A NOVEL ASPECT TAXONOMY AND ASPECT EXTRACTION METHODOLOGY FOR SCHOLARLY BOOK REVIEWS

Wickramarathna Wengappuli Arachchige Chathur Sajeewan Basuru

(158207C)

Degree of Master of Science

Department of Computer Science & Engineering

University of Moratuwa

Sri Lanka

March 2019

A NOVEL ASPECT TAXONOMY AND ASPECT EXTRACTION METHODOLOGY FOR SCHOLARLY BOOK REVIEWS

Wickramarathna Wengappuli Arachchige Chathur Sajeewan Basuru

(158207C)

Thesis/Dissertation submitted in partial fulfillment of the requirements for the degree
Master of Science in Computer Science and Engineering

Department of Computer Science & Engineering

University of Moratuwa

Sri Lanka

March 2019

DECLARATION

“I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date:

Name: W.W.A.C.S. Basuru

The above candidate has carried out research for the master’s dissertation under my supervision.

Signature of the supervisor: Date:

Name of the supervisor: Dr. Surangika Ranathunga

Abstract

Many people decide on the quality of a product based on its online reviews, which is also the most commonly used method when purchasing books from online book stores. Compared to other products, a scholarly book is one of the most difficult products to purchase online since customers have limited access to its internal content. Therefore, a customer has to go through multiple reviews in order to get insight on the book. However, the sheer volume of online reviews makes it difficult for a human to process and extract all the meaningful information in order to make an educated purchase. As a result, a requirement for a sentiment analysis system for scholarly book reviews are much needed at this stage. A more accurate opinion of the book can be obtained through aspect-based summarization. This type of summarization of opinions is critical for scholarly book reviews since content, organization, and other features interpret whether the book can be recommended to a customer at a certain education level.

Compared to sentiment analysis on reviews of products/services such as movies or restaurants, there is no well-defined research in aspect extraction or aspect-based sentiment analysis of scholarly book reviews. Not surprisingly for this domain, there is no well-defined aspect taxonomy or an annotated dataset available to extract aspects or to identify aspect categories. Compared to other domains, identifying aspects of book reviews is difficult since aspects such as the quality of the book or the discussed topics always appear implicitly in reviews.

The main contribution of this research is to identify potential aspects and an aspect taxonomy for scholarly book reviews. We also present a (1.) dependency rule-based unsupervised model for aspect extraction, which works better than state-of-the-art unsupervised methods, and (2.) a clustering-based aspect category identification method. Both of these are important first steps for aspect-based sentiment analysis.

The aspect taxonomy for scholarly book reviews is a hierarchical model. Book and Author have been identified as the first level of the taxonomy. Readability, content, worthiness and price, are the next level of aspect taxonomy under the book aspect category. Author expertise has been identified as an aspect category under author. In order to validate the aspect taxonomy, an unsupervised aspect extraction and clustering algorithm is proposed. An existing dependency rule-based aspect extraction algorithm is improved by adding new rules that extract aspects from book reviews. Two existing clustering algorithms for aspect clustering are merged to obtain a new clustering algorithm to discover the categories of aspect terms. The clustering algorithm is able to find the semantic similarity of aspect terms, while considering the sharing words between aspect terms, and groups similar aspects in to a one cluster. After successfully generating an annotated corpus for the scholarly book reviews in the computer science domain with Cohen’s kappa statistics of 0.76, the dependency rule-based aspect extractor was able to extract both implicit and explicit aspects with precision 76.04%, recall 75.99% and overall F1-score 76.02%. The proposed semantic similarity based aspect clustering algorithm identifies the aspect in the following categories; book, author, readability, content, worthiness, price and author expertise with rand-index 14.41%, V-measure 36.29%, homogeneity 66.18% and completeness 25%.

Keywords: Aspect based sentiment analysis, Dependency rules, Aspect taxonomy, Clustering, Semantic similarity, Stanford dependency parser, GloVe

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my supervisors Dr. Surangika Ranathunga for the continuous support given for the success of this research. This would not have been a success without your tremendous mentorship and advice from the beginning.

I would like to thank all staff from the Department of Computer Science and Engineering for their kindness expressed in all occasions. Special thanks to my loving family: my parents, wife, and sisters. Your encouragement always motivated me to do my best. I would also like to thank all my friends who encouraged me all times.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

LIST OF ABBREVIATIONS

ABSA	Aspect Based Sentiment Analysis
CINAHAL	Cumulative Index to Nursing and Allied Health Literature
CNN	Convolutional Neural Network
CRF	Conditional Random Field
GFST	Generalized Aspect Sentiment Tree
HMM	Hidden Markov Model
IAC	Implicit Aspect Clue
KNN	K-Nearest Neighbors
LDA	Latent Dirichlet Allocation
MEDLINE	Medical Literature Analysis and Retrieval System Online
MLE	Maximum Likelihood Estimation
MV-RNN	Matrix Vector Recurrent Neural Network
NER	Named Entity Recognition
NLP	Natural Language Processing
NPMI	Normalized Pointwise Mutual Information
OWA	Ordered Weighted Averaging
PAS	Positional Aggregation based Scoring
POS	Part of Speech
RNN	Recurrent Neural Network
RNTN	Recursive Neural Tensor Network

SDM	Sequential Dependence Model
SemEval	International Workshop on Semantic Evaluation
SSWE	Sentiment Specific Word Embedding

1 INTRODUCTION

1.1 Background

Amazon.com, Inc. is one of the world’s largest electronic commerce and cloud computing companies. Amazon started as an online bookstore and later diversified their business to include software, video games, electronics, apparel, furniture, and other domains. The overall star rating can be used to get an abstract idea of the quality of any product. However, these star ratings are not good enough to identify hidden information of reviews like the aspects of the products and user opinion towards those aspects.

Book reviews are always very critical since it is very hard to make a decision about a book without looking at the content inside. For the bookstore at Amazon, there are online reviews created every day by customers who have either purchased these books or already read them. The massive volume of reviews for a given entity can often be prohibitive for a potential customer who wishes to read all the relevant information, compare alternatives, and make an informed decision.

Sentiment analysis is the process of computationally identifying and categorizing opinions expressed in a piece of text. It emerged as a sub field of natural language processing to answer this problem, not only for Amazon products, but also for various business domains such as social networks, electronic commerce, hotels, and restaurants. The basic task in sentiment analysis is classifying the polarity of a given text at a document, sentence, or aspect level; whether the expressed opinion is positive, negative, or neutral. Advanced tasks that go beyond polarity sentiment classification looks, for instance, at emotional states such as angry, sad, or happy. Identifying the sentiment polarity of a customer review in general may not provide information about the product features. Aspect based sentiment analysis provides a solution for this by producing a set of relevant aspects, an aggregate score for each aspect, and supporting textual evidence. Aspects can be presented in a user review in two methods; explicit aspects are concepts that explicitly denote the targets in the opinionated sentence, or on the other hand, an aspect can also be expressed indirectly through an implicit aspect clue (IAC) [1]. Explicit aspects are easy to understand, and extract compared to implicit aspects as the opinion expression is a subjective statement. An intelligent sentiment summarizer is capable of identifying both implicit and explicit aspects with good accuracy. In this context, aspect-based sentiment analysis has become the most productive way to identify user opinion on any product or service.

1.2 Problem and Motivation

The usage of book reviews for sentiment analysis is very low compared to other domains such as movie reviews, social media reviews, or restaurant reviews. When it comes to Amazon book reviews, nobody has done a proper analysis on the domain of sentiment analysis. Since Amazon is one of the largest e-commerce platforms in the world, the effect of giving such intelligence to the bookstore will increase the productivity of the business.

Scholarly book reviews play an important role in informing readers about new books and guiding their reading preferences as they explore the Internet and large catalogues provided by the publishers. A good scholarly book review critically evaluates the content, quality, meaning and significance of a book. Therefore, book reviews are an excellent vehicle to inform readers about a new book in the marketplace. Books are relatively expensive, and scholars have limited time to commit to reading. Thus, they may rely upon the book review's evaluative purpose to guide their reading preferences. It is important to use this large volume of user reviews to inform readers on new, innovative and groundbreaking books while being warned of books of poor quality and those that may not relate to their area of interest.

The sentiment classification of book reviews is a constant challenge when compared to other products since diversity in book categories is high and it is difficult to identify aspects. Aspect extraction has become a major challenge for book reviews since the type of aspect totally depends on the perception. One can focus on aspects such as the price and the quality of the paper, while the context of the book can be a main focus for others. So, it is very important to identify the book category and the aspects to be considered.

Hamdan et. al. [2] identified potential aspects for book reviews in the social science domain, but their research was based on Scholarly book reviews written only in French. Therefore, nobody has properly identified the potential aspects for scholarly book reviews written in English and no proper data set is available to be used for aspect-based sentiment analysis. Proper aspect taxonomy is needed to identify the relationship between different aspects and entities. In particular, to date, nobody has properly identified the aspect list for scholarly book reviews, and consequently, aspect level sentiment analysis has not been done for English scholarly book reviews. Because aspects should be identified and extracted prior to identify the polarity towards aspects.

1.3 Research Objective

The objective of this research can be summarized as follows,

- Identify aspect taxonomy for scholarly book reviews
- Scholarly book review corpus annotation
- Aspect extraction and categorization from scholarly book reviews

Aspects pertaining to different book categories vary. Even within scholarly books, depending on the domain, the list of important aspects will differ. Therefore, in this study, scholarly book reviews will be selected from the Computer Science domain. Aspect taxonomy, aspect extraction and categorization methodology are important first steps for aspect based sentiment analysis of scholarly book reviews.

1.4 Research Contribution

This research presents a novel aspect taxonomy for scholarly book reviews. This taxonomy was validated using scholarly book reviews in the Computer Science domain, however the taxonomy is general enough to be applied to other scholarly book categories.

This research also presents a clustering algorithm that groups the extracted aspects in order to identify aspect categories. In order to reduce the noise words, input to the clustering, a dependency rule based aspect extraction method was introduced. New dependency rules have been discovered, and an implicit aspect lexicon for the Computer Science domain is built. This aspect extraction method is able to identify both implicit and explicit aspects. This aspect extraction mechanism was evaluated on scholarly book reviews from the Computer Science domain.

1.5 Structure of Thesis

Chapter 2 contains the literature survey. The final four chapters describe the methodology, experiment results, discussion, and conclusion. Methodology includes data extraction, preparation, creating aspect taxonomy for scholarly book reviews, test corpus annotation process, dependency rule based aspect extraction method, clustering based aspect category identification. The experiment set up and the results present the performance of aspect extraction with respect to the dependency rules and the implicit aspect lexicon, performance of the aspect category identification with respect to the clustering hyperparameters, and performance of the clustering with respect to the performance of the aspect extraction phase. Discussion and conclusion focus on the value to the field and future research potentials based on this research.

2 LITERATURE SURVEY

The literature survey is mainly divided into five sections. The first section reviews the basics of sentiment analysis. The second section focuses on aspect extraction in detail, including existing research related to book reviews. The third section discusses current trends towards sentiment analysis of book reviews. The fourth section discusses the potential aspect categories for scholarly book reviews, while the fifth section briefly discusses the corpus annotation methods.

2.1 Sentiment Analysis

The opinions of others have a significant influence in our daily decision-making process. Sentiment analysis is the computational study of opinions, sentiments and emotions expressed in text [3]. The use of sentiment analysis is becoming more widely leveraged because the information it yields can result in the monetization of products and services. Sentiment analysis can occur at different levels: document level, sentence level and aspect level.

2.1.1 Document Level Classification

The sentiment is extracted from the entire review and the whole opinion is classified based on the overall sentiment of the opinion holder [4]. The goal of document level classification is to classify the review as positive, negative, or neutral. But this method has various limitations since it assumes that the document expresses opinions on a single entity. But this can be false very easily. For example, a product review of a mobile phone can consist of different opinions about its different components. One can give a good sentiment about the battery life and screen but can give negative sentiments on its software and design. But if all these sentiments are grouped as one single entity, sometimes the sentiment can be neutral, and the overall information given to the end user is not accurate.

2.1.2 Sentence Level Classification

Sentence level classification considers each sentence as a separate unit and assumes that a sentence contains only one opinion [5]. This process involves two steps: subjectivity classification of a sentence into one of objective or subjective class, and sentiment classification of subjective sentences in to positive or negative classes. Positive and negative classification is the most basic level of sentiment classification. Sentiment is usually extended to either 3 classes (positive, neutral, negative), or 5 classes (++, +, 0, -, --). Sentence level sentiment analysis does not give very good results for complex sentences with multiple aspects in the same sentence. The sentence will be classified as a neutral sentiment sentence since the ability to identify different sentiments towards different aspects is still not there. Meena and Prabhakar [6] did sentence level sentiment analysis. They used word dependencies and dependency trees to analyze the sentence constructs and also analyzed the effect of WordNet to the accuracy of the results. They identified that conjunctions have a substantial impact on the overall sentiment of a sentence, so presented how atomic sentiments of individual phrases combine together in the presence of conjuncts to decide the overall sentiment of a sentence.

2.1.3 Aspect Level Classification

The goal of aspect level sentiment classification is to identify, and extract object features that have been commented on by the opinion holder and determine whether the opinion is positive, negative, or neutral. Aspect based sentiment summarization can be implemented in three steps: 1. Identify all sentiment laden text fragments in the reviews, 2. Identify relevant aspects for the considered domain that are mentioned in these fragments, 3. Aggregate the sentiment over each aspect, based on the sentiment of the mentions [3].

Aspects can be explicit or implicit. Explicit aspects are aspects that are used by users with explicit words. For example, if you take a review that states the following, *"Price of the book is very high"*, the aspect *price* has been explicitly mentioned. On the other hand, in a review that states: *"This book is very expensive"*, the user is again talking about the *price* aspect but this time no explicit word has been used to express this aspect. The full process of aspect-based sentiment analysis can be summarized as shown in figure 2.1.

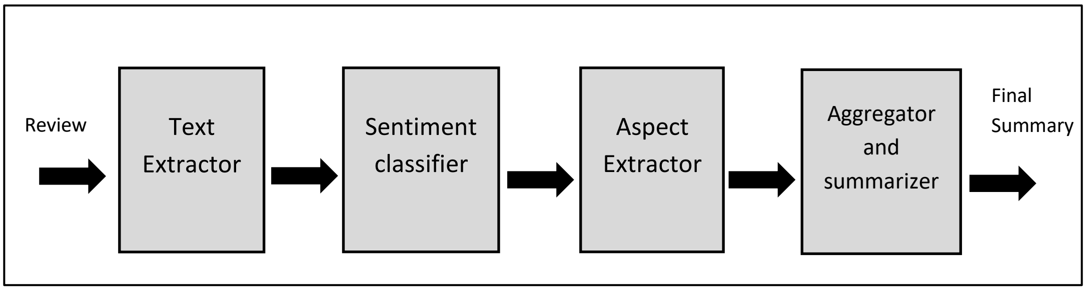

Figure 2.1 : Abstract architecture of Aspect based sentiment classifier [3]

2.2 Defining Aspect Categories and Taxonomy creation

Taxonomy can be done either manually or automatically. Panchendrarajan et al. [7] manually created a hierarchy of aspects for the restaurant domain. They used a random sample of 400 reviews and validated and refined it using another set of 400 reviews. The model consisted of aspects up to four levels, where level one is the restaurant and had six main sub aspects as food, service, ambience, offers, worthiness and other aspects. Each sub aspect was further categorized. For example, aspect service had staff as one of the sub aspects, which also had four other sub aspects; behavior, experience, appearance, and availability.

2.3 Aspect Extraction and Categorization

Once the aspect categories are defined, the next step is to identify the opinion target expressions that also mention aspects, where aspects can either be explicit or implicit. Opinion target expression is a linguistic expression used in a given text to refer to an aspect of the reviewed entity or aspect category [8]. As mentioned earlier, explicit aspects are literally mentioned in the opinion text, whereas implicit aspects are not literally mentioned in the opinion text.

Cheah et al. [8] did a comparative analysis and survey on aspect extraction on sentiment analysis. They summarized all the research papers related to aspect extraction until the end of 2015. More than fifty techniques were summarized for the extraction of explicit aspects and eleven studies were summarized for implicit aspect extraction. Aspect extraction is categorized into three different approaches according to their learning method; unsupervised, semi-supervised and supervised. Cheah et al. [8] identified that unlike the unsupervised and semi-supervised techniques, supervised techniques have been applied in a diverse number of domains. Most of the semi-supervised and unsupervised approaches used product review data.

2.3.1 Unsupervised Methods

Unsupervised aspect extraction techniques do not need any prior training data to extract aspects from reviews, hence, it is very easy to apply to any new domain. Unsupervised techniques have been widely used by researchers for aspect extraction from online reviews. Interestingly, these techniques have been applied on diverse kind of domains and different data sets.

2.3.1.1 Rule-based methods

Rule based aspect extraction looks in to the dependency tree of the sentence and creates handcrafted rules to find the relationship between aspects and opinion words. Sentences that have matches with the defined rules are selected and the aspects are extracted. Bancken et al. [9] introduced an algorithm called 'ASPECTATOR' that works by matching the syntactic dependency path among different words from the sentence. Ten handcrafted dependency paths were defined, and the product aspects and their opinion words were extracted. Further, to cluster aspects, WordNet was used to find synonyms. Poria et al. [1] proposed a rule based approach to extract explicit as well as implicit aspects from reviews. In order to extract implicit aspects, they first identify the implicit aspect clues (IAC) using several dependency rules and used WordNet and SenticNet to identify the synonyms and semantics of each IAC respectively. Explicit aspects were also extracted using dependency rules. The main limitation of the rule-based approach is that it always depends on the grammatical accuracy of the sentence. But dependency rule-based approaches support finding low frequency aspects and are very good in identifying implicit aspects.

2.3.1.2 Clustering based methods

Clustering algorithms always grouped the aspect candidates based on a defined similarity algorithm until stopping criteria is found or similarity threshold is reached, so that no more clusters can be grouped further.

Su et al. [10] proposed a clustering method that utilized the mutual reinforcement association between features and opinion words to find the aspect cluster. The algorithm is capable to handle implicit aspects as well. The algorithm was proposed based on Chinese product reviews. Aspect clustering was done iteratively by fusing both their content information and sentiment link information.

Chen et al. [11] proposed a clustering approach that simultaneously identifies aspects and groups them to aspect categories. The proposed approach extracted both explicit and implicit aspects and does not require any seed terms. Noun and noun phrases were taken as candidates for the explicit aspects and adjectives and verbs were identified as candidates for the IACs. The proposed algorithm grouped similar candidates so that terms referring to the same aspect were put into one cluster. At last, the important aspects were selected from the resulting clusters. Instead of applying agglomerative clustering to all the candidates, it first selects the most frequent candidates for clustering. Reasons for this are two-fold. Firstly, the most frequent candidates are more likely to be the actual aspects that users are interested in. So, by clustering these terms, high quality seed clusters can be generated. Secondly, the clustering algorithm requires pairwise distance between candidates or clusters, which can be very time consuming if the number of candidates is very large. Clustering only the most frequent candidates speeds up the process. The similarity between two cluster candidates was taken as a combination of general semantic similarity and the corpus based statistical association. General semantic similarity was calculated using both WordNet knowledge and statistics from a large web corpus to compute the semantic similarity between words or phrases. Corpus based similarity was calculated using the normalized pointwise mutual information (NPMI). Accuracy of the clustering algorithm always depends on the similarity metric and the dataset used to find the vector representation of the words and phrases. But the biggest strength of using clustering algorithms for aspect categorization is that aspect categories can be mined through clustering to validate a proposed aspect taxonomy.

2.3.1.3 Frequency or statistical methods

Frequency based methods are heavily used at the beginning of aspect extractions where explicit aspects can be extracted using noun or noun phrases. Hu and Liu [12] have used association rule mining to find frequent product features followed by compactness pruning and redundancy pruning of unwanted features. Bafna and Toshniwal [13] further improve the association rule mining based approach by integrating a probabilistic approach. They have used a probabilistic power equation to remove all the nouns that do not represent aspects, although they are frequent. Frequency based approaches are simple and quite effective methods, but they have several limitations. It generates many non-aspect terms and generally misses the low frequency aspects. Manual tuning of parameters like support and confidence for association rule mining, make it difficult to apply the same setup to a different data set.

2.3.1.4 Bootstrapping methods

Bagheri et al. [14] used a bootstrapping algorithm to extract explicit aspects, which require initial seed sets of aspects. The iterative bootstrapping algorithm learns the final list of aspects from a small number of unsupervised seed sets of information. Bootstrapping can be defined as an iterative clustering technique, where in each iteration, the most valuable candidate is chosen to adjust the existing seed set. This is an iterative process until stopping criterion is reached. POS pattern-based heuristics have been used to identify aspects from the reviews. A list of top aspects had been generated using a new matric called A-score, which used the inter relation information among the words. Then the final set of aspects was generated through the bootstrapping algorithm by using the A-score of each aspect to measure the value score. Subset-support pruning, and superset-support pruning have been used to remove unwanted aspects. Bagheri et al. [15] improved the same approach by adding a graph based approach to identify implicit aspects. The opinion words were used in the graph as a node and that node was mapped to a set of aspect nodes. Co-occurrence of aspect and opinion words was assigned as the weight of the edge, connecting two nodes.

2.3.2 Supervised Methods

Supervised learning approaches always need annotated training data for extract aspects. When compared with unsupervised and semi supervised learning approaches, the engineering effort is always high. If there is not enough training data, supervised learning algorithms do not perform very well in the aspect extraction domain.

2.3.2.1 Conditional Random Field (CRF) based Approach

CRF is a discriminative probabilistic sequence model successfully used in various domains such as, part of speech tagging and named entity recognition. CRF falls into the sequence modeling family and is a discriminative classifier. The underlying idea in CRF is defining a conditional probability distribution over label sequences when particular observation sequence is given, and it is not a joint distribution over both label and observation sequence. Hamdan et. al. [2] employed a CRF based approach for aspect extraction in book reviews, which is the only available research on this topic. Once book presentation, problematic, scientific context, scientific method, author's arguments, book organization and judgment about the book have been identified as the aspects for the scholarly book reviews, the CRF based model, with different features, has been used for the experiments. First, they tested the classifier by using only the terms as features. Then, term and POS tagging were used as features. They experimented other features like shape, type, prefix, and suffix, but observed that term and POS features seem to be enough to produce a good result**.**

In general terms, Li et al. [16] proposed skip-chain CRFs and tree CRFs for the extraction of aspects and opinions, which are based on CRF. Skip-chain CRF was used to identify sequential dependencies among continuous words. They learnt that if two words or phrases are connected by the conjunction 'and', then both the words have the same polarity, while if they are connected by 'but', then, the two have opposite polarity. In order to overcome the long-distance dependency, skip-chain CRFs are used to find aspects and opinions.

Tree CRFs were proposed to learn the synthetic structure of sentences in the reviews. Skip-chain CRFs provide the semantic relations with respect to conjunctions, while tree CRFs provide dependency relations among different words in the sentence. Further, skip-tree CRFs were proposed to combine both the methods explained above and use those trees to extract opinion and aspects from movie reviews by giving the list of aspects as input seeds. Chen et al. [17] improved this technique by integrating a self-tagging process in order to minimize manual effort for labeling the data and in order to achieve the desired balance between algorithm complexity and accuracy, by identifying an optimal set of learning functions. The proposed model was compared with other aspect extraction techniques to compare the accuracy of different systems and different levels and using the same dataset on product reviews.

Jakob et al. [18] also used the supervised CRF based techniques to extract the opinion targets from reviews. They observed that the same word may have different representations in different domains. For example, the term *unpredictable* has a positive polarity in movie reviews but a negative polarity in car reviews. So, this method can be used in different domains by keeping in mind that words have different domain probabilities.

Choi et al. [19] used a hierarchical parameter sharing technique using CRFs. They defined the problem as a sequence tagging task to identify both opinions and aspects. This approach not only extracts opinions but also ranks the aspects according to the polarity of their opinion words.

Huang et al. [20] also proposed a CRF-based learning model to extract product aspects. Product aspects were divided into three sub categories for the learning process and certain tagging rules were defined to identify these aspects. After that, similar aspects were categorized using syntactic dependencies and WordNet.

Yang et al. [21] proposed jointly identified opinion related entities, that is opinion expression, opinion targets and opinion holders along with the relations that link the opinions with entities, where these relations were *is-about* and *is-from.* CRF was used to identify opinion and entity words.

Recently, CRF based models were effectively used in SemEval (International Workshop on Semantic Evaluation) Aspect based sentiment analysis tasks. Hamdan

[22] proposed a CRF model for the opinion target extraction in SemEval 2016. This model was ranked second among the nineteen submissions for English restaurant reviews. The CRF based model proposed by Toh et al. [23] for the SemEval 2015 task ranked first in the aspect extraction subtask. The bigram, head word and name list generated using double propagation, have been used as the features for their model.

2.3.2.2 Maximum Entropy Based Approaches

Maximum Entropy classifier is a probabilistic classifier that belongs to the exponential models. It does not assume that the features are conditionally independent of each other. It is based on the principal of maximum entropy and from all the models that fit the training data, selects the one that has the largest entropy. The principal of maximum entropy is a rule that allows to choose the best from a number of different probability distributions that all express the current state of knowledge. This will be the system with the largest remaining uncertainty. Panchendrarajan et al. [7] used a maximum entropy classifier for explicit aspect identification of restaurant reviews. N – grams have been used as features, where n varies from 2 to 5. Research showed impressive results in the restaurant domain.

2.3.2.3 HMM based Approaches

The Hidden Markov model is a finite set of states where each state is associated with a probability distribution. Transition among the states is governed by transaction probabilities and the outcome or observation of the particular state can be generated according to the associated probability distribution. It is only the outcomes, not the state visible to the external observer. The lexicalized HMM based approach naturally integrates linguistic features, such as parts of speech and surrounding contextual clues of words, into automatic learning. This method was previously used for POS tagging and Named Entity Recognition (NER). A lexicalized HMM based model for explicit aspect extraction was proposed for the first time by Jin and Ho [24], and Jin et al. [25]. This work not only identified the product aspects and their opinions, but also identified sentences that contained aspect opinion pairs and categorized the opinion words as either negative or positive. There are two tag sets defined in their model; the first is the basic tag set, which defines different categories of entities, and the second tag set defines patterns for different entities, which is the position of a word in the entity phrase. Then, each sentence was manually tagged with the help of this tag set by representing the pattern between aspects and opinion words. This trained corpus was integrated with the actual tagged data into HMMs. To find an appropriate sequence of manual tags and actual tagged data that maximizes the conditional probability, HMM along with maximum likelihood estimation (MLE) has been used. In the next step, sentences that contain aspect opinion pairs have been identified.

2.3.2.4 Tree based Approaches

The tree-based approach has been used by Jiang et al. [26] where Generalized Aspect-Sentiment Tree (GFST) was used to extract aspects from customer reviews. They defined four different tree kernel spaces to identify aspects from the reviews. The kernel-based methods evaluate the similarity between two trees instead of extracting each individual aspect from each tree.

2.3.2.5 Support Vector Machine based approaches

Support Vector Machines is a discriminative classifier. When labeled training data is given, the algorithm outputs an optimal hyperplane that categorizes the new data. SVM has been successfully used for the aspect category extraction. Pontiki et al. [27] used SVM with linear kernel and unigram features to predict the predefined aspect categories of laptops and restaurants domains. Their research was carried out as a semEval 2015 task and they have performed better than the other 6 teams in the less fine-grained restaurant domain with a 62.68% F1 score. The laptop domain contains 22 entities and 9 attributes, hence 80 aspect category combinations. Pontiki et al. [27] achieved only a 54.10% F1 score where 62.68% was the best out of 9 teams. Alvarez-Lopez et al. [28] also developed a system for aspect category detection based on SVM for the restaurants domain. This system was built for the SemEval 2016 task 5. They used SVM for aspect category identification in subtask 1 and 2. Subtask 1 was sentence level aspect-based sentiment analysis, where an opinionated document about a target entity (e.g. laptop, restaurant or hotel) was given, and the goal was to identify aspect category, opinion target expression and sentiment polarity. Subtask 2 was text level aspect-based sentiment analysis, where a set of customer reviews about an entity like laptop or restaurant was given, and the goal was to identify a set of aspect and polarity tuples that summarized the opinion expressed in each review. Linear SVM classifier combined with word lists has been used to classify the given 12 predefined categories. Words, lemmas, POS tags and bigrams have been used as features and the F score was 67.71% for English language restaurant reviews.

2.3.2.6 Limitations of supervised learning approaches

All the supervised learning approaches have required labeled data for training. Therefore, accuracy of the learned model depends on the accuracy of the training data (how accurately training data is labeled for aspects and not-aspects).

2.3.3 Deep learning for aspect extraction

Deep learning has been a hot topic in the recent past in various domains due to its powerful features. Deep learning models have been effectively used as sentiment analysis algorithms for a few years now. However, it has not been formulated as an aspect extraction algorithm until 2016. Poria et al. [29] presented a deep learning approach to aspect extraction in opinion mining. A seven-layer deep convolutional network was used to tag each word in opinionated sentences as either aspect or non-aspect words. They have overcome the limitations of both Conditional Random Field (CRF) approaches and linguistic pattern-based approaches. CRF is a linear model, so it needs a large number of features to work well. Linguistic patterns need to be crafted by hand and crucially depend on the grammatical accuracy of the sentence. Convolutional neural network (CNN) is a non-linear supervised classifier that can more easily fit the data, hence outperforming other state-of-the-art NLP methods. They also use linguistic patterns to improve the performance of the method. Ruder et al. [30] proposed a robust CNN based model for aspect category extraction in a multilingual setting. The model was successfully experimented on the hotel, laptop, restaurant, phone, and camera domains. They were able to reach a 73.03% F1 score in the English language domain placing 1st or 2nd in both aspect category detection and sentiment polarity detection in each language domain. In contrast to other approaches, their model neither relies on expensive feature engineering, availability of a parser, nor positional information, but solely on a language's input signals. A mini-batch size of 10, maximum sentence length of 100 tokens, GloVe word embedding size of 300, dropout rate of 0.5 and 100 filter maps was used for the experiment. Filter lengths of 3, 4, 5 were used for aspect category extraction.

2.3.4 Semi-Supervised Methods

Semi supervised approaches partially depend upon user input and need initial seeds to start the algorithm. These algorithms also mainly focused on product reviews.

2.3.4.1 Dependency parser based methods

Wu et al. [31] proposed a combination of dependency parsing and SVM for aspect and opinion word extraction. They extracted noun phrases and verb phrases as potential aspects and extracted words with sentiments as potential opinion words. A dependency parser was used to find the relations between opinion words and aspects. A new tree kernel is defined and incorporated with SVM to find the relations between aspects and opinions. The lowest common parent from all possible sub-trees was found out by defining the maximum distance between aspect and opinion word and the threshold was 5. Their main assumption of this model was that opinion words are supposed to be near the aspect word.

2.3.4.2 Lexicon based methods

Wei et al. [32] introduced a semantic based product aspect extraction technique, where the basic idea is that once the aspect candidates are extracted, they provide a lexicon of positive and negative adjectives to identify the opinion word's subjectivity. By using this lexicon, the sentiment-based refinement step finds those aspects that were not actually aspects and prunes such aspects. By giving a list of adjectives and opinions, they identified the opinion irrelevant product aspects. Noun or noun phrases were considered as aspects.

2.3.4.3 Graph based methods

Liu et al. [33] proposed a semi supervised word alignment model to find the association among opinions and their target aspects. Nouns or noun phrases were considered as target aspects of opinion words and the purpose of word alignment was to align opinion words with their potential aspects in a sentence. For these word alignments, syntactic patterns were used to identify the relation between aspects and opinion words. To improve the word alignment and remove expected errors, they used a trained dataset and incorporated it with the alignment process. To identify all the syntactic patterns, a graph-based approach was used. Finally, all the patterns with a confidence below a given threshold, were removed.

2.3.5 Discussion

Supervised aspect extraction approaches have used different datasets that belong to diverse domains and languages. Although most of the work focused on English language datasets, there are a considerable amount of experiments on the domains belonging to the Chinese language. So, the result of one domain cannot be compared with the results of the other domains and there is a similar constraint for the language. Domains and languages of all the approaches need to be identified to conduct a comprehensive and justifiable comparison. Cheah et al. [8] did a comparative analysis on supervised learning approaches for explicit aspect extraction by considering the average precision, recall and F-measure. They identified that most of the research focused on the product domain, therefore, the domain is mentioned explicitly only with those approaches that used domains other than the product domain for the experiments. Most of the research focused on English language datasets, so language was mentioned explicitly on the approaches that used languages other than English when doing the comparative analysis on aspect extraction methods. Their analysis showed that CRF based methods performed well compared to other methods such as tree-based and decision-tree-based methods. Maximum entropy models and deep learning models were not considered in their research.

Deep convolutional neural networks were used for extract aspects of electronic products in several domains such as cameras, DVD, MP3, and Cellphone. This CNN based model generally outperformed the state-of-the-art models in product categories. However, when it comes to the complexity, CRF based models are easier to implement. Maximum entropy models are very easy to implement with good results close to other models. Therefore, CRF and maximum-entropy-based models are practical and efficient classifiers to implement.

Supervised methods are well suited when the aspect taxonomy is well defined and there is a very well annotated data set. Unsupervised algorithms are very useful for aspect extraction in new domains where well defined aspect taxonomy is not available and aspect terms are difficult to identify. Another reason for using unsupervised approaches frequently is the laborious and time-consuming task of training the dataset. Surprisingly, unsupervised learning algorithms also produced good results. Dependency-rule-based aspect extraction methods outperform other unsupervised approaches and even the dependency propagation method with 80-90% recall and precision results. Using only clustering methods for both aspect extraction always shows poor results, but it shows very good results in identifying aspect categories. Clustering based aspect category identification algorithm proposed by Chen et al. [11] has outperformed state-of-the-art semi supervised approach proposed by Zhai et al. [34] and Latent Dirichlet Allocation (LDA) based topic modeling methods [34]. So, a combination of dependency rule-based aspect extraction and clustering-based aspect category identification is an ideal approach for new and unknown domains.

2.4 Current trends towards the sentiment analysis of book reviews

When it comes to sentiment analysis of book reviews, this sub domain was unable to keep up with the momentum compared to other sub domains like restaurant reviews, movie reviews, social media data and electronic product reviews. However, a good number of studies have been done in the area of implementing book recommendation systems. Recommendation systems exploit data mining and information retrieval techniques to predict what item suits the user needs and recommends those items with the largest predicted fit score. There are several technologies used frequently in the concerned field, like association rule based recommendation [35], collaborative filtering [36], web mining techniques [37] and opinion mining techniques [38]. Recommendation systems are information search and filtering tools that provide recommendation for items to be of use to a user. They have become common in helping users to make better choices while searching for books. Web mining has become the most popular technique among researchers due to its effectiveness.

2.4.1 Association rule based techniques

Association rule mining is a method for discovering interesting relations between the variables in the databases. Support is an indication of how frequently the item set appears in the database and confidence is an indication of how often the rule has been found to be true. This method was successfully used for book recommendation systems. Tiwari et al. [35] proposed a book recommendation system for students reading text books. The main objective of this paper was to develop a technique that recommends the most suitable books to students according to their price range and publisher's name. The system-proposed recommendation is stored in the student's web profile and works even when the user remains offline. Features like book category (ex: CSE, Electrical, Civil), sub category (C++, Data structures), publisher name and price were used for associate rule mining.

2.4.2 Collaborative filtering based techniques

Collaborative filtering is a method of making automatic predictions about the interest of a user by collecting preference information from many users. Collaborative filtering systems have many forms, but the general process can be reduced to two steps. The first is to look for users who share the same rating patterns with active users, and the second is to use the ratings from those like-minded users found in the previous step to calculate a prediction for the active user. This is called user based collaborative filtering and the Nearest Neighbor algorithm is a specific application. Item based collaborative filtering is used in a different way. First, build an item-item matrix determining relationships between pairs of items and secondly, infer the tastes of the current user by examining the matrix and matching that user's data. Benkoussas and Hamdan [36] proposed a collaborative filtering method for book recommendation. They tested the combination of the Sequential Dependence Model (SDM) and the use of social information, which takes into account, ratings, tags and customer reviews.

2.4.3 Web mining techniques

Web mining is closely related to data mining, but knowledge is discovered from internet data sources. Sohail et al. [37] proposed a rank based scoring method for the book recommendation system. This research was narrowed down to the area of computer science and 22 Indian universities used it to collect data. The official websites of each university ranked the books in descending order with respect to their curriculum in the area of artificial intelligence. A collection of 41 different books from all 22 ranked universities had been collected for the research. Different scores were received for each book corresponding to the respective university. Then, its aggregation was calculated, which gave rank-based scores with aggregate values. Sohail et al. [39]**,** [40] extended their research by introducing a positional aggregation based scoring (PAS) technique to score the books recommended by the top ranked universities and assigned weights to these scores using fuzzy quantifiers. Ordered Weighted Averaging (OWA) was used as the aggregation operator to find the top books in the artificial intelligence category in the computer science domain. The PAS based technique converts the different ranks that a university suggests for a book into a score between 0 and 1. OWA was used to handle aggregating multiple criterions. The accuracy of the approach was high, compared to their previous research.

2.4.4 Opinion mining and Sentiment analysis at document level

Sohail et. al. [41] also used the opinion mining technique to extract features and analysis of online reviews for the book recommendation system. Feature extraction was done using human intelligence and seven features were identified as important: frequency of occurrence in search engine results, useful content, extraneous content, sufficient material, physical attributes, market availability and price. These features were used to rank a particular book and recommend it to the user. The main limitation of this research is that feature extraction has been done as a complete manual process, where no supervised or unsupervised machine learning algorithm has been used.

Bellan [42] introduced a Bayesian classifier based sentiment analysis framework for historical book reviews. The research mainly focused on two elements: scholarly credibility and writing quality. Scholarly credibility gives an indication of the quality of the research that the reviewer thinks the writer has performed. Thus, scholarly credibility is conceived in terms of the book's academic value, while the writing quality is conceived in terms of the reviewer's assessment of the author's writing style. Although this is a weak approach towards the aspect-based sentiment analysis, this research tried to identify the opportunities and possibilities of aspect-based sentiment analysis of book reviews. They combined a Naïve Bayes classifier with the bag of words and ontology-based approach to get more accurate results. First, they recognize whether or not a sentence is of interest by using a domain ontology about historical books (Interest Detection phase), then a Naïve Bayes classifier with the bag of words model is used to identify the importance of the system. These two approaches gave the ability to only classify sentences containing sentiments about the scholarly credibility and writing quality. Fang et al. [43] used book review data for sentiment analysis of product review data. Instead of using book reviews alone, it analyzed several product categories collected from Amazon.com including beauty, book, home, and electronic. The main drawback is that there is no aspect detection or aspect extraction phase.

Kaggle [44] also published a competition on sentiment analysis on Amazon book reviews. The challenge is to classify positive or negative sentiment at review level. They have selected the reviews of 'Gone Girl' and limited the classifier to the logistic regression. Srujan et al. [45] did a comparative analysis by applying various preprocessing methods and using different classifiers to classify amazon book reviews as either positive or negative. They have compared the accuracy of various classifiers, time elapsed by each classifier and sentiment score of various books. K-Nearest Neighbors (KNN), Random Forest, Naïve Bayes, Decision Tree and Support Vector Machines are the used classifiers. Reviews from eight novels have been used for the experiment: The Martian, The Goldfinch, Fifty Shades of Grey, Gone Girl, The Fault in Our Stars, Unbroken, The Girl on The Train, and The Hunger Games. Classification was done in review level and their experiment showed better results for KNN and Random Forest classifiers. Accuracy was in between 84% to 94%.

2.4.5 Sentiment analysis at aspect level

None of the previous research identified the aspect for book reviews like context, presentation methodology and organization of the content. Hamdan et al. [2] used book reviews for the first time, for aspect based sentiment analysis. French scholarly book reviews have been used for the research. Here, they followed the complete stack for aspect-based sentiment analysis, as mentioned in Figure 1. Book reviews of social and human sciences have been used to enumerate the potential aspects that can be found in a book review. Although this is a good start towards the sentiment analysis of book reviews, the training corpus implemented contains only French book reviews. This limits the opportunity to use this data set for further research opportunities.

2.5 Aspect categories for Scholarly Book reviews

It is easy and straightforward to identify aspect categories for domains like restaurant reviews or electronic product reviews. For example, in the restaurant domain, it can be clearly identified that food, drinks, service, ambiance and location are the most important aspect categories [3]. However, the possible aspects of a book are more ambiguous. The quality of the book, number of pages and discussed topics seem to be good aspects, but it is still not as obvious as in the restaurant domain. In order to overcome this issue, unsupervised learning can be used. It is capable of extracting facts or topics as topic modeling in which we consider each topic related to an aspect. However, it is difficult to evaluate the quality of this method since there can be no correlation between the topic and an aspect in some situations. As mentioned above, Bellan [42] identified two abstract aspect categories; scholarly credibility and writing quality. However, these aspect categories do not have the potential to identify a reviewer's opinion towards the book in detail. Therefore, the most practical method is to use domain experts to identify potential aspect categories.

Hamdan et al. [2] have asked the OpenEdition editorial team, which deals with the book reviews of social and human sciences, to enumerate the potential aspects that may be found in book reviews. They have identified seven aspects: Book presentation, Problematic, Scientific context, Scientific method, Author's arguments, Book organization, and Judgment about the book.

Lee et al. [46] critically discussed how to write a scholarly book review. The purpose of this research was to describe and discuss the processes used to write scholarly book reviews and to provide a recommended strategy and identify potential aspects of a good book review. Research was conducted using three databases: MEDLINE (1950 - 2009) and EMBASE (1980 - 2009) through OVID publishing, which covers the international literature on biomedicine, CINAHAL Plus with Full Text (1937 - 2009) through EBSCO publishing, which is a comprehensive research database for nursing and allied health journals and indexed to Chiropractic Literature (2009 - 2010). In the biomedical literature, there is a number of expert opinion pieces that describe strategies for evaluating books and writing book reviews. Articles were collected using the following criteria: strategies for conducting scholarly book reviews, thematic issues related to the publication of scholarly book reviews, or recommendations on academic writing of which a section pertained to writing scholarly book reviews. Then, the articles that met the inclusion criteria were analyzed by the author and publication type, and narrative information concerning scholarly book reviews and their publication. In order to generate recommendations for conducting book reviews, the author's personal experiences in writing book reviews and acting as a journal editor were used to supplement the evidence gleaned from the articles included in the review. Twelve fine-grained aspects of good scholarly book reviews were identified in this research, as explained in Table 2.1.

However, mentioning all these features in a single review is unlikely to happen. 480 book reviews have been surveyed and it was found that the mean number of features commented on per review was 9.0 ± 2.7. With most reviews spanning 250 to 500 words, it is not possible to include a critique of all appraisal items evaluated. The number of features detected depends on the reviewer and the book. For instance, a student textbook with an index of limited utility is an important finding. However, the same finding in a patient handbook may not deserve a mention. Table 2.1 summarizes all the appraisal items related to a book review. So, it is important to recognize that appraisal item selection is specific to the book under review.

Table 2.1 : Appraisal documentation provided by Lee et al. [46]

Appraisal item	**Description**
Author(s) background and expertise	Author’s qualifications and previous contributions to the topic area to determine the author’s authority.
Book format/ Organization format/Organization	Organization and layout of the book.
Contents - Completeness - Accuracy - Current	Book should be read carefully to evaluate the book for accuracy, completeness, readability, and relevance.
Readability/ Style	The ability to understand the content easily.
Technical features - Table of contents - Chapter layout - Illustration - Typography - Tables - Figures - References - Index - Appendices	
What is unique	
Usefulness to the intended readership	Determine whether or not the contents are appropriate for the readership level. Evaluate aims and objectives from the perspective of the intended readership.
Were the goals of the book achieved	Determine whether or not the author’s intentions, aims and purpose for writing the book is achieved.

Comparison to competitors	Comparison with the key books in the domain.
Comparison to previous editions	Critically review the book compared to its previous edition.
Value of the book to the field of study	Determine the contribution that the book makes to the field
Value for price	Determine whether or not it is worth buying the book
Overall recommendation(s)	Critically identify factual mistakes, shortcomings and convey it in a professional manner. Use descriptive comments instead of conclusions. Provide general evaluation of the book.

This research also suggested that attempts should be made to place a book in a large, broader context to allow judgment of the book against its competitors and to allow for the determination of the book's contribution to its field. They also identified three major issues related to book reviews: conflict of interest, reviewer bias, and time lag in publication of reviews. A major issue that can affect the credibility of a book review is the influence of a conflict of interest. This happens when a reviewer has financial or personal relationships that inappropriately influence his or her actions. Reviewer bias can also influence book reviews. Reviewer bias has the potential to provide an inaccurate representation of the book in question and may negatively influence a readership's perceived value of the book review process. For most academic books, the first one or two years after publication is the period of its greatest sales. Therefore, most balanced reviews can be obtained during the early years. After a few years of publishing, the content might be outdated, and reviews can be biased when compared to new books in the same domain.

2.6 Data Sources and Preprocessing

Aspect sentiment analysis has evolved during the last decade; hence a lot of data sets are freely available for many product domains. Some datasets are annotated with aspects and sentiments while some of them are not annotated. Most of the annotated data sets are from the restaurant, movie, or product domains.

McAuley et al. [47], [48] extracted product reviews and metadata from Amazon, including 143.7 million reviews spanning from May 1996 to July 2014. The extracted dataset includes reviews (ratings, text, helpfulness votes), product metadata (description, category information, price, brand, and image features) and links (also viewed or also bought graphs). It contains information of 11 product categories (books, Cell phones and accessories, clothing, shoes & jewelry, digital music, electronics, grocery & gourmet food, home & kitchen, movies & TV, musical instruments, office products, toys & games). Extracted raw review data contains some duplicate reviews, mainly due to near-identical products whose reviews Amazon merges. For example, VHS and DVD versions of the same documentary can be taken. In the preprocessing stage, those duplications have been removed. In the final stage, duplicates were removed more aggressively by removing duplicates even if they were written by different users. This accounts for users with multiple accounts or even plagiarized reviews. After removing all duplications, the dataset is probably preferable for a sentiment analysis type task. The dataset format is one review per line in JSON. A sample review is shown in figure 2.2 and a description of the data is available in table 2.2.


```
{
"reviewerID": "A2SUAM1J3GNN3B″, "asin": "0000013714",
"reviewerName": "J. McDonald″,
"helpful": [2, 3],
"reviewText": "I bought this for my husband who plays the piano. He is having a
wonderful time playing these old hymns.
The music is at times hard to read because we think the book was published for singing
from more than playing from. Great purchase though!",
"overall": 5.0,
"summary": "Heavenly Highway Hymns", ″unixReviewTime": 1252800000,
"reviewTime": "09 13, 2009"
}
```


Figure 2.2 : Sample review [47] [48]

Table 2.2 : JSON Field description [47] [48]

JSON field	**Description**
reviewerID	ID of the reviewer
Asinz	ID of the product
sreviewerName	Name of the reviewer
Helpful	Helpfulness rating of the review, e.g. 2/3

reviewText	Text of the review
Overall	Rating of the product
Summary	Summary of the review
unixReviewTime	Time of the review (Unix time)
reviewTime	Time of the review (Raw)

SemEval is a rich source of data sets, especially for sentiment analysis. Their focus was twitter sentiment analysis [49] and aspect based sentiment analysis of the reviews in restaurants, laptops, and hotel domains [27], [50]. All the datasets used in SemEval sentiment analysis tasks were annotated with aspect category, opinion target expression and sentiment polarity. Socher et al. [51] used a data set of movie reviews from *rottentomatoes.com* to introduce their deep recursive neural tensor network (RNTN) for sentiment analysis. Hamdan et al. [2] used a French book review data set from the social science domain for aspect extraction and sentiment analysis.

2.7 Vector representation of words

Compared to traditional methods like one-hot encoding or bag-of-words model, word vectors capture information about word's meaning or context. This information is very vital in aspect extraction and aspect category identification where rich representation of words is needed to find the semantic textual similarity.

Perone et al. [52] did a comparative analysis on the performance of sentence embeddings in downstream and linguistic probing tasks. They have evaluated ELMo, FastText, GloVe, Word2Vec, p-mean, Skip-Thought, InferSent and USE(Universal Sentence Encoder) as sentence representation techniques. Evaluated tasks were binary and multi-class classification, entailment and semantic relatedness, semantic and textual similarity, paraphrase detection and caption image retrieval. Skip-Thought, p-mean, InferSent and USE are sentence embedding techniques. ELMo, FastText, GloVe and Word2Vec are word embedding techniques. Hence, traditional bag-of-words averaging was employed to produce the sentence embedding. Overall there was no clear winner and performance of sentence embedding techniques was depended on the nature of the task. When comparing word embedding techniques (ELMo, FastText, GloVe and Word2Vec) in semantic and textual similarity task, it can be observed that FastText has showed best performance after averaging the performance result of 6 corpora.

Devlin et al. [53] introduced a new language representation model called BERT which stands for Bidirectional Encoder Representation from Transformers. BERT considered the contextual representation of word unlike word2vec or GloVe. It considered the context after the word and before the word, hence bidirectional. Their experiment showed that performance of Semantic textual similarity task is better than ELMo. But there was no performance comparison with FastText.

All these variations of word vectors can be effectively used in sentiment analysis as well. Petrolito et al. [54] evaluating the impact of word embedding-based features in sentiment analysis tasks. They have critically evaluated effect of the size of the corpus used to train the embeddings, which text domains are good to train better embeddings and effect of learning method (word or character based word embeddings). Word2Vec and FastText has been used for experiments. They have identified that, regarding word based word embeddings, as the training corpus size increases the accuracy rises. Domain is not important for character-based word embeddings. SVM was used for experiments.

2.8 Summary

Lee et al. [46] critically discussed how to write a scholarly book review. The purpose of this research was to describe and discuss the processes used to write scholarly book reviews and identify potential aspects of a good book review. Research was conducted using three databases in Bio-medicine, nursing and allied health journals. Articles were collected using three criteria: strategies for conducting scholarly book reviews, thematic issues related to the publication of scholarly book reviews, or recommendations on academic writing of which a section pertained to writing scholarly book reviews. Twelve appraisal items of a good scholarly book review were identified: author background and expertise, book format/organization, contents, readability, technical features, what is unique, usefulness to the intended readership, were the goals of the book achieved, comparison to competitors, comparison to previous editions, value of the book to the field of study, value for price and overall.

The first known attempt for aspect-based sentiment analysis of scholarly book reviews was carried out by Hamdan et al. [2]. Based on the opinion of domain experts such as the OpenEdition editorial team that deal with book reviews of social and human sciences, they could identify seven potential aspect categories for French scholarly book reviews in the Social Science domain. These include book presentation, problematic, scientific context, scientific method, author’s arguments, book organization, and judgment about the book. However, the proposed aspect categories have not been properly matched to a taxonomy. Aspect categories such as problematic and scientific methods are defined basically considering the scholarly books in the Social Science domain. The corpus was generated using 200 book reviews in the French language. Each review was segmented to sentences and three annotators were used to extract opinion targets, aspect categories, and their polarities in each sentence. If any annotated opinion target was not under the defined set of aspect categories, they have been separately mentioned during the annotation. However, their Conditional Random Field (CRF) classifier was implemented to identify only explicit aspects, but not aspect categories, or implicit aspects.

Aspect extraction and categorization algorithms can be either supervised, semi supervised or unsupervised. However, the accuracy of the supervised models heavily depends on the accuracy of the labeled data, and engineering effort is always high. If the training data is not enough, supervised learning algorithms do not perform very well in the aspect extraction domain. Unsupervised clustering-based aspect categorization and dependency rule-based aspect extraction methods show promising results and don’t need any annotated or seed data to perform.

3 METHODOLOGY

Unsupervised aspect extraction methods are well suited for the book review domain due to the unavailability of a well-defined aspect taxonomy and difficulty to find an annotated data set. In this research, a hybrid approach that contains a rule-based aspect extraction algorithm and a clustering-based aspect categorization can avoid weaknesses of unsupervised learning approaches. Selected baseline clustering algorithm proposed by Chen et al. [11] for aspect category identification shows better results compared to state-of-the-art semi supervised and topic modeling based approaches.

Research can be divided in to five phases: 1. Building an aspect taxonomy for scholarly book reviews, 2. Scholarly book review corpus annotation, 3. Dependency rule-based aspect extraction, 4. Clustering based aspect categorization, 5. Evaluation of results.

3.1 Data extraction and preparation

In order to extract fine-grained aspects, reviews were extracted from the Computer Science domain. Amazon categorizes books in several ways. It has 37 categories and some of them are Art & Photography, Audible Audiobooks, Biographies & Memoirs, Books on CD, Business & Money, Science & Math, Children's books, Christian books & Bibles, Comics & Graphic novels, and Computer & Technology. The Computer & Technology category again divides in to 18 sub categories and computer science is one of them. A custom java-based data extractor was written to extract data from the Amazon website. The data extractor sends requests to each of book's URLs under the computer science category and extracts the HTML content of that page. Then, it preprocesses the content and filters out the unwanted fields. Then, the review text, overall rating, helpfulness score, reviewer name, product ID and summary from each review, is extracted. The helpfulness score and star rating were the main selection criteria for selecting the top 1000 reviews containing all the aspect categories. Helpfulness is the main indicator that shows how important that review is for purchase decisions. In the Computer Science domain, there are eight book categories: AI and Machine Learning, Bioinformatics, Computer Simulation, Cybernetics, Human-

Computer interaction, Information theory, Robotics, and System Analysis & Design. A total of 4480 reviews from AI and machine learning were extracted by sending requests to the Amazon website. Then, the data was divided in to five buckets according to the star rating, and the top reviews were selected from each bucket based on the helpfulness score (how many users have marked the review as helpful). The top 1000 reviews from AI and Machine Learning category were extracted for the experiments. Data distribution is showed in Figure 3.1.

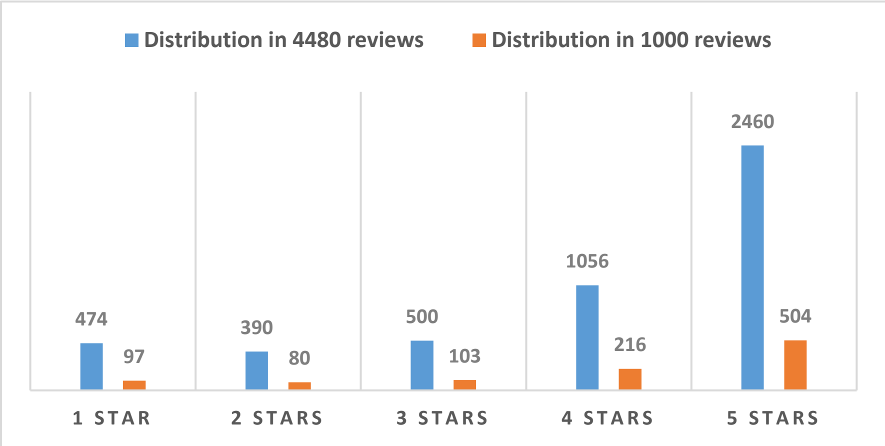

Figure 3.1 : Distribution of star rating in AI and Machine learning amazon book reviews

It can be seen that 50% of the extracted reviews are 5-star reviews. So, extracted reviews contain more positive reviews compared to negative reviews.

3.2 Aspect taxonomy for scholarly Book reviews

By considering the aspects defined by Hamdan et al. [2] and Appraisal items of a good book review discussed by Lee et al. [46], aspect taxonomy for scholarly book reviews is manually defined. Aspect taxonomy defined by Hamdan et al. [2] has only one level and seven aspect categories were defined specifically for the social science domain: Book presentation, Problematic, Scientific context, Scientific method, Author's arguments, Book organization and Judgment about the book. By looking at appraisal items mentioned by Lee et al. [46], it can be seen that scientific context and scientific method are related to the content of the book. Book presentation and book organization are mainly aligned with the readability of the book, and the judgment of the book mainly discusses aspects related to the worthiness of the book. Based on these observations, a new aspect taxonomy is derived.

Reviews mainly discuss about the book and the author. So, the book and author have been taken as the top two entities of the aspect taxonomy. Detailed aspect taxonomy is shown in Figure 3.2, where taxonomy is defined for three levels. Taxonomy is more focused on content and the quality of the book, since evaluating aspects like the quality of the paper and back cover does not seem to be very important when it comes to online purchases.

Under readability, there are four aspects defined: style, content, book structure and organization, book format and design. Readability contains more fine-grained aspects such as semantic elements like vocabulary and syntactic elements like sentence structure. Propositions, Organization, and Coherence are the three main aspects considered under content. Book structure and organization focus on the aspects such as table of contents, chapter layout, headings, navigation, references, index, and appendices. Typography, illustration, figures, and tables are some of the fine-grained aspects considered under book format and design.

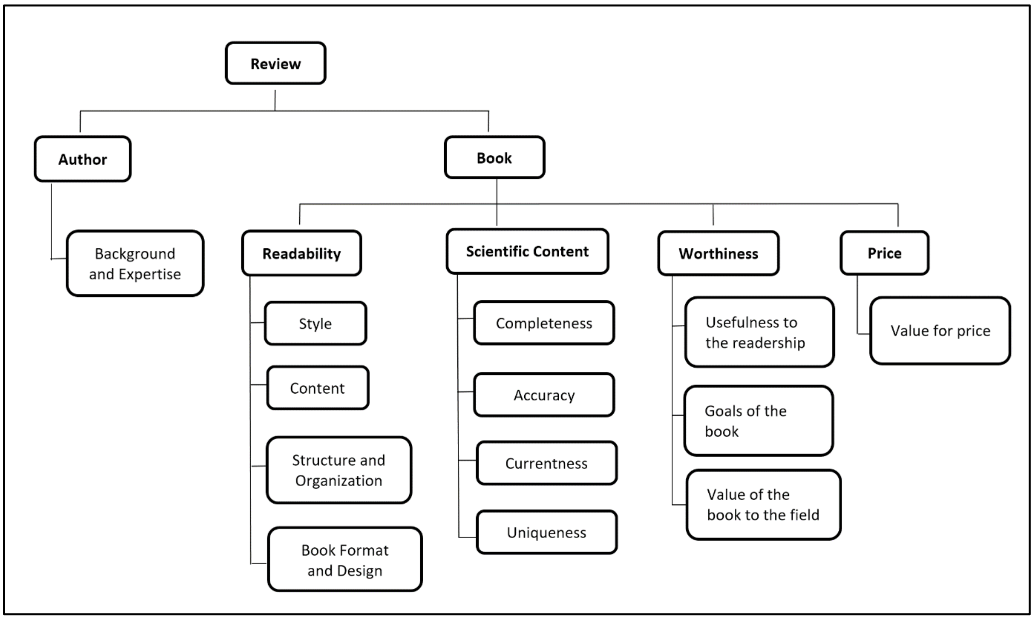

Figure 3.2 : Hierarchy of Aspects for scholarly book reviews

3.3 Scholarly book review corpus annotation

In order to validate the newly derived aspects, the corpus was annotated only with the first and second levels (Author: background and expertise, book: readability, scientific content, worthiness, price) of the taxonomy. Since this research problem is completely new to the sentiment analysis domain, identifying first and second level aspects with higher accuracy is the focus of this research. Without gaining higher accuracy at the top levels of the taxonomy, it is impossible to predict more granular aspects. When there was an ambiguity about the more fine-grained aspect, the sentence was tagged with the top-level aspect. For example, if a given aspect is not aligned with readability, scientific content, worthiness, or price, it will be tagged as "Book", which means that the review explains a general aspect of the book.

The top 100 reviews (based on the helpfulness score) from the extracted 1000 reviews from AI and machine learning were used for the tagging. Reviews were sampled so that it preserves the distribution of the original data set by considering the overall star rating of reviews. The original review corpus was divided according to the overall star review and the top 100 were selected according to the helpfulness score (how many people have marked the review as helpful) of the review. Figure 3.3 shows the distribution of the tagged data set.

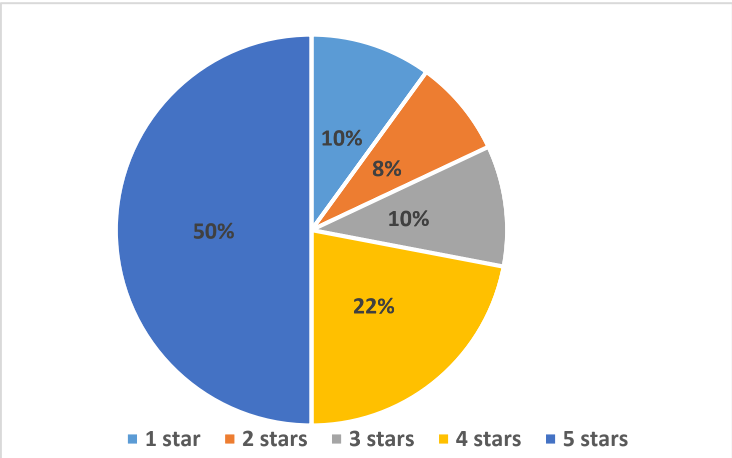

Figure 3.3 : Distribution of test data (100 reviews)

Reviews were tagged in two levels. At the first iteration, all the sentence and sentence phrases were tagged where both aspects and sentiment words are present. If there is no sentiment towards a particular aspect, it will not be useful at the sentiment analysis. Brat rapid annotation tool [55] was used for tagging and Figure 3.4 shows an example tagged review that identifies the sentence and sentence phrases containing opinions and aspects. Figure 3.4 only shows the first iteration of tagging whereas second level aspects of the aspect taxonomy are mentioned in the format *'<first level aspect>_<second level aspect>'*.

In the second iteration, all the explicit aspect terms and Implicit Aspect Clues (IAC) were tagged. Noun and noun phrases are considered as candidates for explicit aspects. For example, a sentence like *"There's little/no attempt to demystify concepts to the newcomer, and the exposition is all over the map"* explicitly mentions the aspect category *"Scientific Content"* using the aspect expressions "concepts" and *"exposition"*. Noun, noun phrases, verbs, and, adjectives can be candidates for IACs. IACs can either be single words or multiple word expressions and a common approach for IAC identification is to assume that sentiments or polarity words are good candidates for IACs, which might not be true all the time.

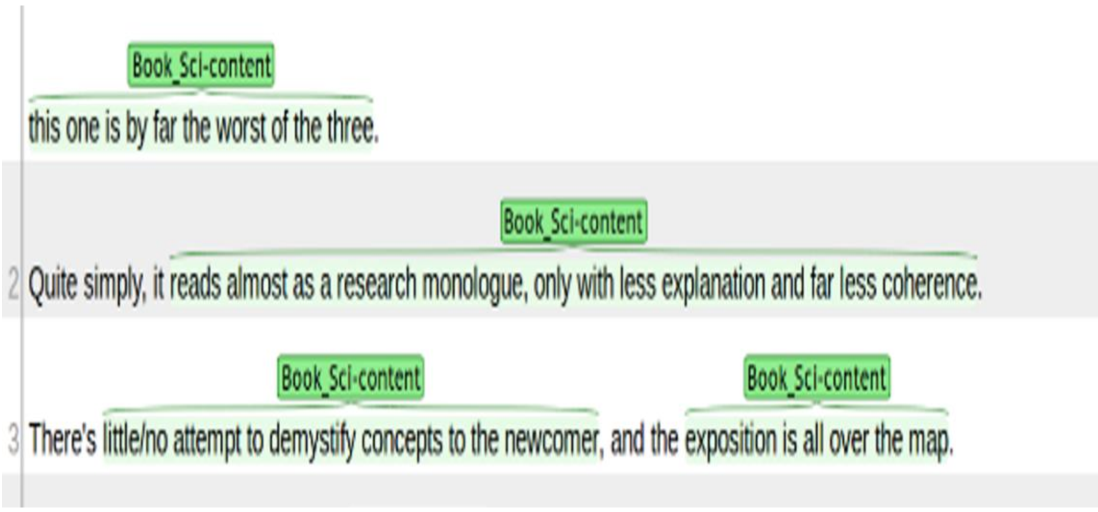

Figure 3.4 : Annotated review for identifying sentence phrases containing aspects and opinions using Brat annotation tool

For example, take the sentence *"they tried to squeeze every bit of information*. The word *"squeeze"* implies the readability of the book, which is a verb and not a sentiment or polarity word. But in a sentence like *"There simply isn't a clear, coherent path that the authors set out to go on in writing a given chapter of this text"*, the adjectives *"clear"* and *"coherent"* implicitly mention the readability of the book. But in the following examples, IACs are nouns or noun phrases: *"Ray Kurzweil wrote a thick volume combining 50's style naive technology optimism", "There are certainly good models out there", "Kurzweil wins Olympic gold in name-dropping with the singularity is near"*, *"Always suspicious the use of quotations of old or dead wise men to cover up the lack of content in a book"*. Sometimes complex phrases can be taken as candidates for an IAC. In the sentence, *"Sandra Blakeslee was much better and more logical with apt experiments documented to highlight every point"* there is an implicit mention of the aspect scientific content using the phrase *"highlight every point"*. End to end corpus annotation is explained in Figure 3.5.

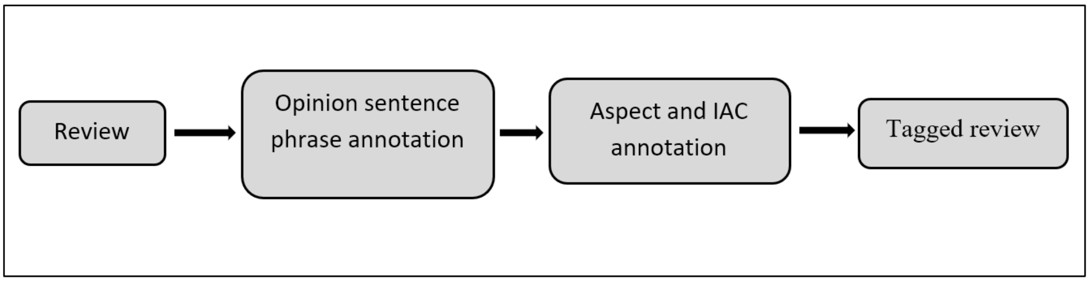

Figure 3.5 : Aspect annotation process

Two independent annotators tagged 100 reviews and compared the kappa values for the accuracy of the taxonomy and quality of the annotation. Annotation was carried out, not to train a model, but to validate the results of the rule-based aspect extraction and clustering-based aspect categorization. The average inter-annotator agreement on aspect annotation was $k = 0.764$ according to the Cohen's kappa statistics [56] The total 100 reviews were taken for tagging. Sentences with aspects can have single or multiple aspects in the same sentence. Tagging statistics of the sentences are given in Table 3.1.

Table 3.1 : Statistics about sentences with aspects

Domain	Sentences with n aspects			
AI and Machine learning	n = 0	n ≥ 1	n ≥ 2	n ≥ 0
	475	1243	172	1718

Sentences were tagged under seven aspect categories (Book, author, author expertise, book readability, book scientific content, book worthiness, book price) and they can be further categorized as implicit aspects and explicit aspects. Distribution of aspect categories over annotated data is shown in Figure 3.6.

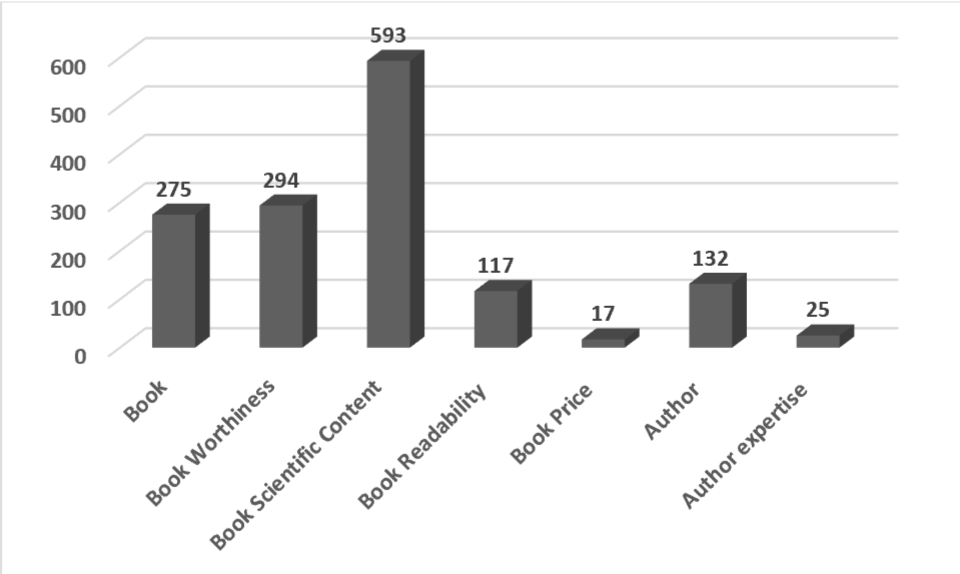

Figure 3.6 : Aspect distribution over each aspect category (total aspects=1453)

3.4 End to end system architecture of Aspect extraction and clustering

Data extraction, preprocessing, aspect extraction, and clustering create a single pipeline as shown in Figure 3.7. Aspect extraction is improved by introducing new rules to the dependency rule-based aspect extraction introduced by Poria et al. [1]. Aspect clustering is implemented by extending the work of Chen et al. [11]. The original algorithm extracts noun, noun phrases, adjectives, and verbs as candidates for the aspect terms and the important aspects are selected from the resulting clusters. But this approach extracts a huge amount of unwanted terms, and it has a substantial computational complexity, which they tried to solve by identifying seed clusters first and merging other aspects to those clusters. The dependency rule-based aspect extractor reduces the noise introduced to the clustering algorithm by extracting only the most potential candidates for aspects, hence no seed cluster identification is needed.

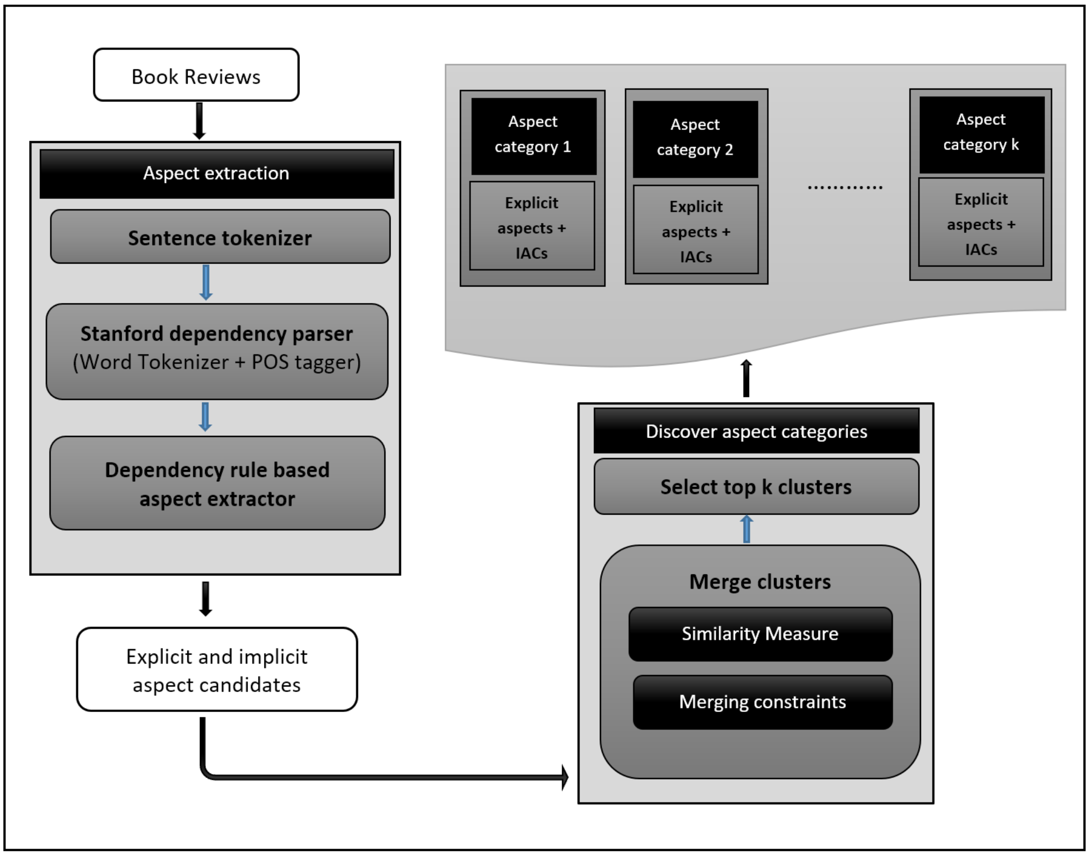

Figure 3.7 : Pipeline for extracting explicit and implicit aspects from Scholarly book reviews.

One review is considered as a single document and sentences are extracted from each review. The aspect extractor module extracts the explicit and implicit candidates and keeps track of the location of the original review as well. Extracted aspect candidates are fed in to the aspect clustering algorithm, where group aspects have higher probability to be in the same aspect category, while pruning unwanted aspect terms are extracted during the aspect extraction phase. These aspect clusters also have back reference to all the reviews and locations of the occurrences of each aspect term.

3.5 Dependency rule-based aspect extraction

Dependency rules proposed by Poria et al. [1] were used as the baseline for aspect extraction from book reviews. Four new rules were also introduced. The rules defined by Poria et al. [1] did not perform as expected for the book review domain and their performance results were only mentioned for explicit aspects. The implicit aspect lexicon was built prior to aspect extraction.

Stanford dependency parser is used to generate the sentence dependency tree. The proposed aspect parser is based on two general rules: Rules for sentences having subject verb and rules for the sentences that do not have a subject verb. The complete set of rules is discussed below. Rules which were defined by Poria et al. [1] were not changed and kept as it is. All the newly added rules are mentioned in Section 3.5.3.

3.5.1 IAC Lexicon

The top 100 annotated book reviews from AI and machine learning domain were taken and Implicit Aspect Clues(IACs) were extracted from each review. The main candidates for the IAC were, adjectives, adverbs, and verbs. Then, for each IAC in the list, synonyms and antonyms were obtained from pretrained word2vec model and added back to the IAC lexicon. Word2vec pretrained model was trained on google news corpus of 3 billion running words and model has 3 million 300-dimension English word vectors. Thus, a lexicon of 3337 IAC candidates was built. This IAC lexicon is used for some of the aspect extraction rules. SenticNet 5 is used as a concept-level opinion lexicon. The common-sense knowledge base contains around 100000 multi-word expressions labeled by their polarity scores.

3.5.2 Dependency rules proposed by Poria et al. [1]

If an active token is found to be a syntactic subject of a token and if the active token, *h* is in a subject noun relation with word, *t* then,

1. *t* is extracted as an aspect, if *t* has any adverbial or adjective modifier and the modifier exists in SenticNet [1].

Example: *"Many important concepts are skimmed over way too quickly."*

Here, the word '*concepts'* is in a subject relation with *skimmed* and modified by the adverb modifier *quickly*, so *skimmed* is extracted as the aspect.

2. If the sentence does not have any auxiliary verbs (*is, was, would, should, could, etc.*) then,

2.1. If the verb, *t* is modified by an adjective or an adverb or it is in *adverbial clause modifier* in relation with another token, both *h* and *t* are extracted as aspects [1].

Example: *"The book naturally will attract some buzz."*

Here, *book* is in a subject relation with *attract* and *attract* is modified by the adverb *naturally*, hence both the aspects *book* and *attract* are extracted.

2.2. If *t* has any direct object relation with a token, *n* and the POS of the token is Noun and *n* is not in SenticNet, then *n* is extracted as an aspect [1].

Example:

"In my own field, neurobiology, he mistakes models with complete, reverse engineered, functional reproductions of neural systems."

Here, *mistakes* have direct object relation with *models* and it is not in SenticNet, hence, the word "*models"* is extracted as the aspect.

2.3. If *t* has any direct object relation with a token, *n* and the POS of the token *n* is Noun and *n* exists in SenticNet, then the token, *n* is extracted as an aspect term. In the dependency parse tree of the sentence, if another token n_1 is connected to *n* using any dependency relation and the POS of n_1 is Noun, then n_1 is extracted as an aspect [1].

Example:

"Always suspicious the use of quotations of old or dead wise men to cover up the lack of content in a book."

Here, *suspicious* is in direct object relation with *use* and *use* connects to *quotations* using a dependency relation. Hence, both *use* and *quotations* are extracted as aspects.

2.4. If t is in open clausal complement relation with a token t_1, then the aspect t-t_1 is extracted if t-t_1 exists in the opinion lexicon. If t_1 is connected with a token t_2 whose POS is noun, then t_2 is extracted as an aspect [1].

Example: "*I'd hesitate to trust this author and publisher again*"

Here, *hesitate* is in open clausal complement relation with *trust* and *hesitate to trust* (t-t_1) is in opinion lexicon. Hence, *hesitate to trust* is extracted as an aspect.

3. A copula is the relation between the complement of a copular verb and the copular verb. If the token, t is in copula relation with a copular verb and the copular verb exists in the implicit aspect lexicon, then t is extracted as the aspect term.

   Example: *"This book is expensive"*

   Here, *"expensive"* is extracted as an aspect.

4. If the token, t is in copula relation with a copular verb and the POS of h is Noun, then h is extracted as an explicit aspect [1].

   Example: *"There's little attempt to demystify concepts to the newcomer, and the exposition is all over the map."*

   Here, the noun *exposition (h)* is subject relation with *all (t)* and all is in copula relation with *is*. Hence, *exposition* is extracted as an explicit aspect.

5. If the token, t is in copula relation with a copular verb and the copular verb is connected to a token, t_1 using any dependency relation and t_1 is a verb, then both t_1 and t are extracted as implicit aspect terms, if they exist in the implicit aspect lexicon.

   Example: "*The text is repetitive, confused, and often doesn't match up with the code and data sets to which it refers.*"

Here, *repetitive* is in copula relation with *is* and *confused* and *repetitive* connect with each other by a dependency relation. Hence, both *repetitive* and *confused* are extracted as implicit aspect terms.

When there is no subject noun relation in the dependency parse tree,

6. If an adjective or adverb or verb, *h* is in infinitival or open clausal complement relation with a token, *t* and *h* exists in the implicit aspect lexicon, then *h* is extracted as an aspect.

   Example: *"This book was satisfying and thought provoking, and I highly recommend it to anyone interested in the mysteries of the very large and the very small."*

   Here, *satisfying* is in open clausal complement relation with *provoking* and both the words exist in the implicit aspect lexicon, so *provoking* is extracted as an aspect.

7. If a token, *h* is connected to a noun, *t* using a prepositional relation, then both *h* and *t* are extracted as aspects [1].

   Example: *"Love the content of this book"*

   Here, *"content"* is extracted as an aspect.

8. If a token, *h* is in a direct object relation with a token *t*, *t* is extracted as the aspect [1].

   Example: *"Take the history and the personalities and ignore the analysis."*

   Here, *ignore* is in direct object relation with *analysis*, hence *analysis* is extracted as an aspect.

9. For each aspect term extracted above, if an aspect term *h* is in co-ordination or conjunct relation with another token *t*, then *t* is also extracted as an aspect [1].

   Example:

   "This book is amazing and easy to read"

   Here, *"amazing"* is extracted as the aspect first. As *"amazing"* is in conjunct relation with *"easy"*, then *"read"* is also extracted as an aspect.

10. A noun compound modifier of an NP is any noun that serves to modify the head noun. If *t* is extracted as an aspect and t has the noun compound modifier *h*, then the aspect *h-t* is extracted and *t* is removed from the aspect list [1].

Example:

"I ordered this Machine learning book, but It has many typos"

Here, *"Machine learning"* and *"book"* are in noun compound modifier relation, and only *"Machine learning book"* is extracted as an aspect.

3.5.3 Proposed new rules

Four new rules were added to the existing rules proposed by Poria et al. By looking at the annotated data set and their dependency graph, it can be verified that the following dependency rules are not covered by Poria et al.

11. If an active token is found to be a syntactic subject of a token, and if this active token, h is in a subject noun relation with word, t, and if t is in an open clausal complement relation with another token t_1, and t exists in SenticNet , then t is extracted as an aspect.

    Example: *"This book is <u>relevant</u> on too many levels to thoroughly <u>list</u>, but just a few include psych, engineering, algorithms, computational complexity, machine learning, AI, dynamic systems, education, consciousness, neurology, math."*

    Here, *"relevant"* is in open clausal complement relation with *"list"* and *"relevant"* exists in SenticNet, hence *"relevant"* is extracted as an aspect.

12. If the sentence does not have any auxiliary verbs (is, was, would, should, could, etc.) then, if t is in clausal complement relation with a token t_1 then the aspect $t - t_1$ is extracted if t, t_1 exists in the IAC lexicon. If t_1 is connected with a token t_2 whose POS is noun, then t_2 is extracted as an aspect.

    Example:

    "I must <u>point</u> out that the book is very math <u>heavy</u>."

    Here, *"point"* is in clausal complement relation with *"heavy"*, hence *"point out that the book is very math heavy"* is extracted as an aspect.

13. If an active token is found to be a syntactic subject of a token, and if the active token, *h* is in a subject noun relation with word, *t*, and if the sentence does not have any auxiliary verbs (is, was, would, should, could, etc.), and if *t* is in an open clausal complement relation with another token, then both *h* and *t* are extracted as aspects.

    Example: *"There is an established literature on mind design and <u>author</u> has <u>contributed</u> very <u>little</u> to it."*

    Here, "*contributed"* is in an open casual complement relation with token *"little"*, hence both *"author"* and *"contributed"* are extracted as aspects.

14. If there is no subject noun relation in the dependency parse tree, and if an adjective or adverb or verb, *h* is in dependent or clausal complement relation with a token, *t* and *h* exists in the implicit aspect lexicon, then h is extracted as an aspect.

    Example: *"Too <u>lightweight</u> for a practitioner to <u>learn</u> much from it other than the ML World of Pedro Domingos."*

    Here, "*lightweight"* is in dependent relation with *"learn"* and *"lightweight"* exists in the implicit aspect lexicon, hence extracted as an aspect.

This rule-based method is fully unsupervised and depends on the accuracy of the dependency parser, implicit aspect lexicon and the SenticNet, rather than a training corpus and supervised learning accuracy. Extracted aspects are fed in to the clustering algorithm, which further prune the unwanted aspects term during the clustering process and finds out possible aspect categories.

3.6 Clustering-based aspect categorization

The basic concept for the aspect clustering algorithm is taken from the algorithm proposed by Chen et al. [11], which shows superior results compared to other unsupervised aspect categories. The novel clustering algorithm is introduced based on the concepts proposed by Chen et al. [11] and Zhai et al. [34]. The clustering algorithm can be explained in four steps: vector representation of aspect terms, similarity measure calculation, clustering algorithm, and merging constraints.

3.6.1 Vector representation of aspects

Instead of using words in their surface form like in WordNet, a vector representation of words was used. Word vectors or word embeddings are multi-dimensional meaning representations of word. Based on the study by Handler [57], it could be identified that word embeddings can contains a much larger set of similar words compared to WordNet, which is only 118000. WordNet is a hand-crafted ontological representation of relationship between words. Its representations are symbolic and only words having ontological relations can be related. So, word embeddings shows better results compared to wordNet. Word vectors can be generated using algorithms like FastText, ELMo, BERT, word2vec or GloVe [52]. FastText [58] is a word vector representation model implemented by the Facebook AI research lab, which is actually an extension to the word2vec model.

GloVe has overcome limitations of Word2Vec by capturing the global context. The global context is captured by the statistics of word co-occurrences in a corpus while still capturing sematic and syntactic meaning as in Word2Vec. Although ELMo performs better than GloVe, its pre-trained models are trained on Wikipedia and news crawl data which is not much related to our research problem. ELMo vectors are either 3072-dimension or 1024-dimension which computationally complex compared to 300-dimension GloVe vectors and trained on pre-trained word embeddings from GloVe. FastText performs better since it can handle rare and out of vocabulary words since each word is taken as character of N-grams and BERT due to its ability to learn contextual relation between words and sub words. But pre trained model is required to find the word vectors and further fine tuning of the model if required. After evaluating the performance and complexity, GloVe and FastText can be considered as best models to find word vectors. Baseline clustering algorithm proposed by Chen et al. [11] used both wordNet and co-occurrence based similarity metric to find the similarity between aspect terms. Since GloVe training also performed on aggregated global word-word co-occurrence statistics from a corpus, it is more close representation to the original algorithm.

The Spacy NLP library has been used to find the vector representation of the aspect term, which is implemented on top of the GloVe algorithm. GloVe is an unsupervised learning algorithm to find vector representation for words, and training is performed on aggregated global word-word co-occurrence statistics from a corpus. That basically count show frequently a word appears in a context.

Word vectors were found by training the GloVe model on Common Crawl data, which contains 1.1 million unique 300-dimension vectors. This common crawl data is text written from blogs, news or comments. These pre-trained vectors are used to find the word embeddings of aspect terms. But in order to get optimal results GloVe model has to be trained with a domain specific corpus.

3.6.2 Similarity measure calculation

Similarity measure calculation is done in two steps. First, pairwise semantic similarity of aspect terms has been calculated. Then, each term has a similarity value compared to other terms in the extracted aspect list. Assume G is a *n x n* semantic similarity matrix, where G_{ij} is the cosine similarity between 300-dimension GloVe word vectors of x_i and x_j, $G_{ij} \in [0,1]$, $G_{ij} = 1$ when $i = j$, and $G_{ij} = G_{ji}$. A candidate, x_i can be represented by the i^{th} row in G meaning $g_{i:}$ represent x_i in terms of the semantic similarity with other aspect terms. Figure 3.8 explains this scenario.

$$
\begin{array}{r}
\\
\\
\\
g_{i:} \longrightarrow \\
\\
\end{array}
\begin{bmatrix}
G_{11} & G_{12} & \cdots & G_{1j} \cdots & G_{1n} \\
G_{21} & G_{22} & & G_{2j} \cdots & G_{2n} \\
\vdots & & \ddots & & \vdots \\
G_{i1} & G_{i2} & \cdots & G_{ij} \cdots & G_{in} \\
G_{n1} & G_{n2} & & G_{nj} \cdots & G_{nn}
\end{bmatrix}
$$

Figure 3.8: Semantic similarity matrix

Assume the word vectors of *aspect_term_i* and *aspect_term_j* is $\vec{u}$ and $\vec{v}$, then cosine similarity between *aspect_term_i* and *aspect_term_j* can be calculated as in the following equation (1).

$$
G_{ij} = cosine(\vec{u}, \vec{v}) = \frac{\vec{u}.\vec{v}}{\|\vec{u}\|\|\vec{v}\|} = \frac{\sum_{i=1}^{n} u_i v_i}{\sqrt{\sum_{i=1}^{n} u_i^2}\sqrt{\sum_{i=1}^{n} v_i^2}} \qquad (1)
$$

Then, vector $g_{i:}$ can be represented as $\langle G_{i1}, G_{i2}, \dots, G_{ij}, G_{in} \rangle$. In order to calculate the semantic similarity between aspect terms x_i and x_j, cosine similarity between respective vectors $g_{i:}$ and $g_{j:}$ is taken as in equation (2).

$$sim_g(x_i, x_j) = 1 - \ cosine(g_{i:}, g_{j:}) \qquad (2)$$

$sim_g(x_i, x_j)$ calculates the comparison between $g_{i:}$ and $g_{j:}$. Similar row vectors in G indicate similar semantic meanings of two terms (e.g. "price" and "inexpensive"). Based on the GloVe based similarities between the candidates, two distance measures are defined for clustering as in equation (3) and (4). Since $G_{ij} \in [0,1]$, the values of $sim_g(x_i, x_j)$ also range from 0 to 1. When $i = j$, all the similarity metrics between x_i and x_j is 0.

$$dist_{avg}(C_l, C_m) = \frac{\sum_{x_{i'} \in C_l} \sum_{x_{j'} \in C_m} (1 - sim(x_{i'}, x_{j'}))}{|C_l| \ x \ |C_m|} \qquad (3)$$

$$dist_{rep}(C_l, C_m) = sim(argmax_{x_{i'} \in C_l} f(x_{i'}), argmax_{x_{j'} \in C_m} f(x_{j'})) \qquad (4)$$

$dist_{avg}(C_l, C_m)$ calculates the average candidate distances between cluster C_l and C_m. $dist_{rep}(C_l, C_m)$ calculates the distance between the most frequent terms (representative terms) of the two clusters. Two clusters describing a same aspect should be close to each other in terms of both average distance and representative distance. Therefore, the final distance is defined as the maximum of average distance and the representative distance, as in equation (5).

$$dist\ (C_l, C_m) = \max(dist_{avg}(C_l, C_m), dist_{rep}(C_l, C_m)) \qquad (5)$$

Sharing words is an important clue that can be used to cluster similar aspects, as explained by Zhai et al. [34]. Many aspect expressions are phrases consisting of multiple words, e.g., "Machine Intelligence", "Machine Learning", and "Machine Vision", share "Machine" as their common word. Aspect terms sharing some words are likely to belong to the same group or cluster. But this constraint can be violated in some occasions. Hence, this is introduced as an exponentially decaying function to the clustering algorithm, as in equation (6).

$$sim\left(x_i, x_j\right) = sim_g\left(x_i, x_j\right) e^{\frac{-w_{i\cap j}}{w_{i\cup j} - w_{i\cap j}}} \quad (6)$$

Each aspect term has a list of words that have been found in other aspect terms. These sharing words are considered after removing stop words. So $w_{i\cap j}$ is the sharing words common to both aspect terms x_i and x_j, and $w_{i\cup j}$ is the union of the sharing word of aspect terms x_i and x_j. This ensures that when there are no sharing words common to x_i and x_j, $sim\left(x_i, x_j\right) = sim_g\left(x_i, x_j\right)$ and when all the sharing words are common to both aspects, $sim\left(x_i, x_j\right) = 0$.

3.6.3 Merging constraints

Merging constraints proposed by Chen et al. [11] are been used without any further changes. Merging constraints further improve the clustering performance. Two clusters cannot be merged if they violate any of the following merging constraints.

1. The distance between two clusters should be less than the given threshold μ (Algorithm 3.1)
2. There must be at least one noun or noun phrase existing in one of the two clusters. Most of the time, noun and noun phrases represent explicit aspects, hence merging aspects with only IACs will be less as it is very difficult to identify an aspect category without an explicit aspect term.
3. The sum of the frequencies of the candidates from two clusters co-occurring in the same sentence must be higher than the sum of frequencies of them co-occurring in the same document but difference sentences. because there is a higher probability of different aspects being in different sentences in a review and the same aspect in a small window like in the same sentence.

3.6.4 Clustering Algorithm

Chen et al. [11] have extracted noun, noun phrases, adjectives and verbs as candidates for the aspect terms and important aspects are selected from the resulting clusters. But this approach extracts a huge amount of unwanted terms and computational complexity is very high. This problem is answered by this research by introducing a dependency rule-based aspect extractor before the clustering algorithm. The dependency rule-based aspect extractor reduces the noise introduced to the clustering algorithm by extracting the most potential candidates for aspects.

INPUT: $X = \{x_1, x_2, \ldots.., x_n\}, k, \mu$
OUTPUT: $\{A\}_{j=1}^{k}$
[1] Set $C_1 \leftarrow \{x_1\}, \ldots \ldots, C_n \leftarrow \{x_n\}$;
[2] Set $\emptyset \leftarrow \{C_1, \ldots \ldots, C_n\}$;
[3] **while** *there exists a pair of mergeable clusters* from $\emptyset$ **do**
Select a pair of closest clusters C_l and C_m such that VIOLATE-CONSTRAINTS *(C_l, C_m, μ)* is false;
$C_m \leftarrow C_l \cup C_m$
$\emptyset \leftarrow \emptyset - \{C_l\}$
[4] $\{A\}_{j=1}^{k} \leftarrow SELECT(\emptyset, k)$

Algorithm 3.1 : Clustering for Aspect Discovery

Algorithm 3.1 illustrates the clustering process. Aspect candidates generated by the dependency rule-based aspect extractor is input to the clustering algorithm as a set X that contains *n* candidate terms, a natural number, *k* indicating the number of aspects, and real number, μ indicating the upper bound of the distance of two mergeable clusters.

The clustering process starts with every term x_i in its own cluster C_i and $\emptyset$ is the set of all clusters. Then at each iteration, a pair of clusters (C_l and C_m), that are most likely to represent the same aspect category are merged together. A newly derived GloVe based domain-specific similarity measure, and merging constraints proposed by Chen et al. [11] are used to determine the similarity between two clusters. Merging constraints further ensure that terms from different aspects are not merged. The clustering process stops when there are no pairs of clusters satisfying the constraints. Finally, *k* clusters are selected from $\emptyset$ as potential aspects. Those *k* clusters are selected so that the frequencies of the members in the cluster has the highest sum.

3.6.5 Novelty of the proposed algorithm

The proposed new algorithm improves the algorithm proposed by Chen et al. [11] in three directions.

1. Replaced the WordNet and NPMI based semantic similarity calculation of aspect terms by GloVe based word vector representation. Instead of using words in their surface form, a vector representation of words was used. 300-dimensional word vectors are a better representation compared to wordNet.
2. Chen et al. [11] identified seed clusters based on the frequency of occurrence of aspect terms and the main assumption behind that approach is that frequently mentioned terms are more likely the actual features of customer interests. The second reason is that the clustering algorithm requires pairwise distances between candidates and it could be very time consuming if there is a large number of candidates. But in this research, we already filter a large number of false positives at dependency rule-based aspect extraction, so time complexity is not high compared to the original algorithm where all the noun, noun phrases, adjectives, and verbs are taken as candidates for the clustering. The other reason is that there can be aspects and aspect categories with low frequency and Chen's algorithms can't capture them during the seed cluster generation process. So, the new algorithm does agglomerative clustering on the entire aspect candidate set.
3. Introduced a soft constraint to include the effect of sharing words. When aspect expressions consist of multiple words, sharing words is an important clue and in aspect expressions, some sharing words are most likely to belong to the same aspect category or cluster.

4 EXPERIMENTAL SETUP AND RESULTS

Experiments were carried out to validate the performance of the aspect extraction and the performance of the aspect categorization algorithm.

4.1 Evaluations on Aspect extraction

Performance of the aspect extraction is calculated using the gold standard dataset. Precision, recall, and F-score are calculated for the rules proposed by Poria et al. [1] and newly added rules are listed in Table 4.1. The table compares the cumulative performance of both explicit and implicit aspect extraction.

Table 4.1 : Aspect extraction performance

Algorithm	**Precision (%)**	**Recall (%)**	**F-Score (%)**
Poria et al.	74.40	83.83	78.83
Poria et al. + New dependency rules	73.98	**88.44**	**80.56**

Since there is no previous research carried out in the book review domain, comparative evaluation of experimental results is not performed here. But proposed new rules improved the recall considerably while having negligible decrease in precision. This suggests that the new rules are able to identify the true positives accurately, without extracting false positives.

The dependency rules defined by Poria's et al. [1] were unable to identify five dependency relations,

1. A term that has an open clausal complement relationship with another token and the aspect term exists in SenticNet.
2. If a term has clausal complement relation with another term and it exists in the IAC lexicon, the complete phrase between those two tokens is considered as an aspect. The sentence should not have any auxiliary verbs.

3. If a term, t has a clausal complement relation with another term, t_1 and that term also has connected to another term, t_2, which is a noun, then that noun term, t_2 is also extracted as an aspect. The sentence should not have any auxiliary verbs.
4. If a term, h is in a subject noun relationship with another term, t and the sentence does not have any auxiliary verbs, then if term, t is in an open clausal complement relation with another term, both h and t are extracted as aspects.
5. If there is no subject noun relation in the dependency parse tree, and if there is an adjective or adverb or verb, h in a dependent or clausal complement relation with another token, t and h exists in the IAC lexicon, then h is extracted as an aspect.

These five dependency rules increase the recall by 5.5% and the F-score by 2.2% while decreasing precision only by 0.56%. Table 4.2 shows some of the explicit and implicit aspects extracted by the dependency rules.

Table 4.2 : Extracted explicit and implicit aspects

Explicit Aspects	**Implicit Aspects**
book, volume, arguments, exposition, errors, symbols, scientist, extrapolations, audience, subject, readers, charts, index, answers, thesis, strengths, topic, biography, notation, explanations, equations, ideas, content, figures, insights, theory, introduction, references, knowledge, style, writer, examples, derivations, methods, concepts, tips, answers, classification, principles, exercises, textbook, predictions, forecasts	skimmed, readable, bothered to buy, illustrate, mentions, wasted, organized, worth, veracity, recommend, make you aware, valuable, disappointed, applied, explain, biased, hard to read, emphasis, described, edits, sound very helpful, make you think, needs to sharpen, analyzes, demonstrates, structures of the question, covers, gives the exact algorithm to implement, recommend, surprised how author even know, want to use, willing to learn, code, spends, quoted, readable, easier, essential reading, going to benefit, easy to read, hard to read, satisfying, covers texturing, coherent, try to slog, wasted hours studying, known

Table 4.3 shows a set of the aspects mentioned in Table 4.2 that can be explained using the dependency rules associated with them. Dependency rules 3, 5, 6, 12 and 14 are using the IAC lexicon to extract aspects. No aspect has been extracted from

dependency rules 3, 6, and 7. So only rule 5, 12 and 14 actively use the IAC lexicon to extract the aspect. The same aspect term can be extracted from a different dependency rule as well. Therefore, the algorithm filters and merges all the duplicate aspects at the end.

Table 4.3: Dependency rule relation with extracted aspects

Aspect term	**Dependency rule**	**Sentence dependency relation**
Known	1	This book is **very** well ***known***
Readers	2.1	There are so many more, great, stories that most ***readers*** will **still** ***enjoy*** the book.
Errors	2.2	The 4th printing coming out this month will surely **fix** some ***errors***, but there are just too many.
Volume	2.3	Ray kurzweil ***wrote*** a thick ***volume*** combining 50 's style naive technology optimism, uncritical extrapolation of current trends and somewhat more than half knowledge of biology.
Hard to read	2.4	It was about 10 years ago when I first found turing's original paper on internet and thought it wouldn't be so ***hard*** *to* ***read*** and understand it.
Technique	4	First, it provides enough theory to allow a potential user to understand the essential insights that motivate specific techniques and to evaluate the situations in which those ***technique*** **are appropriate**.
Understand	5	Theory is there to aim the reader as to ***understand*** the purpose and the r labs at the end of each chapter are as ***valuable*** than the end of chapter exercises.
Nltk library	8	**Using** the ***nltk library*** or plan to do so.
Useful	9	Segaran has done an excellent job of explaining complex algorithms and

		mathematical concepts with clear examples and code that is both **easy to read** and ***useful***.
Computer science	10	the book is so readable that I usually forget I'm reading a very technical book that goes in to very core of **computer *science***.
Wasted	11	I ***wasted*** hours **studying** irrelevant information.
believes that computer modeling of brain functioning will yield	12	More importantly, he ***believes** that computer modeling of brain functioning will **yield*** the ***algorithms*** we need in order to eventually achieve an artificial general intelligence.
Contributed	13	there is an established literature on mind design and **kurzweil** has ***contributed*** very **little** to it.
Lightweight	14	too ***lightweight*** for a practitioner to **learn** much from it other than the ml world of pedro domingos.

Performance of each dependency rule sows in Table 4.4. Since rule 9 and 10 are supporting rule for other rules, it haven't been considered to calculate number of aspects extracted. Table 4.4 shows that, 65% of the aspects are extracted from second dependency rule (Section 3.5.2, 3.5.3) which does not use the IAC lexicon. It uses subject noun relation, adjective/adverbial modifier, adverbial clause modifier, direct object relation and open clausal complement relation as dependency relations to create rules. Only 12.6% of the extracted aspects have been used IAC lexicon.

Table 4.4: Number of aspects extracted by each dependency rule

Dependency Rule	True positives (TP)	Fales Positives (FP)	Precision (%)
1	119	105	53.13
2.1	326	80	80.30
2.2	138	38	78.41
2.3	212	80	72.60
2.4	160	4	97.56

3	1	0	100
4	124	46	72.94
5	80	7	91.95
6	0	0	-
7	0	0	-
8	17	20	45.95
11	6	53	10.17
12	37	5	88.10
13	64	14	82.05
14	1	0	100

4.2 Evaluations on discovering aspect categories

The performance of the aspect clustering is also calculated using the annotated dataset. Rand index, homogeneity, completeness, and V-measure are the measures used to evaluate the clustering algorithm. Clustering performance was measured in two steps. First, aspects of the gold standard dataset were input to the clustering algorithm where there were no false positives and the performance of aspect category identification was measured. Selecting top clusters based on the frequency of occurrence of aspects in the corpus is not considered since there are no false positives to filter. Number of clusters in the gold standard dataset is sever.

Secondly, the output of the feature extraction algorithm was input to the clustering algorithm where false positives are also there. There are 1453 aspects in the gold standard dataset and only 1285 are extracted at the aspect extraction phase. 452 false positives are added on top of that. Therefore, 1737 terms were input to the clustering algorithm. All the false positives are considered as one cluster when calculating rand index, homogeneity and completeness. The top cluster selection step in algorithm 3.1 is implemented. The quality of the extracted aspects (precision, recall, rand index) are calculated after selecting the top clusters to validate the performance of the end-to-end flow.

Clustering performance is calculated over different combinations of clustering hyperparameters to identify the best combination of parameters. The upper bound of distance (*μ)*, was changed to find how they impact the final results of aspect clustering. Figure 4.1, 4.2, 4.3, and 4.4 show how the Rand Index, V-measure, Homogeneity and completeness of the gold standard dataset and aspect candidates were extracted from the dependency rule-based aspect extractor changes to different distance upper bound values(*μ)*. Based on the experiment results $\mu = 0.03$ is taken as the optimum distance upper bound for the gold standard dataset and aspect candidates extracted from dependency rules. Table 4.5 shows the aspect clustering results for both the gold standard dataset and aspects extracted from dependency rule-based aspect extraction. Since the generated clusters are much higher than the condensed aspect categories, completeness and rand index are expected to be low, but homogeneity shows the quality of the generated clusters.

As shown in Figure 4.1, at first, the rand index increases slowly since the similarity algorithm starts clustering similar aspect terms. But there is a rapid decrease in performance parameters when $\mu > 0.03$. This is because more clusters are allowed to be merged as we increase the distance threshold, which is good at first, but then it introduces more noise by allowing wrong clusters to merge.

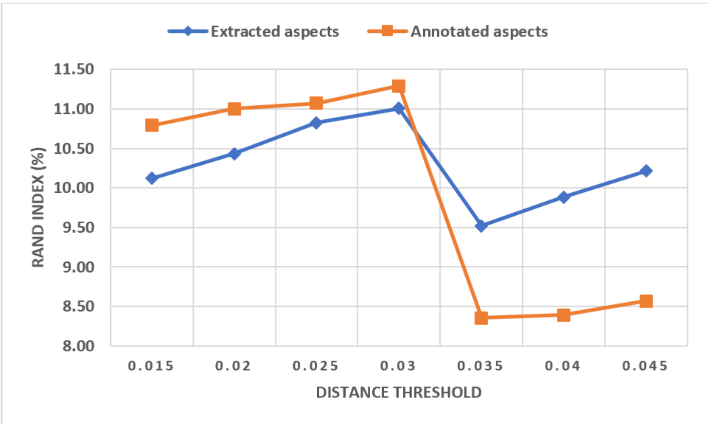

Figure 4.1 : Rand Index of aspect clustering for different distance thresholds (μ)

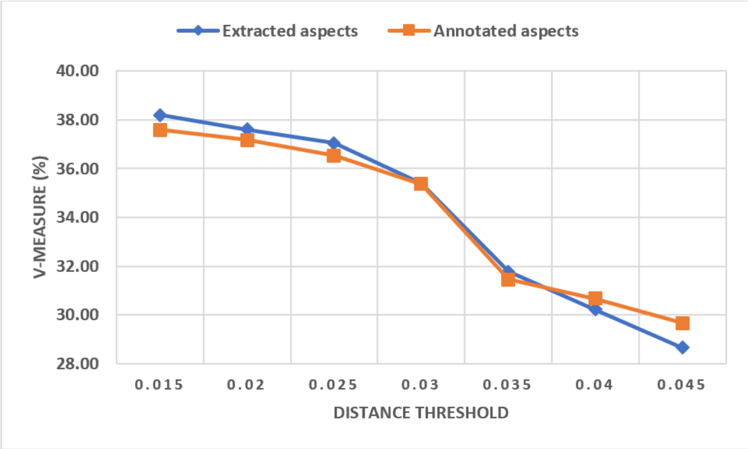

Figure 4.2 : V-measure of aspect clustering for different distance thresholds (µ)

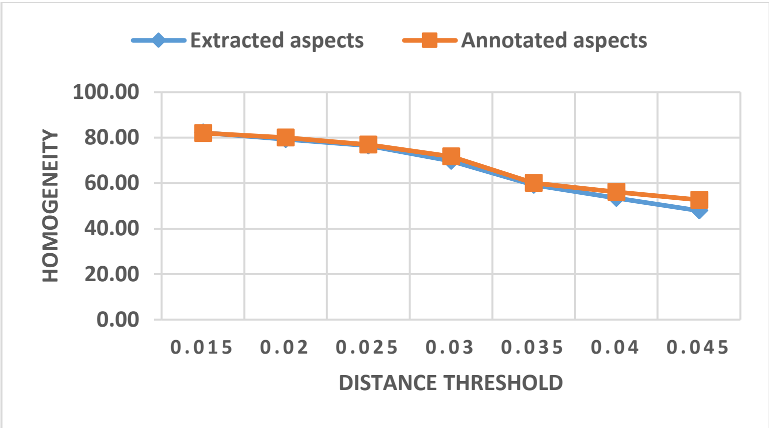

Figure 4.3 : Homogeneity of aspect clustering for different distance thresholds (µ)

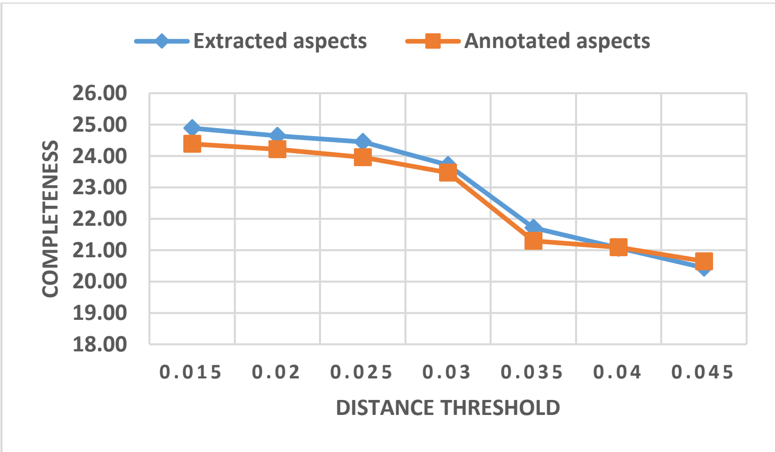

Figure 4.4: Completeness score of aspect clustering for different distance thresholds (µ = 0.03)

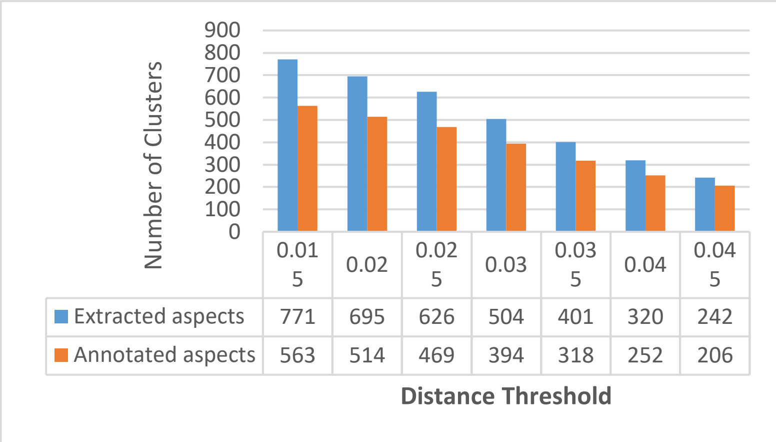

Figure 4.5 : Cluster distribution for different distance thresholds

Table 4.5: Aspect clustering performance for different distance threshold values (DT-distance threshold, N- no of clusters, RI-rand Index, Ho-homogeneity, Co-completeness, V- V measure)

DT	**Aspects extracted from dependency rules (N=1737, Aspect categories = 8)**					**Gold standard aspects (N = 1453, Aspect Categories = 7)**				
	N	**RI**	**Ho**	**Co**	**V**	**N**	**RI**	**Ho**	**Co**	**V**
0.015	771	10.12	82.19	24.89	38.20	563	10.79	82.04	24.38	37.59
0.02	695	10.43	79.33	24.64	37.60	514	11.00	79.94	24.22	37.17
0.025	626	10.82	76.51	24.45	37.05	469	11.07	76.91	23.96	36.54
0.03	**504**	**11.01**	**69.72**	**23.71**	**35.39**	**394**	**11.29**	**71.72**	**23.47**	**35.37**
0.035	401	9.52	59.16	21.71	31.76	318	8.36	60.11	21.29	31.45
0.04	320	9.89	53.41	21.07	30.22	252	8.39	56.17	21.09	30.67
0.045	242	10.22	47.89	20.44	28.65	206	8.57	52.63	20.65	29.66

When selecting the top clusters, different percentages of clusters generated from the dependency rule-based algorithm were selected. Figure 4.6 shows precision and recall changes for different cluster percentages. Based on the precision recall tradeoff, 60% is selected as the optimal percentage. The top 60% of clusters having the highest sum of members' frequencies of occurrence, are selected. Table 4.6 shows the precision, recall, and F-score after selecting the top clusters for each distance threshold at the clustering phase.

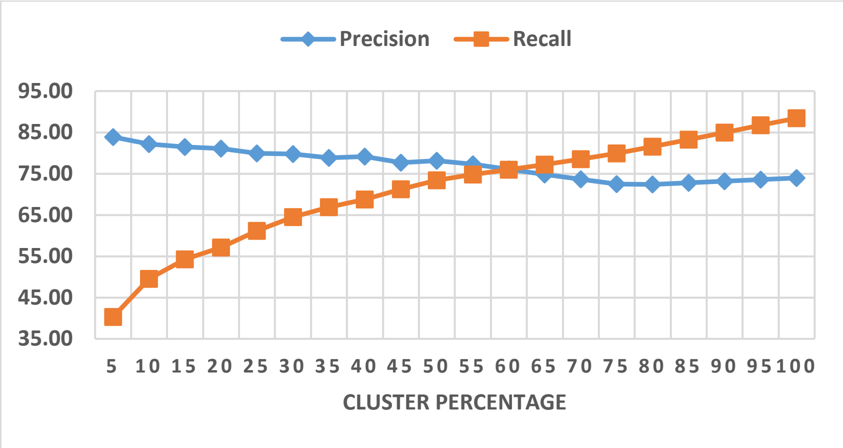

Figure 4.6 : Precision-Recall curve for different cluster percentages (μ = 0.03)

Table 4.6 : Quality of dependency rule-based aspects after aspect clustering

Distance Threshold	**Top 60% clusters**	**Aspect extraction performance (%)**			**Aspect clustering performance (%)**			
		Pre	**Re**	**F1**	**Ho**	**Com**	**V**	**RI**
0.015	463	75.37	69.35	72.23	78.68	27.02	40.22	15.40
0.02	417	76.08	71.34	73.63	75.76	26.56	39.33	15.23
0.025	376	76.46	73.12	74.75	72.93	26.13	38.48	15.16
0.03	**303**	**76.04**	**75.99**	**76.02**	**66.18**	**25.00**	**36.29**	**14.41**
0.035	241	75.90	78.49	77.18	55.51	22.44	31.96	11.81
0.04	192	76.07	80.53	78.24	50.52	21.63	30.29	11.75
0.045	146	75.20	82.38	78.62	45.18	20.75	28.44	11.55

based on the results of different distance threshold values, $\mu = 0.03$ is selected as the optimum distance threshold. After the top clusters are selected, precision of the aspect extraction improved by 2.79%, while the F-measure declined by 5.64%. Clustering V-measure improved by 2.54% and the rand index improved by 30.88%. The clustering algorithm was able to remove false positives and noisy clusters from the extracted aspect candidates so that the precision and rand index improved significantly.

Exponentially decaying function (equation (6)), introduced to absorb the significance of having sharing words between aspects terms for similarity calculation, also improved the results significantly in every performance measure. It improved the rand index by 18.6% and the F score by 1%, after selecting the top 60% clusters. Table 4.7 shows the improvement of all the performance parameters.

Table 4.7: Aspect extraction and categorization performance after selecting top 60% of the generated clusters (μ = 0.03)

	All Clusters		**Top 60% clusters**	
	sim_{gs}	***sim_g***	***sim_{gs}***	***sim_g***
Clusters	504	537	303	323
Homogeneity	69.72	64.48	66.18	60.01
Completeness	23.71	22.06	**25.00**	22.98
V-measure	35.39	32.87	**36.29**	33.25
Rank Index	11.01	9.11	**14.41**	12.15
Precision	73.98		**76.04**	75.51
Recall	88.44		75.99	75.07
F score	80.56		76.02	75.29

Table 4.8 shows the sample clusters that were generated. In a majority of the aspects, the aspect terms did not match with the aspect category and false positives are explicitly mentioned.

Table 4.8: Example aspect clusters generated from the clustering algorithm (μ = 0.03, top 60% clusters)

Book Scientific Content: Knowledge, Strengths, approach, emphasis, concept, makes difficult mathematical concepts accessible, sense knowledge, level approach, approaches
Book Scientific Content: Machine learning, write machine learning, machine, learning machines, support vector machines, machine learning techniques

Book Scientific Content: presented, described, proposed, introduced, adopted, represented, follows
Book Readability: Chapter, chapters, sample chapter
Book Worthiness: Recommend reading, recommend purchasing, recommend
False Positives : gotten, kept, happened, stuck, felt, gone, took, came, got, missed(readability), felt(Scientific content), guy(Author)

5 DISCUSSION

5.1 Aspect taxonomy

Aspect taxonomy is created considering four factors:1. Readers should be able to identify new, innovative and ground-breaking books, 2. Be warned of books of poor quality and those that may not be related to the field, 3. Scholarly books are expensive, 4. Scholars have limited time to commit to reading and 5. Readers should be informed about expert authors in the field. These factors can play a large role in influencing the development of future editions as well. Scientific content, worthiness, price, readability, and author expertise are the aspect categories proposed in this research to measure sentiment towards each of these factors in the book review.

According to the annotated dataset, 40.81% of the annotated aspects are related to scientific content of the book, and 20.23% of aspects are related to worthiness and value of the book to the field. This is because most of the scholar book reviews are written to evaluate the content of the book since many readers want to identify new, innovative, and ground-breaking books. 18.92% of the aspects are related to other aspects about the book, except worthiness and scientific content. Most of them are general sentiments about the book mentioning the book as an entity in the review. 8.05% of the aspects are about the readability of the books, including table of contents, typos, headings, references, illustrations, figures, and tables. 9.08% of the aspects are about the author and 1.72% of the aspects specifically relate to the author's expertise in the field. The authors have invested a lot of time and effort in writing their books and it is not surprising that an author would be curious as to how readers perceive their books. So, many reviewers provide the recognition and appreciation they deserve. Only 1.17% of the aspects are about the price of the book. This is expected for scholar book reviews since readers normally don't compare the price of the book when they need to select one book out of several books published on the same subject content. This is again justifying the reason for having 61.05% of the aspects on the content of the book and worthiness to the field. 71.99 % of the aspects are annotated with 2nd level aspects of the aspect taxonomy and only 28.01% aspects are annotated with the 1st level aspects (book or author) since their aspect category is not defined in the aspect taxonomy explicitly or they are just high-level sentiment expressions about the book or author. This shows that the proposed aspect taxonomy is suitable for scholarly book reviews with kappa statistics 0.76.

5.2 Aspect extraction

The proposed aspect extraction algorithm is fully unsupervised and heavily depends on the accuracy of the dependency parser and the opinion lexicon. Hence, non-grammatical sentences can be parsed without properly generating the dependency graph and aspects might not be extracted properly. The proposed framework only leverages on common-sense knowledge and on the dependency structure of the sentences. 5.6% increase in recall, 2.4% increase in F-score while only losing precision by 0.55% compared to rules mentioned by Poria et al. [1] ensures that the rationale behind new rules match with the domain and are able to identify the true positives accurately. The accuracy of the dependency rules 3, 5, 6, and 14 depend on the accuracy of the implicit aspect lexicon, which is created using a 100 book reviews in the computer science domain. But in our data set, only dependency rule 5 and 14 have been used. So, this is a clear example that importance and priority of dependency rules can be changed from domain to domain. Instead, specifically looking at the potential implicit aspect terms, adjectives, adverbs, and verbs of annotated sentences are taken as seeds to create the lexicon. Then, unnecessary terms are included in the lexicon. Therefore, more accurate lexicon generation by manually annotating IACs in a larger corpus will improve the accuracy of the results. More dependency rules can be discovered to increase the recall as well.

5.3 Aspect clustering

Similarity measurement of the aspect clustering is based on the semantic similarity-based distance metric of the aspect terms, and common words shared between the aspect terms. So, the number of clusters generated from the clustering algorithm is 303, which is considerably high compared to the seven fine-grained aspect categories. This is due to two reasons; first there are aspects terms in the same aspect category, but semantic similarity distance can be different. For example, implicit aspect *“described”* and explicit aspect *“support vector machines”* are both under the aspect category *“Book content”* but they are in two different clusters, as mentioned in Table 4.8. The other reason is that the algorithm operates in a very small distance threshold range from 0.015 to 0.045. Increasing the distance threshold will decrease the number of clusters but performance will dramatically decrease since clusters from different aspect categories will merge together.

Vector representation of aspect terms were taken from a pre-trained GloVe word vector, which has been built using written texts on blogs, news, and comments. There are one million 300-dimension word vectors in that pre-trained model. But this is more of a common vector representation of words. By using a domain-specific corpus to generate word vectors, the accuracy of the clustering algorithm can be increased.

6 CONCLUSION AND FUTURE WORK

This research shows promising results from three directions. There was no well-defined aspect taxonomy for book reviews, and this research was able to derive and validate an aspect taxonomy for scholarly book reviews. Only two levels of the aspect taxonomy are validated in this research. The data set can be annotated for more fine-grained lower aspect categories and can experiment the accuracy in a future research. Further, the annotated gold standard dataset can be used for future research in this domain.

This research illustrated an unsupervised dependency rule-based aspect extraction algorithm for extracting both explicit and implicit aspects from book reviews, with high accuracy. The proposed new dependency rules were able to improve the performance. Discovering more dependency rules and combining existing rules for complex aspect extraction are future efforts. Complex aspect extraction will be able to extract longer phrases as aspects. Creating a noise free IAC lexicon will improve the accuracy of the aspect extraction further. The proposed aspect clustering algorithm is able to identify the aspect categories while removing noise words extracted at the aspect extraction phase. Clusters are created based on semantic similarities and merging constraints, then, the top aspect clusters are selected. Both implicit and explicit aspects can be clustered accurately, and it does not need any seed words to perform. Since the number of generated clusters is high compared to the actual aspect categories, better vector representation of aspects using domain specific corpus for aspect clustering algorithm or a supervised/deep learning approach can be applied to identify aspect categories in the future. Detail analysis on how word embeddings can improve the performance of clustering also essential. Instead of using frequency to filter unwanted clusters, domain knowledge can be used to improve the selection.

REFERENCES

[1] A. Poria, Soujanya and Cambria, Erik and Ku, Lun-Wei and Gui, Chen and Gelbukh, “A rule-based approach to aspect extraction from product reviews,” *Proc. Second Work. Nat. Lang. Process. Soc. Media*, pp. 28–37, 2014.

[2] H. Hamdan, P. Bellot, and F. Bechet, “Sentiment Analysis in Scholarly Book Reviews,” *CoRR*, vol. abs/1603.0, 2016.

[3] S. Blair-goldensohn, K. Hannan, R. Mcdonald, T. Neylon, G. A. Reis, and J. Reynar, “Building a Sentiment Summarizer for Local Service Reviews,” in *WWW Workshop on NLP in the Information Explosion Era (NLPIX)*, 2008.

[4] A. Yessenalina, Y. Yue, and C. Cardie, “Multi-level Structured Models for Document-level Sentiment Classification,” in *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 2010, pp. 1046–1056.

[5] V. K. Bongirwar, “A Survey on Sentence Level Sentiment Analysis,” *Int. J. Comput. Sci. Trends Technol.*, vol. 3, no. 3, pp. 110–113, 2013.

[6] A. Meena and T. V Prabhakar, “Sentence level sentiment analysis in the presence of conjuncts using linguistic analysis,” *Adv. Inf. Retr.*, vol. 4425, no. 2, pp. 573–580, 2007.

[7] R. Panchendrarajan, M. N. N. Ahamed, B. Murugaiah, S. Prakhash, S. Ranathunga, and A. Pemasiri, “Implicit Aspect Detection in Restaurant Reviews,” in *Proceedings of NAACL-HLT 2016*, 2016, pp. 128–136.

[8] T. Rana and Y.-N. Cheah, “Aspect extraction in sentiment analysis: comparative analysis and survey,” *Artif. Intell. Rev.*, vol. 45, no. 1, pp. 1–25, 2016.

[9] W. Bancken, D. Alfarone, and J. Davis, “Automatically detecting and rating product aspects from textual customer reviews,” in *Proceedings of the 1st International Conference on Interactions Between Data Mining and Natural Language Processing*, 2014, pp. 1–16.

[10] Q. Su, X. Xu, H. Guo, Z. Guo, X. Wu, X. Zhang, B. Swen, and Z. Su, “Hidden

sentiment association in chinese web opinion mining," in *Proceeding of the 17th international conference on World Wide Web*, 2008, pp. 959–968.

[11] L. Chen, J. Martineau, D. Cheng, and A. Sheth, "Clustering for Simultaneous Extraction of Aspects and Features from Reviews," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016, pp. 789–799.

[12] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004, pp. 168–177.

[13] K. Bafna and D. Toshniwal, "Feature based summarization of customers' reviews of online products," *Procedia Comput. Sci.*, vol. 22, pp. 142–151, 2013.

[14] A. Bagheri, M. Saraee, and F. de Jong, "An Unsupervised Aspect Detection Model for Sentiment Analysis of Reviews," in *Natural Language Processing and Information Systems, NLDB 2013*, 2013, pp. 140–151.

[15] A. Bagheri, M. Saraee, and F. De Jong, "Care more about customers: Unsupervised domain-independent aspect detection for sentiment analysis of customer reviews," *Knowledge-Based Syst.*, vol. 52, pp. 201–213, Nov. 2013.

[16] F. Li, C. Han, M. Huang, X. Zhu, Y.-J. Xia, S. Zhang, and H. Yu, "Structure-aware Review Mining and Summarization," in *Proceedings of the 23rd International Conference on Computational Linguistics*, 2010, pp. 653–661.

[17] L. Chen, L. Qi, and F. Wang, "Comparison of feature-level learning methods for mining online consumer reviews," *Expert Syst. Appl.*, vol. 39, no. 10, pp. 9588–9601, 2012.

[18] N. Jakob and I. Gurevych, "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*, 2010, pp. 1035–1045.

[19] Y. Choi and C. Cardie, “Hierarchical Sequential Learning for Extracting Opinions and their Attributes,” in *Proceedings of the ACL 2010 Conference Short Papers*, 2010, pp. 269–274.

[20] S. Huang, X. Liu, X. Peng, and Z. Niu, “Fine-grained product features extraction and categorization in reviews opinion mining,” in *Proceedings of the 2012 IEEE 12th International Conference on Data Mining Workshops*, 2012, pp. 680–686.

[21] B. Yang and C. Cardie, “Joint Inference for Fine-grained Opinion Extraction,” in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 2013, pp. 1640–1649.

[22] H. Hamdan, “SentiSys at SemEval-2016 Task 5 : Opinion Target Extraction and Sentiment Polarity Detection,” in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval -2016)*, 2016, pp. 350–355.

[23] Z. Toh, F. Way, and F. Way, “NLANGP : Supervised Machine Learning System for Aspect Category Classification and Opinion Target Extraction,” in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 2015, pp. 719–724.

[24] W. Jin and H. H. Ho, “A Novel Lexicalized HMM-based Learning Framework for Web Opinion Mining,” in *Proceedings of the 26th Annual International Conference on Machine Learning*, 2009, pp. 1–8.

[25] W. Jin, H. H. Ho, and R. K. Srihari, “OpinionMiner: A Novel Machine Learning System for Web Opinion Mining and Extraction,” in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2009, pp. 1195–1204.

[26] P. Jiang, C. Zhang, H. Fu, Z. Niu, and Q. Yang, “An approach based on tree kernels for opinion mining of online product reviews,” in *2010 IEEE International Conference on Data Mining*, 2010, pp. 256–265.

[27] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, “SemEval-2015 Task 12: Aspect Based Sentiment Analysis,” in *Proceedings of the 9th International Workshop on Semantic*

Evaluation (SemEval 2015), 2015, pp. 486–495.

[28] T. Àlvarez-López, J. Juncal-Martínez, M. Fernández-Gavilanes, E. Costa-Montenegro, and J. F. González-Castaño, “GTI at SemEval-2016 Task 5: SVM and CRF for Aspect Detection and Unsupervised Aspect-Based Sentiment Analysis,” in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016, pp. 306–311.

[29] S. Poria, E. Cambria, and A. Gelbukh, “Aspect Extraction for Opinion Mining with a Deep Convolutional Neural Network,” *Knowledge-Based Syst.*, vol. 108, pp. 42–49, 2016.

[30] S. Ruder, P. Ghaffari, and J. G. Breslin, “INSIGHT-1 at SemEval-2016 Task 5: Deep Learning for Multilingual Aspect-based Sentiment Analysis,” in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016, pp. 330–336.

[31] Y. Wu, Q. Zhang, X. Huang, and L. Wu, “Phrase dependency parsing for opinion mining,” in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 2009, vol. 3, pp. 1533–1541.

[32] C.-P. Wei, Y.-M. Chen, C.-S. Yang, and C. C. Yang, “Understanding what concerns consumers: a semantic approach to product feature extraction from consumer reviews,” *Inf. Syst. E-bus. Manag.*, vol. 8, no. 2, pp. 149–167, 2010.

[33] K. Liu, L. Xu, and J. Zhao, “Opinion target extraction using word-based translation model,” in *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 2012, pp. 1346–1356.

[34] Z. Zhai, B. Liu, H. Xu, and P. Jia, “Clustering product features for opinion mining,” in *Proceedings of the fourth ACM international conference on Web search and data mining - WSDM ’11*, 2011, pp. 347–354.

[35] K. Priyanka and A. shanker Tewari, “Book recommendation system based on collaborative filtering and association rule mining for college students,” in *proceedings of International conference on Contemporary Computing and Informatics (IC3I)*, 2014, pp. 135–138.

[36] C. Benkoussas and H. Hamdan, "Collaborative Filtering for Book Recommendation," in *Conference and Labs of the Evaluation Forum*, 2014, pp. 501–507.

[37] S. S. Sohail and R. Ali, "Book Recommendation Technique Using Rank Based Scoring Method," in *National Conference on Recent Innovations & Advancements in Information Technology*, 2014, pp. 140–146.

[38] P. Bhure and N. Adhe, "Book Recommendation System Using Opinion Mining Technique," *Int. J. Res. Eng. Technol.*, vol. 04, no. 01, pp. 332–334, 2015.

[39] S. S. Sohail, J. Siddiqui, and R. Ali, "Ordered ranked weighted aggregation based book recommendation technique: A link mining approach," in *proceedings of 2014 14th International Conference on Hybrid Intelligent Systems (HIS)*, 2014, pp. 309–314.

[40] S. S. Sohail, J. Siddiqui, and R. Ali, "OWA based book recommendation technique," *Procedia Comput. Sci.*, vol. 62, pp. 126–133, 2015.

[41] S. S. Sohail, J. Siddiqui, and R. Ali, "Feature Extraction and Analysis of Online Reviews for the Recommendation of Books using Opinion Mining Technique," *Perspect. Sci.*, vol. 8, pp. 754–756, 2016.

[42] M. Van Bellen, "Sentiment Analysis on historical book reviews with a Bayesian Classifie," University of Amsterdam, 2010.

[43] X. Fang and J. Zhan, "Sentiment analysis using product review data," *J. Big Data*, vol. 2, no. 5, pp. 1–14, 2015.

[44] "Sentiment analysis on Amazon book review," *Kaggle*, 2017. [Online]. Available: https://www.kaggle.com/c/sentiment-analysis-on-amazon-book-review.

[45] K. S. Srujan, H. R. Rao, K. Karthik, B. S. Harish, and H. M. K. Kumar, *Classification of Amazon Book Reviews Based on Sentiment Analysis*, vol. 672, no. March. Springer Singapore, 2018.

[46] A. D. Lee, B. N. Green, C. D. Johnson, and J. Nyquist, "How to write a

scholarly book review for publication in a peer-reviewed journal: a review of the literature.," *J. Chiropr. Educ.*, vol. 24, no. 1, pp. 57–69, 2010.

[47] J. McAuley, C. Targett, Q. Shi, and A. Van Den Hengel, "Image-based Recommendations on Styles and Substitutes," in *Proceeding of 38th ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '15)*, 2015, pp. 43–52.

[48] J. McAuley, R. Pandey, and J. Leskovec, "Inferring Networks of Substitutable and Complementary Products," in *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'15)*, 2015, pp. 785–794.

[49] P. Nakov, A. Ritter, S. Rosenthal, V. Stoyanov, and F. Sebastiani, "SemEval-2016 Task 4: Sentiment Analysis in Twitter," in *Proceedings of the 10th International Workshop on Semantic Evaluation*, 2016, pp. 1–18.

[50] T. Hercig, T. Brychcín, L. Svoboda, and M. Konkol, "UWB at SemEval-2016 Task 5: Aspect Based Sentiment Analysis," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016, pp. 342–349.

[51] R. Socher, A. Perelygin, and J. Wu, "Recursive deep models for semantic compositionality over a sentiment treebank," *Proc. 2013 Conf. Empir. Methods Nat. Lang. Process. EMNLP '13*, pp. 1631–1642, 2013.

[52] C. S. Perone, R. Silveira, and T. S. Paula, "Evaluation of sentence embeddings in downstream and linguistic probing tasks," *CoRR*, vol. abs/1806.0, 2018.

[53] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *CoRR*, vol. abs/1810.0, 2018.

[54] R. Petrolito and F. Dell'orletta, "Word Embeddings in Sentiment Analysis," *CLiC-it*, 2018.

[55] P. Stenetorp, S. Pyysalo, G. Topi, T. Ohta, S. Ananiadou, and J. Tsujii, "BRAT : a Web-based Tool for NLP-Assisted Text Annotation," in

Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL '12), 2012, pp. 102–107.

[56] A. J. Viera and J. M. Garrett, "Understanding Interobserver Agreement: The Kappa Statistic," *Fam. Med.*, vol. 37 5, pp. 360–363, 2005.

[57] A. Handler, "An empirical study of semantic similarity in WordNet and Word2Vec," University of New Orleans, 2014.

[58] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," *Trans. Assoc. Comput. Linguist.*, vol. 5, no. 135–146, 2016.