AraMAMS Restaurants Reviews Corpus for Aspect-Based Sentiment Analysis

Subjects: Computer Science, Cybernetics
Contributor: Alanod AlMasaud, Heyam H. Al-Baity

AraMA comprises 10,750 Google Maps reviews for restaurants in Riyadh, Saudi Arabia. It covers four aspect categories—food, environment, service, and price—along with four sentiment polarities: positive, negative, neutral, and conflict. All AraMA reviews are labeled with at least two aspect categories. A second version, named AraMAMS, includes reviews labeled with at least two different sentiments, making it the first Arabic multi-aspect, multi-sentiment dataset. Aspect-based sentiment analysis (ABSA) is a field of SA that goes one step further than SA by automatically assigning sentiments to certain features or aspects in the text.

Keywords: sentiment analysis; aspect-based sentiment analysis; sustainability

1. Introduction

With the growth of social media usage recently, it is essential to discover and reap the benefits of online user-generated information to enhance a product or service, and help to create more effective marketing efforts. For instance, analyzing consumers’ feelings and opinion data from reviews on e-commerce platforms is very important, as it provides insight into customers’ satisfaction levels. This type of data analysis can provide businesses with valuable insights into customer sentiment, brand perception, market trends, and investment opportunities. By leveraging these insights, businesses can enhance customer satisfaction, brand reputation, market competitiveness, and financial performance. Overall, assessing customer reviews is a critical component of constructing a strong and equitable infrastructure that promotes economic growth and improves the quality of life and wellbeing.

However, analyzing this opinion data manually would be impossible, given the enormous volume of textual content. As a result, the field of sentiment analysis (SA) has emerged as an AI tool that allows automatic extraction of the knowledge about opinions, emotions, and attitudes concealed within unstructured texts. Yet, SA only provides a view of what people like or dislike. It basically classifies a given text into positive, negative, and neutral sentiments. Aspect-based sentiment analysis (ABSA) is a field of SA that goes one step further than SA by automatically assigning sentiments to certain features or aspects in the text. The primary goal of ABSA is to extract the relevant aspects and then classify them into different sentiment polarities. This entails breaking down text data into smaller fragments in order to obtain deeper, more granular insights. As such, all relationships between the entities involved must be appropriately identified and linked to the conveyed sentiment. Thus, the main challenge of this task is to distinguish between different opinion contexts for different aspects or targets. Of late, ABSA has become one of the most important tasks of SA, since it can extract a deeper insight from text to ensure that the right decisions are made, and provide a clearer image of weaknesses.

ABSA can play a significant role in supporting the Sustainable Development Goals (SDGs) of the 2030 Agenda, which were adopted by the United Nations to address global challenges and promote sustainable development. This can be accomplished by providing insights into the sustainability performance of businesses. For instance, ABSA for customer reviews can help in assessing the sustainability efforts of restaurants, hotels, retail stores, or service providers by analyzing sentiments related to specific aspects of sustainability, and this allows for a comprehensive assessment. Analyzing restaurant customer reviews is critical for constructing high-quality, long-lasting, and robust infrastructure to promote economic development and human wellbeing. If restaurant owners analyze customer reviews, they will be more aware of their needs, enabling them to direct efforts and money to improve these aspects more quickly and with less effort. As a result, client happiness and loyalty rise, resulting in higher revenue and economic progress. Furthermore, restaurant owners can focus more on enabling cheap and equitable access to high-quality food and improve eating experiences for all consumers in order to obtain more positive comments which, in turn, will increase the number of visitors to the restaurant. Overall, assessing customer reviews is a critical component of constructing a strong and equitable infrastructure that promotes economic growth and the wellbeing of all people.
In recent years, a number of researchers have carried out a great deal of work in the field of SA and its application. However, ABSA studies are still scarce compared to SA research, especially in the Arabic language. This is for two main reasons: the lack of labelled dataset resources in Arabic, and the complexity of the Arabic language. There are three different varieties of the Arabic language, Classical Arabic (CA), which is used in the Holy Qur’an of Islam; Modern Standard Arabic (MSA), which is used in official contexts such as newspapers and education; and Dialectical Arabic (DA), which is used in daily conversation and in most social media content. DA also differs from one Arab nation and the next, and does not have standard orthographies.

2. Arabic Multi-Aspect, Multi-Sentiment Restaurants Reviews Corpus for Aspect-Based Sentiment Analysis

There are several studies in the Arabic language for ABSA. Table 1 provides a summary of the datasets used regarding their domain, size, Arabic language type, publicity, predefined aspect categories, sentiment polarity, and, if applicable, the platform used during annotation. For a comprehensive review of the available Arabic ABSA research, referring the reader to

<table>
<thead>
<tr>
<th>Paper</th>
<th>Domain</th>
<th>Size</th>
<th>Language</th>
<th>Published</th>
<th>Platform</th>
<th>Polarities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al-Samadi [9]</td>
<td>Book</td>
<td>1513 Arabic book reviews</td>
<td>MSA</td>
<td>Yes</td>
<td>N/A</td>
<td>positive, negative, neutral, or conflict</td>
</tr>
<tr>
<td>Abd-Elhamid [10]</td>
<td>Multiple domains</td>
<td>200 reviews</td>
<td>DA</td>
<td>No</td>
<td>N/A</td>
<td>positive, neutral</td>
</tr>
<tr>
<td>Al-Samadi, AlAyyoub [11]</td>
<td>News posts related to the Gaza conflict</td>
<td>2265 news posts</td>
<td>MSA</td>
<td>No</td>
<td>BRAT</td>
<td>positive, neutral, or conflict</td>
</tr>
<tr>
<td>Al-Samadi, Qwasmeh [12]</td>
<td>Hotel</td>
<td>2291 hotel reviews</td>
<td>MSA, DA</td>
<td>Yes</td>
<td>N/A</td>
<td>positive, neutral, or conflict</td>
</tr>
<tr>
<td>Ashi et al. [13]</td>
<td>Saudi Airline service-related</td>
<td>5000 tweets</td>
<td>MSA, DA (Saudi dialect)</td>
<td>No</td>
<td>N/A</td>
<td>positive or negative</td>
</tr>
<tr>
<td>Alshammari and AlMansour [14]</td>
<td>Telecommunication companies in Saudi Arabia</td>
<td>1098 tweets</td>
<td>DA</td>
<td>No</td>
<td>DataTracking</td>
<td>positive, negative and natural</td>
</tr>
<tr>
<td>Allassaf and Qamar [15]</td>
<td>Educational</td>
<td>7934 tweets</td>
<td>DA</td>
<td>No</td>
<td>N/A</td>
<td>negative and not negative</td>
</tr>
<tr>
<td>Masadeh and Saad AlAzzam [16]</td>
<td>Book</td>
<td>1000 Arabic book reviews</td>
<td>MSA</td>
<td>No</td>
<td>N/A</td>
<td>positive or negative</td>
</tr>
<tr>
<td>Areed [17]</td>
<td>Government</td>
<td>2071 Arabic reviews</td>
<td>MSA, DA</td>
<td>No</td>
<td>GARSA</td>
<td>ranging from −1.0 to 1.0</td>
</tr>
</tbody>
</table>

Note: Table contains the paper’s author, dataset domain, size, language, whether it is published or not, the platform used during annotation, and sentiment polarity tags.

In 2015, Ref. [10] provided the first research benchmark dataset for ABSA. The authors created the human-annotated Arabic dataset for book reviews (HAAD) [14]. It consists of 1513 annotated book reviews taken from the large Arabic book review corpus (LABR) dataset that was essentially created for SA [13]. The authors annotated aspect terms, aspect categories, and sentiment polarities. In Ref. [6], published in the same year, the authors collected 200 reviews from forums, Facebook, YouTube, and Google search. Then, they extracted aspects using part-of-speech tagging (POS) and manually annotated sentiment polarity.

In 2016, Al-Sarhan et al. [11] collected 2265 Arabic news posts related to the Gaza conflict, associated comments from Al Jazeera and Al Arabiya (well-known Arabic news networks), and related posts on Facebook. They annotated the posts’
aspect categories, aspect terms, and sentiment polarities. Additionally, they annotated comment categories and sentiment polarities. They chose the most dominant aspect category only for both posts and comments. In the same year, Semantic Evaluation (SemEval) launched a workshop to create Arabic hotel reviews for ABSA; it has been used as a benchmark until now. The dataset includes a total of 2291 annotated review sentences, of which 1839 were used for training and 452 for testing. The sentences were gathered from hotel booking websites such as Booking.com, (accessed on 23 June 2023) and TripAdvisor.com, (accessed on 23 June 2023). The selected reviews belong to hotels from different Arabian cities such as Dubai, Mecca, Amman, Beirut, etc. They annotated aspect terms, aspect categories, and sentiment polarities. The aspect category annotations were more detailed; annotators were required to identify entities and attributes. Further, the category field needed to be completed using the syntax of (Entity # Attribute). Entities were predefined as hotel, rooms, room_amenities, facilities, service, location, food&drinks. Attributes were defined as general, prices, design&features, cleanliness, comfort, quality, style&options, and miscellaneous. For example, in the sentence “the rooms are comfortable”, the entity is the room, and the attribute is comfort. Thus, the category field would be defined as value (room#comfort).

A much simpler corpus was subsequently created. The authors of compiled a total of 5000 tweets related to the service on Saudi airlines. They annotated aspect categories and sentiment polarities. Additionally, in Ref. customers' sentiments were extracted—using machine learning and deep learning approaches—from 1098 tweets, collected by the authors, regarding the Saudi telecommunication companies STC, Mobile, and Zain. The paper was part of an ongoing project. They extracted available aspects such as internet, customer services, network, billing, packages, and general. For annotation, they manually annotated sentiment polarities using the DataTracking website.

In 2020, a total of 7934 tweets related to Qassim university were collected by the author of. Annotators labelled aspect categories and sentiment polarities. In, 1000 Arabic book reviews were selected and annotated from an LABR dataset. Annotators labelled aspect terms and sentiment polarity terms. Since a review can contain more than one aspect, they added one sentiment for an entire review, much like SA. Additionally, in, a total of 2071 Arabic reviews were selected from the Apple Store and Google Play. Using these reviews, 60 different mobile apps were created by the United Arab Emirates’ government. Annotation was carried out with a specially designed computer application named “GARSA”. Annotators labelled aspect terms, sentiment words, and aspect categories.

References


Retrieved from https://encyclopedia.pub/entry/history/show/109076