What do you like in movies? Aspect-based sentiment analysis for movie reviews

Thomas Mayr

Abstract

This report investigates aspect-based sentiment analysis (ABSA) in movie reviews, focusing on filmmaking aspects such as acting and cinematography. Using a dataset of 83,530 reviews from Letterboxd, the study employs a pipeline combining Word2Vec-based aspect detection, associating aspects with sentiments with SpaCy's dependency parser and employing VADER for sentiment analysis. While effective at capturing aspect-specific sentiments, the approach highlights challenges and short-comings of a mainly rule-based approach to ABSA.

1 Introduction

This project explores the field of aspect-based sentiment analysis (ABSA), aiming to analyze a dataset of movie reviews from the popular movie platform Letterboxd. The objective was to determine the sentiment associated with various aspects of filmmaking, such as 'acting' and 'cinematography.' The project comprises two main components: an aspect detection mechanism and a sentiment classification system. The aspect detection mechanism identifies references to predefined aspects within the reviews, while the sentiment classifier determines the polarity and intensity of the associated opinions.

Aspect-based sentiment analysis (ABSA) is an active and growing field within natural language processing, focusing on the fine-grained analysis of opinions toward specific aspects of an entity [1]. ABSA encompasses two primary sub-tasks: aspect extraction and sentiment analysis. Aspect extraction aims to identify and classify aspects, whether explicit (directly mentioned) or implicit (implied), which often presents challenges such as ambiguity and domain dependency. Sentiment analysis then is concerned with assigning polarity scores and classifying sentiments, characterized by the need to process complex interactions and dependencies between aspects, sentiments, and context.

A popular family of techniques in ABSA are rule-based approaches which rely on predefined linguistic or domain-specific rules to identify and extract aspects or sentiments from the text [2]. These approaches use explicit patterns, heuristics, or dependency parsing techniques to process and analyze text data in contrast to other approaches which often use elaborate statistical or machine learning models. The approach used in this project is mainly based on such rule-based techniques, but incorporates the benefits of a machine learning algorithm for the aspect extraction phase.

Key challenges addressed in this project include the reliable detection of aspects, the accurate association of sentiments with those aspects, and the selection of techniques for assigning sentiment polarities. After careful consideration a semantic similarity-based approach leveraging Word2Vec was ultimately chosen to create a vocabulary for each aspect label. This vocabulary was subsequently used for aspect detection, with sentiments extracted using SpaCy and analyzed for polarity using the rule-based sentiment analysis tool VADER.

2 Research Question and Methodology

The primary objective of this project is to analyze user-generated movie reviews to understand the sentiment polarity associated with specific predefined aspects of movies. By focusing on aspects such as acting or story, this project seeks to identify sentiments (positive, negative, or neutral) and opinions linked to these aspects.

The research question driving this project is:

How can aspects be accurately detected in movie reviews, and how can aspect-specific sentiments be extracted and mapped to such aspects?

3 Experimental Results

3.1 Data

The data used in this project was sourced from Kaggle[1] and contains 83530 reviews (after preprocessing) from popular movie platform Letterboxd (www.letterboxd.com) including the review itself, a rating of up to five stars and the name of the movie. The data features the 250 top rated movies on Letterboxd with around 360 reviews for each film.

3.2 Aspect Detection

Aspect detection is a critical task in ABSA, involving the identification of relevant aspects within textual data. Unlike traditional aspect extraction, which is mostly concerned with discovering a priori unknown aspects, this project focuses on predefined, well established aspects fundamental to filmmaking. Five aspects were selected for analysis: **directing, story, acting, cinematography, and music**. These aspects were chosen based on their clarity, broad relevance, and distinctiveness to minimize overlap.

Detecting an aspect solely by its direct mention (e.g., 'acting') would be insufficient, therefore, related terms were included to capture broader references to each

[1] https://www.kaggle.com/datasets/riyosha/letterboxd-movie-reviews-90000

aspect. Various approaches, described below, for building these aspect vocabularies were considered.

3.2.1 WordNet

Ontology-based methods leverage external lexical resources to capture semantic relations. WordNet, a comprehensive lexical database of English, was evaluated for this purpose.

WordNet organizes words into synonym sets (synsets) with hierarchical relationships such as hypernyms (broader terms) and hyponyms (narrower terms). This structure is theoretically well-suited for discovering semantically related terms. For example, the synset 'filming.n.01' corresponds to 'cinematography,' defined as 'the act of making a film'.

Related terms were identified by exploring synonyms, hypernyms, hyponyms, and other semantic relations. Despite this, the approach often resulted in vocabularies that were either too narrow (missing key terms) or too broad (including loosely related words). Therefore, this method would have required extensive manual curation to balance false positives and negatives, limiting its scalability and effectiveness. Consequently, relying solely on WordNet was deemed impractical.

3.2.2 Word2Vec

Since the ontology-based WordNet approach did not yield satisfactory results, we instead pursued a distributional semantic similarity approach using Word2Vec. Word2Vec is a neural embedding algorithm that learns dense vector representations of words, capturing both semantic and syntactic relationships based on the contexts in which words appear.

In contrast to WordNet's ontology-based structure—which may not be well-tailored for a specific dataset—Word2Vec can be trained directly on our corpus of movie reviews, ensuring that the learned word similarities reflect real usage in that domain.

The model was trained on the review dataset, and for each aspect, the top 25 most similar words were identified using cosine similarity. Counts of these words within the dataset were also calculated to assess their relevance.

This approach produced more satisfactory vocabularies compared to the ontology-based WordNet method. However, the automatically generated candidate word sets still required some manual refinement. In particular, many terms that had strong semantic relations to the domain of movies in general, but not to the respective aspect had to be pruned. Moreover, sometimes crucial terms were missing from candidate sets and were added manually (done sparingly to avoid overly distorted results). For instance, for the aspect 'directing,' essential terms like 'director' and 'filmmaker' were manually added to the final vocabulary, as they were not among the top candidates but are clearly central to the concept. For the the final vocabularies and corresponding word distributions see figure 1 and table 1 in the appendix.

While ultimately effective, this method's reliance on manual curation limits its generalizability and reproducibility for other datasets. In section 4, we discuss how more sophisticated approaches might be able to retain or even increase accuracy without extensive manual intervention.

Fig. 1 Distribution of aspect word frequencies

3.3 Aspect-Sentiment Analysis

The strategy chosen for the analysis of aspect-specific sentiments in the reviews combines linguistic dependency parsing with rule-based sentiment analysis. The implemented pipeline consists of two stages: opinion phrase extraction, and sentiment classification.

3.3.1 Opinion Phrase Extraction

First, we leveraged the final aspect vocabulary obtained using the Word2Vec approach, described in the preceding section, to identify tokens corresponding to each aspect. SpaCy's dependency parser was used to process sentences, enabling token-level analysis. Opinion phrases surrounding aspect tokens were extracted using linguistic dependency patterns. The goal was to identify modifiers (e.g., adjectives and adverbs) and negations directly linked to aspect tokens. For instance, if the aspect token 'acting' appears in the sentence, phrases like 'superb acting' or 'not convincing acting' are extracted. Specific patterns, such as adjectival complements of linking verbs (e.g., 'is excellent'), were also considered to ensure nuanced sentiment capture.

3.3.2 Sentiment Classification

Extracted opinion phrases were analyzed using the VADER sentiment analysis tool, which provides polarity scores (positive, negative, neutral) and a compound score representing overall sentiment intensity. VADER's rule-based approach is particularly effective for short texts and incorporates heuristics for negations and intensifiers. Polarity labels were assigned based on the compound score thresholds: positive (≥ 0.05), negative (≤ -0.05), and neutral (otherwise). For reviews containing multiple mentions of the same aspect, an aggregation mechanism was applied. Compound scores were averaged to determine overall sentiment, with precedence rules ensuring that negative sentiments were not overshadowed by neutral or positive scores.

To demonstrate the short-comings of our model we examine the output of our sentiment analyzer on a sentence level for two sentences from a review for the film "There Will Be Blood":

Sentence: *A brilliantly bleak portrait of capitalism, There Will Be Blood uses perfect performances to demonstrate that greed corrupts everyone, regardless of nobility or intentions.*

Aspect Sentiments: {*'acting': ('positive', 0.5719)*}

Here we see a classic example of the sentiment association working as intended. It correctly picks up the acting aspect and associates a strong positive polarity with it (*perfect performances*).

Sentence: *Daniel Day Lewis' gradual descent into madness is the obvious highlight here, but I'm equally impressed with Paul Dano's dual performances, as Eli falls farther than anyone else in this film, and is believable and emotional at every step of his journey.*

Aspect Sentiments: {*'acting': ('neutral', 0.0)*}

In this example we see a more complicated aspect sentiment structure and the detector fails to associate the obvious positive sentiment with the acting aspect and instead falsely outputs a neutral sentiment. This exemplifies the need for a more refined sentiment association process.

Moving to the general results of our analysis, figure 2 illustrates the counts of detected aspects at the review level, where multiple mentions of the same aspect in a review are counted as one. The aspects story and acting dominate, appearing in approximately a quarter of all reviews ($n = 83,530$) at least once. The other aspects, direction, music, and cinematography are each detected in nearly 10,000 reviews.

In figure 3, which displays the distribution of sentiment scores by aspect, several aspects (story, direction, and music) show sentiment scores tightly clustered around 0. This indicates a high frequency of neutral mentions. Although this clustering does not necessarily imply issues with our detection or sentiment association process, it may warrant further investigation to see if the majority of these are actually neutral or if we fail to detect associated sentiments like in the example sentence.

When examining the broader sentiment distribution by category ("positive," "neutral," "negative") in figure 4, we again observe that neutral sentiments are most prevalent, followed by positive sentiments. This skew toward positive sentiment is expected, given that the dataset only features highly rated movies. To evaluate whether our sentiment association process performs equally well for both positive and negative sentiment, testing on a dataset with a more balanced review distribution would be necessary.

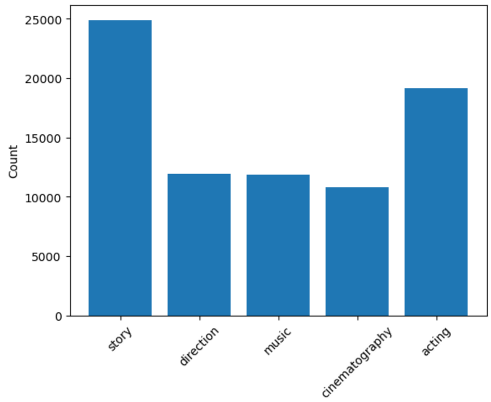

Fig. 2 Distribution of aspect mentions per review

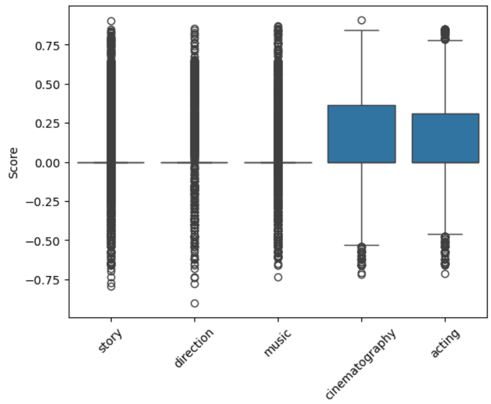

Fig. 3 Distribution of sentiment scores by aspect

4 Concluding Remarks

The experimental approach successfully captured aspects and associated sentiments, leveraging both syntactic relationships and sentiment heuristics. However, it has limitations:

- Dependency on manual curation of aspect dictionaries may hinder scalability and introduces a somewhat arbitrary and subjective element to the process.
- Rule-based sentiment analysis might struggle with highly contextual language or sarcasm.
- Averaging compound scores may oversimplify complex sentiments expressed in multifaceted reviews.

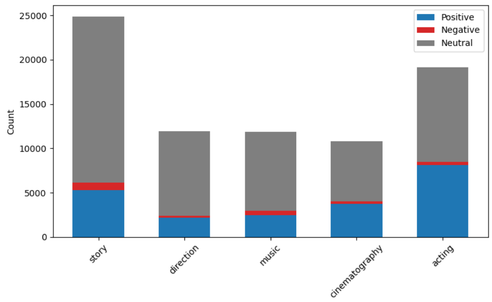

Fig. 4 Distribution of sentiment categories by aspect

Building on this approach, we could potentially improve upon the aspect detection mechanism by combining multiple methods of word embedding (see [3], [4]) or developing a way to synthesize word sets found with word embedding and the ontology-based approach [5]. Doing so could potentially remove the need for any manual intervention.

Furthermore, by closely examining the results of the opinion phrase extraction and sentiment association phase, we might find ways to capture more elaborate dependency patterns between aspect tokens and sentiments.

In addition, to systematically improve upon the methods used here, a reliable set of metrics should be developed to compare and evaluate models. A comparison between the results of this rule-based sentiment analysis and a more sophisticated transformer-based model like BERT could also provide meaningful insights.

With regard to practical use cases, future work could use the capabilities of this model and integrate external information about the characteristics of movies in question and try to discover aspect sentiment patterns for specific genres or time periods of movies. We might also try to combine results with a film recommendation system, e.g. let users search for "movies with strong acting and great cinematography".

References

[1] Schouten, K., Frasincar, F.: Survey on aspect-level sentiment analysis. IEEE Transactions on Knowledge and Data Engineering **28**(3), 813–830 (2016) https://doi.org/10.1109/TKDE.2015.2485209

[2] Rana, T.A., Cheah, Y.-N.: Aspect extraction in sentiment analysis: comparative analysis and survey. Artificial Intelligence Review **46**(4), 459–483 (2016) https://doi.org/10.1007/s10462-016-9472-z

[3] Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543 (2014)

[4] Pham, D.-H., Le, A.-C.: Exploiting multiple word embeddings and one-hot character vectors for aspect-based sentiment analysis. International Journal of Approximate Reasoning **103**, 1–10 (2018) https://doi.org/10.1016/j.ijar.2018.08.

003

[5] De Kok, S., Frasincar, F.: Using word embeddings for ontology-driven aspect-based sentiment analysis. In: Proceedings of the 35th Annual ACM Symposium on Applied Computing, pp. 834–842 (2020)

Appendix

Aspect	Word	Frequency	Total
acting	performance	8338	36042
	performances	7668	
	cast	4337	
	actors	3676	
	chemistry	867	
	ensemble	579	
	casting	556	
	actor	3066	
	actress	1355	
	performer	173	
	acting	5427	
cinematography	visuals	1809	14798
	imagery	1345	
	lighting	1102	
	camerawork	962	
	framing	879	
	photography	701	
	composition	645	
	scenery	329	
	choreography	297	
	cinematography	6729	
direction	filmmaking	3094	16862
	directing	1231	
	director	7047	
	filmmaker	1772	
	direction	3718	
music	score	5365	19686
	sound	3732	
	soundtrack	1672	
	song	1424	
	musical	1301	
	songs	494	
	piano	447	
	music	5251	
story	narrative	5256	50834
	plot	5241	
	script	2804	
	storytelling	2532	
	screenplay	2010	
	storyline	555	
	dialogue	4476	
	writing	3202	
	dialogues	593	
	narration	560	
	plotline	71	
	story	23534	

Table 1 Vocabulary words, individual frequencies, and total frequencies per aspect.