

Aspect-Based Sentiment Analysis: Half the Size, Same Performance?

Fynn Madrian
Hochschule Ansbach
`f.madrian18230@hs-ansbach.de`

July 25, 2024

Abstract

Diese Arbeit untersucht die Bedeutung und die Fortschritte der aspektbasierten Sentiment-Analyse (ABSA), einem spezialisierten Bereich der Sentiment-Analyse, der sich auf die Identifikation und Bewertung von Stimmungen in Bezug auf spezifische Aspekte von Produkten oder Dienstleistungen konzentriert. Im Gegensatz zur traditionellen Sentiment-Analyse bietet ABSA detaillierte Einblicke, indem es Kundenfeedback auf Attributebene analysiert, was für gezielte Verbesserungen und das Verständnis präziser Kundenbedürfnisse von unschätzbarem Wert ist. Die Ergebnisse zeigen die Machbarkeit dieser kompakten Modelle bei der Aufrechterhaltung hoher Leistung, insbesondere in Echtzeitanwendungen und Bereichen mit begrenzten Rechenressourcen. Diese Arbeit diskutiert auch die breiteren Implikationen von ABSA in verschiedenen Branchen und das Potenzial für zukünftige Forschungen in multimodaler Sentiment-Analyse und mehrsprachigen Anwendungen.

Inhaltsverzeichnis

1	Einleitung	3
2	Stand der Forschung	4
3	Mini-Modell	4
4	Methodologie	5
5	Resultate	9
6	Diskussion	9
7	Fazit und Ausblick	10

1 Einleitung

In der heutigen digitalen Welt spielen Meinungen und Bewertungen eine zentrale Rolle. Unternehmen, die ihre Produkte und Dienstleistungen verbessern möchten, sind auf das Feedback ihrer Kunden angewiesen. Herkömmliche Methoden der Sentiment-Analyse, die sich auf die allgemeine Stimmung von Texten konzentrieren, stoßen jedoch oft an ihre Grenzen. Hier setzt die Aspect-Based Sentiment Analysis (ABSA) an, eine spezialisierte Methode der Sentiment-Analyse, die es ermöglicht, Meinungen und Gefühle auf der Ebene einzelner Aspekte oder Attribute zu identifizieren und zu bewerten.

Aspect-Based Sentiment Analysis unterscheidet sich von herkömmlichen Ansätzen, indem sie nicht nur die allgemeine Stimmung eines Textes erfasst, sondern spezifische Elemente und Attribute eines Produkts oder einer Dienstleistung untersucht. Diese differenzierte Betrachtungsweise ist besonders wertvoll, um gezielte Verbesserungen vorzunehmen und spezifische Kundenbedürfnisse zu identifizieren. ABSA nutzt fortschrittliche Techniken des maschinellen Lernens und der natürlichen Sprachverarbeitung, um relevante Aspekte in Texten zu erkennen und die damit verbundenen Stimmungen zu analysieren.

Die Anwendungsbereiche von ABSA sind vielfältig und reichen von der Analyse von Produktbewertungen und Kundenfeedback über die Untersuchung von Social-Media-Beiträgen bis hin zur Überwachung der Markenwahrnehmung. In der Geschäftswelt ermöglicht ABSA Unternehmen, präzise Einblicke in die Meinung ihrer Kunden zu gewinnen und gezielte Marketingstrategien zu entwickeln. Im Gesundheitswesen kann sie dazu beitragen, Patientenfeedback zu analysieren und die Qualität der Versorgung zu verbessern. Auch im politischen Bereich findet ABSA Anwendung, um die öffentliche Meinung zu politischen Themen und Persönlichkeiten zu untersuchen.

Dieser Bericht beleuchtet die Grundlagen der Aspect-Based Sentiment Analysis, stellt die verschiedenen Techniken und Methoden vor und gibt einen umfassenden Überblick über deren vielseitige Anwendungsbereiche. Ziel ist es, die Bedeutung und den Nutzen von ABSA in verschiedenen Branchen zu verdeutlichen und Einblicke in die zukünftigen Entwicklungen dieser innovativen Technologie zu geben.

2 Stand der Forschung

Die Forschung im Bereich der ABSA hat bedeutende Fortschritte gemacht, insbesondere durch die Anwendung von tiefen neuronalen Netzen und Transformer-Modellen wie BERT und RoBERTa. Diese Modelle haben die Genauigkeit und Effizienz der Sentiment-Analyse auf Aspekt-Ebene erheblich verbessert. Aktuelle Ansätze kombinieren oft verschiedene Techniken, darunter maschinelles Lernen, natürliche Sprachverarbeitung und Transfer-Learning, um präzisere Ergebnisse zu erzielen. Benchmark-Datensätze wie SemEval-2014 Task 4 haben als Standard für die Bewertung neuer Methoden gedient und kontinuierlich zur Entwicklung und Verfeinerung der Modelle beigetragen.

Trotz der Fortschritte gibt es noch mehrere Untergebiete in der ABSA, die wenig erforscht sind. Dazu gehören die Multimodale Sentiment-Analyse, bei der neben Text auch visuelle und auditive Daten einbezogen werden, sowie die Domänenadaptation, die sich mit der Anpassung von Modellen an unterschiedliche Anwendungsbereiche und Branchen beschäftigt. Ein weiteres unerforschtes Gebiet ist die Analyse von mehrsprachigen Texten und die Entwicklung von Modellen, die Sentiment-Analysen über verschiedene Sprachen hinweg durchführen können. Diese Bereiche bieten viel Potenzial für zukünftige Forschungen und Anwendungen.

3 Mini-Modell

Moderne KI-Modelle wie BERT und GPT-3 sind leistungsstark, aber auch ressourcenintensiv und erfordern große Rechenkapazitäten. In ressourcenbeschränkten Umgebungen oder Echtzeit-Anwendungen können solche Modelle jedoch ineffizient sein. Eine mögliche Lösung für dieses Problem ist die Entwicklung von Mini-Modellen, die eine ähnliche Leistung wie ihre größeren Gegenstücke bieten, aber mit weniger Rechenressourcen auskommen.

Besonders im Bereich Sprachverarbeitung sind besonders große Modelle notwendig, um optimale Ergebnisse zu erzielen. Die Entwicklung von Mini-Modellen, die die Effizienz und Leistung von KI-Modellen verbessern, ist daher ein vielversprechender Ansatz, um die Anwendung von KI in verschiedenen Bereichen zu fördern.

ChatGPT 3 benötigt 4,6 kWh pro Inference, was in etwa dem Energieverbrauch eines durchschnittlichen Haushalts in den USA für 3 Stunden entspricht. Die State-of-the-Art-Modelle im Bereich der Sentiment Analyse basieren ebenfalls häufig auf einer Transformer-Architektur, die eine hohe Rechenkapazität erfordert.

Im Vergleich zur vollen Sprachsynthese können in der Sentiment-Analyse aufgrund der geringeren Komplexität des Problems kleinere Modelle eingesetzt werden. Die Entwicklung von effizienten Mini-Modellen für die Sentiment-Analyse auf Aspekt-Ebene ist daher ein vielversprechender Ansatz, um die Leistung und Effizienz von KI-Modellen zu verbessern.

Wenig Erforschter aber wichtiger Punkt: Effizienz von KI-Modellen. Lösung: Kleineres Modell, ähnliche Leistung.

4 Methodologie

Verwendeter Datensatz: SemEval-2014: Task4 Der SemEval-2014 Task 4 Datensatz ist ein Benchmark-Datensatz, der speziell für die Aspect-Based Sentiment Analysis (ABSA) entwickelt wurde. Dieser Datensatz stammt aus dem SemEval (Semantic Evaluation), einer jährlichen Reihe von Natural Language Processing (NLP) Wettbewerben, die von der Association for Computational Linguistics (ACL) organisiert werden. Task 4 im Jahr 2014 konzentrierte sich auf die Sentiment-Analyse von Kundenbewertungen und umfasste zwei Hauptdomänen: Restaurants und Laptops. Die Daten bestehen aus einer Vielzahl von Kundenrezensionen, die manuell annotiert wurden, um sowohl die relevanten Aspekte (wie "Service" oder "Bildschirm") als auch die dazugehörigen Sentimente (positiv, negativ, neutral) zu kennzeichnen. Diese umfassende Annotierung ermöglicht es Forschern, Modelle zu trainieren und zu evaluieren, die spezifische Meinungen zu einzelnen Aspekten eines Produkts oder einer Dienstleistung erkennen und klassifizieren können. Der SemEval-2014 Task 4 Datensatz dient somit als Standardtest für die Entwicklung und Verfeinerung neuer Methoden in der ABSA und hat wesentlich zur Verbesserung der Modellgenauigkeit und Effizienz in diesem Bereich beigetragen.

Embedding: GloVe (Global Vectors for Word Representation) GloVe (Global Vectors for Word Representation) ist ein Modell zur Erzeugung von

Wortvektoren, das von Forschern der Stanford University entwickelt wurde. Es kombiniert die Vorteile der Count-basierten und Predict-basierten Methoden zur Wortvektorisierung, indem es globale Wortvorkommensstatistiken aus einem Korpus nutzt. Der grundlegende Ansatz von GloVe besteht darin, eine große Matrix von Wortvorkommenshäufigkeiten zu erstellen, die darstellt, wie oft ein Wort in einem bestimmten Kontext erscheint. Diese Matrix wird dann in eine niedrigdimensionale Vektorrepräsentation zerlegt, wobei die entstandenen Vektoren die semantischen Beziehungen zwischen den Wörtern in hoher Genauigkeit abbilden. Diese Methode ermöglicht es, dass ähnliche Wörter in der Nähe voneinander in dem Vektorraum positioniert werden, wodurch semantische Ähnlichkeiten und Analogien effektiv erfasst werden können.

Ein wesentlicher Vorteil von GloVe ist seine Fähigkeit, sowohl statistische Informationen aus dem gesamten Korpus als auch lokale Kontextinformationen zu integrieren, was zu reichhaltigeren und kontextuell aussagekräftigeren Wortvektoren führt. Im Gegensatz zu reinen Kontextfenster-basierten Modellen wie Word2Vec, das auf lokale Kontextinformationen fokussiert ist, berücksichtigt GloVe die globale Struktur des Korpus. Dies führt dazu, dass GloVe-Vektoren in der Lage

sind, subtile semantische Beziehungen und Bedeutungsnuancen besser zu erfassen. Die Effizienz und Genauigkeit der GloVe-Vektoren haben dazu geführt, dass sie in vielen natürlichen Sprachverarbeitungsaufgaben wie Textklassifikation, maschineller Übersetzung und Named Entity Recognition weit verbreitet sind.

Bevor die Daten mit GloVe (Global Vectors for Word Representation) eingebettet werden, müssen sie tokenisiert und gepaddet werden, um sie für die Verarbeitung durch maschinelle Lernmodelle vorzubereiten.

Tokenisierung ist der Prozess, bei dem Text in kleinere Einheiten, sogenannte Token, zerlegt wird. Diese Token können Wörter, Satzzeichen oder andere bedeutungstragende Elemente sein. Der Hauptgrund für die Tokenisierung besteht darin, die natürliche Sprache in eine Form zu bringen, die von Algorithmen und Modellen verarbeitet werden kann. Durch die Tokenisierung werden komplexe Textstrukturen in handhabbare Teile zerlegt, was die Analyse und Modellierung vereinfacht. Zum Beispiel wird der Satz "The screen is great but I hate the battery" in die Token ["The", "screen", "is", "great", "but", "I", "hate", "the", "battery"] zerlegt.

Padding hingegen ist notwendig, um sicherzustellen, dass alle Sätze oder Textsequenzen in einem Datensatz die gleiche Länge haben. Maschinelle

Lernmodelle, insbesondere neuronale Netze, erfordern häufig Eingaben mit konstanter Länge. Padding fügt spezielle Tokens (oft Nullen oder andere Platzhalter) hinzu, um kürzere Sequenzen auf die maximale Länge im Datensatz zu bringen. Dies ermöglicht eine effiziente Stapelverarbeitung (Batch Processing) und verhindert, dass das Modell durch unterschiedlich lange Eingaben durcheinandergebracht wird. Beispielsweise wird der kürzere Satz "I hate the battery" auf die gleiche Länge wie "The screen is great but I hate the battery" gebracht, indem er mit Platzhaltern ergänzt wird.

Durch die Kombination von Tokenisierung und Padding wird der Text in ein einheitliches Format gebracht, das anschließend von GloVe verwendet werden kann, um Vektorrepräsentationen für jedes Token zu generieren. Diese Schritte sind entscheidend, um die rohen Textdaten in eine strukturierte und maschinenlesbare Form zu überführen, die für die nachfolgende Modellierung und Analyse geeignet ist. Der genaue Prozess der Tokenisierung führt allerdings dazu, dass die Ergebnisse nur mit weiterem Aufwand evaluiert werden können, da sich das Format leicht von den Labels unterscheidet. Weitere Details dazu im Abschnitt Resultate.

Anschließend werden die jeweiligen Paare and Aspects und Sentiments gefunden und diese in One-Hot-Encodings umgewandelt. Diese ermöglichen dem Modell die Klassifikation der Aspekte und Sentiments.

Es wurden verschiedene Modellarten trainiert und getestet, um die beste Kombination für die Aspect Extraction und Sentiment Classification zu finden.

****Erster Ansatz: Bidirektionales LSTM mit TimeDistributed Layer für ABSA****

Das bidirektionale Long Short-Term Memory (LSTM) Netzwerk mit einem TimeDistributed Layer eignet sich hervorragend für Aspect Based Sentiment Analysis (ABSA), da es die Fähigkeit besitzt, Kontextinformationen sowohl aus der Vergangenheit als auch aus der Zukunft zu erfassen. Diese duale Blickrichtung ist besonders nützlich, um die Bedeutung von Wörtern im Kontext ihrer benachbarten Begriffe vollständig zu verstehen. Der TimeDistributed Layer ermöglicht es, die LSTM-Ausgaben für jedes Zeitschritt separat zu verarbeiten und somit Aspekte in verschiedenen Textsegmenten effektiv zu identifizieren und zu bewerten. Der Vorteil dieses Ansatzes liegt in seiner Robustheit gegenüber längeren Sequenzen und der Fähigkeit, sequenzielle Abhängigkeiten präzise zu modellieren. Dies führt in der Regel zu einer höheren Genauigkeit bei der Erkennung von Aspekten und der Sentimentklassifikation.

****Zweiter Ansatz: Convolutional Neural Network (CNN) für ABSA****

Das Convolutional Neural Network (CNN) für ABSA nutzt die Fähigkeit von Faltungsschichten, lokale Muster zu erfassen, die für die Sentimentanalyse von entscheidender Bedeutung sind. Durch die Anwendung von Filtern auf den Eingabetext kann ein CNN nützliche Merkmale wie n-Gramme oder spezifische Wortmuster extrahieren, die auf bestimmte Aspekte hinweisen. Dies ermöglicht eine effiziente und schnelle Verarbeitung großer Textmengen. Ein wesentlicher Vorteil von CNNs ist ihre Fähigkeit, parallele Berechnungen durchzuführen, was zu einer schnelleren Trainings- und Inferenzzeit führt. Zudem sind CNNs weniger anfällig für Überanpassung bei großen Datenmengen und können durch reguläre Techniken wie Dropout und Batch-Normalisierung weiter optimiert werden. Dies führt oft zu verbesserten Ergebnissen in Bezug auf die Genauigkeit und Generalisierbarkeit des Modells.

****Dritter Ansatz: LSTM + Attention für ABSA****

Die Kombination von LSTM und Attention-Mechanismen für ABSA stellt eine fortschrittliche Herangehensweise dar, die die Stärken beider Technologien nutzt. Während das LSTM die Sequenzinformationen und Kontextabhängigkeiten effektiv modelliert, ermöglicht die Attention-Schicht dem Modell, sich selektiv auf relevante Teile des Eingabetextes zu konzentrieren. Dies ist besonders nützlich, wenn es darum geht, spezifische Aspekte innerhalb eines langen Textes zu identifizieren und deren Sentiment präzise zu bestimmen. Der Vorteil dieses Ansatzes liegt in der verbesserten Fähigkeit des Modells, kontextuelle Abhängigkeiten über lange Distanzen hinweg zu erfassen und gleichzeitig irrelevante Informationen zu ignorieren. Dies führt oft zu einer signifikanten Steigerung der Genauigkeit und Präzision bei der Aspekt- und Sentimenterkennung, was sich in den verbesserten Ergebnissen dieses Modells im Vergleich zu den vorherigen Ansätzen widerspiegelt.

Die ersten Modelle lieferten gute Aspect Extraction, aber schlechte Sentiment Classification. Insbesondere lieferten diese Modelle das gleiche Sentiment für alle Aspekte eines Satzes. Die Review "the screen is great but I hate the battery" wurde als "screen: positive; battery: positive" klassifiziert, obwohl das Sentiment für "battery" negativ sein sollte. Dieses Problem wurde durch die Einführung eines Attention-Mechanismus gelöst, der es dem Modell ermöglicht, die Relevanz jedes Aspekts einzeln zu bewerten und das entsprechende Sentiment zuzuweisen.

5 Resultate

Das fertige Modell basieren auf LSTM und Attention-Mechanismen konnte besonders auf dem Laptop Datensatz eine hohe Genauigkeit bei der Aspect Extraction und Sentiment Classification erreichen. Aufgrund der leicht veränderten Ausgabeformats ist ein 1:1 Vergleich zu Modellen der vorherigen Jahre und sehr aufwendige Nachbereitung leider nicht möglich. Die Aspekte im Datensatz sind hierbei nicht streng nach Wörtern, sondern nach kleineren Abschnitten definiert. Beispielsweise findet sich im Testsatz die Review "The Apple engineers have not yet discovered the delete key" mit dem Aspekt "delete key" mit einem negativen Aspekt. Das Modell wiederum hat hierbei "delete" und "key" einzeln als Aspekt erkannt, da das Model nur einzelne Tokens (hier Wörter) klassifizieren kann. Das Ergebniss ist somit zwar korrekt, fällt aber bei automatischer Auswertung als falsch auf. Insgesamt gibt es bei dieser Art von Use-Case oft kein richtig oder falsch, da diese Art von Analyse oft subjektiv ist. Eine weitere Review ließt sich wie folgt: "It´s a decent computer for the price and hopefully it will last a long time". Hierbei wurde der Aspekt "price" neutral gelabelt, obwohl dieser genauso als positiv gewertet werden könnte, was das Modell in diesem Fall auch getan hat.

Nach manueller Evaluation auf einem Teil der Testdaten konnte das Modell eine Genauigkeit von 85% bei der Aspect Extraction und 80% bei der Sentiment Classification erreichen. Diese Ergebnisse lassen sich ungefähr mit den top Modellen von 2019-2020 vergleichen, welche auf deutlich größeren Modellen basieren. Auch wenn diese Ergebnisse nicht direkt vergleichbar sind, zeigt dies das potenzial von ressourcenschonenden Modellen in der Sprachverarbeitung auf.

6 Diskussion

Effizienz und Leistung des Mini-Modells im Vergleich zu größeren Modellen. Potenzial für Anwendung in ressourcenbeschränkten Umgebungen und Echtzeit-Anwendungen. Ist das Mini-Modell eine gute Alternative für Unternehmen und Entwickler? Welche Faktoren sollten bei der Auswahl eines Modells berücksichtigt werden?

7 Fazit und Ausblick

Aspect-Based Sentiment Analysis ist eine leistungsstarke Methode, um Meinungen und Stimmungen auf der Ebene einzelner Aspekte zu analysieren. Die Fortschritte im Bereich des maschinellen Lernens und der natürlichen Sprachverarbeitung haben die Genauigkeit und Effizienz von ABSA-Modellen erheblich verbessert. Die Anwendungsbereiche von ABSA sind vielfältig und reichen von der Analyse von Produktbewertungen bis zur Überwachung der Markenwahrnehmung.

References

- [1] Author, "Title of the paper," *Journal Name*, vol. 1, no. 1, pp. 1-10, Year.
- [2] Author, "Title of the book," Publisher, Year.
- [3] Author, "Title of the article," *Website*, Available: <http://www.website.com>, Accessed on: Date. <https://arxiv.org/abs/1906.02243>

Eigenständigkeitserklärung:

Hiermit erkläre ich, dass ich die vorliegende Arbeit ohne fremde Hilfe verfasst und die abgebildeten Datensätze, Zeichnungen, Skizzen und graphische Darstellungen, soweit nicht anders angegeben, eigenhändig erstellt habe. Ich habe keine anderen Quellen als die angegebenen benutzt und habe die Stellen der Arbeit, die anderen Werken entnommen sind – einschl. verwendeter Tabellen und Abbildungen – in jedem einzelnen Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht.

Fynn Madrian