Expert Systems With Applications 171 (2021) 114604

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Analysis of sentiment expressions for user-centered design

Yi Han, Mohsen Moghaddam *

Department of Mechanical and Industrial Engineering, Northeastern University, Boston, MA, USA

A R T I C L E I N F O

Keywords:
Sentiment analysis
Information extraction
Natural language processing
User-centered design

A B S T R A C T

Devising intelligent systems capable of identifying the idiosyncratic needs of users at scale and translating them into attribute-level design feedback and recommendations is a key prerequisite for successful user-centered design processes. Recent studies show that 49% of design firms lack systems and tools for monitoring external platforms, and only 8% have adopted digital, data-driven approaches for new product development despite acknowledging them as a high priority. The state-of-the-art attribute-level sentiment analysis approaches based on deep learning have achieved promising results; however, these methods pose strict preconditions, require manually labeled data for training and pre-defined attributes by experts, and only classify sentiments intro predefined categories which have limited implications for designers. This article develops a rule-based methodology for extracting and analyzing the sentiment *expressions* of users on a large scale, from myriad reviews available on social media and e-commerce platforms. The methodology further advances current unsupervised attribute-level sentiment analysis approaches by enabling efficient identification and mapping of sentiment expressions of individual users onto their respective attributes. Experiments on a large dataset scraped from a major e-commerce retail store for apparel and indicate 74.3%–93.8% precision in extracting attribute-level sentiment expressions of users and demonstrate the feasibility and potentials of the developed methodology for large-scale need finding from user reviews.

1. Introduction

A major barrier to user-centered design is the relative absence of formal mechanisms for translating the “voices” of individual users into the design of differentiated product attributes along which preferences diverge (Salvador, De Holan, & Piller, 2009). Current mechanisms for engaging users in the design process revolve primarily around surveys and focus-group studies (Fogliatto & da Silveira, 2008; Griffin & Hauser, 1993) and web-based configurators (Felfernig, 2007; Franke, Schreier, & Kaiser, 2010). These solutions merely target an inherently biased fraction of users and product instances, and leave an entire swath of pertinent user behavior, preferences, and opinions not captured. The inability of some users to identify and directly express their true preferences (Franke, Keinz, & Steger, 2009) further exacerbates this gap. This is while the rapidly growing social media and e-commerce platforms provide an unprecedented source of knowledge about the preferences of much larger and more diverse populations of users and on a multitude of product instances within the same class/family. This article presents a novel methodology for the analysis of sentiment expressions (ASE) of users to draw attribute-level design insights and recommendations from unstructured, online user reviews. ASE contributes to the broader field of aspect-based sentiment analysis (ABSA) in natural language processing (NLP) by enabling the extraction of sentiment *expressions* in a semi-supervised fashion.

The majority of existing ABSA methods classify user sentiments into predefined categories (*e.g.*, “positive”, “negative”, “neutral”) using labeled training data and for some experts pre-defined aspects (García-Pablos, Cuadros, & Rigau, 2018; Ma et al., 2017; Peng, Ma, Li, & Cambria, 2018; Pontiki et al., 2015; Pontiki, Galanis, & Papageorgiou, 2015; Rietzler, Stabinger, Opitz, & Engl, 2019; Sun & Huang, 2019; Tamchyna & Veselovská, 2016). The proposed ASE methodology aims at addressing two practical limitations associated with the current state of ABSA literature. First, such sentiment classification approaches may leave important user feedback and insights uncaptured. Second, the supervised nature of most ABSA methods limits their practical use due to the crucial need for laborious labeling and annotation of online review data for training (*e.g.*, Lee & Bradlow, 2011; Pang & Lee, 2006; Ravi & Ravi, 2015; Tang, Tan, & Cheng, 2009; Zhang, Wang, & Liu, 2018a; Suryadi & Kim, 2019; Zhou, Jiao, & Linsey, 2015). The proposed ASE methodology goes beyond identification of sentiment polarity and

* Corresponding author.
E-mail addresses: han.yi1@northeastern.edu (Y. Han), mohsen@northeastern.edu (M. Moghaddam).

https://doi.org/10.1016/j.eswa.2021.114604
Received 21 November 2019; Received in revised form 8 January 2021; Accepted 11 January 2021
Available online 15 January 2021
0957-4174/

rather extracts the information about the very same sentiment phrases expressed by users (Fig. 1), without the need for labeled/annotated training data. The developed methodology has been tested and validated on a large dataset of scraped from an online retail store for sneakers and athletic apparel.

1.1. Motivation

Our motivation stems from the growing abundance of user-generated feedback and the lack of advanced computational frameworks and techniques for turning data into new design knowledge and insights. Recent studies show that 49% of design firms lack systems and tools for monitoring external platforms, and only 8% have adopted digital, data-driven approaches for new product development despite acknowledging them as a high priority (Oracle, 2019). The proposed methodology aims at addressing the need for effective mechanisms for translating the preferences of individual users expressed through online reviews into attribute-level design feedback and recommendations (Griffin & Hauser, 1993; Salvador et al., 2009; Wagner & Majchrzak, 2006; Zhang, Wang, & Liu, 2018b). The foundational capabilities to realize this vision include (Salvador, de Holan, & Piller, 2009): (A) solution space development to gather user preferences, (B) robust process design to enhance manufacturing flexibility, and (C) choice navigation to minimize the burden of choice. Current approaches to mass-personalization predominantly emphasize Capability B to accommodate the uncertainty and diversity of user preferences. Examples include research on reconfigurable manufacturing systems, outsourcing (Xu, 2012), modular product design and postponement (Koren, Hu, Gu, & Shpitalni, 2013), and coordination of operations along supply chains (Salvador, Rungtusanatham, & Forza, 2004).

These approaches are essentially “reactive” in that they treat individual preferences as mostly unknown and intractable, and instead attempt to prepare for the subsequent uncertain behavior at the downstream manufacturing stage. Even more proactive, solution-space development approaches (Capability A) such as focus-groups or web-based configurators (Felfernig, 2007; Franke et al., 2009, 2010; Griffin & Hauser, 1993) are not fully effective due to targeting small groups of individuals and limited product instances that are hard to scale up. These limitations have hindered the practical adoption of mass-personalization due to extensive economic and operational limitations (Fogliatto, Da Silveira, & Borenstein, 2012). Further, neglecting the voices of the majority of users, the gap between what is designed and what is desired lingers notwithstanding the knowledge and expertise of the design team. There is therefore a critical need for a preference modeling mechanism built upon direct feedback from users that leverages the unprecedented wealth of knowledge embedded in freely available online reviews.

Google reports 600% increase, between years 2015–2017, in the amount of time people spend exploring others’ experiences before making a decision on a product or service (Google, 2017). A more recent study (*Local Consumer Review Survey | Online Reviews Statistics & Trends*,2019) reports that 86% of consumers read reviews for local businesses, 80% of 18–34 year-olds have written online reviews and 91% of 18–34 year-olds trust online reviews as much as personal recommendations. These statistics pinpoint the significant impact of online reviews through e-commerce and social media platforms (1) on users’ preferences and choices, especially millennials, and (2) on the prosperity of small and domestic design and manufacturing startups. The ASE methodology aims at leveraging this unprecedented opportunity to align the design of customizable product attributes with the heterogeneous preferences of users that are increasingly expressed in unstructured text format. Not meeting this need will limit the industrial adoption of mass-personalization due to excessive inventories, financial losses, and poor delivery performances (Fogliatto & da Silveira, 2008), hindering the potential transformative impacts on various other industries such as food, nutrition, orthopedics, electronics, and home-building in terms of identifying user segments and tailoring the design, production, and service to the specific needs of each segment.

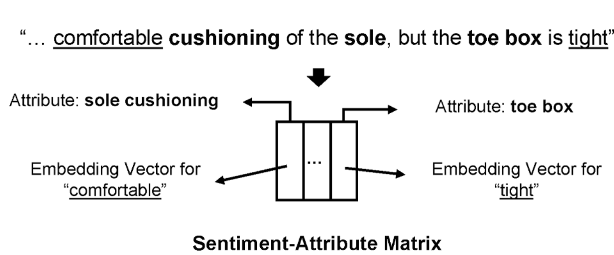

Fig. 1. Proposed attribute-level sentiment analysis.

1.2. State-of-the-Art

The sentiments of users about individual attributes of a product are significantly more informative than their overall sentiment about the product (Wang et al., 2014) for gauging their idiosyncratic preferences and choices. The finer-granularity of attribute-level sentiments enables the extraction of more informative design feedback and recommendations, specifically because online reviews typically involve considerable diversity in terms of size, structure, and the way emotions and feelings are expressed (Fogliatto et al., 2012; Tamchyna & Veselovská, 2016). Sentiment analysis is the computational study of people’s opinions, preferences, emotions, or attitudes towards the attributes of a product, service, event, topic, or even individual (Bing, 2015). Since the early 2000′s, sentiment analysis has become a central area of research in NLP, conducted at three different levels: document, sentence, and attribute (García-Pablos et al., 2018; Tamchyna & Veselovská, 2016).

Unlike document- and sentence-level analysis which merely output overall polarity of user sentiment, attribute-level sentiment analysis carries out attribute extraction, entity identification, sentiment description, and attribute-sentiment mapping (Zhang et al., 2018a). Attribute-level sentiment analysis is still a developing field of research. Only recently have studies been conducted on mapping user needs onto design specifications through sentiment analysis (Wang, Mo, & Tseng, 2018); however, the proposed methods are limited to document-level sentiment classification which are unable to capture attribute-level information. Further, existing attribute-level sentiment analysis solutions typically focus on sentiment polarity (*e.g.*, positive/negative/neutral), and leave the detailed description of individuals’ opinions about different attributes uncaptured (García-Pablos et al., 2018; Peng et al., 2018; Pontiki et al., 2016; Sun & Huang, 2019; Tamchyna & Veselovská, 2016). ASE aims at addressing these gaps by focusing on attribute-level sentiment expressions rather than polarity. ASE is built upon the rationale that true user-centered design can only be achieved if the voices of all (or at least majority of) users are heard across several related product types and classes. A detailed review of state-of-the-art sentiment analysis methods is provided in Section 2.

1.3. Objectives & outline

The overarching goal of this research is to build and test an efficient computational methodology for extracting attribute-level sentiment expressions of individual users from online reviews. Shifting the focus from sentiment polarity to sentiment expression, the methodology can shed light on the diversity and similarity of user preferences at scale: Users may converge into a number of segments, where at least one feasible design alternative exists that would satisfy an entire segment while no such alternative exists for more than one segment. This article builds the foundation for achieving this goal by accomplishing the following research objectives:

- *Objective 1: Identify attribute-level sentiment expressions of individual users*. This objective revolves around investigating methods for lexicon building and word embedding, mapping sentiment expressions onto differentiated attributes, and extracting customizable attributes along which user preferences diverge. To this end, two rule-based methods for ASE are developed and tested in this article.
- *Objective 2: Measure ASE-based user similarities across various product types*. This objective is motivated by Arrow's theorem that proves any social welfare function which satisfies the transitivity, unanimity, and independence of irrelevant alternatives conditions is a dictatorship (Arrow, 1950). The implication of this theorem for this research is that it is impossible to generate a design alternative that maximizes the utility of User A without reducing the utility of User B, if A and B have significantly different preferences. In order to increase the utility of all individual users in terms of satisfying their idiosyncratic needs, it is therefore to necessary to generate at least as many design alternatives as the number of distinct user types or segments. This objective builds on the results of ASE to identify potential patterns in user preferences.
- *Objective 3: Build ASE and validate its performance through a large dataset*. Objective leverages the knowledge and experience from the authors' preliminary work and industry collaboration to set up a pilot study on *sneakers* based on a dataset retrieved from an e-commerce platform.

The remainder of this article is organized as follows. Section 2 presents a detailed background on sentiment analysis, particularly ABSA (readers familiar with sentiment analysis may skip this section). Section 3 elaborates on the developed ASE methodology including the rule-based algorithms and the similarity measurement approach. Section 4 presents the case study and discusses the experimental results, analyses, and limitation. Section 5 concludes the article and provides directions for future research.

2. Background

Sentiment analysis is the computational study of people's opinions, preferences, emotions, or attitudes towards the attributes of something like a product, a service, an individual, an event, a topic, and so forth (Bing, 2015). Sentiment analysis has recently gained tremendous attention due to its unprecedented ability to quantitatively evaluate the success of a product or a service in terms of performance and user satisfaction (see, *e.g.*, Law, Kwok, & Ng, 2016; Li, Bing, Li, Lam, & Yang, 2018; Mirtalaie, Hussain, Chang, & Hussain, 2017; Ureña, Chiclana, & Herrera-Viedma, 2018; Ureña, Kou, Dong, Chiclana, & Herrera-Viedma, 2019). This relatively young area of natural language processing (Zhang et al., 2018a) may be the breakthrough path to the designers' long-standing *holy grail*: bridging the gap between what is designed and what is desired. This section provides an overview of ABSA, with methods categorized based on the need for labeled training data.

2.1. Attribute-Level sentiment analysis with labeled dataset

Attribute-level sentiment analysis, also known as aspect-based sentiment analysis (ABSA), is the process of specifying users' opinions from online reviews, forum discussions, or social media, usually by classifying them into three categories: positive, negative, and neutral. Therefore, a key requirement for these methods is generating labeled training data (Liu, 2010a); once a labeled dataset is generated, any supervised learning method can be applied to solve the problem. In previous studies on sentiment analysis of user reviews, researchers mostly focused on the *polarity* of users' opinions (*i.e.*, whether or not the user liked the product) (Chen, Yan, & Wang, 2017; Dhaoui, Webster, & Tan, 2017; Fernández-Gavilanes, Álvarez-López, Juncal-Martínez, Costa-Montenegro, & Javier González-Castaño, 2016; Yi, Nasukawa, Bunescu, & Niblack, 1889; Ma & Chen, 2017; Pham & Le, 2018; Wu, Wu, Chang, Wu, & Huang, 2019; Zhai, Liu, Xu, & Jia, 2011; Zheng, Lin, Wang, Lin, & Song, 2014). These methods generate predictive models given labeled datasets, based on a variety of machine learning and deep learning techniques, with the objective of classifying sentiment (Alqaryouti, Siyam, Monem, & Shaalan, 2019; Karimi, Rossi, Prati, & Full, 2020; Li, Bing, Li, Lam, & Yang, 2018; Soh, Yu, Narayanan, Duraisamy, & Chen, 2019; Suryadi & Kim, 2019; Toh & Su, 2015; Xu, Zhang, Xin, & Yang, 2019; Xu, Liu, Shu, & Yu, 2018; Xu, Lin, Wang, Yin, & Wang, 2017; Yang, Zhang, Jiang, & Li, 2019; Zhang et al., 2018a). Supervised attribute-level sentiment analysis has been extensively investigated using several well-established labeled dataset benchmarks (Pontiki et al., 2015; Pontiki, Galanis, & pavlopoulos, 2015; Tamchyna & Veselovská, 2016). Deep learning-based methods have recently been proven to achieve state-of-the-art performance in sentiment classification tasks (Do, Prasad, Maag, & Alsadoon, 2019). In most recent studies, large pre-trained language models such as BERT (Devlin, Chang, Lee, & Toutanova, 2018), in conjunction with efficient fine-tuning strategies, have been proven highly efficient and optimal in several sentiment classification tasks (Hoang, Bihorac, & Rouces, 2019; Sun & Huang, 2019).

A major limitation of current attribute-level sentiment analysis techniques based on deep learning applied for user-centered design, however, is the critical need for laborious labeling and annotation of online review data for training (*e.g.*, (Suryadi & Kim, 2019; Zhou et al., 2015)). This requirement hinders their practical use due to the difficulty of replicating those methods and experiments, and the lack of scalability and transferability across different domains. Another limitation of these methods is that all product attributes must be pre-defined by experts and/or researchers building the models (Alqaryouti et al., 2019; Karimi et al., 2020; Li et al., 2018; Pontiki et al., 2016, 2015; SemEval., 2014 *The 8th International Workshop on Semantic Evaluation Proceedings of the Workshop Dublin, Ireland*, 2014; Soh et al., 2019; Suryadi & Kim, 2018; Tamchyna & Veselovská, 2016; Toh & Su, 2015; Xu et al., 2018, 2017; Yang et al., 2019), instead of identifying and analyzing the attributes concurrent with the sentiment analysis process. The proposed ASE methodology tackles these limitations by taking an semi-supervised, rule-based approach to this problem, as described in Section 3. Before presenting the methodology in detail, an overview of related work on unsupervised attribute-level sentiment analysis is presented next.

2.2. Attribute-Level sentiment analysis with unlabeled dataset

Research on attribute-level sentiment analysis with unlabeled dataset is still at an early stage compared to the analysis with labeled dataset. Classic unsupervised methods such as clustering have been applied in previous unlabeled sentiment analysis studies (Fernández-Gavilanes, Álvarez-López, Juncal-Martínez, Costa-Montenegro, & Javier González-Castaño, 2016; Hu, Tang, Gao, & Liu, 2013; Pandarachalil, Sendhilkumar, & Mahalakshmi, 2015; Singh, Piryani, Uddin, & Waila, 2013; Suresh & Gladston, 2016; Vashishtha & Susan, 2019) for extracting clusters of users that belong to certain similar emotion range. These methods are powerful for macro-level analysis of customers; however, they may not provide sufficient elaborated insights for the engineering design community (Zhou et al., 2015). Further, defining proper performance benchmark for unsupervised sentiment analysis methods is more challenging. In this context, adjusted rand index and normalized mutual information are two extensively used benchmarks (Santos & Embrechts, 2009; Vinh, Epps, & Bailey, 2010), mainly concerned with the quality of clusters (*e.g.*, the distance between clusters instead of users). Several studies have also investigated semi-supervised method for attribute-level sentiment analysis (Da Silva, Coletta, Hruschka, & Hruschka, 2016; Usha & Indra Devi, 2013; Wang et al., 2014; Zhai et al., 2011), in which, domain knowledge, syntax rules, and semantic similarities are used in conjunction with clustering methods for clustering users based on their sentiments. To draw more elaborate insights, researchers have developed a variety of unsupervised, rule-based methods in different domains (Chen & Yao, 2010; Lakkaraju, Bhattacharyya, Bhattacharya, & Merugu, 2011; Liu, Yao, & Wu, 2005; Vashishtha & Susan, 2019; Wan, 2008). However, none of those studies were related to either sentiment expressions or design-related topics.

2.3. Implications and limitations for product design

Attribute-level sentiment analysis has become a key enabler for large-scale need-finding from myriad users via e-commerce and social media platforms. A significant knowledge gap in current literature, however, is the critical need for laborious labeling and annotation of online review data for training. The supervised learning approaches in this context allow to train deep neural networks on large, labeled datasets to predict the polarity of users' opinions for certain, predefined aspects of a product (Pontiki, Galanis, et al., 2015; Rosenthal et al., 2015), such as the screen or battery of a cellphone. Such aggregation of user's sentiments into a small number of labels may not fully utilize the information inherent in online product reviews for informing the product design process. Specifically, it is desired to extract user sentiments at a finer level of granularity; and ideally shift the focus from sentiment classification to extraction of brief expressions of their sentiments. Thus, existing supervised models are not the best candidates for tackling this problem. When it comes to unsupervised learning, the majority of existing models are focused on aspect extraction, aspect clustering, and sentiment polarity clustering (Cranenburgh, 2019; Desai et al., 2015; García-Pablos et al., 2018; García Pablos, Cuadros, & Rigau, 2015; Hercig, Brychcín, Svoboda, Konkol, & Steinberger, 2016; Suresh & Gladston, 2016).

To better serve the engineering design community, the outcomes of sentiment analysis must go beyond sentiment polarity and sentiment clustering and extract more elaborate information on users' opinions (Tetsuya Nasukawa, 1956; Wang et al., 2014). Both information extraction and sentiment analysis are intended to extract some desired information from text. However, sentiment analysis is more concerned with finding the polarity of users' *emotions* based on syntax and semantic rules, while information extraction covers a broader scope through extracting the *relationships* between different elements of a sentence (Batista, Martins, & Silva, 2015; Chen, Ji, Tan, & Niu, 2007; Gupta, Roth, & Schütze, 2018). The goal of the proposed ASE methodology is to track the information related to the users' perception of a particular attribute of a product or a service. To accomplish Objective 1 which is to extract attribute-level sentiment expressions, this article builds and tests a methodology at the intersection of attribute-level sentiment analysis and information extraction techniques, as described next.

3. Methodology

This section presents the details of the ASE methodology (Fig. 2) including the lexicon building method and the semi-supervised, rule-based algorithms for identifying attribute-level sentiment expressions of individual users (Objective 1) as well as a product similarity measure that enables the aggregation of user reviews across an entire product class (Objective 2). The case of sneakers is used as a running example to better explain the methodology.

3.1. Lexicon building

The motivation behind generating the lexicons lies in the differences between concept generation and expression by different individuals (Cao, Li, & Ramani, 2011). For a given product (*e.g.*, sneakers), an attribute must satisfy one of the following relationships (Yi, Nasukawa, Bunescu, &Niblack, n.d.): *a part-of* the product (*e.g.*, outsole), *a feature-of* the product (*e.g.*, comfort), or *an aspect-of a feature of* the product (*e.g.*, color). A certain attribute may be expressed using different synonymous terms. It is therefore necessary to categorize all the different expressions that refer to the same attribute into one group. For further analysis, the input text corpora can be converted into embeddings to represent each word/phrase in the corpora via a unique vector obtained from a shallow neural network. Since the word vectors are encoded close to their neighbor words in a large labeled corpus, a more accurate representation can be obtained.

Two separate lexicons are developed, one for attributes and the other for sentiments. The attribute lexicon includes manually picked descriptions of products (*e.g.*, sneakers) extracted from online product catalogs and glossaries (*e.g.*, (*A Complete List of Sneakerhead Terminology From A to Z* | KickBackz, 2019; *Sneaker Genius – Glossary of Sneaker Terms* | Genius, n.d.; Sneaker Terminology, 2019; *Sneaker Terminology* | SneakerFiles, n.d.; *The Essential Guide to Sneaker Vocabulary* | Sole Collector, n.d.). In the case of sneakers, the manually generated attribute lexicon comprises a total of roughly 500 descriptions grouped into 23 main attributes (*e.g.*, color, energy return, permeability, weight, stability, durability). For the sentiment lexicon, the lexicon created by Liu and

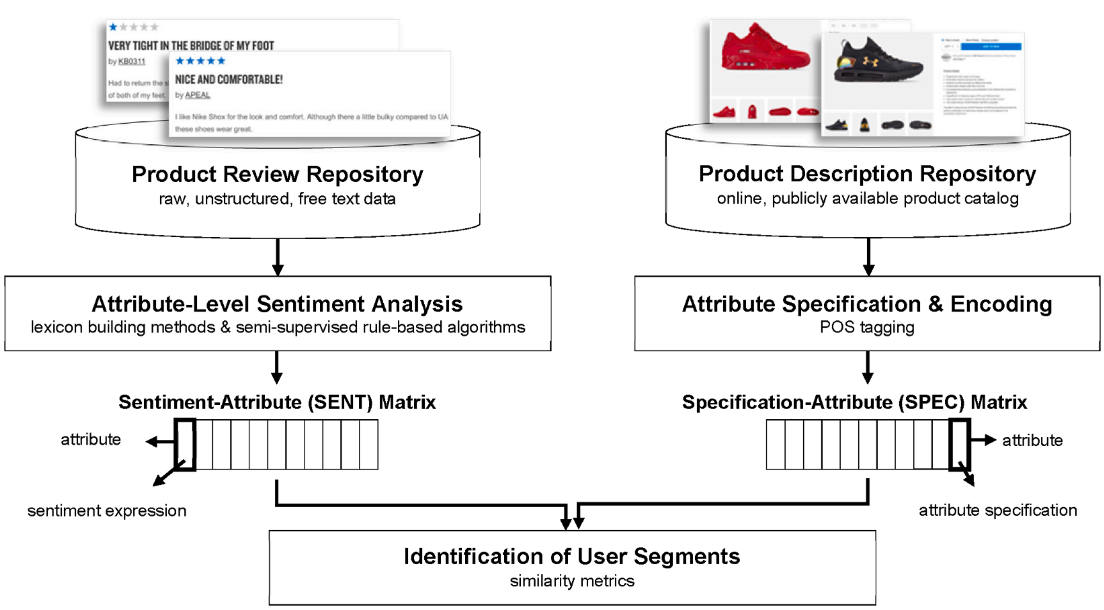

Fig. 2. Overview of the ASE methodology.

Hu (Hu, Liu, & Street, 2004) has been adopted which includes manually picked sentiment words from a dictionary with the vocabulary size of over 6000 words. In this section, attribute-level product descriptions are also generated and encoded as a specification-attribute matrix (SPEC matrix, for short), where each column represents the specification of the respective attribute as an embedding vector (see Fig. 2). The SPEC matrix will be used for measuring similarities between pairs of products (*e.g.*, Nike Air Max vs. Adidas Ultraboost) within the same class (*e.g.*, sneakers).

3.2. Semi-Supervised, Rule-based algorithms

Two rule-based algorithms are developed for identifying and extracting attribute-level expressions of user sentiments from online reviews in a semi-supervised fashion, as described next.

3.2.1. Part of speech with window set (PWS) algorithm

The PWS algorithm (Algorithm 1) sets a window centering the attribute with a parametric window-size if more than one attributes are present in the sentence; otherwise, the entire sentence is taken into consideration. The nearest adjective, verb or noun is then identified as the descriptive part of the attribute. Next, all of the words inside the window are paired with the sentiment lexicon and the nearest sentiment word is selected as the respective sentiment. The algorithm searching for potential *adversatives* inside the window (*e.g.*, 'not', 'however') to ensure accurate identification of the polarity of the sentiment description. Previous studies show that the outcomes of an NLP task with window sliding techniques are usually sensitive to the size of the window (Zhai et al., 2011). In PWS, the notion of window size setting is inspired by previous studies (Masood, Abbasi, & Keong, 2020; Xianghua, Guo, Yanyan, & Zhiqiang, 2013). For sentiment relationship extraction, it is argued that window size should be between 3 and 10 (Desai & Mining, 2015; Lakkaraju, Socher, & Manning, 2014). Without loss of generality, our experiments on the PWS algorithm use a window size of 5.

The PWS algorithm is built upon two assumptions associated with syntactic expression:

1) The description of an attribute is adjacent or near the attribute. For example, in the phrase "… the edge of the heel", "heel" and "edge" are respectively the attribute and the description with two words between them. According to this assumption, the critical element of the PWS algorithm is the window size. Without loss of generality, the window size is set to 5 in this article.
2) The description and sentiment expression associated to an attribute can be tracked by their part-of-speech (POS) tag. For extracting the description and the sentiment expression from a window, a number of syntactic rules are defined; *e.g.*, the nearest "NN" represents a noun description of an attribute. For example, in the phrase "… the edge of the heel", the word "edge" is identified as an "NN" POS describing "heel".

Algorithm 1: Part of Speech with Window Set (PWS)

```
input: POS
  output: (attribute + description + sentiment)
  set window = 5
  word window = list
  output = list
  for word in sentence:
    if wordin attribute lexicon:
      word window = word +- 5 around words
      for words in word window:
        description = argmin length (nn/adj – word)
        if words in sentiment lexicon:
          sentiment = argmin length (words – word)
        end if
      end for
```

(*continued on next column*)

(*continued*)

Algorithm 1: Part of Speech with Window Set (PWS)

```
        output.append (word, description, sentiment)
      end if
  end for
  print (output)
```

3.2.2. Parsing with syntactic rules (PSR) algorithm

The PSR algorithm (Algorithm 2) is built on the Stanford parser (The Stanford Natural Language Processing Group, n.d.) to analyze syntactic rules and trees in three steps: (1) pair all words with the attribute lexicon, (2) track the descriptive part in the sentence, and (3) extract the sentiment expression. Penn (Klein & Manning, 2007) is applied to find the descriptive part in the sentence according to the syntax tree. The PSR algorithm is built upon two assumptions:

1) The description of an attribute is either a noun or an adjective (*e.g.*, in "height of the heel": "heel" is the attribute, "height" is the description). Several variants of nouns and adjectives such as proper nouns, plural nouns, comparative adjectives, and superlative adjectives (Mirtalaie, Hussain, Chang, & Hussain, 2018) are thus explored.
2) The sentiment expressions appear according to these rules: (a) the closest verb (*e.g.*, in "I like the soft heel", "like" is the sentiment expression); (b) subsequent expression (*e.g.*, in "I think the soft heel is great", "great" is the sentiment expression).

Building on these heuristic rules, the PSR algorithm utilizes the *dependency tree* to identify the relationships between words within a sentence. The dependency tree is the syntax parsing results provided by the Stanford parser. Through the syntax parser, a single sentence can be recast as a tree with a *root*, *branches*, and *leaves*. On the dependency tree, each word is represented by a position label, and every two words in a sentence are assigned a relationship label. In this article, 16 relationship types are applied to identify the description of an attribute; *e.g.*, 'acomp' (adjectival complement) which defines an adjectival complement of a verb. To identify sentiment expressions, several marks are also applied related to emotions; *e.g.*, 'amod' (adjectival meaning modifier) and 'rcmod' (relative clause modifier).

Although parsers are typically trained at the sentence level, in this study, all punctuations are removed and each review is treated as one sentence. The main motivation behind this decision is that many users write their reviews in a totally unstructured and informal manner, and misuse punctuations (*e.g.*, use comma to end a sentence) or not even use them. For instance, the longest review in our dataset contains 248 words and is written in a single sentence. Thus, breaking a single review down into sentences may cause the risk of information loss if done automatically (*i.e.*, according to the used punctuations) or may require laborious data cleaning of the data if done manually. Nevertheless, treating the entire review as one single sentence may also cause information bias to some extent. Since the purpose of the PSR algorithm is to extract the user's sentiment about certain attributes of a product, however, the use of Stanford parser with no punctuations yields satisfactory performance in this study, because only the relationships that contain an attribute word are taken into consideration (see Algorithm 2). In case multiple descriptions or sentiment words are related to the same attribute, the PSR algorithm breaks the tie by choosing the one with the closest proximity. Future research must address the issue of "mispunctuation" in unstructured and informal user reviews, and develop methods for automated correction of punctuations to ensure more precise results from parser-based algorithms such as PSR.

Algorithm 2: Parsing with Syntactic Rules (PSR)

```
input: dependency results
  output: (attribute + description + sentiment)
```

(*continued on next page*)

(*continued*)

Algorithm 2: Parsing with Syntactic Rules (PSR)

```
for word in review:
  if word in attribute lexicon:
    phrase = previous VP + word + next VP
    for words in phrase:
      description = argmin length ('acomp', 'amod', 'rcmod', etc. relationships
that contains the attribute word)
      sentiment = argmin length ('nsubj', 'dobj', etc. relationships that contains
the at tribute word)
    end for
    phrase sentiment = next NP
      sentiment2 = argmin length (word - adj)
    output.append (word + description + sentiment + sentiment2)
  end if
end for
print (output)
```

3.2.3. Summary of algorithms

Table 1 presents a comparative summary of the two rule-based algorithms presented in this section. The outcome of the rule-based algorithms is a sentiment-attribute matrix (SENT matrix; see Fig. 2) for each individual product review. The benefit of the ASE methodology is that the SENT matrix can be decoded into descriptive text anytime throughout the rest of the analyses. This helps the designer access detailed information about each data-point (*i.e.*, user), on-demand.

3.3. Product similarity metric

Since the analysis of sentiment expressions is conducted on various products within the same class, it is necessary to develop a measure to quantify the degree of similarity between each pair of product instances. Devising such a measure is important because it enables comparisons among all users who have purchased and rated some product within a large class of products (*e.g.*, everyone who has purchased and rated a running shoe). The developed *product similarity metric* utilizes the SPEC matrices as input and calculates pairwise similarities based on the weighted sum of cosine similarity between the embedding vectors describing different attributes of the products. Let **A** and **B** denote the SPEC matrices of products a and b, respectively. By definition, the i^{th} column of these matrices therefore represents the specification of the i^{th} attribute encoded as an embedding vector. The degree of similarity between products a and b can thus be computed as $\Delta(a,b) = \mathscr{w}^{\intercal}\cdot\theta$, where

$$\theta_i = \arccos\left(\frac{\mathbf{A}_{*i}\cdot\mathbf{B}_{*i}}{\|\mathbf{A}_{*i}\|\|\mathbf{B}_{*i}\|}\right), \forall i, \tag{1}$$

and $\mathscr{w}$ is a weights vector. Fine-tuning the similarity measure and identifying the best weight values are challenging tasks that requires further investigation. The product similarity measure can then be applied to aggregate the reviews associated with the entire class of products.

Table 1
Algorithms summary and comparison.

Algorithm	PWS	PSR
POS tagging method	NLTK (Natural Language Toolkit)	NLTK
Embedding method	Word2VEC	Word2VEC
Assumptions	The description and sentiment expressions of an attribute appear near the attribute word	The relationships of description and sentiment expressions with their attributes can be classified into limited categories
Base unit	Word (word-by-word analysis)	Word
Limitations	The performance (precision) is highly sensitive to the window size	The performance (precision) highly depends on the parsing results
Requirements	Attribute lexicon, sentiment lexicon, POS tags, window size	Attribute lexicon, sentiment lexicon, Stanford parsing tool, POS tags

3.4. Evaluation metrics

Four metrics are utilized to analyze the performance of the ASE rule-based algorithms in successfully extracting sentiment expressions and mapping them onto their respective attributes: recall, precision, redundancy, and F-score. The metrics need to be calculated using a labeled data as the benchmark. In this article, 200 reviews were labeled by several mechanical engineering undergraduate students. The evaluation metrics are as follows:

1) Recall (γ): The measure of how many words are correctly selected in a unit (i):

$$\gamma = \frac{1}{N}\sum_i\left(\frac{|c_i \cap o_i|}{o_i}\right), \tag{2}$$

where c_i and o_i are the predicted and observed labels for the ith word, respectively. A *unit* may refer to a sentence, a paragraph, a corpus, or any portion of them. Recall is used to measure the performance of the ASE methodology in successfully identifying attributes using the attribute lexicon.

2) Precision (ζ*)*: The measure of how many of the selected words are correctly labeled in a unit i (for $i = 1, \cdots, N$):

$$\zeta = \frac{1}{N}\sum_i\left(\frac{|c_i \cap o_i|}{c_i}\right) \tag{3}$$

The sentiment-attribute mapping precision metric measures the success of the PWS and PSR algorithms in accurately mapping each attribute to its sentiment expression.

3) Redundancy (κ): The ratio of the number of unnecessary words (u_i) to the total number of words (m_i) available in unit i (for $i = 1, \cdots, N$):

$$\kappa = \frac{1}{N}\sum_i\left(\frac{u_i}{m_i}\right) \tag{4}$$

A selected word is treated as unnecessary if it is not one of followings: attribute, description, sentiment expression.

4) F-score (ϕ): The balance of precision and recall providing a more realistic measure of the model performance. It is measured as the weighted harmonic mean of precision and recall:

$$\phi = \frac{(\gamma^{-1} + \zeta^{-1})^{-1}}{2}. \tag{5}$$

4. Experiments

To test and validate the performance of the ASE methodology in extracting attribute-level sentiment expressions of users from online reviews (Objective 3), numerical experiments have been conducted on a large dataset scraped from an online retail store for sneakers and athletic apparel.

4.1. Preliminaries

4.1.1. Data

The dataset comprises all user reviews for sneakers posted on the online retail store before 7/31/2019. The total number of user reviews is 23564. A relatively smaller dataset was also collected from another online store to create a contrast group. The main dataset provides information about multiple aspects including the user location, gender, overall ratings or stars, the title and the body of the review. In this case study, only the review titles and bodies were captured, and the rest of the available data was discarded. The raw data included several typos, grammar errors, and slangs, which were directly imported into the algorithms without correction. The input data is available at: https://github.com/hanyidaxia/NER_BERT.

4.1.2. Lexicon building for attributes and sentiments

The attribute lexicon of sneakers was built manually, which includes 23 main attributes (*e.g.*, color, energy return, permeability, weight, stability, durability). Several various synonyms of the main attributes were collected from 10,000 reviews, which add up to 500 attribute words in total. The descriptions of the attributes were then tracked using NLTK (Natural Language Toolkit) post tagger, which translates phrases/sentences into POS tags. For example, “I love the classic style” is translated into (“I”, “PRP”), (“love”, “VBP”), (“the”, “DT”), (“classic”, “JJ”), (“style”, “NN”). The attribute lexicon is one of the most critical elements of ASE, since it highly effects the performance in terms of finding attributes in user reviews and their respective sentiment expressions. The sentiment lexicon was adopted from an existing online sentiment lexicon (Liu, 2010b), which was enhanced by adding some domain-specific sentiment expressions relation to sneakers. Both rule-based algorithms (PWS and PSR) utilize the same attribute and sentiment lexicons, which they use as inputs along with user reviews, POS tags, and dependency trees (PSR).

4.1.3. Word embedding

Word2Vec (Mikolov, Chen, Corrado, & Dean, 2006) was adopted for word embedding, which encodes each word into a unique number vector so the computer can comprehend and operate on them. Word2-Vec is one of the most popular language models is pre-trained model—a two-layer neural network that transforms text into vectors. The idea behind Word2Vec embedding is straightforward. The model is trained by a skip-gram model with hierarchical softmax output and can transfer words into *n* dimensional vectors, where *n* is another training parameter (*e.g.*, the size of the hidden layer).

4.2. Results and analyses

Table 2 presents the results with respect to the metrics developed in Section 3D. The evaluation was conducted on 200 reviews selected at random. As suggested by the results, the PWS algorithm yields acceptable performance in successfully identifying sentiment expressions; however, the redundancy rate is a bit high. Further, the analyses reveal a highly variable performance for PWS in different situations—in some reviews, it could perform at a 100% precision, while some others, its precision drops to less than 20%. The comparison between PWS and PSR indicates a much better performance from PWS compared to PSR in terms of both recall and precision, and thus the F1-score. However, PSR outperforms PWS in terms of redundancy rate with much less length relative to PWS. The results reflect the inner logic of the algorithms to some extent. Both algorithms identify attributes with respect to the predefined lexicon. The main difference between the two algorithms stems from the way they associate descriptions and sentiments to the identified attributes: PWS selects the closest description and sentiment words in the window surrounding the attribute while PSR locates those two words using a syntax dependency tree. This has resulted in the significant performance differences between the two algorithms as shown in Table 2; although more rigorous investigation is required because the presented results are based on 200 manually labeled test data points. The 200 manually labeled dataset was only used for testing the performance of the two algorithms. These 200 reviews were randomly collected from the raw dataset, excluding those used in the attribute lexicon building.

Table 2
Quantitative comparison between the semi-supervised, rule-based algorithms.

Metric	PWS	PSR
Recall (γ)	79.3%	77.6%
Sentiment-attribute mapping precision (ζ)	93.8%	74.3%
Redundancy (κ)	55.6%	37.3%
F1-score (ϕ)	86.6%	76.0%

Since we use two lexicons in the first algorithm and one lexicon with one parser in the second algorithm, both algorithms were limited to single word outputs. There are several word combinations in the lexicon, yet through POS-tagging of the algorithm, the combinations in the reviews were divided into independent words. This results in single-word outputs for attributes, descriptions, and sentiments. The authors acknowledge this as a limitation of the proposed algorithms, and propose this problem as a direction for future research.

The anecdotal findings presented in Table 2 shed light on key features of the PWS and PSR algorithms. The PWS algorithm highly depends on the quality of the attribute lexicon and the window size. If the lexicon is accurate enough, this algorithm can identify all attributes in the raw dataset. However, the window is the biggest challenge for this algorithm, which has contributed to the relatively high redundancy value in the tests (55.6%). People *usually* express their feelings for something within a sentence, but not *always*. PWS works best on reviews with short sentences where the probabilities of attributes, sentiments, and descriptions appearing in the same window are high. One can maintain the performance of PWS for longer sentences by increasing the window size; however, that comes at the price of increasing the redundancy of the results. The PSR algorithm, on the other hand, has a different logic. It is not sensitive to the sentence length—as long as the dependency tree is able to identify the target relationships, the sentiment and description words can be extracted effectively. However, PSR highly depends on the accuracy of the relationship identifier; *i.e.*, the dependency tree. The dependency tree has better performance in more formally written text (Chen & Manning, 2014). The user review data includes incredible amounts of grammar errors, unstructured sentences, and typos, which in turn make it difficult for the PSR algorithm to compete with the PWS algorithm in terms of precision and recall. Nevertheless, the use of dependency trees prevents the PSR algorithm from generating redundant outcomes, as validated by the results presented in Table 2.

The overarching goal of the ASE methodology, as discussed in Section 1, is to extract attribute-level sentiment expressions of users from online reviews. Three examples are provided below to illustrate how this task is accomplished by the PWS and PSR algorithms:

> *“The uncaged ultra Boost offers a fit goes perfectly with the foot. Also the sock upper really makes it very easy to slip them on and gives them a very nice look.”*

PWS outputs: attribute: “fit”, sentiment: “perfectly”, description: “uncaged”, attribute: “upper”, sentiment: “easy”, description: “sock”, and attribute: "look”, sentiment: “nice”. The inner logic of PWS is as follows. First identify the attribute, which is “fit”. Then check the previous 5 words and following 5 words, and identify the closest word that in the sentiment lexicon which is “perfectly”, and the closest adj/nn, which is is “uncaged”. The second output set and examples follow the same logic.

PSR outputs: attribute: “fit”, sentiment: “perfectly”, description: “Boost”, attribute: “upper”, sentiment: “easy”, description: “sock”, and attribute: ”look”, sentiment: “nice”, description: ”upper”. The inner logic of PSR is as follows. The first step is the same with PWS, the attribute in the lexicon was “fit”, from the Stanford parser we could get that the word “fit” and “perfectly” gives the relationship as “advmod”; “fit” and “Boost” have the relationship as “rcmod”. The second set follows the same structure. The second output set and other examples presented below follow this structure as well.

> *“Love these shoes. The stretchy fabric and the neck on the shoes are amazing and very easy to slip into.”*

PWS outputs: attribute: "fabric", sentiment: "", description: "stretchy", attribute: "neck", sentiment:" amazing", description: "stretchy". Obviously, the PWS algorithm has a limitation in this kind of situation, and the PSR algorithm can provide more meaningful results. As presented in the first example, in this example, the word been identified first was "fabric", and in a size-5 window, PWS is not able to find a meaningful sentiment word; however, a description word "stretchy" has been found. Then the second attribute was identified with the window sliding.

PSR outputs: attribute: "fabric", sentiment: "love", description: "stretchy" and attribute: "neck", sentiment: "amazing", description: "slip". Same as the first algorithm, the word "fabric" was identified first, then a failry strange result shows that the word "love" and "stretchy" constitute a relationship "dobj". Apparently, the word "love" and "stretchy" are not in the same sentence, but Stanford parser can only manipulate relationships in one sentence, all punctuations are required to be removed before utilizing the parser; this justifies the output that "love" corresponds to the word "stretchy". The PSR output in this case does not represent the exact intention of the user; however, it captures the emotion to some extent. In English syntax rules, neighboring words usually have similar objects (Jurafsky & Martin, 2018); in this case, the user first discusses her overall opinion about the shoe, and then specifically praises the fabric of this shoe. Although the object of discussion changes, the emotion remains the same. The PSR algorithm is astricted by the adoped parser; the results may be enhanced in the future by updating the parser.

> *"I am a female 47 years old runner. These shoes are the best running shoes I've had in a long time. They are perfect for my flat feet I usually struggle finding the perfect shoe, but these were absolutely amazing I can't wait to buy them in every color."*

PWS outputs: attribute: "fit", sentiment: "perfect", description: "flat", attribute: "color", sentiment: "amazing". After identifying the word "fit" and "color", PWS found the set: {"perfect", "flat"} and {"amazing", ""} inside the window.

PSR outputs: attribute: "fit", sentiment: "perfect", description: "flat", attribute: "color", sentiment: "amazing". The relationship found in this case was "acomp" for {"fit", "perfect"} and "advcl" for {"fit", "flat"}. In the second set, the relationship "amod" was detected in the set {"color", "amazing"}.

By looking at the outputs of the ASE algorithms, the design team can quickly conclude that the shoes are a perfect fit for the flat feet of the user and that she finds the offered colors amazing. The sentiment similarity of the first two users for the attribute "fit" is 0.78, since one of them uses the word "perfect" and the other uses the word "amazing" to express their sentiments. In the proposed methodology, those similarities construct the attribute-level similarity between the users' sentiments. The SENT matrices generated by the PWS algorithm were used to measure attribute-level similarities between pairs of users (who have expressed sentiments about that particular attribute) by calculating the cosine similarity of the respective word embeddings of their sentiment expressions modified with respect to the product similarity measure (Eq. (1)). The results for a number of common attributes are depicted in Fig. 3, which signify the existence of the anticipated user segments-although these preliminary results must be fully investigated and verified through a rigorous cluster analysis methodology (see Yi et al., 1889; Zhai et al., 2011). These results shown may also be aggregated to represent the overall (product-level) set of user segments. In the network representations presented in Fig. 3, the nodes, links, and colors represent individual users, pairwise similarities, and user clusters with close similarity, respectively.

5. Conclusions and future research directions

This article presented new semi-supervised, rule-based algorithms for extracting sentiment expressions of users from online reviews, which can serve as an intelligent system augmenting the performance of design teams in identifying user needs on a large scale and translating them into new design concepts. The proposed ASE methodology tackles two limitations of current ABSA literature associated with the need for labeled

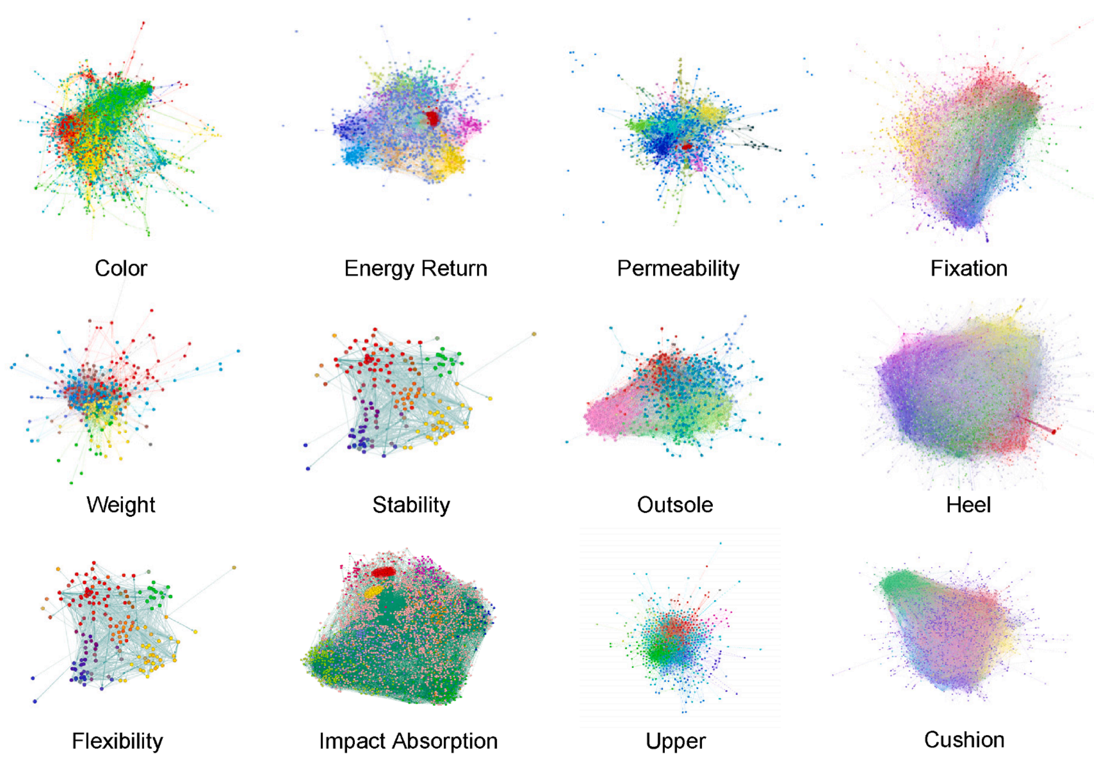

Fig. 3. Clusters of users identified based on the similarity of their expressed sentiments about attributes of different sneakers. Each node represents a user and the length of each link is inversely proportional to the pairwise similarity of the user sentiments.

training data and the focus on sentiment polarity. Successful deployment of ASE, however, requires accurate and comprehensive attribute lexicons which is challenging for two reasons. First, different products may use different terms to refer to the same attribute, and users may refer to those attributes using a variety of synonymous terms. The authors believe that it is possible to develop methods based on pretrained lexical databases and embeddings along with manual labeling and computational methods for pattern recognition and classification in order to identify and group these variations as a single, representative term for each attribute. Second, some attributes and sentiments comprise multiple words that are individually meaningless (*e.g.*, water resistant). The authors propose to address this challenge by enriching the attribute and sentiment lexicons to identify single-word synonyms of the mentioned phrases. User clustering based on sentiment analysis results may also encounter challenges associated with finding appropriate representation for similarity and dealing with inconsistent clustering results from different models or subsets of attributes. Nevertheless, it is possible to tackle these challenges through an ensemble method (Yi, Nasukawa, Bunescu, Niblack, 1889; Zhai et al., 2011) where multiple perspectives on the same data are taken into consideration in order to reach the best performance possible.

This article contributes to the current state of ABSA literature for augmenting the designers' performance in eliciting users' need on a large scale by building and testing two semi-supervised, rule-based algorithms for extracting attribute-level sentiment expressions of users from online reviews. The majority of current attribute-level sentiment analysis methods focus on predicting predefined labels of given attributes, which requires laborious labeling of training data and expert definition of attributes. The algorithms offer clear logic, strong adaptability, and ease of use. This article also introduced a new attribute lexicon for sneaker design, which is universal and can be expanded and used by any rule-based sentiment analysis project on sneaker/shoe data. Further, the proposed methodology enables large-scale comparative analysis of user sentiments and thus clustering of users accordingly. The experiments compared the sentiment similarities of over 20,000 real users with respect to several attributes of sneakers. The visualization of the clustering results shown in Fig. 3 is a useful tool for identifying key product attributes for designers in terms of the number of user reviews (*i.e.*, network size) and the existence of user clusters associated with each attribute (*e.g.*, "color"," "heel", "energy return" in Fig. 3).

The developed rule-based algorithms have several limitations related to their inner logic which must be addressed in future research:

- The PWS algorithm is built upon a basic assumption that the description and sentiment expressions of an attribute appear near the attribute itself. Hence, the performance of this algorithm is highly dependent on the window size, and thus, the window-size parameter may function as a destabilizing factor.
- The critical step in the PSR algorithm is parsing using the Stanford parser, which may be limited by the precision of the Stanford parser itself along with the dependency tree results. Since the mapping of descriptions and sentiment expressions is based on syntactic rules, the parser may not properly function on some unstructured reviews with several slangs or grammar errors.
- Both algorithms search for sentiment expressions and attribute words based on the attribute lexicon and the sentiment lexicon. Therefore, the values of the attribute recall (γ) and sentiment-attribute mapping precision (ζ) metrics are very sensitive to the design of those two lexicons. This property also limits the algorithm into the product class under study (*e.g.*, sneakers), and therefore, generalization to other product classes requires significantly different lexicons, especially for attributes.
- Neither algorithm takes into account the uncertainty stemming from sentence information imbalance. The forms in which users express their preferences vary. Some users express their opinion in an *explicit* fashion, by making it clear which attributes of the product they like or dislike. Other users, however, tend to express their opinions in an *implicit* fashion, which makes it difficult to identify and interpret the attribute they are referring to and/or their sentiments about it.

The proposed research can potentially be applied to a broad range of product design settings where novice designers can improve their design skills by working on the outcomes of the ASE algorithms that lays the foundation for modeling the preferences of user segments and suggests design ideas based on myriad online product reviews. A related example is the recent collaboration between IBM, Tommy Hilfiger, and FIT on enhancing the learning experience of designers through an AI tool that captures and suggests design ideas based on online design images (Rachel Arthur, n.d.). Future research will incorporate the perspectives of large crowds of users and designers in a new design to effectively and accurately align the solution space with what diverse user segments need. This regulatory loop is expected to create a positive, indirect synergy between various designers and users for better alignment of user needs with design. Future research will also expand the research direction to other data such as location, gender, culture, stars, and so forth. It is also possible to draw connections between the notion of review-oriented user segmentation and community building on social media.

CRediT authorship contribution statement

Yi Han: Data curation, Formal analysis, Investigation, Methodology, Validation, Software, Writing - original draft. **Mohsen Moghaddam:** Conceptualization, Formal analysis, Investigation, Methodology, Writing - original draft, Project administration, Resources, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank the anonymous reviewers for their constructive comments that improved the quality and clarity of our work. Research reported in this article was supported by the TIER 1: Seed Grant/Proof of Concept Program of Northeastern University.

References

Alqaryouti, O., Siyam, N., Monem, A. A., & Shaalan, K. (2019). Aspect-based sentiment analysis using smart government review data. *Applied Computing and Informatics*. https://doi.org/10.1016/j.aci.2019.11.003

Arrow, K. J. (1950). A difficulty in the concept of social welfare. *Journal of Political Economy, 58*(4), 328–346. https://doi.org/10.1086/256963

Batista, D. S., Martins, B., & Silva, M. J. (2015). Semi-Supervised Bootstrapping of Relationship Extractors with Distributional Semantics. September, 499–504. https://doi.org/10.18653/v1/d15-1056.

Bing, L. (2015). Sentiment analysis: Mining opinions, sentiments, and emotions. *The Cambridge University Press, 38*, 41–51. https://doi.org/10.1162/COLI

Cao, D., Li, Z., & Ramani, K. (2011). Ontology-based customer preference modeling for concept generation. *Advanced Engineering Informatics, 25*(2), 162–176. https://doi.org/10.1016/j.aei.2010.07.007

Chen, D., & Manning, C. D. (2014). A fast and accurate dependency parser using neural networks. EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, i, 740–750. https://doi.org/10.3115/v1/d14-1082.

Chen, M., & Yao, T. (2010). Combining dependency parsing with shallow semantic analysis for Chinese opinion-element relation identification. 2010 4th International Universal Communication Symposium, IUCS 2010 - Proceedings, 299–305. https://doi.org/10.1109/IUCS.2010.5666009.

Chen, J., Ji, D., Tan, C. L., & Niu, Z. (2007). Relation extraction using label propagation based semi-supervised learning. July, 129–136. https://doi.org/10.3115/1220175.1220192.

Chen, L., Yan, D., & Wang, F. (2017). User perception of sentiment-integrated critiquing in recommender systems. International Journal of Human Computer Studies, 121 (January 2017), 4–20. https://doi.org/10.1016/j.ijhcs.2017.09.005.

Cranenburgh, A. Van. (2019). Embarrassingly Simple Unsupervised Aspect Extraction.

Da Silva, N. F. F., Coletta, L. F. S., Hruschka, E. R., & Hruschka, E. R. (2016). Using unsupervised information to improve semi-supervised tweet sentiment classification. *Information Sciences, 355–356*, 348–365. https://doi.org/10.1016/j.ins.2016.02.002

Desai, J. M., & Mining, A. D. (2015). Sentiment analysis Approach to adapt a shallow parsing based sentiment lexicon. 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 1–4. https://doi.org/10.1109/ICIIECS.2015.7193160.

Desai, J. M., Andhariya, S. R., Giannakopoulos, A., Musat, C., Hossmann, A., Baeriswyl, M., … Baeriswyl, M. (2015). Unsupervised aspect term extraction with B-LSTM & CRF using automatically labelled datasets. *ArXiv, 37*(SemEval), 180–188.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. ArXiv.

Dhaoui, C., Webster, C. M., & Tan, L. P. (2017). Social media sentiment analysis: Lexicon versus machine learning. *Journal of Consumer Marketing, 34*(6), 480–488. https://doi.org/10.1108/JCM-03-2017-2141

Do, H. H., Prasad, P. W. C., Maag, A., & Alsadoon, A. (2019). Deep learning for aspect-based sentiment analysis: A comparative review. *Expert Systems with Applications, 118*, 272–299. https://doi.org/10.1016/j.eswa.2018.10.003

Felfernig, A. (2007). Standardized configuration knowledge representations as technological foundation for mass customization. *IEEE Transactions on Engineering Management, 54*(1), 41–56. https://doi.org/10.1109/TEM.2006.889066

Fernández-Gavilanes, M., Álvarez-López, T., Juncal-Martínez, J., Costa-Montenegro, E., & Javier González-Castaño, F. (2016). Unsupervised method for sentiment analysis in online texts. *Expert Systems with Applications, 58*, 57–75. https://doi.org/10.1016/j.eswa.2016.03.031

Fernández-Gavilanes, M., Álvarez-López, T., Juncal-Martínez, J., Costa-Montenegro, E., & González-Castaño, F. J. (2016). Corrigendum to "Unsupervised method for sentiment analysis in online texts" (Expert Systems with Applications (2016) 58 (57–75)(10.1016/j.eswa.2016.03.031)(S0957417416301300)). Expert Systems with Applications, 61, 394. https://doi.org/10.1016/j.eswa.2016.06.007.

Fogliatto, F. S., & da Silveira, G. J. C. (2008). Mass customization: A method for market segmentation and choice menu design. *International Journal of Production Economics, 111*(2), 606–622. https://doi.org/10.1016/j.ijpe.2007.02.034

Fogliatto, F. S., Da Silveira, G. J. C., & Borenstein, D. (2012). The mass customization decade: An updated review of the literature. *International Journal of Production Economics, 138*(1), 14–25. https://doi.org/10.1016/j.ijpe.2012.03.002

Franke, N., Keinz, P., & Steger, C. J. (2009). Testing the value of customization: When do customers really prefer products tailored to their preferences? *Journal of Marketing, 73*(5), 103–121. https://doi.org/10.1509/jmkg.73.5.103

Franke, N., Schreier, M., & Kaiser, U. (2010). The "I Designed It Myself" effect in mass customization. *Management Science, 56*(1), 125–140. https://doi.org/10.1287/mnsc.1090.1077

García Pablos, A., Cuadros, M., & Rigau, G. (2015). V3: Unsupervised aspect based sentiment analysis for SemEval2015 Task 12. *SemEval*, 714–718. https://doi.org/10.18653/v1/s15-2121.

García-Pablos, A., Cuadros, M., & Rigau, G. (2018). W2VLDA: Almost unsupervised system for aspect based sentiment analysis. *Expert Systems with Applications, 91*, 127–137. https://doi.org/10.1016/j.eswa.2017.08.049

Google. (2017). 6X growth in mobile watch time of travel diaries and vlogs in the past two years - YouTube Data, U.S.

Griffin, A., & Hauser, J. R. (1993). The voice of the customer. *Marketing Science, 12*(1), 1–27. https://doi.org/10.1287/mksc.12.1.1

Gupta, P., Roth, B., & Schütze, H. (2018). Joint Bootstrapping Machines for High Confidence Relation Extraction. 26–36. https://doi.org/10.18653/v1/n18-1003.

Hercig, T., Brychcín, T., Svoboda, L., Konkol, M., & Steinberger, J. (2016). Unsupervised methods to improve aspect-based sentiment analysis in czech. Computacion y Sistemas, 20(3), 365–375. https://doi.org/10.13053/CyS-20-3-2469.

Hoang, M., Bihorac, O. A., & Rouces, J. (2019). Aspect-Based Sentiment Analysis using {BERT}. Proceedings of the 22nd Nordic Conference on Computational Linguistics, 187–196. https://www.aclweb.org/anthology/W19-6120.

Hu, M., Liu, B., & Street, S. M. (2004). Mining and Summarizing Customer Reviews. https://doi.org/10.1002/j.1532-2149.2013.00312.x.

Hu, X., Tang, J., Gao, H., & Liu, H. (2013). Unsupervised sentiment analysis with emotional signals. WWW 2013 - Proceedings of the 22nd International Conference on World Wide Web, 607–617. https://doi.org/10.1145/2488388.248844.

Yi, J., Nasukawa, T., Bunescu, R., & Niblack, W. (1889). Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. Notes and Queries, s7-VIII(201), 349. https://doi.org/10.1093/nq/s7-VIII.201.349-g.

Jurafsky, D., & Martin, J. H. (2018). Speech and Language Processing.

Karimi, A., Rossi, L., Prati, A., & Full, K. (2020). Adversarial Training for Aspect-Based Sentiment Analysis with BERT. http://arxiv.org/abs/2001.1131.

A Complete List of Sneakerhead Terminology From A to Z | KickBackz. (n.d.). Retrieved April 15, 2019, from https://kickbackz.myshopify.com/pages/a-complete-list-of-sneakerhead-terminology-from-a-to-z.

Klein, D., & Manning, C. D. (2007). Accurate unlexicalized parsing. 423–430. https://doi.org/10.3115/1075096.1075150.

Koren, Y., Hu, S. J., Gu, P., & Shpitalni, M. (2013). Open-architecture products. *CIRP Annals – Manufacturing Technology, 62*(2), 719–729. https://doi.org/10.1016/j.cirp.2013.06.001

Lakkaraju, H., Bhattacharyya, C., Bhattacharya, I., & Merugu, S. (2011). Exploiting coherence for the simultaneous discovery of latent facets and associated sentiments. Proceedings of the 11th SIAM International Conference on Data Mining, SDM 2011, May, 498–509. https://doi.org/10.1137/1.9781611972818.43.

Lakkaraju, H., Socher, R., & Manning, C. D. (2014). Aspect specific sentiment analysis using hierarchical deep learning. *NIPS WS on Deep Neural Networks and Representation Learning*, 1–9.

Law, M., Kwok, R. C. W., & Ng, M. (2016). An extended online purchase intention model for middle-aged online users. *Electronic Commerce Research and Applications, 20*, 132–146. https://doi.org/10.1016/j.elerap.2016.10.005

Lee, T. Y., & Bradlow, E. T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research, 48*(5), 881–894. https://doi.org/10.1509/jmkr.48.5.881

Li, C.-C., Dong, Y., Xu, Y., Chiclana, F., Herrera-Viedma, E., & Herrera, F. (2018). An overview on managing additive consistency of reciprocal preference relations for consistency-driven decision making and Fusion: Taxonomy and future directions. *Information Fusion.* https://doi.org/10.1016/j.inffus.2018.12.004

Li, X., Bing, L., Li, P., Lam, W., & Yang, Z. (2018). Aspect term extraction with history attention and selective transformation. IJCAI International Joint Conference on Artificial Intelligence, 2018-July, 4194–4200. https://doi.org/10.24963/ijcai.2018/583.

Liu, J., Yao, J., & Wu, G. (2005). Sentiment classification using Information Extraction technique. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 3646 LNCS, 216–227. https://doi.org/10.1007/11552253_20.

Liu, B. (2010a). Handbook of Natural Language Processing (Second). http://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf.

Liu, B. (2010b). Sentiment Analysis and Subjectivity. 1–38.

Local Consumer Review Survey | Online Reviews Statistics & Trends. (2019). https://www.brightlocal.com/research/local-consumer-review-survey.

Ma, B., & Chen, H. (2017). A Chinese product feature extraction method based on KNN algorithm. *Open Journal of Social Sciences, 05*(10), 128–138. https://doi.org/10.4236/jss.2017.510012

Ma, R., Wang, K., Qiu, T., Sangaiah, A. K., Lin, D., & Liaqat, H. Bin (2017). Feature-based compositing memory networks for aspect-based sentiment classification in social internet of things. *Future Generation Computer Systems, 92*, 879–888. https://doi.org/10.1016/j.future.2017.11.036

Masood, M. A., Abbasi, R. A., & Keong, N. W. (2020). Context-aware sliding window for sentiment classification. *IEEE Access, 8*, 4870–4884. https://doi.org/10.1109/ACCESS.2019.2963586

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2006). Distributed representations of words and phrases and their compositionality. *CrossRef Listing of Deleted DOIs, 1*, 1–9. https://doi.org/10.1162/jmlr.2003.3.4-5.951

Mirtalaie, M. A., Hussain, O. K., Chang, E., & Hussain, F. K. (2017). A decision support framework for identifying novel ideas in new product development from cross-domain analysis. *Information Systems, 69*, 59–80. https://doi.org/10.1016/j.is.2017.04.003

Mirtalaie, M. A., Hussain, O. K., Chang, E., & Hussain, F. K. (2018). Extracting sentiment knowledge from pros/cons product reviews: Discovering features along with the polarity strength of their associated opinions. *Expert Systems with Applications, 114*, 267–288. https://doi.org/10.1016/j.eswa.2018.07.046

ORACLE. (2019). The Digital Transformation of Product Design: How are design teams using and planning for design technology fueled by data? Engineering.Com.

Pandarachalil, R., Sendhilkumar, S., & Mahalakshmi, G. S. (2015). Twitter sentiment analysis for large-scale data: An unsupervised approach. *Cognitive Computation, 7*(2), 254–262. https://doi.org/10.1007/s12559-014-9310-z

Pang, B., & Lee, L. (2006). Opinion mining and sentiment analysis. *Foundations and Trends in Informatio Retrieval, 1*(2), 91–231. https://doi.org/10.1561/1500000001

Peng, H., Ma, Y., Li, Y., & Cambria, E. (2018). Learning multi-grained aspect target sequence for Chinese sentiment analysis. *Knowledge-Based Systems, 148*, 167–176. https://doi.org/10.1016/j.knosys.2018.02.034

Pham, D. H., & Le, A. C. (2018). Learning multiple layers of knowledge representation for aspect based sentiment analysis. Data and Knowledge Engineering, 114(June 2017), 26–39. https://doi.org/10.1016/j.datak.2017.06.001.

Pontiki, M., Galanis, D., & Papageorgiou, H. (2015). SemEval-2015 Task 12: Aspect Based Sentiment Analysis.

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-smadi, M., Al-ayyoub, M., Zhao, Y., Qin, B., Clercq, O. De, Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., Jiménez-zafra, S. M., & Eryiğit, G. (2016). SemEval-2016 Task 5: Aspect Based Sentiment Analysis. 19–30.

Pontiki, M., Galanis, D. D., Pavlopoulos, J., Papageorgiou, H. H., Androutsopoulos, I., Manandhar, S., Al-smadi, M., Al-ayyoub, M., Zhao, Y., Qin, B., Clercq, O. De, Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., Jiménez-zafra, S. M., & Eryiğit, G. (2015). SemEval-2014 Task 4: Aspect based sentiment analysis. *SemEval, 27–35*. https://doi.org/10.3115/v1/s14-2004

Rachel Arthur. (n.d.). Artificial Intelligence Empowers Designers In IBM, Tommy Hilfiger And FIT Collaboration. Retrieved February 6, 2019, from https://www.forbes.com/sites/rachelarthur/2018/01/15/ai-ibm-tommy-hilfiger/#2b92a31178ac.

Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems, 89*, 14–46. https://doi.org/10.1016/j.knosys.2015.06.015

Rietzler, A., Stabinger, S., Opitz, P., & Engl, S. (2019). Adapt or Get Left Behind: Domain Adaptation through BERT Language Model Finetuning for Aspect-Target Sentiment Classification. http://arxiv.org/abs/1908.11860.

Rosenthal, S., Nakov, P., Kiritchenko, S., Mohammad, S., Ritter, A., & Stoyanov, V. (2015). SemEval-2015 Task 10: Sentiment analysis in Twitter. *SemEval*, 451–463. https://doi.org/10.18653/v1/s15-2078.

Salvador, F., de Holan, P. M., & Piller, F. (2009). Cracking the code of mass customization. *MIT Sloan Review*, 71–78.

Salvador, F., De Holan, P. M., & Piller, F. T. (2009). Cracking the code of mass customization | MIT sloan management review. *MIT Sloan Management Review, 50315*.

Salvador, F., Rungtusanatham, M., & Forza, C. (2004). Supply-chain configurations for mass customization. *Production Planning & Control, 15*(4), 381–397. https://doi.org/10.1080/0953728042000238818

Santos, J. M., & Embrechts, M. (2009). On the use of the adjusted rand index as a metric for evaluating supervised classification. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 5769 LNCS(PART 2), 175–184. https://doi.org/10.1007/978-3-642-04277-5_18.

SemEval 2014 The 8th International Workshop on Semantic Evaluation Proceedings of the Workshop Dublin , Ireland (Issue SemEval). (2014).

Singh, V. K., Piryani, R., Uddin, A., Waila, P., & Marisha. (2013). Sentiment analysis of textual reviews; Evaluating machine learning, unsupervised and SentiWordNet approaches. 2013 5th International Conference on Knowledge and Smart Technology (KST), 122–127. https://doi.org/10.1109/kst.2013.6512800.

Sneaker Genius – Glossary of Sneaker terms | Genius. (n.d.). Retrieved April 15, 2019, from https://genius.com/Sneaker-genius-glossary-of-sneaker-terms-annotated.

Sneaker terminology. (n.d.). Retrieved April 15, 2019, from https://www.complex.com/sneakers/2012/09/a-beginners-guide-to-sneaker-terminology/aglets.

Sneaker Terminology | SneakerFiles. (n.d.). Retrieved April 15, 2019, from https://www.sneakerfiles.com/sneaker-terminology/.

Soh, C., Yu, S., Narayanan, A., Duraisamy, S., & Chen, L. (2019). Employee profiling via aspect-based sentiment and network for insider threats detection. *Expert Systems with Applications, 135*, 351–361. https://doi.org/10.1016/j.eswa.2019.05.043

The Essential Guide to Sneaker Vocabulary | Sole Collector. (n.d.). Retrieved April 15, 2019, from https://solecollector.com/news/guide-to-sneaker-terms/.

C. Sun L. Huang X. Qiu Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence 2019 http://arxiv.org/abs/1903.09588.

Suresh, H., & Gladston, R. S. (2016). An unsupervised fuzzy clustering method for twitter sentiment analysis. 2016 International Conference on Computation System and Information Technology for Sustainable Solutions, CSITSS 2016, 80–85. https://doi.org/10.1109/CSITSS.2016.7779444.

Suryadi, D., & Kim, H. (2018). A systematic methodology based on word embedding for identifying the relation between online customer reviews and sales rank. *Journal of Mechanical Design, Transactions of the ASME, 140*(12), 1–12. https://doi.org/10.1115/1.4040913

Suryadi, D., & Kim, H. M. (2019). A data-driven approach to product usage context identification from online customer reviews. *Journal of Mechanical Design, 141*(12). https://doi.org/10.1115/1.4044523

Tamchyna, A., & Veselovská, K. (2016). UFAL at SemEval-2016 task 5: Recurrent neural networks for sentence classification. SemEval 2016 - 10th International Workshop on Semantic Evaluation, Proceedings, 367–371. https://doi.org/10.18653/v1/s16-1059.

Tang, H., Tan, S., & Cheng, X. (2009). A survey on sentiment detection of reviews. *Expert Systems with Applications, 36*(7), 10760–10773. https://doi.org/10.1016/j.eswa.2009.02.063

Tetsuya Nasukawa, J. Y. (1956). Sentiment analysis: Capturing favorability using natural language processing. *Biological Reviews, 31*(1), 1–29. https://doi.org/10.1111/j.1469-185X.1956.tb01550.x

The Stanford Natural Language Processing Group. (n.d.). Retrieved April 11, 2019, from https://nlp.stanford.edu/software/lex-parser.shtml.

Toh, Z., & Su, J. (2015). NLANGP: Supervised machine learning system for aspect category. *Classification and Opinion Target Extraction., 14*(SemEval), 496–501.

Ureña, R., Chiclana, F., & Herrera-Viedma, E. (2018). A new influence based network for opinion propagation in social network based scenarios. *Procedia Computer Science, 139*, 329–337. https://doi.org/10.1016/j.procs.2018.10.282

Ureña, R., Kou, G., Dong, Y., Chiclana, F., & Herrera-Viedma, E. (2019). A review on trust propagation and opinion dynamics in social networks and group decision making frameworks. *Information Sciences, 478*, 461–475. https://doi.org/10.1016/j.ins.2018.11.037

Usha, M. S., & Indra Devi, M. (2013). Analysis of sentiments using unsupervised learning techniques. 2013 International Conference on Information Communication and Embedded Systems, ICICES 2013, 241–245. https://doi.org/10.1109/ICICES.2013.6508203.

Vashishtha, S., & Susan, S. (2019). Fuzzy rule based unsupervised sentiment analysis from social media posts. *Expert Systems with Applications, 138*. https://doi.org/10.1016/j.eswa.2019.112834

Vinh, N. X., Epps, J., & Bailey, J. (2010). Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research, 11*, 2837–2854.

Wagner, C., & Majchrzak, A. (2006). Enabling Customer-Centricity Using Wikis and the Wiki Way. *Journal of Management Information Systems, 23*(3), 17–43. https://doi.org/10.2753/MIS0742-1222230302

Wan, X. (2008). Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis. EMNLP 2008 - 2008 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference: A Meeting of SIGDAT, a Special Interest Group of the ACL, October, 553–561. https://doi.org/10.3115/1613715.1613783.

Wang, T., Cai, Y., Leung, H. F., Lau, R. Y. K., Li, Q., & Min, H. (2014). Product aspect extraction supervised with online domain knowledge. *Knowledge-Based Systems, 71*, 86–100. https://doi.org/10.1016/j.knosys.2014.05.018

Wang, Y., Mo, D. Y., & Tseng, M. M. (2018). Mapping customer needs to design parameters in the front end of product design by applying deep learning. *CIRP Annals, 67*(1), 145–148. https://doi.org/10.1016/j.cirp.2018.04.018

Wu, S., Wu, F., Chang, Y., Wu, C., & Huang, Y. (2019). Automatic construction of target-specific sentiment lexicon. *Expert Systems with Applications, 116*, 285–298. https://doi.org/10.1016/j.eswa.2018.09.024

Xianghua, F., Guo, L., Yanyan, G., & Zhiqiang, W. (2013). Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon. *Knowledge-Based Systems, 37*, 186–195. https://doi.org/10.1016/j.knosys.2012.08.003

Xu, X. (2012). From cloud computing to cloud manufacturing. *Robotics and Computer-Integrated Manufacturing, 28*(1), 75–86. https://doi.org/10.1016/J.RCIM.2011.07.002

Xu, L., Lin, J., Wang, L., Yin, C., & Wang, J. (2017). Deep convolutional neural network based approach for aspect-based sentiment. *Analysis., 143*(Ast), 199–204.

Xu, H., Liu, B., Shu, L., & Yu, P. S. (2018). Double embeddings and cnn-based sequence labeling for aspect extraction. ACL 2018–56th Annual Meeting of the Association for Computational Linguistics. *Proceedings of the Conference (Long Papers), 2*, 592–598.

Xu, F., Zhang, X., Xin, Z., & Yang, A. (2019). Investigation on the Chinese text sentiment analysis based on convolutional neural networks in deep learning. *Computers, Materials and Continua, 58*(3), 697–709. https://doi.org/10.32604/cmc.2019.05375.

Yang, C., Zhang, H., Jiang, B., & Li, K. (2019). Aspect-based sentiment analysis with alternating coattention networks. *Information Processing and Management, 56*(3), 463–478. https://doi.org/10.1016/j.ipm.2018.12.004

Yi, J., Nasukawa, T., Bunescu, R., & Niblack, W. (n.d.). Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques. Third IEEE International Conference on Data Mining, 427–434.

Zhai, Z., Liu, B., Xu, H., & Jia, P. (2011). Clustering product features for opinion mining. Proceedings of the 4th ACM International Conference on Web Search and Data Mining, WSDM 2011, 347–354. https://doi.org/10.1145/1935826.1935884.

Zhang, L., Wang, S., & Liu, B. (2018b). Deep Learning for sentiment analysis: a survey. ArXiv:1801.07883 [Cs, Stat].

Zhang, L., Wang, S., & Liu, B. (2018a). Deep Learning for Sentiment Analysis: A Survey. CoRR, abs/1801.0.

Zheng, X., Lin, Z., Wang, X., Lin, K. J., & Song, M. (2014). Incorporating appraisal expression patterns into topic modeling for aspect and sentiment word identification. *Knowledge-Based Systems, 61*, 29–47. https://doi.org/10.1016/j.knosys.2014.02.003

Zhou, F., Jiao, R. J., & Linsey, J. S. (2015). Latent customer needs elicitation by use case analogical reasoning from sentiment analysis of online product reviews. *Journal of Mechanical Design, Transactions of the ASME, 137*(7). https://doi.org/10.1115/1.4030159