Aspect Based Sentiment Analysis in Music: a case study with Spotify

Giovanni Maria Biancofiore*
Polytechnic University of Bari
Bari, Italy
giovannimaria.baincofiore@poliba.it

Tommaso Di Noia
Polytechnic University of Bari
Bari, Italy
tommaso.dinoia@poliba.it

Eugenio Di Sciascio
Polytechnic University of Bari
Bari, Italy
eugenio.disciascio@poliba.it

Fedelucio Narducci
Polytechnic University of Bari
Bari, Italy
fedelucio.narducci@poliba.it

Paolo Pastore†
Polytechnic University of Bari
Bari, Italy
paolo.pastore1@poliba.it

ABSTRACT

Nowadays, there are more and more social networks and Web platforms that give their users the opportunity to share their opinions and tastes on items of different types. This inevitably led to a growth of data relating to the subjective sphere of each individual. This information is extremely useful for several purposes, such as providing personalized recommendation services or understanding opinions conveyed through text. Sentiment Analysis provides helpful methods to analyze these textual opinions (e.g. reviews) from a global point of view. In case we want a more detailed representation of the opinion represented in a text, Aspect-based Sentiment Analysis identifies a valuable option thanks to its fine-grained level of text analysis.

In this paper, we have designed a processing pipeline aimed to extracting domain-related aspects from text by means of an unsupervised approach. We formally define Aspect Terms and Aspect Categories as well as Aspect-based Sentiment Embedding, an approach of representing documents by computing aggregated sentiment scores for each aspect. We perform experimental evaluations on the Spotify dataset to prove the utility of our technique in predicting elements strictly related to emotions and feelings. Our results show improvements on the regression task for sentiment-related features compared to the classical semantic-based representations.

CCS CONCEPTS

• **Information systems** → **Content analysis and feature selection**; **Data encoding and canonicalization**; *Language models*; **Sentiment analysis**; • **Computing methodologies** → *Dimensionality reduction and manifold learning*;

*Corresponding author
†Corresponding author

SAC'22, April 25 –April 29, 2022, Brno, Czech Republic

ACM ISBN 978-1-4503-8713-2/22/04…$15.00
https://doi.org/xx.xxx/xxx_x

KEYWORDS

Natural Language Processing, Sentiment Analysis, Machine Learning, Aspect Based Sentiment Analysis

ACM Reference Format:
Giovanni Maria Biancofiore, Tommaso Di Noia, Eugenio Di Sciascio, Fedelucio Narducci, and Paolo Pastore. 2022. Aspect Based Sentiment Analysis in Music: a case study with Spotify. In *Proceedings of ACM SAC Conference (SAC'22).* ACM, New York, NY, USA, Article 4, 8 pages. https://doi.org/xx.xxx/xxx_x

1 INTRODUCTION

The increasing amount of user-generated content on blogs, social networks, and e-commerce websites has pushed many companies in further exploring data mining technologies to exploit this source of subjective information. One of the most prolific research areas in the Natural Language Processing is Sentiment Analysis, which aims to identify and extract user opinions from text. In fact, Sentiment Analysis has the goal to identify how sentiments are expressed in texts and whether the expressions indicate positive (favorable) or negative (unfavorable) opinions toward the subject [27]. For instance, a sentence like *"The latest Apple Macbook has finally arrived and is amazing!"* tweeted by a user can give to the Apple Company a valuable opinion about its product just released. Another example might be related to the detection of the emotional state of a customer during a dialog with a chatbot. The efficacy of an automatic customer service can be improved by moving the dialogue to human operator in case the conversation starts to be frustrating for the user. Sentiment Analysis techniques are designed to cover all those situations, finding and measuring the subjectivity behind words that express a Sentiment.

Sentiment Analysis can be performed at three different levels: document level, sentence level and entity level. While at both document and sentence level the goal of sentiment analysis is to find the overall opinion expressed in a text, entity-based sentiment analysis aims to discover the opinion of a user with respect to a specific entity extracted from the raw text. This task is typically known as Aspect-Based Sentiment Analysis (ABSA). More specifically, ABSA mines opinions from text about specific entities and their aspects [31]. It is usually divided into two main sub-tasks: the Aspect Term Extraction (ATE), which extracts all the words that refers to an aspect, and the Opinion Term Extraction (OTE), that aims to find all the expressions that convey the user opinion with respect to the extracted aspect terms (i.e. assign a sentiment score to a specific aspect term).

The task of ABSA is not trivial because of different reasons: there are not objective and formal definitions of *Aspect* and, in general, possible aspects differ from a domain to another. Moreover, users could refer to the same concept using different words or expressions. For example, in the Semeval 2014 laptop dataset [32] there are a lot of training samples that express opinions about the aspect *"Technical Support"*, but reviewers refer to it with expressions like *"technical service"*, *"tech support"* or *"staff"*. Additionally, there are not consistent and labeled datasets for the ABSA task, pushing the majority of work in this field to adopt unsupervised approaches.

Nevertheless, a huge amount of data about sentiment conveyed by text is available given the high interest of the community in automatically understanding subjective opinions and emotions from written documents. Among several domains, datasets belonging to the music field undoubtedly offer the richest collection of texts. On this line, a good number of work started to analyze and exploit that data for several purposes, like recommending music based on users emotional states [1] as well as analyzing how songs characteristics match user personalities [26].

In this work we aim to answer and give empirical evidence to the following research questions:

- *Is it possible to group Aspect Terms extracted through unsupervised approaches into coherent Domain Aspect Categories relevant in a ABSA scenario?*
- *Can a Sentiment-based text representation improve the performances of Sentiment-related tasks?*

In order to reach these goals, the unsupervised paradigm chosen for the aspect detection task exploits dependency trees and syntactic relations between words to extract not only *Aspect Terms* but also *Opinion Words*, allowing us to assign a sentiment score to them. Therefore, we adopt a Word2Vec representation to group together all the similar *Aspect Terms* and define the *Aspect Categories* they belong to.

As main contribution, we introduce an unsupervised method to extract aspects from text which are relevant for Sentiment Analysis. Compared to other state-of-the-art solutions, which detect the *Aspect Terms* but compute a sentiment score for the overall document, our solution groups all the found *Aspect Terms* into *Aspect Categories*, each of them characterized by a specific sentiment value. To highlight these distinctions, we also provide formal definitions about *Aspect Term* and *Aspect Category*. Moreover, we introduce the concept of *Aspect-Based Sentiment Embedding*, a representation strategy of textual documents based on domain aspects and their related sentiment scores. We demonstrate the utility of this embedding technique by designing an experiment aiming to prove that our approach grants improved performances on predicting textual features related to human feelings against the well-known semantic approaches.

Our intuition is that a sentiment-based embedding could be more representative than a semantic one for text rich of sentiment. To the best of our knowledge, this is the first attempt that introduces the concept of numerical representation based on sentiment and exploits it to improve performances of sentiment-related tasks.

Since song lyrics are rich of human emotions [7], we choose the music domain to evaluate our sentiment-based representation on a prediction task. In particular, Spotify, one of the most renowned music streaming services, released a dataset composed by a set of features that best fit our needs paired with song lyrics. To prove our hypothesis, we train several models for predicting different features on a given lyrics representation, which can be based on sentiment (Aspect Based Sentiment Embedding) or semantic vectors (i.e. Doc2Vec). This comparison is made because while the latter rely more on semantic features like word co-occurrence or context that embed the semantic meaning of each word, the former enclose information strictly related to subjective sentiment and opinions. In other words, the semantic embedding identifies what the user means while the sentiment one points out what the user feels. Therefore, our proposed sentiment-based embedding strategy is more suitable to assess features strictly related to human emotions. Finally, we assessed statistical significance of our results by a T-Student test.

2 RELATED WORK

Most recent works in the Sentiment Analysis field exploit supervised learning techniques that have proven to be very effective for different text classification tasks. In particular, among different supervised approaches, there has been a lot of interest in Deep Learning models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNN) with all their variants (i.e. LSTMs and GRUs), or Language Models like ELMo [29] and BERT [8].

CNNs have shown remarkably strong performances on the important task of sentiment classification [15–17]. The promising results achieved with CNN-based models have driven researchers to further explore the Deep Learning field. For instance, in [11] authors proposed a joint CNN and LSTM framework that takes the features extracted by the CNN as input for the LSTM to perform sentiment analysis on short texts by means of pre-trained word embedding model. Similarly, in [3] Behera et al. investigated a hybrid approach of CNN and LSTM called Co-LSTM for the analysis of consumer reviews posted on social media, aiming to be domain independent and highly adaptable in examining big social data while keeping scalability. Recently, there has been a lot of interest in the combination of classical Deep Learning architectures and Attention mechanisms, whose goal is to mimic cognitive attention in finding relevant contextual elements in a sentence [35]. Many studies [2, 5, 38] have demonstrated that the combination of the previous techniques have improved performances on the text classification task. The advent of powerful language models such as BERT defined a new trend in the NLP research field. In particular, there has been a lot of interest in training custom models for specific languages that can be exploited for solving different research problems, i.e. Sentiment Analysis [30, 34].

As opposed to extracting the general sentiment expressed in a piece of text, Aspect-Based Sentiment Analysis (ABSA) aims to extract both entities described in the text (e.g. attributes or components of a product or service) and the writers opinions about such entities. As a result, there has been an increased interest by the NLP community on ABSA and Aspect Term Extraction (ATE), so more resources have been made available for these tasks. Early works

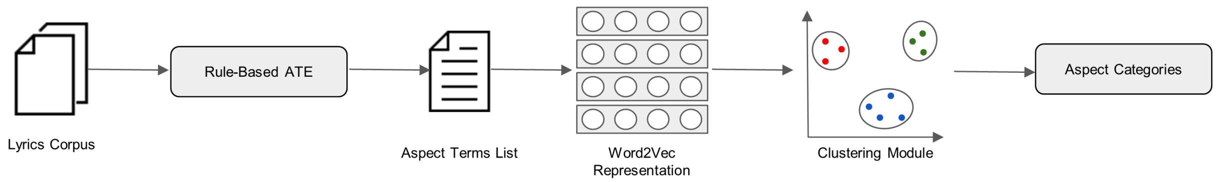

Figure 1: Aspect Term Extraction and Clustering

deal with the extraction of Aspect Terms and the related sentiment by using association rules [13], Conditional Random Fields (CRF) [14], knowledge-based topic modeling [6], or double propagation [21, 22]. Recently, the research focus has moved on using Deep Learning models for ATE. For instance, in [33] authors made different experiments with Convolutional Neural Networks (CNNs) and several word embedding strategies. The main problem of CNNs is that they are not able to handle long-term dependencies. This is the reason why recent works started to adopt Recurrent architectures, such as LSTMs. The latter are used by Li et al. [20] where aspects detection history and opinion summary are used to enhance the ATE model.

Other Deep Learning based approaches also tried to exploit the information in the dependency trees to enrich the sentence representations as input to the ATE models. Ye et al. [40] proposed a model based on CNNs, that uses Tree-Based Convolution to incorporate dependency relationships. Luo et al. [23] used a Bi-directional LSTM to learn a representation of the dependency tree for each review.

In our work, the main focus is on extracting aspect terms and sentiment information from raw text in an unsupervised way. In literature, different approaches have been proposed to implement unsupervised ABSA models. In [10] authors developed a framework able to perform both supervised and unsupervised ATE in large review datasets, based on Bi-Directional LSTM networks and CRF. The idea is to automatically annotate a dataset in an unsupervised manner by only considering nouns and noun phrases as candidate aspects. The resulting labeled dataset is then used for training the Deep Learning model. In [37] Vargas et al. proposed a simple approach called SUAEx for unsupervised aspect extraction, which relies solely on the similarity of word embeddings. He et al. [12] introduced a novel neural approach with the aim of discovering coherent aspects exploiting the distribution of word co-occurrences through the use of neural word embeddings. In addition, they implemented an attention mechanism to de-emphasize irrelevant words during training, further improving the coherence of aspects. In [4] authors designed a two-step hybrid model by combining linguistic patterns with Deep Learning techniques to improve the ATE task. The first step exploits a rule-based approach to extract aspect terms, then used as input for the second step that consists in training an attention based Deep Learning model. Other works, instead of using rule-based approaches, focused on the implementation of domain-independent Deep Learning ATE models trained on multiple datasets (e.g. restaurant, laptops and hotel reviews) in order to improve their robustness and generality [28].

3 UNSUPERVISED ASPECTS EXTRACTION

One of the main issues we try to address is that aspect terms extracted through state-of-the-art approaches are not grouped into categories since they are not inferred. Another problem is that these works do not leverage all the fine-grained sentiment information enclosed in each aspect. In our work, we address the problem of unsupervised ABSA for extracting relevant aspects and the related sentiment in order to understand if features related to human feelings can be better inferred with a representation more based on sentiment data instead of semantic one. For this purpose, we implement an unsupervised ABSA model that has been tested on the Spotify dataset, creating a representation of each song based both on the extracted aspects and the related sentiment. These songs encodings are then used for predicting different scores like danceability, valence, mode and others that will be discussed in the next sections.

To the best of our knowledge, there are not works in literature that exploit aspects and their related sentiment for a regression task like the one described before.

Finding relevant aspects in a specific domain is a not trivial task since they are not known a priori and because of their intrinsic subjectivity. These concepts that typically describe features of items like instruments or cars could be inferred by exploiting several Natural Language Processing techniques if they have not already been provided.

To reach this objective, one possibility is to collect all the nouns that are directly related to words that carry a sentiment with them (e.g. *unconditional love*, *great party*, *beautiful sun*, etc.). The main intuition behind this assumption is that items aspects are generally evaluated by the reviewers who expose opinions on them [10, 12].

Without loss of generality, it is possible to widen this assumption to any written text belonging to any domain. For example, in the music domain aspects can be identified with topics treated by the song-writer. Therefore, depending on the topics and their sentiment polarity we can infer some peculiar features that are more related to subjective emotions instead of the semantic meaning.

Given a set of documents, we define the pipeline based on an unsupervised approach depicted in Figure 1, able to identify all those aspects that characterize a specific domain. This choice has been made given the absence of a complete and consistent dataset designed for this task, as well as to keep the aspect extraction algorithm as less domain dependant as possible. To better explain how the proposed method reaches the final goal of detecting the domain aspects, we found helpful to introduce two formal definitions about the *Aspect Category* and the *Aspect Term*.

Definition 1 (Aspect Category). *Given a domain D, an Aspect Category A, or more briefly Aspect, is either a topic or concept which meticulously describe a specific domain characteristic s.t. a domain is portrayed by a set of N Aspects* $D = \{A_0, A_1, ...A_N\}$.

As an example we may consider the domain of movies. Some Aspect Categories which best portray film industry products are undoubtedly the ones evaluated by the Academy of Motion Picture Arts and Sciences assigning the Academia Award, like *Picture*, *Visual Effect*, *Director*, etc. Then, for the movie domain and the corresponding Aspects are $D = \{Picture, Director, Visual\_Effect\}$

Definition 2 (Aspect Term). *Let A be an Aspect for the domain D and w a generic natural language word. Then w is an Aspect Term for D if w is a token used to denote the concept or the topic of the Aspect A. We denote this relation with* $w \triangleleft A$.

Following the previous example, a movie can be entirely described by talking about its computer graphic, the plot and the actors performances as well, all words that belong to the Aspect Terms sets of the movie domain. For instance:

$$(light, color, saturation, landscape) \triangleleft Picture$$

$$(monster, fire, explosion, dragon) \triangleleft Visual\_Effect$$

$$(adaptation, framing, casting) \triangleleft Director$$

Therefore, given the sentence: *"Interstellar is a film with incredible landscapes, where realistic space and planets are the context for great actors and an impeccable soundtrack."*, we can infer that:

$$D = \{Picture, Visual\_Effect, Actor, Music\}$$

$$(landscape) \triangleleft Picture \land (soundtrack) \triangleleft Music$$

$$(space, planets) \triangleleft Visual\_Effect \land (actor) \triangleleft Actor$$

The main goal of our pipeline is the Aspect Term Extraction. It bases on a sequence of processing steps that starting from textual data, is able to identify *Aspect Categories* together with the corresponding *Aspect Terms*. Following the process depicted in Figure 1, we have a sequence of steps:

1. **Candidate Aspect Term Detection:** all the words that have an high probability to be an *Aspect Term* are here identified by means of the *Rule-Based Aspect Term Extractor*. It searches entities on which opinions are expressed exploiting some syntactic rules. In details, nouns usually refer to entities, while adjectives, adverbs, and some verbs usually carry a sentiment information that define the opinion polarity [39]. Thus, only those nouns that are strictly linked to words having a non neutral sentiment score are selected. To this purpose, a Dependency Parse Tree is generated for all the sentences in each document of a collection since dependency relations reveal existing connections between word and possible sentiment inflections.
   All nouns that have dependencies with adjectives and adverbs with the adjectival modifier (amod) and adverbial modifier (advmod) relation respectively are considered as *Candidate Aspect Term*, including also nouns that have some connections with verbs characterized by a non-neutral sentiment score, obtained from the Sentiment Lexicon SentiWords [9]. This because some expressions like *"The colors makes you love the film"* give a positive opinion on the aspect *Picture* with the composed verb *makes - love* without exploiting adjectives neither adverbs. It is worth noticing that words enclosing opinions are considered only as a filter condition to select candidate terms of interest, without storing any information about them, while the frequency of each unique term on the entire collection of documents is kept.
2. **Aspect Term List Generation:** From the *Candidate Aspect Terms*, all the words that have the same *lemma* are collapsed via lemmatization. For each term who share the same lemma we sum its relative number of occurrences, then we delete all terms whose counter is lower then a specific threshold (10 in our case), resulting in the final *Aspect Terms List* as outlined in Figure 1. This choice is for avoiding to take into account aspects that are not highly informative for inferring *Aspect Categories*, thus generating noise in the following steps.
3. **Word2Vec Representation**: aiming to define which are the main *Aspects* that qualify the domain of interest, we take advantage of semantic vector representations of terms like Word2Vec [25]. Due to its known capability of shaping categorical feature like words into numerical representations with an intrinsic semantic spatial distribution, we found Word2Vec particularly fits our goal. In details, we compute for each lemma that populates the *Aspect Terms List* its Word2Vec embedding.
4. **Aspect Categories Extraction**: all the Aspect Terms embeddings are grouped in clusters by means of the K-means algorithm [24]. It allows us to define those categories which best assemble the *Aspect Terms* in a semantically consistent way. The *Aspect Categories* are thus identified with *centroids* of the retrieved clusters, and each of them keeps track of a list of *Aspect Terms* to which they belong to.

To clarify the process of Aspect Term Extraction and Clustering, shall we consider two example tracks from the music domain: *"My Father's Eyes"* by Eric Clapton and *"Civil War"* by Guns 'N Roses. From these two songs our approach is able to extract some aspects such that:

$$D = \{Family, Love, War, Humanity\}$$

More in depth, to define the previous *Aspect Categories*, the Rule-Based ATE module receives as input a set of lyrics from which we present the following excerpts: *"When I look in my father's eyes, my father's eyes, then the light begins to shine and I hear those ancient lullabies"* and *"I don't need your civil war, it feeds the rich while it buries the poor, your power hungry sellin' soldiers, in a human grocery store"* and then, it extracts some candidate *Aspect Terms* like *"father"*, *"lullabies"*, *"war"*, *"power"* and *"soldiers"*. After removing not frequent terms like *"lullabies"* and *"power"*, we obtain that the remaining words belong to the Aspect Term List and Word2Vec representations are computed for them. These representations are then used by the Clustering module, together with all the other aspect terms extracted from the lyrics, to define the previous Aspect Categories, reaching:

$$(father) \triangleleft Family \land (war, soldier) \triangleleft War$$

It is worth noticing that our *Aspect Extraction* approach is not domain dependant, so it can be applied without loss of generality to

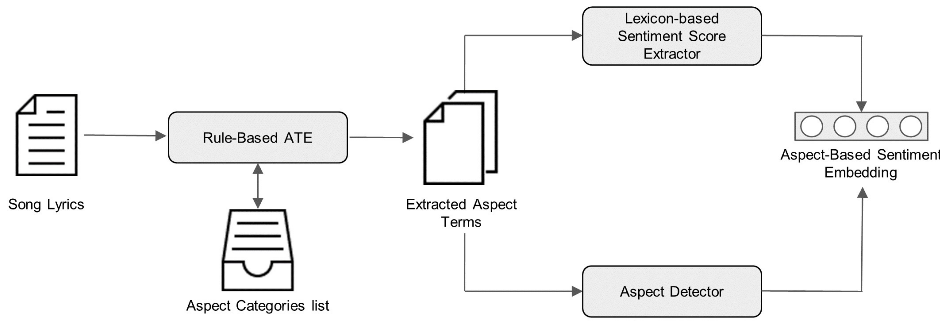

Figure 2: Aspect Based sentiment representation strategy

several objectives in many different domains, like reviewing items for e-commerce platform or detecting sentiment features in a songs collection.

4 SENTIMENT REPRESENTATION

The main idea behind our work is to identify a numerical sentiment-based representation of documents, which could deal with features strictly related to emotions or opinions in a more accurate way.

To define this new kind of embedding, we found useful to set the *Aspect Categories* as dimensions of the vector representation, in the sense that each domain aspect match with a specific vector position whose element or value enclose a sentiment information about that Aspect. For instance, assuming that we found fifty *Aspect Categories* in the music doamin:

$$D = \{People, Friendship, ..., Love, ..., Time\}$$

Sentiment embeddings of songs lyrics will have the first position referred to the aspect *People*, the second one related to *Friendship*, the forty-eighth to *Love* and the last to *Time*. Each of them could have a non zero value depending whether the song-writer expresses an opinion on the aspect or not, as also highlighted in Figure 3. More formally:

Definition 3 (Aspect-Based Sentiment Embedding). *Let A be a set of Aspect Categories, N the cardinality of A and s a sentiment score s.t.* $s \in [-1, 1]$. *Given a document D, composed by different sentences, we define the Aspect-Based Sentiment Embedding* $\mathbf{e}$ *of D the vector:*

$$\mathbf{e} = (s_0, s_1, ..., s_N)$$

where $s_i, i = 0, 1, ..., N$ *is the sentiment score computed on the i-th Aspect Category.*

Documents that treat the same aspects are similar in their representations. Moreover, depending on the opinions expressed for each aspect, the computed sentiment value is an aggregation of sentiment scores obtained from each adjective, adverb, and verb linked to that aspect.

More in depth, we base our sentiment based representation on the results produced by the *Aspect Term Extractor* presented in Section 3. Starting from the list of *Aspect Categories* that our model found over a domain, our approach takes as input a document written in Natural Language and searches for aspects and the related sentiment scores to shape the aspect-based sentiment embeddings. The designed pipeline is outlined in Figure 2.

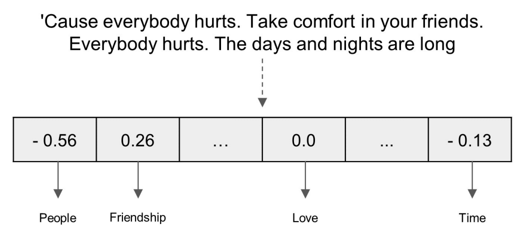

Figure 3: Aspect-Based Sentiment Embedding example

To evaluate our system, we choose to build the aspect-based sentiment embeddings over documents belonging to the music domain. This because songs lyrics carry a more emotional meaning instead a semantic one, so they particularly fit to test performances in predicting sentiment based features. However, our approach is not dependant on the music domain, thus relying on a domain independent method that exploits an *Aspect Category* list to identify the numerical representations.

The proposed method receives in input a document, that is a song in this case, and searches for *Extracted Aspect Terms* by means of the *Rule-Based Aspect Term Extractor* already adopted in the *Aspect Term Extractor* outlined in Figure 1. Here, terms with sentiment inflections are collected together with adjectives, adverbs, and non neutral verbs to specify their sentiment value. For each aspect term we go back to its lemma and then we compute its Word2Vec representation.

As a consequence, the *Aspect Detector* module, shown in Figure 2, estimates the cosine similarity between all the aspect terms and the centroids of each aspect category stored in the *Aspect Category List*. In this way, we are able to assign each aspect word to a specific *Aspect Category* based on the highest similarity value, while we delete those terms whose similarity scores are lower than a certain threshold. As a result, we produce a list of *Aspect Terms* and related sentiment words assigned to specific *Aspect Categories*.

We are now able to build the *Aspect-Based Sentiment Embedding*, whose dimension is equal to the number of categories composing the *Aspect Categories* list. In fact, each position of the embedding vector matches with a specific *Aspect Category*, while the value is computed based on the sentiment score related with that aspect. To this scope, we implement a Lexicon-based Sentiment Score Extractor which takes advantage of the Sentiment Lexicon SentiWords, a high coverage resource containing roughly 155,000 English words associated with a sentiment score included between -1 and 1 [9], to retrieve the sentiment scores related to those words that carry the opinion of the writer. It follows that the overall sentiment value for an aspect is an aggregation of scores obtained for each aspect terms belonging to that aspect category. In our case, the aggregation function we adopted is the simple summing function that sums all the sentiment of adjectives, adverbs and non neutral verb related to each aspect terms belonging to one aspect category.

The Aspect-Based Sentiment Embedding built we this approach follows an intuition similar to the one that characterizes the one-hot encoding representation. While in the one-hot encoding embedding each word identifies a specific position in the vector, in our approach the aspect categories define the space of the embedding vectors. In this way, documents that treat similar aspects with a similar sentiment share the same emotional meaning.

5 EXPERIMENT

This section describes the in-vitro experiments that we have run in order to understand whether some music track features in the Spotify dataset, typically related to the feelings conveyed through songs, can be better inferred using a representation based on the sentiment instead of a classical based on semantics.

We have used a free dataset on Kaggle[1] that contains various types of information over more than 18,000 Spotify songs including artist, album, audio features (e.g. loudness), lyrics, the language in which they are written, genres, and sub-genres. The dataset contains many songs in a period from 1921 to 2020.

From the dataset we have selected only songs with lyrics in English, discarding the instrumental ones because not exploitable by our model. We have split the dataset using the following proportions: 60% for the training set, 20% for the validation set and 20% for the test set. In order to create a sentiment based representation for songs, the first step was to feed the training to the pipeline described in Section 3. In this way we were able to extract all the aspect terms from lyrics and to group them into aspect categories identifying which are the main aspects for the music domain. For this purpose we have adopted a Word2Vec representation of aspect terms exploiting the model trained on Google-News dataset[2] and then we have used the K-means algorithm to cluster the aspect terms into their categories, setting the number of clusters to 75. This choice has been made using the Elbow method [36] which demonstrated that the best number of clusters to model our aspect categories is not greater than 75.

Once the aspect categories have been defined, we have proceeded with the online step in order to estimate the Sentiment-Based Aspect Embedding for each song in the dataset.

As stated in section 4, song lyrics were given as input to the Rule-Based ATE module for extracting candidate *Aspect Terms* and related opinion words from text. For each candidate, we compute similarity scores with the centroids of all the *Aspect Categories*, keeping only those values greater than a certain threshold, whose value is empirically set to 0.4. The highest score identifies the cluster to which the candidate belongs. If all the similarities scores were discarded in the previous step, then the candidate is not a pertinent as domain *Aspect Term*. At the same time the related sentiment was calculated using the SentiWords sentiment lexicon. Following this approach we were able to represent a song as a vector of 75 elements in which the element at the position i represents the sentiment expressed in the text with respect to the i-th *Aspect Category*, according to the Definition 3.

The main objective of our work is to answer the following research question: *"What is the best representation for each Spotify feature? The semantic or the sentiment-based one?"* To this purpose, we decide to use Doc2Vec [19] as our baseline for the semantic-based representation of songs lyrics. For validating our hypothesis, we have trained regression models on both sentiment and semantic songs representations to predict the following six key-features in the Spotify dataset (for a total of 12 models), whose values are in a range from 0 to 1:

- **Valence** describes the musical positiveness conveyed by a track.
- **Mode** indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived.
- **Speechiness** detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value.
- **Acousticness** is a confidence measure of whether the track is acoustic
- **Liveness** detects the presence of an audience in the recording.
- **Danceability** describes how suitable a track is for dancing based on a combination of musical elements

For the sake of completeness we have trained also the regression models for Speechiness, Acousticness, Liveness and Danceability even if those variables are not related to the sentiment. In fact, Speechiness represents how much the track is wordy, while the others are associated to the technical audio characteristics. On the contrary, Valence and Mode are more linked to the sentiment conveyed by song texts. The Valence is the positiveness of a song, in the sense that tracks with high valence score sound more positive (happy, cheerful, euphoric), while tracks with low valence value sound more negative (sad, depressed, angry). The Mode is the likelihood of a song of being in a major or minor scale. Even though this description seems unrelated to sentiment, different studies demonstrated that minor scales are typically used for songs that convey a negative mood while major ones are used for songs that communicate a positive sentiment [18]. The model's architecture adopted for solving the regression task is a Fully Connected Neural Network with 3 hidden layers with ReLU activation and an output layer with linear activation function. In order to optimize the training and avoid overfitting, we used the early stopping strategy by monitoring the Mean Absolute Error (MAE) and Mean Squared Error (MSE). We have also trained a Doc2Vec model on the song lyrics to obtain those semantic-based representations to use as input in our model. Table 1 and Table 2 show the evaluated performances of our models in terms of MSE and MAE. Table 1 contains the results

[1]https://www.kaggle.com/imuhammad/audio-features-and-lyrics-of-spotify-songs
[2]https://drive.google.com/file/d/0B7XkCwpI5KDYNlNUTTlSS21pQmM/

Spotify Feature	Representation	MSE	MAE
Valence	doc2vec	0.0542	**0.1891**
	sentiment	**0.0530**	0.1920
Mode	doc2vec	0.2694	0.4686
	sentiment	**0.2507**	**0.4682**
Speechiness	doc2vec	**0.0082**	**0.0619**
	sentiment	0.0124	0.082
Acousticness	doc2vec	0.0574	**0.1618**
	sentiment	**0.0488**	**0.1618**
Liveness	doc2vec	0.0256	**0.1122**
	sentiment	**0.0234**	0.1140
Danceability	doc2vec	**0.0270**	**0.1295**
	sentiment	0.0307	0.1416

Table 1: Evaluation results of several model, one for each Spotify feature, trained on Sentiment Representations and the Doc2Vec embeddings, that achieve the best value of MAE in the validation set. Values in bold identify the best performing model on the related feature.

Spotify Feature	Representation	MSE	MAE
Valence	doc2vec	0.0542	**0.1891**
	sentiment	**0.0524**	0.1911
Mode	doc2vec	0.2597	0.4803
	sentiment	**0.2470**	**0.4725**
Speechiness	doc2vec	**0.0082**	**0.0619**
	sentiment	0.0129	0.0741
Acousticness	doc2vec	**0.0508**	**0.1632**
	sentiment	0.0517	0.1907
Liveness	doc2vec	0.0256	**0.1122**
	sentiment	**0.0235**	0.1158
Danceability	doc2vec	**0.0268**	**0.1287**
	sentiment	0.0307	0.1423

Table 2: Evaluation results of several model, one for each Spotify feature, trained on Sentiment Representations and the Doc2Vec embeddings, that achieve the best value of MSE in the validation set. Values in bold identify the best performing model on the related feature.

that we reached with both our semantic-based embeddings and the doc2vec representation with the model that achieved the lowest value of MAE on the validation set. Similarly, Table 2 shows the differences in performance of the models obtained with the lowest value of MSE on the validation set.

The results presented in the tables confirm our hypothesis about the target variables. The MSE smaller than 0.0018 for the Valence with the Sentiment-Based embeddings demonstrates that the model commits less consistent errors with respect to the Doc2Vec approach. As regards the Mode, we reach a proximity value for large errors equal to 0.137 on average. This when the sentiment embedding is used, reflecting improved results also on the MAE. As a further prove of our intuition, the performances on Speechiness suggests that the semantic-based representation are more suitable in predicting targets which depend solely on textual features. The results for the other target variables seem to be quite unclear, which depends on the fact that these features are not connected to the lyrics of a song but they are computed on the audio characteristics.

We performed also a T-Student test for comparing the predicted values with the real targets and in all the settings we obtained a *p-value* < 0.01. Therefore, we can conclude that the presented results are statistically reliable.

6 CONCLUSIONS

In this work, we have introduced a new unsupervised approach for extracting sentiment aspects from documents on a specific domain by detecting *Aspect Terms* and finding *Aspect Categories* which they belong to. Furthermore, we have formally defined *Aspect Term* and *Aspect Category* highlighting their differences, as well as the *Aspect-Based Sentiment Embedding*, a new kind of document representation based on sentiment values computed for each *Aspect Category*. We have also provided a preliminary use-case of our sentiment representation technique, motivating its benefits by performing an in-vitro experimental evaluation on the Spotify's features regression task. We have compared our results with those obtained from the usage of semantic-based embeddings (i.e. Doc2Vec) on Spotify songs lyrics, and we have demonstrated little improvements on the Mode and Valence features prediction.

To statistically validate our experiments, we have also performed a T-Student test on the predicted data that confirms the reached results. Despite the basic configuration of our method, we have proved that a sentiment representation could be more informative than semantic one in a sentiment related task.

As Future Work, we plan to improve performances by adopting different clustering methods either choosing for a more complex aggregation function to compute the aspect sentiment value.

Moreover, we will perform an in depth analysis of Aspect Categories retrieved by several clustering algorithms, testing the utility of the Aspect-Based Sentiment Embeddings on more domains and with more baselines, as well as find new applications on which our Sentiment representation could be useful.

REFERENCES

[1] Pedro Álvarez, A. Guiu, José Ramón Beltrán, J. García de Quirós, and Sandra Baldassarri. 2019. DJ-Running: An Emotion-based System for Recommending Spotify Songs to Runners. In *Proceedings of the 7th International Conference on Sport Sciences Research and Technology Support, icSPORTS 2019, Vienna, Austria, September 20-21, 2019*, João Vilas-Boas, Pedro Pezarat Correia, and Jan Cabri (Eds.). ScitePress, 55–63.

[2] Mohammad Ehsan Basiri, Shahla Nemati, Moloud Abdar, Erik Cambria, and U. Rajendra Acharya. 2021. ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for sentiment analysis. *Future Gener. Comput. Syst.* 115 (2021), 279–294.

[3] Ranjan Kumar Behera, Monalisa Jena, Santanu Kumar Rath, and Sanjay Misra. 2021. Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data. *Inf. Process. Manag.* 58, 1 (2021), 102435.

[4] Ganpat Singh Chauhan, Yogesh Kumar Meena, Dinesh Gopalani, and Ravi Nahta. 2020. A two-step hybrid unsupervised model with attention mechanism for aspect extraction. *Expert Syst. Appl.* 161 (2020), 113673.

[5] Peng Chen, Zhongqian Sun, Lidong Bing, and Wei Yang. 2017. Recurrent Attention Network on Memory for Aspect Sentiment Analysis. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, Martha Palmer, Rebecca Hwa, and Sebastian Riedel (Eds.). Association for Computational Linguistics, 452–461.

[6] Zhiyuan Chen, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malú Castellanos, and Riddhiman Ghosh. 2013. Exploiting Domain Knowledge in Aspect Extraction. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*. ACL, 1655–1667.

[7] Jinhyuck Choi, Jin-Hee Song, and Yanggon Kim. 2018. An Analysis of Music Lyrics by Measuring the Distance of Emotion and Sentiment. In *19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2018, Busan, Korea (South), June 27-29, 2018*. IEEE Computer Society, 176–181.

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, Jill Burstein, Christy Doran, and Thamar Solorio (Eds.). Association for Computational Linguistics, 4171–4186.

[9] Lorenzo Gatti, Marco Guerini, and Marco Turchi. 2016. SentiWords: Deriving a High Precision and High Coverage Lexicon for Sentiment Analysis. *IEEE Trans. Affect. Comput.* 7, 4 (2016), 409–421.

[10] Athanasios Giannakopoulos, Claudiu Musat, Andreea Hossmann, and Michael Baeriswyl. 2017. Unsupervised Aspect Term Extraction with B-LSTM & CRF using Automatically Labelled Datasets. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, WASSA@EMNLP 2017, Copenhagen, Denmark, September 8, 2017*, Alexandra Balahur, Saif M. Mohammad, and Erik van der Goot (Eds.). Association for Computational Linguistics, 180–188.

[11] Abdalraouf Hassan and Ausif Mahmood. 2018. Convolutional Recurrent Deep Learning Model for Sentence Classification. *IEEE Access* 6 (2018), 13949–13957.

[12] Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2017. An Unsupervised Neural Attention Model for Aspect Extraction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, Regina Barzilay and Min-Yen Kan (Eds.). Association for Computational Linguistics, 388–397.

[13] Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, August 22-25, 2004*, Won Kim, Ron Kohavi, Johannes Gehrke, and William DuMouchel (Eds.). ACM, 168–177.

[14] Niklas Jakob and Iryna Gurevych. 2010. 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, EMNLP 2010, 9-11 October 2010, MIT Stata Center, Massachusetts, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*. ACL, 1035–1045.

[15] Rie Johnson and Tong Zhang. 2015. Effective Use of Word Order for Text Categorization with Convolutional Neural Networks. In *NAACL*.

[16] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A Convolutional Neural Network for Modelling Sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Baltimore, Maryland, 655–665.

[17] Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, 1746–1751.

[18] Artemy Kolchinsky, Nakul Dhande, Kengjeun Park, and Yong-Yeol Ahn. 2017. The Minor Fall, the Major Lift: Inferring Emotional Valence of Musical Chords through Lyrics. *CoRR* abs/1706.08609 (2017). arXiv:1706.08609

[19] Quoc V. Le and Tomás Mikolov. 2014. Distributed Representations of Sentences and Documents. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014 (JMLR Workshop and Conference Proceedings)*, Vol. 32. JMLR.org, 1188–1196.

[20] Xin Li and Wai Lam. 2017. Deep Multi-Task Learning for Aspect Term Extraction with Memory Interaction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, Martha Palmer, Rebecca Hwa, and Sebastian Riedel (Eds.). Association for Computational Linguistics, 2886–2892.

[21] Qian Liu, Zhiqiang Gao, Bing Liu, and Yuanlin Zhang. 2015. Automated Rule Selection for Aspect Extraction in Opinion Mining. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, Qiang Yang and Michael J. Wooldridge (Eds.). AAAI Press, 1291–1297.

[22] Qian Liu, Bing Liu, Yuanlin Zhang, Doo Soon Kim, and Zhiqiang Gao. 2016. Improving Opinion Aspect Extraction Using Semantic Similarity and Aspect Associations. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*, Dale Schuurmans and Michael P. Wellman (Eds.). AAAI Press, 2986–2992.

[23] Huaishao Luo, Tianrui Li, Bing Liu, Bin Wang, and Herwig Unger. 2019. Improving Aspect Term Extraction With Bidirectional Dependency Tree Representation. *IEEE ACM Trans. Audio Speech Lang. Process.* 27, 7 (2019), 1201–1212.

[24] J. MacQueen. 1967. Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability - Vol. 1*, L. M. Le Cam and J. Neyman (Eds.). University of California Press, Berkeley, CA, USA, 281–297.

[25] Tomás Mikolov, Quoc V. Le, and Ilya Sutskever. 2013. Exploiting Similarities among Languages for Machine Translation. *CoRR* abs/1309.4168 (2013). arXiv:1309.4168

[26] Martijn Millecamp, Nyi Nyi Htun, Yucheng Jin, and Katrien Verbert. 2018. Controlling Spotify Recommendations: Effects of Personal Characteristics on Music Recommender User Interfaces. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization, UMAP 2018, Singapore, July 08-11, 2018*, Tanja Mitrovic, Jie Zhang, Li Chen, and David Chin (Eds.). ACM, 101–109.

[27] Tetsuya Nasukawa and Jeonghee Yi. 2003. Sentiment analysis: capturing favorability using natural language processing. In *Proceedings of the 2nd International Conference on Knowledge Capture (K-CAP 2003), October 23-25, 2003, Sanibel Island, FL, USA*, John H. Gennari, Bruce W. Porter, and Yolanda Gil (Eds.). ACM, 70–77.

[28] Paolo Pastore, Andrea Iovine, Fedelucio Narducci, and Giovanni Semeraro. 2021. A General Aspect-Term-Extraction Model for Multi-Criteria Recommendations (Long paper). In *Joint Workshop Proceedings of the 3rd Edition of Knowledge-aware and Conversational Recommender Systems (KaRS) and the 5th Edition of Recommendation in Complex Environments (ComplexRec) co-located with 15th ACM Conference on Recommender Systems (RecSys 2021), Virtual Event, Amsterdam, The Netherlands, September 25, 2021 (CEUR Workshop Proceedings)*, Vito Walter Anelli, Pierpaolo Basile, Tommaso Di Noia, Francesco Maria Donini, Cataldo Musto, Fedelucio Narducci, Markus Zanker, Himan Abdollahpouri, Toine Bogers, Bamshad Mobasher, Casper Petersen, and Maria Soledad Pera (Eds.), Vol. 2960. CEUR-WS.org.

[29] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. *CoRR* abs/1802.05365 (2018). arXiv:1802.05365

[30] Marco Polignano, Pierpaolo Basile, Marco de Gemmis, Giovanni Semeraro, and Valerio Basile. 2019. AlBERTo: Italian BERT Language Understanding Model for NLP Challenging Tasks Based on Tweets. In *Proceedings of the Sixth Italian Conference on Computational Linguistics, Bari, Italy, November 13-15, 2019 (CEUR Workshop Proceedings)*, Raffaella Bernardi, Roberto Navigli, and Giovanni Semeraro (Eds.), Vol. 2481. CEUR-WS.org.

[31] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia V. Loukachevitch, Evgeniy V. Kotelnikov, Núria Bel, Salud María Jiménez Zafra, and Gülsen Eryigit. 2016. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2016, San Diego, CA, USA, June 16-17, 2016*, Steven Bethard, Daniel M. Cer, Marine Carpuat, David Jurgens, Preslav Nakov, and Torsten Zesch (Eds.). The Association for Computer Linguistics, 19–30.

[32] Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation, SemEval@COLING 2014, Dublin, Ireland, August 23-24, 2014*, Preslav Nakov and Torsten Zesch (Eds.). The Association for Computer Linguistics, 27–35.

[33] Soujanya Poria, Erik Cambria, and Alexander F. Gelbukh. 2016. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowl. Based Syst.* 108 (2016), 42–49.

[34] Marco Pota, Mirko Ventura, Rosario Catelli, and Massimo Esposito. 2021. An Effective BERT-Based Pipeline for Twitter Sentiment Analysis: A Case Study in Italian. *Sensors* 21, 1 (2021), 133.

[35] Minchae Song, Hyunjung Park, and Kyung-shik Shin. 2019. Attention-based long short-term memory network using sentiment lexicon embedding for aspect-level sentiment analysis in Korean. *Inf. Process. Manag.* 56, 3 (2019), 637–653.

[36] Robert L Thorndike. 1953. Who belongs in the family? *Psychometrika* 18, 4 (1953), 267–276.

[37] Danny Suarez Vargas, Lucas Rafael Costella Pessutto, and Viviane Pereira Moreira. 2020. Simple Unsupervised Similarity-Based Aspect Extraction. *CoRR* abs/2008.10820 (2020). arXiv:2008.10820

[38] Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016. Attention-based LSTM for Aspect-level Sentiment Classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, Jian Su, Xavier Carreras, and Kevin Duh (Eds.). The Association for Computational Linguistics, 606–615.

[39] Chuhan Wu, Fangzhao Wu, Sixing Wu, Zhigang Yuan, and Yongfeng Huang. 2018. A hybrid unsupervised method for aspect term and opinion target extraction. *Knowl. Based Syst.* 148 (2018), 66–73.

[40] Hai Ye, Zichao Yan, Zhunchen Luo, and Wen-Han Chao. 2017. Dependency-Tree Based Convolutional Neural Networks for Aspect Term Extraction. In *Advances in Knowledge Discovery and Data Mining - 21st Pacific-Asia Conference, PAKDD 2017, Jeju, South Korea, May 23-26, 2017, Proceedings, Part II (Lecture Notes in Computer Science)*, Jinho Kim, Kyuseok Shim, Longbing Cao, Jae-Gil Lee, Xuemin Lin, and Yang-Sae Moon (Eds.), Vol. 10235. 350–362.