed a real-time system to extract and classify the YouTube cooking recipes reviews automatically.

The emotions could be easily associated with an interesting application of human-computer interaction, where a system identifies that the user is upset or annoyed, the system could change the user interface to a different mode of interaction as in [21]. In the works [3, 4, 12] and [28] the authors have used a lexicon of the most used emoticons and injections.

Aspect-based sentiment analysis is another important task of opinion mining. Aspect or feature extraction is the critical subtask of aspect-based sentiment analysis. This subtask aims to extract fine-grained aspects from online users' reviews [30]. The early works focused on the extraction of subjective and objective aspects. The authors in [16] have proposed RubE-unsupervised rule-based methods that extract both subjective and objective features from online consumer reviews. In [1] the authors have proposed a state-of-the-art research for aspect-based sentiment analysis of Arabic Hotels' reviews using two implementations of long short-term memory (LSTM) neural networks. The work in [11] has described W2VLDA, an almost unsupervised system based on topic modelling that performs aspect category classification, for any given domain and language. In [2] the research has presented an enhanced approach for Aspect-Based Sentiment Analysis (ABSA) of Hotels' Arabic reviews using supervised machine learning. In [24] and [31] works the authors improve aspect extraction using Aspect Frequency and Semantic Similarity-Based Approach. The work in [30] proposes a sequential pattern-based approach to detect objective aspects. The work in [29] has proposed a rules-based model that uses rules defined on the basis of sequential patterns mined from customer reviews. Few studies care about aspect-based sentiment analysis in cooking and food recipes. The authors in [24] have attempted to characterize the cuisine type and words for arrangement from the aspect of ingredients and cooking actions. The work in [5] has presented the PREFer food recommender system, apt to provide users with personalized and healthy menus, taking into account both user's short/long-term preferences and medical prescriptions.

In our case, we covered all these points to obtain a real-time aspect-based sentiment analysis system of YouTube cooking recipes, which is used to make a decision.

3 The Proposed System Overview

This section describes the overall system (show Fig. 1).

The proposed system works based on the following steps.

- In the proposed system, the users enter the cooking recipe name and then click on the button search.
- Data collection: using the YouTube APIs from (Google developers), this system retrieves automatically the URLs of YouTube videos of the cooking recipe entered in the field (all the recipes with names which have a strong similarity with the text

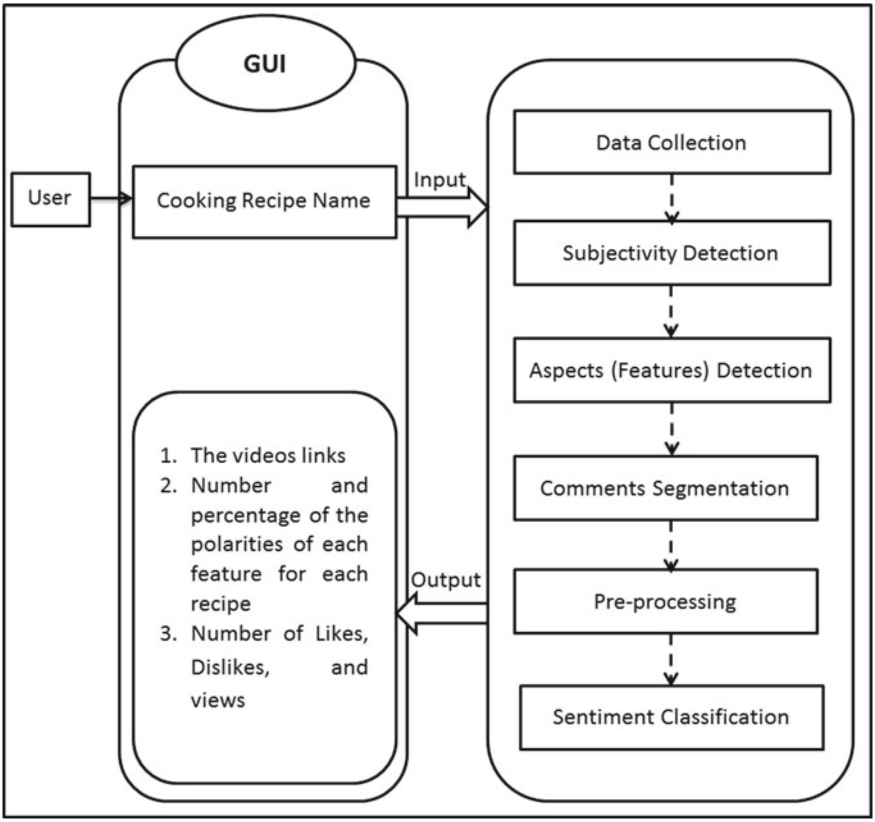

Fig. 1 The proposed system using process

in the request). Then, the system collects the generated comments on these recipes videos and stores them in the database.

- Subjectivity detection: in this step, the system filter automatically opinions "reviews" from the generated comments, and eliminating texts that bear no opinion by classifying the generated comments in the classes (opinion, or other).
- Aspect detection: in this step, the system detects the aspect (attribute or feature) of each recipe mentioned in the reviews.
- Comment segmentation: split the comments (opinions) into fragments according to the features detected in the previous step.
- Pre-processing: this step shows the considered text features and the developed algorithms which are applied to the dataset.
- Sentiment classification: in this stage, the system extracts the polarity of each fragment of opinion (according to the features) and also the recipe itself.
- Getting the results: the system count and view the number of positive comments, and negative according to each feature for each recipe. It then displays the percent-

Table 1 Examples of the top used emoticons [4]

Positive emoticon	Negative emoticon
B-)	X-(
_	:-#
:*), :*	</3
:p, =p	O.o
8-)	:-(, :(
:), =), :-)	:_(
<3	:'(
XP, X-P	:/
XD, :D, =D, :-D	

age and comments for each recipe video, and then it shows the overall percentage, and the number of the Likes, Dislikes, and Views of each recipe video.

The details of these steps are shown in the next sub-sections.

4 Methodology

This section presents the methodology followed in this work.

4.1 Pre-processing

This section includes all analysis and developed algorithms used in the preprocessing step.

4.1.1 Bags Development

This step focuses on the informal language of the comments about cooking recipes video on YouTube.

Following our previous study in [4], we reuse the two types of bags developed in [4]: bag for emoticons and bag for interjections. After a deep analysis, we have concluded the results, which are more than 60 emoticons and 130 injections. Some examples are shown in (Tables 1 and 2).

Emoticons' bag: We create a bag of the used emoticons in YouTube comments, whether they express a positive or a negative sentiment.

Injections' bag: we create a bag of the injections in YouTube comments, whether they express a positive or a negative sentiment.

Table 2 Examples of the top used injections [4]

Positive injection	Negative injection
Wow, waw	Oh dear
Miam, miamy, miamiam	No way
Haha, hehe, hihi	Argh
Thank you	Boo, booh
Oy	Brr, brrr
Ahh, ahhh	Oops
Mmm, emm, hum, huumm	Bekhg, buk

Table 3 The developed algorithms for the subjectivity classification [4]

Algo1. Replace all injection with the same string "INJ"	Algo2. Replace all emoticon with the same string "IMO"
W ← Corpus	*W ← Corpus*
I ← set of Injections	*E ← set of Emoticons*
Foreach w W	*Foreach w W*
Foreach i I	*Foreach e E*
If w = i	*If w = e*
w ← "INJ"	*w ← "EMO"*
EndIf	*EndIf*
EndForeach	*EndForeach*
EndForeach	*EndForeach*

4.1.2 Algorithms Development

In this section, we show all algorithms developed by us in previous study [4] and used them to improve our classifications. The social media users utilize emoticons and injections in their social text. Knowing exactly the injection/emoticon used in the text is not important. The important is the sentiment reflected by those users. The classifier takes each emoticon or injection as a different word. But if we replace all (positive emoticon by PosEMO, positive injection by PosINJ, negative emoticon by NegEMO and negative injection by NegINJ) in sentiment classification, and all (emoticons by EMO and injections by INJ) in subjectivity classification. The classifier takes them as the same word (Tables 3 and 4).

4.1.3 Textual Features

The aim of this step is to clean the dataset by:

- Data pre-processing 1
  - Keeping stopwords;
  - Removing numbers and punctuations;

Table 4 The developed algorithms for the sentiment classification [4]

Algo3. Replace all positive injection by "PosINJ" and negative injection by "NegINJ"	Algo4. Replace all positive emoticon by "PosEMO" and negative emoticon by "NegEMO"
W ← Corpus	*W ← Corpus*
PI ← set of positive Injections	*PE ← set of Positive Emoticons*
NI ← set of negative	*NE ← set of Negative*
Injections	*Emoticons*
Foreach w ∈ W	*Foreach w ∈ W*
Foreach pi ∈ PI	*Foreach pe ∈PE*
If w = pi	*If w = pe*
w ← "PosINJ"	*W ← "PosEMO"*
EndIf	*EndIf*
EndForeach	*EndForeach*
Foreach ni ∈ NI	*Foreach ne ∈ NE*
If w = ni	*If w = ne*
w ← "NegINJ"	*W ← "NegEMO"*
EndIf	*EndIf*
EndForeach	*EndForeach*
EndForeach	*EndForeach*

  – Removing all word appears less than 5 times;
  – Stemming: removing prefix and suffix finding the stem or the root of the word. The lovinsStemmer [22] is a well-known like stemming algorithm.

• Data pre-processing 2

  – Removing stopwords like 'of', 'and', 'my' that don't have an influence on sentiment classification.
  – Removing numbers and punctuations.
  – Removing all word appears less than 3 times.
  – Stemming: removing prefix and suffix finding the stem or the root of the word. The lovinsStemmer [22] is a well-known like stemming algorithm.

• Feature selection

  – Using Term frequency inverse document frequency TF-IDF [34].
  – Part of Speech (POS) tagging: consists of tagging a word in a text to a particular part of speech based on its context and its definition. In English, it has nine parts of speech: noun, verb, article, adjective, preposition, pronoun, adverb, conjunction and interjection [28]. We use POS as features just for show which parts appear in which category more than the other.

After the pre-processing and representing phase, various supervised algorithms can be applied. We have chosen two among the popular ones: Naïve Bayes (NB), NB is a family of simple probabilistic classifiers based on applying Bayes' theorem, and Support Vector Machine (SVM). SVM has achieved great success in text classification, especially in the binary classification. SVMs are supervised learning models which analyze data and recognize patterns which are used for classification and regression analysis [3].

4.2 Subjectivity Detection

For Opinion filtering classification, our approaches are as follow:

- **NB** (Baseline): using the classifier Naïve bayes + the textual features as proposed in Pre-processing1.
- **SVM** (Baseline): using the classifier SVM + the textual features as proposed in Pre-processing1.
- **Naïve Bayes + Algo1 + Algo2**: using the classifier Naïve bayes + the textual features as proposed in Pre-processing1 + Algo1 + Algo2.
- **SVM + algo1 + Algo2**: using the classifier SVM + the textual features as proposed in Pre-processing1 + Algo1 + Algo2.

4.3 Feature Extraction from YouTube Cooking Recipes Reviews

This section deals with the aspect-based sentiment analysis, which concerns the extraction of cooking recipes aspects (Features). Thus we are not limited to extract the explicit aspects but we are also interested in detecting and extracting the implicit aspects. As we have mentioned above we are interested in the list of six feature categories which is limited by analyzing the collected corpora. The aspects are (Duration, Difficulty, Decoration, Cost, Healthy and Taste).

This system aims to extract features for the given cooking recipe from parsed review data. The resource is *the users' comments*. We have found that users might explicitly mention features related information in their reviews in addition to their opinions. For example, in a sentence "nice recipe but there it is so difficult" from a recipe review, the phrase "it is so difficult" reflects a negative opinion about the potential target feature difficulty. Moreover, one review can include one or more aspects of about the cooking recipe, where each part of the comment (expressing an opinion about a special aspect) carries a polarity (positive opinion or negative). The explicit features are those features that occur explicitly in the review sentence, for example the sentence: "the taste is very delicious" indicate directly the feature taste. The implicit features are those referred to in the comments but are not occurred explicitly, in this case the feature is called the implicit aspect. We take for example the sentence “yummy, so delicious” in this review the feature ‘taste’ is not occurred, the comment indicate that the cooking taste is delicious, and this comment refers to the ‘taste’ feature by using sentiment words that reflect the person’s opinion.

As mentioned above, one of our objectives in this study is to extract cooking recipe aspects from the associated comments, for this purpose we adopt a rule-based approach proposed in [27]. This approach is modified to adequate cooking recipe peculiarities. The proposed approach use a set of conditional rules that relying on rules for the sentences having subject verb and for sentences that do not have subject verb [27].

We collected possible implicit features based on cooking recipe features described by a lexicon of sentiment words from the reviews for each aspect category. We extracted the sentences that include implicit aspects after that, we extracted the Implicit Aspect Clues (IACs) for each sentence, following the category that they belong to. We take for example the following sentence “The ingredients are too expensive” the IAC here is expensive which refers to the category cost.

The users use similar sentiment words to describe the same cooking recipe features. We add synonyms and antonyms for the IAC to enrich each aspect category. For example, delicious and its synonym tasty both have the same category taste. Also expensive and its anonym inexpensive belong to the same category price. The rule-based approach require pre-processing tasks, this latter consists of two successive steps: first, the sentence is analyzed by Stanford-dependency-parser to obtain the sentence dependency tree. Then, a Stanford lemmatizer is used to lemmatize each word in each sentence.

After that, we apply aspect parser on the sentences, in the example illustrated below three rules are applied, where the extracted aspects are recipe, delicious and use.

The parser extract all the aspect occurred in the sentences, however, we focus on six features. Therefore, we select only the features that we interest on.

Some of these rules are shown in the Table 5 with concrete examples.

4.4 Sentiment Classification

For the sentiment classification, our approaches are as follow:

- ***NB (baseline)***. using the classifier Naïve bayes + the textual features as proposed in Pre-processing1.
- ***SVM (baseline)***. using the classifier SVM + Use textual features as proposed in Preprocessing1.
- ***NB + Algo3 + Algo3***. using the classifier Naïve bayes + preprocessing1 + Algo3 + Algo4.
- ***SVM + Algo3 + Algo3***. using the classifier SVM + preprocessing1 + Algo3 + Algo4.

Table 5 Some of these rules with concrete examples

Rules	Example
If the word *w* has any adverbial or adjective modifier and the modifier exists in one of our lexicons, then *w* is extracted as an aspect	Nice <u>decoration</u>
If the sentence does not have auxiliary verb, and if the verb w is modified by an adjective or an adverb or it is in adverbial clause modifier relation with another token, then both h and *w* are extracted as aspects	These <u>ingredients costs</u> less
If the sentence does not have auxiliary verb, and if *w* has any direct object relation with a token n and the POS of the token is Noun and n is not in our lexicons then n is extracted as an aspect	I like the <u>look</u> of this cake.
A copula is the relation between the complement of a copular verb and the copular verb. If the token w is in copula relation with a copular verb and the copular verb exists in the implicit aspect lexicon, then w is extract as aspect term	The main dish is <u>expensive</u>.
If an adjective or adverb w is in infinitival or open clausal complement relation with a token w and h exists in the implicit aspect lexicon, then h is extracted as an aspect	Very <u>hard</u> to cook

- ***SVM + MD***. using the classifier Naïve bayes + preprocessing1 + meta-data of the cooking recipe video in the comments about recipe.
- ***SVM + Algo3 + Algo4 + MD***. using the classifier SVM + preprocessing1 + Algo3 + Algo4 + meta-data of the cooking recipe video in the comments about recipe.

5 Experiments

5.1 Data Collection and Analysis

The objective of this chapter deals with opinion mining and sentiment analysis including subjectivity detection, Aspect extraction and sentiment classification. In addition, this work aims to study the problem of the short text with special characteristics typically found on social media and to show how much it is important to consider these characteristics in the pre-processing phase.

YouTube comments are perfect for these due to their abundance and short length. Moreover, YouTube is a popular video social network with a great diversity of users,

Table 6 YouTube video metadata

YouTube video meta-data	Description
Video description	Text associated with the video, this text is posted by the video owner
Views	Number of times that video is seen
Likes and dislikes	Number of persons whom like\don't like the video
Comments	The comments posted by viewers

which means that collecting a sufficiently large dataset with those characteristics on various topics is feasible.

According to YouTube structure, we can easily get the information about any YouTube Video food recipes 'meta-data' (recipe description, Views, Likes & dislikes and comments). All Likes, Dislikes and comments are performed manually by real people, using unique high-quality YouTube accounts. Based on the comments analysis, we select seven important features related to cooking recipes which are (duration, cost, difficulty, Decoration, healthy and taste.), we take in consideration the comments containing the aforementioned features and ignore the other comments which don't concern the recipes itself such as "you are my best cook" this comment express an opinion about the cook and not about the recipe. The YouTube Video Meta-data is summarized in Table 6.

These texts were annotated manually by three human annotators as following:

- For creating the training model of the opinion filtering (5000 subjectives and 5000 objectives) [4]
- For creating the training models of the aspects sentiment classification we collected 10,000 comments, where one comment can bear one or more aspects. The details are shown in the Table 7.

Table 7 Number of positive and negative aspect-based opinion

Aspects	# of positive	# of negative
Duration	1108	914
Decoration	963	898
Difficulty	968	1024
Healty	1061	1352
Cost	1244	1197
Taste	1047	914
Recipe	2109	1634

5.2 Evaluation

We now evaluate the performance of our system. Firstly, we have to evaluate the performance of machine learning training models, F-Measure (F) is one of the standard metrics employed for evaluating our machine learning models, and this metric includes two fundamental factors, i.e., precision (P) and recall (R), which are obtained from the following relations [41].

$$P = \frac{true\,positive}{true\,positive + true\,negative}$$

$$R = \frac{true\,positive}{true\,positive + false\,negative}$$

$$F = \frac{2 \times P \times R}{P + R}$$

And evaluating the Rule-based aspects-based sentiment analysis approaches using precision (P) and recall (R).

6 Results and Discussion

We have to experiment the two training models, using 10-fold cross-validation with WEKA [40] where the used machine learning algorithms are already implemented.

6.1 Experiment with the Subjectivity Detection Training Model

We started our experiment with showing which parts of speech are the most appeared in each class (subjective and objective) [4]. The result is showing in Fig. 2.

According to Fig. 2, we note that injections and emoticons appear only in subjective text, which it contains more adjectives and adverbs, and this due to the nature of the subjective text contrary to the objective text which express usually facts.

After analyzing the dataset, we show and compare the obtained results of the subjectivity detection using the different machine learning algorithms and applying the different algorithms based on the injections and emoticons in the preprocessing step.

Comparing the both results (see Table 8), the better results were got using SVM classifier and applying the algorithms Algo1 and Algo2. The improvement of the accuracy results after using Algo1 and Algo2 is almost 14.7%. The same goes with

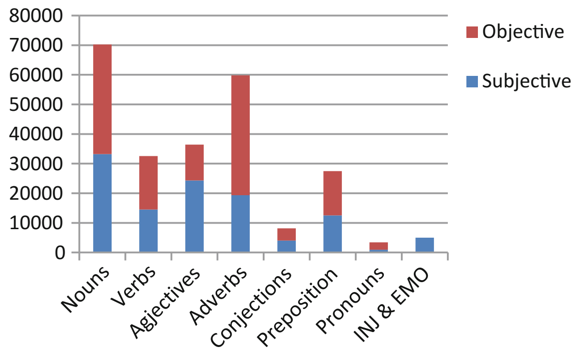

Fig. 2 The distribution of parts of speech in each category [4]

Table 8 Subjectivity detection results

Features	Precision	Recall	F-measure
Naïve Bayes	0.772	0.749	0.761
Naïve Bayes + Algo1 + Algo2	0.844	0.912	0.877
SVM	0.787	0.787	0.787
SVM + algo1 + Algo2	0.936	0.935	0.9355

Table 9 results based on V2 data version using Algorithm 1 and Algorithm 2

Version of data	SVM		
	Recall	Precision	F-measure
Subjective	0.903	0.963	0.932
Objective	0.966	0.909	0.937

the precision, recall and the F-measure. Table 9 shows the details of the results of each class (subjective and objective) using V2.

Table 9 calculates the performance results for the classification of the binary classifier at the stage of using V2.

6.2 *Evaluation on Aspect Extraction*

We carried out experiments on both our dataset and Semeval 2014 aspect based sentiment analysis data obtained from http://alt.qcri.org/semeval2014/task4/index.php?id=data-and-tools. Results are shown in Table 10.

Comparing the results (Table 10), our method based on the constructed lexicon build using cooking recipe reviews outperform the work in [27] based on SenticNet lexicon by overall 2.69% of precision and 4.15 of recall using Semeval dataset and 8.9% of precision and 13.5% of recall using our dataset.

Table 10 Results on the Semeval 2014 dataset and our dataset

Method	Semeval dataset (restaurant)	Our dataset	P (%)	R (%)
Work in [27] (SenticNet)	✓		85.21	88.15
Work in [27] (SenticNet)		✓	85.3	81.2
Our (based our lexicon)	✓		87.9	92.3
Our (our lexicon)		✓	94.2	94.7

Table 11 Aspect-based sentiment classifications results

Aspects	NB			NB + Algo3 +Algo4			SVM			SVM + Algo3 + Algo4		
	P	R	F	P	R	F	P	R	F	P	R	F
Duration	0.765	0.691	0.726	0.779	0.783	0.781	0.826	0.826	0.826	**0.953**	**0.942**	**0.947**
Decoration	0.829	0.762	0.794	**0.929**	0.867	0.897	0.824	0.824	0.824	0.904	**0.902**	**0.903**
Difficulty	0.688	0.86	0.765	0.835	0.846	0.84	0.853	0.812	0.832	**0.860**	**0.912**	**0.885**
Healty	0.661	0.743	0.7	0.714	0.753	0.733	0.717	0.74	0.729	**0.819**	**0.897**	**0.856**
Cost	0.795	0.816	0.805	0.851	0.842	0.844	0.834	**0.893**	0.863	**0.957**	0.87	**0.911**
Taste	0.853	0.812	0.832	0.884	0.928	0.905	0.879	0.923	0.901	**0.959**	**0.962**	**0.96**

Table 12 Cooking recipe sentiment classifications results

Method	P	R	F
NB	0.772	0.749	0.761
NB + Algo3 + Algo4	0.929	0.867	0.897
SVM	0.896	0.918	0.907
SVM + Algo3 + Algo4	0.948	0.950	0.949
SVM + MD	0.944	0.947	0.946
SVM + Algo3 + Algo4 + MD	**0.969**	**0.952**	**0.961**

6.3 Evaluation on Sentiment Classifications

We evaluate the following methods for aspect sentiment classification:

We train the Naïve Bayes (NB) and Support Vector Machines (SVMs) with the kernel RBF.

Tables 11 and 12 show the obtaining results with applying the different methods.

Results and Analysis. It can be observed from the results presented in Table 12 that using our proposed SVM + Algo3 + Algo4 + MD method for recipe sentiment classification outperforms the other methods using only textual features based on emoticons and injections or using only the meta-data associated to the YouTube cooking recipe video (*Likes, Dislikes and views information*). In particular, the improvement is more significant with the SVM RBF kernel where over 19.7% of precision 20.3% of recall and 20% of F-Measure comparing to NB (baseline) and 7.3% of precision of 3.7% recall and 5.4% of F-Measure comparing to SVM (baseline) of improvement is observed.

According to the Table 11 we remark that adding (Algo3 and Algo4) algorithms based on emoticons and injections to the textual features (Baselines) improve the both Baselines in all aspects sentiment classifications by: 1.4% of precision using NB (baseline) and 12.7% of precision using SVM (baseline) in duration. 10% of precision using NB (baseline) and 8% of precision using SVM (baseline) in dressing. 14.7% of precision using NB (baseline) and 0.7% of precision using SVM (baseline) in difficulty. 5.3% of precision using NB (baseline) and 10.2% of precision using SVM (baseline) in healthy. 5.6% of precision using NB (baseline) and 12.3% of precision using SVM (baseline) in cost, and 3.1% of precision using NB (baseline) and 8% of precision using SVM (baseline) in taste.

We remark that using SVM with RBF kernel outperforms NB classifier in the almost Aspect sentiment classifications.

However, additionally incorporating textual features based on emoticons and injections (Algo3 and Algo4) into our proposed methods gives marginal improvements compared to baselines. In addition, incorporating YouTube cooking recipe video meta-data (Likes, Dislikes and Views) improve the performance of the classifiers.

7 Conclusion and Future Works

This chapter deals with opinion mining and sentiment analysis fields. It proposed a real-time system for opinion extraction, and classification automatically the cooking recipes YouTube based on their aspects (implicit and explicit). Firstly, the proposed system collects YouTube videos meta-data (likes, dislikes, views and comments) about cooking recipes. Next step is to filter undesired texts (objective texts) using SVM classifier. Then we have constructed a lexicon for cooking recipes reviews and illustrated a method for extracting both explicit and implicit aspects from opinionated text. Then, classifying these texts (opinions) into a positive or negative class using the model built by SVM classifier. Our system can classify cooking recipes into ‘recommended” or “not recommended” and also compare between recipes.

In future works, we will propose new approaches based on other social text characteristics, and new aspect-based sentiment analysis methods.

References

1. Al-Smadi, M., Talafha, B., Al-Ayyoub, M., Jararweh, Y.: Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews. Int. J. Mach. Learn. Cybern. 1–13 (2018)
2. Al-Smadi, M., Al-Ayyoub, M., Jararweh, Y., Qawasmeh, O.: Enhancing aspect-based sentiment analysis of Arabic hotels' reviews using morphological, syntactic and semantic features. Inf. Process. Manage. (2018)
3. Benkhelifa, R., Laallam, F.Z.: Facebook posts text classification to improve information filtering. In: Proceedings of the 12th International Conference on Web Information Systems and Technologies, pp. 202–207. Rome, Italy (2016)
4. Benkhelifa, R., Laallam, F.Z.: Opinion extraction and classification of real-time youtube cooking recipes comments. In: International Conference on Advanced Machine Learning Technologies and Applications, pp. 395–404. Springer, Cham (2018)
5. Bianchini, D., De Antonellis, V., De Franceschi, N., Melchiori, M.: PREFer: a prescription-based food recommender system. Comput. Standards Interf. **54**, 64–75 (2017)
6. Chaturvedi, I., Ragusa, E., Gastaldo, P., Zunino, R., Cambria, E.: Bayesian network based extreme learning machine for subjectivity detection. J. Franklin Inst. (2017)
7. Choi, Y., Cardie, C.: Hierarchical sequential learning for extracting opinions and their attributes. In: Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2010), pp. 268–274 (2010)
8. Dave, K., Lawrence, S., Pennock, D.: Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In: Proceedings of the 12th International Conference on World Wide Web, ACM, New York, NY, USA, WWW '03
9. Dey, K., Shrivastava, R., Kaushik, S.: Twitter stance detection-a subjectivity and sentiment polarity inspired two-phase approach. In: SENTIRE Workshop, ICDM, Nov 2017
10. Durant, K.T., Smith, M.D.: Mining sentiment classification from political web logs, WEBKDD '06. Philadelphia, Pennysylvania, USA, ACM 1-59593-4448 (2006)
11. García-Pablos, A., Cuadros, M., Rigau, G.: W2VLDA: almost unsupervised system for aspect based sentiment analysis. Expert Syst. Appl. **91**, 127–137 (2018)
12. Hamouda, S.B., Akaichi. J.: Social networks' text mining for sentiment classification: the case of facebook' statuses updates in the "Arabic Spring" Era". Int. J. Appl. Innov. Eng. Manage. (2013)
13. Höpken, W., Fuchs, M., Menner, T., Lexhagen, M.: Sensing the online social sphere using a sentiment analytical approach. In: Analytics in Smart Tourism Design, pp. 129–146. Springer International Publishing (2017)
14. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, NY, USA, KDD (2004)
15. Kamal, A.: Subjectivity classification using machine learning techniques for mining feature-opinion pairs from web opinion sources, New Delhi—110025, India
16. Kang, Y., Zhou, L.: RubE: rule-based methods for extracting product features from online consumer reviews. Inf. Manag. **54**(2), 166–176 (2017)
17. Kübler, S., Liu, C., Sayyed, Z.A.: To use or not to use: feature selection for sentiment analysis of highly imbalanced data. Nat. Lang. Eng. **24**(1), 3–37 (2018)
18. Liu, B.: Sentiment analysis and opinion mining. Synth. Lect. Human Lang. Technol. **5**(1), 1–167 (2012)
19. Liu, C., Guo, C., Dakota, D., Rajagopalan, S., Li, W., K¨ubler, S.: My curiosity was satisfied, but not in a goodway: predicting user ratings for online recipes. In: Proceedings of the Second Workshop on Natural Language Processing for Social Media (SocialNLP), pp. 12–21. Dublin, Ireland, 24 Aug 2014
20. Liu, Q., Gao, Z., Liu, B., Zhang, Y.: Automated rule selection for aspect extraction in opinion mining. In: Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI '15, pp. 1291–1297. AAAI Press (2015)

21. Lisa Hankin, L.: The effects of user reviews on online purchasing behavior across multiple product categories. Master's Final Project Report, UC Berkeley School of Information (2007)
22. Lovins, J.B.: Development of a stemming algorithm. Mech. Trans. Comput. Linguist. (1968)
23. Manek, A.S., Shenoy, P.D., Mohan, M.C., Venugopal, K.R.: Aspect term extraction for sentiment analysis in large movie reviews using Gini index feature selection method and SVM classifier. World Wide Web **20**(2), 135–154 (2017)
24. Ozaki, T., Gao, X., Mizutani, M.: Extraction of characteristic sets of ingredients and cooking actions on cuisine type. In: 2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA), pp. 509–513. IEEE (2017)
25. Pang, A., Lee, L.: A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts In: Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL '04), pp. 271–278. Barcelona, ES (2004)
26. Piryani, R., Madhavi, D., Singh, V.K.: Analytical mapping of opinion mining and sentiment analysis research during 2000–2015. Inf. Process. Manage. **53**(1), 122–150 (2017)
27. Poria, S., Cambria, E., Ku, L.W., Gui, C., Gelbukh, A.: A rule-based approach to aspect extraction from product reviews. In: Proceedings of the Second Workshop on Natural Language Processing for Social Media (SocialNLP), pp. 28–37. Association for Computational Linguistics and Dublin City University (2014)
28. Pugsee, P., Niyomvanich, M.: Suggestion analysis for food recipe improvement. In: Proceeding of the 2015 International Conference on Advanced Informatics: Concepts, Theory and Application (ICAICTA) (2015)
29. Rana, T.A., Cheah, Y.N.: A two-fold rule-based model for aspect extraction. Expert Syst. Appl. **89**, 273–285 (2017)
30. Rana, T.A., Cheah, Y.N.: Exploiting sequential patterns to detect objective aspects from online reviews. In: 2016 International Conference on Advanced Informatics: Concepts, Theory And Application (ICAICTA), pp. 1–5. IEEE, Aug 2016
31. Rana, T.A., Cheah, Y.N.: Improving aspect extraction using aspect frequency and semantic similarity-based approach for aspect-based sentiment analysis. In: International Conference on Computing and Information Technology, pp. 317–326. Springer, Cham, July 2017
32. Rao, S., Kakkar, M.: A rating approach based on sentiment analysis. In: 2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence, pp. 557–562. IEEE, Jan 2017
33. Rokicki, M., Herder, E., Trattner, C.: How editorial, temporal and social biases affect online food popularity and appreciation. In: ICWSM, pp. 192–200 (2017)
34. Sebastiani, A.: Machine learning in automated text categorization. ACM Comput. Surv. **34**, 1–47 (2002)
35. Singh, A., Shukla, N., Mishra, N.: Social media data analytics to improve supply chain management in food industries. Transp. Res. Part E Logist. Transp. Rev. (2017)
36. Tan, S.S., Na, J.C.: Mining semantic patterns for sentiment analysis of product reviews. In: International Conference on Theory and Practice of Digital Libraries, pp. 382–393. Springer, Cham, Sept 2017
37. Verma, S., Bhattacharyya, P.: Incorporating semantic knowledge for sentiment analysis. In: Proceedings of International Conference on Natural Language Processing (2009)
38. Raut, V.B., et al.: Survey on opinion mining and summarization of user reviews on web. (IJCSIT) Int. J. Comput. Sci. Inf. Technol. **5**(2), 1026–1030 (2014)
39. Wilson, T.: Fine-grained subjectivity and sentiment analysis: recognizing the intensity, polarity, and attitudes of private states, University of Pittsburgh (2008)
40. Witten, H.A., Frank, E.: Data mining: practical machine learning tools and techniques with java implementations, Morgan Kaufmann (2000)
41. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco (2005)
42. Zhang, X., Zhu, F.: The influence of online consumer reviews on the demand for experience goods: the case of video games. In: 27th International Conference on Information Systems (ICIS). Milwaukee, AISPress (2006)

43. Zeng, L., Li, F.: A classification-based approach for implicit feature identification. In: Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. 12th China National Conference, CCL 2013 and First International Symposium, NLP-NABD 2013, Suzhou, China, 10–12 Oct 2013, Proceedings, volume 8202 of Lecture Notes in Computer Science, pp. 190–202 (2013)
44. Zhen, H., Chang, K., Kim, J.: Implicit feature identification via co-occurrence association rule mining. In: Computational Linguistics and Intelligent Text Processing. 12th International Conference, CICLing 2011, Tokyo, Japan, 20–26 Feb 2011. Proceedings, Part I, volume 6608 of Lecture Notes in Computer Science, pp. 393–404 (2011)