Computers & Industrial Engineering 169 (2022) 108186

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

Towards Lean Automation: Fine-Grained sentiment analysis for customer value identification

Yan Xiao [a,b], Congdong Li [a,c,d,*], Matthias Thürer [e], Yide Liu [a], Ting Qu [d,e,f]

[a] Macau University of Science and Technology, School of Business, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macao, PR China
[b] Chongqing University of Technology, College of Mechanical Engineering, Chongqing University of Technology, Chongqing 400054, PR China
[c] Jinan University, Management School, Jinan University, Guangzhou 510632, PR China
[d] Jinan University, Institute of Physical Internet, Jinan University (Zhuhai Campus), 519070 Zhuhai, PR China
[e] Jinan University, School of Intelligent Systems Science and Engineering, Jinan University (Zhuhai Campus), 519070 Zhuhai, PR China
[f] Jinan University, Institute of the Belt and Road & Guangdong-Hong Kong-Macao Greater Bay Area, Jinan University, Guangzhou 510632, PR China

ARTICLE INFO

Keywords:
Product Development
Fine-grained Sentiment Analysis
Deep Learning
Case Study

ABSTRACT

That customer value should drive product development and production is a basic tenet of the Toyota Production System and Lean. Traditional means to extract what the customer wants often focus on customer surveys. But surveys are time consuming and costly. At the same time, there exists a large amount of customer comments in online reviews that is easily accessible, whilst the advances of data science, for example as part of Lean Automation, provide new means to extract information from this data. In this context, a new approach to fine-grained sentiment analysis of Chinese consumer data is developed. The new approach integrates pre-training language model, conditional random field model and linguistic knowledge model. The new approach is shown to outperform traditional approaches in a comparison experiment, while an ablation experiment shows that our new approach is parsimonious, i.e., all three constituting components are needed. Finally, a use case is presented that exemplifies how our new approach can support managers in identifying customer value (through positive evaluations), and most importantly guide Lean improvement, through detailed information on characteristics that are evaluated negatively, ranked according to customer importance. Findings have important implications for research and practice.

1. Introduction

Kiirchiro Toyoda, the founder of the Toyota Motor Cooperation, is quoted to have said "*Study what customers want and reflect that in your product*" (Hino, 2005). That it is customer value that should determine what a company produces and, consequently, how a company produces, remains a fundamental principle of the Toyota Production System (Ohno, 1988; Thomas, 1989; Yazdani, 1994), Lean Production (Womack et al., 1990), and the more recent Lean Automation (Bittencourt et al., 2020; Tortorella et al., 2020), which seeks to integrate Lean with Industry 4.0. In this Lean Automation context, this study proposes a new data mining approach to distil what Chinese language customer want, to provide input for product development and process improvement.

The first and most fundamental principle in lean thinking is "*identifying customer value*" (Jasti & Kodali, 2015). Customer value should guide product development, to ensure only valued product characteristics are provided and overprocessing waste in the production process avoided. However, customer value can only be defined by the customer (Jasti & Kodali, 2014). Conducting interviews is a traditional method to capture customer requirements (Ghezzi, 2019; Maheshwari et al.), and large-scale consumer surveys and market analysis are widely applied (Bicen & Johnson, 2015; Bortolini et al., 2018; Ghezzi, 2019; Peralta et al., 2020). Other techniques that are traditionally used include Ethnographical Studies, Contextual Inquiry, Customer Defection, Prototypes, Minimum Viable Product, and Hypotheses Test (Peralta et al., 2020), whilst it is also argued that interviews can be enhanced by observing customers that actual use the product (Bortolini et al., 2018). Meanwhile, Quality Function Deployment (QFD) includes means to rank product characteristics from the customer perspective (Bamford et al., 2015; Gautam & Singh, 2008; Khurum et al., 2014), as does the Kano

* Corresponding author at: Macau University of Science and Technology, School of Business, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macao, PR China (C. Li).
E-mail addresses: licd@jnu.edu.cn (C. Li), matthiasthurer@workloadcontrol.com (M. Thürer), ydliu@must.edu.mo (Y. Liu), quting@jnu.edu.cn (T. Qu).

https://doi.org/10.1016/j.cie.2022.108186
Received 8 June 2021; Received in revised form 30 December 2021; Accepted 13 April 2022
Available online 16 April 2022

model (Tontini et al., 2012) or other forms of prioritization matrices (Jasti & Kodali, 2015; Khurum et al., 2014).

But conducting surveys and collecting primary data is time consuming and expensive. In fact, all of the traditional techniques to distil what the costumer want listed above, are typically restricted to small samples. New technologies, specifically the advancement of data science, together with the vast availability of consumer data, arguably provides a cheaper and faster alternative (Sheng et al., 2017). In fact, today large amounts of consumer reviews are available in twitters, blogs and product reviews (Chae, 2015). These reviews can be used to reveal consumer preferences and opinions on certain aspects of products, thereby identifying what consumers care the most about (Jin et al., 2016; Zhang et al., 2019).

In this context, fine-grained sentiment analysis has received extensive attention (Cambria, 2016). Fine-grained sentiment analysis is important since coarse-grained sentiment analysis can only judge the overall sentimental tendency of the text, i.e., in relation to the product itself. Text-level and sentence-level sentiment analysis encounter problems when analysing multiple aspects of the review object (Zhang et al., 2019). But the product-centred sentimental tendency of the product review text may be inconsistent with the sentimental tendency in relation to the different aspects of the product. In other words, there may be differences in customer sentiments in relation to the product and its characteristics. For product development and process improvement the latter is the more important. It is therefore necessary to identify aspect words, emotional words, and the relationship between them. For example, a text with a negative sentiment tendency does not mean that the reviewer has a negative sentiment in relation to the review object and all its aspects in the text. It is consequently necessary to separately identify the emotional expressions on objects and their aspects.

The aim of fine-grained sentiment analysis is to extract fine-grained opinion elements that reveal customer preferences in relation to specific product aspects and corresponding emotional words (Do et al., 2019). It thus seems appropriate to enhance Lean, providing a cheaper and faster approach to distil what the customer wants in terms of specific customer requirements. But traditional fine-grained sentiment analysis is typically operationalized through statistical machine learning methods of feature engineering, such as topic models, part-of-speech rules and WordNet dictionaries, domain ontology conceptual space models, etc. These methods depend on manually constructed features and have limitations when encountering different application scenarios (LeCun et al., 2015) (Goodfellow et al., 2016; Xiao et al., 2021). Deep learning provides a means to overcome these shortcomings, benefiting fine-grained sentiment analysis specifically through the powerful learning abilities of deep neural networks. But existing neural network models usually use the coding capability of the network structure to extract traits of text, disregarding linguistic knowledge contained in the language itself, such as part-of-speech knowledge and word segmentation knowledge. This may result in incomprehensive trait extraction. Further, the existing neural network models employ conventional sequence models to encode the sequence features of text when modelling user features, which may lead to inadequate trait interaction since the context information of users is not fully captured.

In response, and following the guidelines in (Denyer et al., 2008), this paper starts with the following design proposition: A new mechanism for fine grained sentiment analysis of Chinese consumer data is developed that by focusing on product features provides information on value (positive) and process improvement possibilities (negative).

The new approach extends existing literature by integrating pretraining language model, conditional random field model and linguistic knowledge model to extract the sentiment polarity of users from historical data containing user preferences and product characteristics. While there has been previous literature on identifying product weaknesses (Zhang et al., 2012) and to-be-improved product features (Zhang et al., 2019), this literature typically focuses on the sentence level. In contrast, our approach goes one level deeper. It focuses on the noun phrase level, which in Chinese is the main noun, and any words or phrases that describe or modify the main noun. The focus on Chinese consumer data is another contribution given that most literature on fine-grained sentiment analysis focuses on English (Zhang et al., 2009). There are restrictions on the application of deep learning models across different language families (Ma & Hovy, 2016), for example Indo-European and Sino-Tibetan, and techniques developed for one language family are typically not directly transferable to other language families. Research on sentiment analysis in English language has undergone major developments in recent years, but Chinese sentiment analysis research has not evolved significantly despite the exponential growth of Chinese e-business and e-markets (Peng et al., 2017).

To verify its effectiveness and accuracy, we then compare the new approach against six alternative approaches, before we conduct an ablation experiment to prove that our approach is parsimonious, i.e., all three constituting components are needed. To provide pragmatic validity, a use case is presented that exemplifies how our approach can support managers in identifying customer value (through positive evaluations). Most importantly, the use case also highlights how the new approach can be used to guide Lean improvement. Our new approach provides detailed negative evaluated characteristics ranked according to customer importance. These negative characteristics can be matched with a firm's process statistics to effectively link customer value to improvement efforts.

The remainder of this paper is structured as follows. The relevant literature on fine grained sentiment analysis is first reviewed in Section 2, before our new fine-grained sentiment analysis approach is outlined in Section 3. The results from comparison and ablation experiments are then provided in Section 4. The use case is presented in Section 5. Finally, Section 6 summarizes main conclusions, including limitations and venues for future research.

2. Literature review

Sentiment analysis can be divided into coarse-grained sentiment analysis and fine-grained sentiment analysis. The application of coarse-grained sentiment analysis to customer reviews refers to the overall sentimental inclination classification of the text. First it is judged whether texts are tendentiousness, and then texts are sorted according to polarity classification, e.g., into positive and negative, or commendatory, neutral and derogatory, etc. Coarse-grained analysis includes text-level and sentence-level sentiment analysis. For example, Yu & Hatzivassiloglou, (2003) employ a Bayesian classification model to classify positive and negative opinions of news. Three different supervised machine learning methods (Naïve Bayes, Maximum Entropy and Support Vector Machine) are applied by Pang et al., (2002) to classify movie reviews. Meanwhile, a five-fold cross-validation approach to classifying tweets subjectively and objectively is used by Jiang et al., (2011), whilst Mcdonald et al. (2007) propose a hierarchical sequence learning model for sentence-level and text-level sentiment classification. Above studies highlight the potential of coarse-grained sentiment analysis for overall text classification. However, to guide product development and process improvement efforts, more detail on specific characteristics is needed. To distil users' sentimental attitudes towards a specific review object and its related attributes, fine-grained sentiment analysis, which uses information extraction technology to distil aspects of products and opinion targets from the review text (Pontiki et al., 2018; Schouten & Frasincar, 2016), arguably provides a better solution. The literature on fine-grained sentiment analysis can be divided into rule-based fine-grained sentiment analysis, fine-grained sentiment analysis based on traditional machine learning, and fine-grained sentiment analysis based on deep learning. All three will be reviewed next.

2.1. Rule-based Fine-grained sentiment analysis

Rule-based fine-grained sentiment analysis mainly analyses the syntactic structure information and linguistic rules of the text. It obtains the sentiment polarity of the different aspects of products through rule matching. For example, Long et al. (2015) propose a series of identification rules to extract sentiment evaluation units via a dependency syntax, whilst Mazur et al. (2017) assess the weight ratio of aspects of objects by lexical analysis. Mazur et al. (2017) first extract the aspects of objects, then use pre-defined rules to filter the aspects of object words, and then conduct aspect-level sentiment analysis. Meanwhile, Yan et al. (2015) utilize the Pagerank algorithm to extract the characteristic attributes of the product, before a multi-rule combination is used to conduct a fine-grained emotional orientation analysis.

Apart from studies focusing on traditional rule-based methods, more recent studies also focus on the foundation of the external knowledge base, e.g., emotional dictionaries. For example, Mishra et al., (2019) propose a fine-grained sentiment analysis for microblog comments that integrates emotional dictionary and semantic rules. This expands diversified emotional dictionaries, such as emoticons, network terms, and single-character emotional words, and integrates semantic rules to establish a microblog sentiment analysis model.

Rule-base sentiment analysis is a powerful tool to capture the structural component information in a sentence given the rules of linguistics. However, statistical characteristics and semantic information of the entire sentence or comment text can typically not be captured. This led to the development of alternative approaches that use traditional machine learning and deep learning.

2.2. Fine-grained sentiment analysis based on traditional Machine learning

Fine-grained sentiment analysis on the basis of traditional machine learning typically uses statistical features in the text, such as Term Frequency–Inverse Document Frequency (TF-IDF; e.g. Agarwal et al., (2020); Quan & Ren, (2014)), Hidden Markov Model (HMM; e.g. W. Jin & Ho, (2009)), or Conditional Random Field (CRF; e.g. F. Li et al., (2010)), to extract aspects of the product. It then uses traditional machine learning algorithms to classify the emotional inclination for these aspects, such as Support Vector Machine (SVM, e.g. (Wagner et al., 2014)), logistic regression (Rezaeian et al., 2020), or Naive Bayes (Liu et al., 2020).

For example, the TF-IDF algorithm utilized by Abirami & Askarunisa, (2016) extracts the evaluation objects (and their corresponding opinions) to construct feature opinion pairs from review data. It then uses a sentiment dictionary with specific sentiment intensity to analyse the sentiment tendency of the opinion words. Meanwhile, Miao et al., (2010) uses CRF to extract the aspects of products by typical rules, before dividing these aspects into multi-classified emotions through the Naive Bayes method.

Fine-grained sentiment analysis that uses traditional machine learning can make the utmost of the statistical characteristics of texts, while capturing the dependent lexical relation. It also is less labour intensive than rule-based analysis because it does not require a template definition, and because it can more easily be automated. However, traditional machine learning methods struggle with semantic features in texts, which in turn may lead to a loss in implicit semantic information. To better use the lexical and semantic information in the sentence, analysis methods based on deep learning were developed.

2.3. Fine-grained sentiment analysis on the premise of deep learning

Fine-grained sentiment analysis based on deep learning provides an end-to-end analysis (Li et al., 2020) by transforming aspects of objects extraction and sentiment classification problems into sequence labelling problems that are then solved by use of deep neural networks, such as Recurrent Neural Network (RNN, e.g. Tang et al., (2016)), Convolutional Neural Network (CNN; e.g. Collobert et al., (2011)), or self-attention mechanism (e.g. Song et al., (2019)).

RNN has the inherent advantage of capturing long sequence text features. For example, Tang et al., (2015) use Long Short Term Memory (LSTM) networks, a special kind of RNN, to capture the semantic for input information features. This is followed by use of the self-attention mechanism to match the information and complete the emotional classification. Meanwhile, Khattak et al. (2020) use LSTM to semantically model the input text information, and additional employ CRF to model the dependent relationship between words to complete the mutual constraints between tags.

Although CNN is not well suited for capturing the features of timing sequences, it has been successfully applied for text classification in the field of Nature Language Processing (NLP). For example, Xu et al. (2020) combined word embedding and CNN. Word2Vec is first used to construct a product feature word list and noise vocabulary, before the noise vocabulary is utilized to extract the feature words. Finally, CNN is used to classify fine-grained emotions at the product aspect-level. Meanwhile, Li et al., (2020) combine CNN with Bidirectional Long-term and Short-term Memory Network (BLSTM) into a new CNN-BLSTM sentiment analysis. This new CNN-BLSTM integrates the phrase features of CNN and the timing sequence features of LSTM.

Both RNN and CNN suffer from limited memory, and consequently struggle with the semantics of long sentences. This weakness can be overcome by self-attention. For example, Song et al., (2019) employs the self-attention mechanism to encode the input text information, whilst Li et al. (2019) fine-tune the end-to-end model of fine-grained sentiment analysis by use of the Bidirectional Encoder Representations from Transformers (BERT) pre-training model. In the above-mentioned deep learning model, the BERT model can entirely capture the semantic information of the text, producing a better final emotion classification. However, the BERT model lacks the statistical characteristics and POS knowledge of traditional machine learning.

2.4. Discussion of the literature

A large body of literature on fine-grained sentiment analysis emerged. Although existing models provided good results in some contexts, they often suffer from an incomprehensive extraction of user characteristic. For example, existing neural network models usually employ network structures to extract and encode text features, neglecting linguistic knowledge. This may lead to incomplete characteristic extraction. Furthermore, existing neural network models capture user representations using conventional sequence models to encode the user-related sequence characteristics, such as RNN or LSTM. This does not fully capture the context information of users, thus contributing to inadequate feature interaction. These issues impose restrictions on the performance of deep learning models.

In response, this study presents a new approach to fine-grained sentiment analysis. To address the problem of incomplete user feature extraction, and to capture richer text features, the new model incorporates external linguistic knowledge, lexical information, and word segment information in the text coding. To solve the issue of inadequate user feature interaction, the pre-training language model BERT is adopted to encode user features. This allows for more efficiently and fully exploring the latent needs and value recognitions of users. Finally, CRF is employed to decode the optimal sequence of user preference labels to enhance the text representation and improve the accuracy of the model.

3. Method development

3.1. Definition of the task

The goal of the proposed model is to learn a prediction function of customer needs and values, which predicts the sentiment polarity of a user for a product with which the user has interacted based on historical data. More specifically, the task is to predict the sentiment label y of user h for a new product x according to the input historical text X_h and the corresponding sentiment label Y_h, which can be formalized as.

$$y = f_\theta(x|X_h, Y_h),$$

Where f refers to the prediction model, θ is a parameter set in the model, $X_h = \{x_1, x_2, \cdots, x_n\}$ is the historical data of product interactions, $Y_h = \{y_1, y_2, \cdots, y_n\}$ the corresponding sentiment polarity, $x_i = \{w_1, w_2, \cdots, w_k\}$, $y_k = \{z_1, z_2, \cdots, z_k\}$, $w_k(k = 1, 2, \cdots)$ denotes the k th individual word in the text x_i, and $z_k \in \{B-POS, I-POS, B-NEG, I-BEG, B-NEU, I-NEU, O\}, k = 1, 2, \cdots$ refers to the kth sentiment polarity label in the sentiment label sequence y_i.

Fine-grained sentiment analysis is generally divided into two tasks: aspects of product extraction and sentiment analysis. Considering the inconsistency between the overall sentimental tendency of item review data and the sentimental tendency of each aspect of items, both are transformed into a sequence labeling task, which converts the two-step task into an end-to-end task to comprehensively capture the corresponding relationship between semantic information and the lexical produced in texts.

As illustrative example, the following review: "*I have received the integrated stove from your store and have used it for a while. I really like its exterior design, which gives it a modern appearance, and its noise is no big deal, but its energy consumption is a bit too high. I hope there will be follow-up improvements.*" The user in this example has a positive attitude towards the appearance of the integrated stove, a moderate attitude towards its noise, and a negative attitude towards its high energy consumption. Note that we provide an English example here given English readership, but the proposed model is developed for Chinese. Chinese and English are quite different. For example, Chinese does not segment words by spaces in sentences; rather a string of Chinese text is made up of equally spaced graphemes that are called characters (Peng et al., 2017; Zhang et al., 2009). A 'word' or concept is typically composed of two or more characters. Chinese texts processing consequently starts with word segmentation. Meanwhile, words in Chinese are semantically and syntactically ambiguous. Most characters have far more than one meaning, and the combination of characters further increases complexities. This makes the computation of sentiment polarity a difficult task (Peng et al., 2017). Finally, Chinese is an analytical language, which means it has a very high syntactic dependency.

Our model assigns a label to each word in the text, giving a POS label to the product characteristics to which users show positive attitudes, a NEG label to the characteristics to which users have negative attitudes, and a NEU label to neutral characteristics. Since product characteristics are normally composed of multiple characters, the BIO template is employed to mark the beginning and end of the word to prevent overlap, that is: the beginning of a word is marked as B, the subsequent word is marked as I, and the irrelevant word is marked as O. Combined with the above three emotional tags, 7 types of labels are obtained as listed in Table 1.

In our illustrative example, the Chinese character term "Outer（外）" in Chinese word term "Outer Appearance（外观）" should be marked with B-POS label, the Chinese character term "Appearance (观)" should be marked with I-POS label; the Chinese character term "Noise（噪）" in "Noisy sound（噪音）" should be marked with B-NEU label, and the Chinese character term "Sound（音）" should be marked with I-NEU label; the Chinese character term "Energy (能)" in "Energy consumption (能耗)" should be marked with the B-NEG label, and the Chinese character term "Consumption (耗)" should be marked with the I-NEG label. All other characters that are not related to product characteristics are marked as O.

Table 1
Summary of the Seven Types of Tags Used.

Types	Tags
Start character with positive features	B-POS
Subsequent character with positive features	I-POS
Start character with negative features	B-NEG
Subsequent character with negative features	I-NEG
Start character with neutral features	B-NEU
Subsequent character with neutral features	I-NEU
Non-feature-character	O

3.2. Model structure

The framework diagram for the model construction is given in Fig. 1. The model mainly consists of an input layer (Step 1), an encoding layer (Step 2) and a decoding layer (Step 3). The input layer is used to input the user's historical data, the encoding layer is employed to encode the user's features and learn the user representation containing abundant information, and the decoding layer is used to decode the user's demand label sequence. This model is described in detail next.

3.2.1. Input layer

Since texts in natural language apply discrete coding, it is difficult to conduct the semantic modeling. For example, the Unicode coding scheme does not take the semantic meaning of the text into account. Therefore, word embedding technology is adopted, mapping discrete words to continuous and high-dimensional real vector space. Given the input text $\boldsymbol{x} = \{w_1, w_2, \cdots, w_n\}$, the character vector of it should be reflected as:

$$h_i^{(char)} \in R^{d_w}, i = 1, 2, \cdots, n \, \mathrm{h}_\mathrm{i}^{(\mathrm{char})} \in \mathrm{R}^{\mathrm{d_w}}, \mathrm{i} = 1, 2, \cdots, \mathrm{n} \tag{1}$$

where d_w is the word vector dimension.

In addition to considering the semantic meaning of each word, the position of each character in the text also has a crucial impact on its semantics. Hence, our model also assigns a position code to each character, which shows the distance between the first word of the text and each character. Word embedding technology is adopted to map the position code to real vector:

$$h_i^{(p)} \in R^{d_p}, i = 1, 2, \cdots, n \tag{2}$$

where d_p is the dimension of the position vector.

Product features are typically expressed as nominal phrases, or phrases related to appearance, price, performance, etc. Linguistic POS knowledge of each word in the text can enhance the model's ability to recognize product features. The POS vector is therefore used in addition to the traditional character and position vector. The LTP tool is employed to conduct the POS analysis on the input text and assign a POS label to each word. There are 28 types of POS labels in LTP. Each POS label is mapped to the vector space by word embedding technology:

$$h_i^{(pos)} \in R^{d_{pos}}, i = 1, 2, \cdots, n \tag{3}$$

where d_{pos} is seen as the dimension of the POS vector.

Apart from POS information, product features are usually fixed phrases. The introduction of linguistic word segmentation knowledge can provide the model with rich boundary information of characters and reduce label errors. The LTP tool is first employed to segment the input text. After receiving the result, the BIO template is used to indicate the boundary of the phrase. If a word is the beginning of an expression, it will be marked as B, and the subsequent character will be marked as I; if it is not a phrase, it will be marked as O. After obtaining the word segmentation labels, Word embedding technology is employed to map the three kinds of word-segmentation labels to the vector space:

$$h_i^{(seg)} \in R^{d_{seg}}, i = 1, 2, \cdots, n \tag{4}$$

where d_{seg} is the dimension of word segmentation labels.

Consider the following illustrative example: "*This computer has a good performance, but the keyboard feels a little terrible*". Then, x represents the set of words in the sentence, the text feature $h_i^{(char)}$ refers to the vector representation of each word in the sentence (e.g., the vector of "this"),

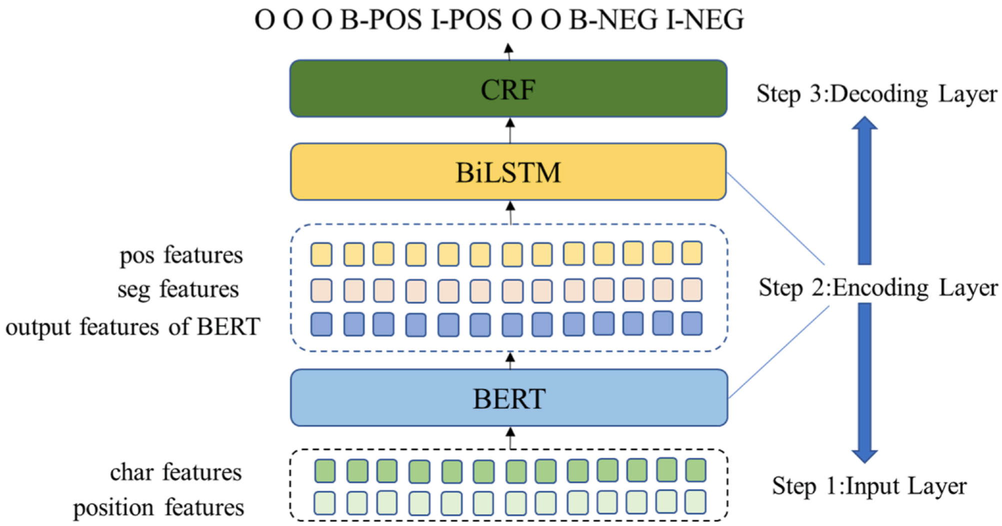

Fig. 1. Model Structure.

the position feature $h_i^{(p)}$ is a vector representation of the position corresponding to each word in the sentence (e.g., the vector corresponding to position 0 for "this"), the part-of-speech feature $h_i^{(pos)}$ is a vector representation of the part-of-speech corresponding to the word (e.g., the vector of the part-of-speech "pronoun" corresponding to "this"), and the word segmentation feature $h_i^{(seg)}$ refers to the label corresponding to the word segmentation in the sentence, e.g., the Chinese character for computer corresponds to the labels B and I. So, four different kinds of mapping are completed, from discrete text to continuous semantic vector space.

3.2.2. Coding layer

Given input text $x = \{w_1, w_2, \cdots, w_n\}$, our model first uses the word embedding technology to gain the word vector of each word $\left\{h_i^{(char)}\right\}_{i=0}^{n}$ and the position vector $\left\{h_i^{(p)}\right\}_{i=0}^{n}$, before combining them together:

$$h_i^{(input)} = \left[h_i^{(char)}; h_i^{(p)}\right], i = 1, 2, \cdots, n \tag{5}$$

$$h_i^{(input)} \in R^{d_w + d_p}$$

The pre-trained language model BERT is then employed to encode the input vector and to capture contextual semantic information of the text. In order to better encode text information and learn richer semantic representations, the pre-training language model BERT in this paper uses a multi-layer Transformer encoder, where the input of each layer is the output of the previous layer The computational process is abbreviated in this paper as follows:

$$\left\{\mathrm{h}_i^{(\mathrm{BERT})}\right\}_{i=1}^{n} = \mathrm{BERT}(\mathrm{h}_1^{(\mathrm{input})}, \mathrm{h}_2^{(\mathrm{input})}, \cdots, \mathrm{h}_n^{(\mathrm{input})}) \tag{6}$$

where $\mathrm{h}_i^{(\mathrm{BERT})} \in \mathrm{R}^{d_{\mathrm{BERT}}}$ is the input of the Transformer encoder at layer *i*.

After obtaining the uniform representation of user features $E = \left[h_1^{(input)}; h_2^{(input)}; \cdots; h_n^{(input)}\right] \in R^{(d_{char}+d_p)\times n}$, BERT encodes the information through applying the Transformer structure of self-attention systems, which can obtain the similarity of any two elements in the text sequence by matrix calculation. This creates the connection between words within the same sentence.

The first step is to calculate the user's feature representation through the self-attention mechanism H:

$$H = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_w + d_p}}\right)V \tag{7}$$

where the softmax operation normalizes the vector representation to a probability distribution, being Q, K, V the query representation (Query), key representation (Key) and value representation (Value), respectively. The above self-attention mechanism can be described as a set of query representation Q and key-value pair representation K, V mirrored to the user's representation, and the weight assigned to V is calculated by the similarity function between Q and the corresponding K, i.e., $\mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_w+d_p}}\right)$. Q, K, V are calculated as follows:

$$Q = W_Q H + b_Q \tag{8}$$

$$K = W_K H + b_K \tag{9}$$

$$V = W_V H + b_V \tag{10}$$

where $\mathrm{W_Q}$, $\mathrm{W_K}$ and $\mathrm{W_V}$ are the weight matrices and $\mathrm{b_Q}$,$\mathrm{b_K}$ and $\mathrm{b_V}$ are the biases, i.e., the query representation Q and the key-value pair representation K, V are obtained by linear projection.

Having obtained the user feature representation H, all neurons in an intermediate layer are normalized to speed up the training process of the model and to make it converge faster. The second step is consequently to calculate the layer-normalized representation, which is obtained through scaling and translation transformations:

$$H = \mathrm{LayerNorm}(H) = \alpha \odot \left(\frac{H-\mu}{\sigma}\right) + \beta \tag{11}$$

where $\odot$ is element multiplication, μ is the mean of H, σ is the variance of H, α, β are learnable parameters, and $\frac{\mathrm{H}-\mu}{\sigma}$ is the standard normalized result of the user representation H.

Finally, the third step is to calculate the linear transformation:

$$H = \mathrm{ReLU}(WH + b) \tag{12}$$

where W, b are learnable parameters, and ReLU is an Activation Function.

Although BERT can model contextual semantic information, it may encounter problems with the linguistic knowledge contained in the text. We therefore introduce POS information and word segmentation information into the output representation of BERT. After obtaining the output expression of BERT, the approach concatenates the POS vector and word segmentation vector of each word within the text:

$$h_i^{(fuse)} = \left[h_i^{(BERT)}; h_i^{(pos)}; h_i^{(seg)}\right], i = 1, 2, \cdots, n \tag{13}$$

The obtained text representation is then encoded using a bidirectional LSTM encoder to get a final representation for each word that contains both contextual semantic information, linguistic part-of-speech information, and word segmentation information:

$$h_i^{(output)} = BiLSTM(h_1^{(fuse)}, h_2^{(fuse)}, \cdots, h_n^{(fuse)}) \tag{14}$$

$$h_i^{(output)} \in R^{out}$$

3.2.3. Decoding layer

After the eigenvector of each character in the text is obtained through the coding layer, the corresponding sentimental polarity label must be decoded according to the eigenvector. Considering that the sentiment polarity corresponding to each word in the text is often related to the sentiment polarity of the context, the CRF model is used. When executing the sentiment polarity label decoding, the CRF model considers not only the probability of each character being marked with a different label, but also the dependency between labels. The probability that the text $\boldsymbol{x} = \{w_1, w_2, \cdots, w_n\}$ finally decodes its corresponding sentiment polarity label sequence $\boldsymbol{y} = \{\mathrm{y}_1, \mathrm{y}_2, \cdots, \mathrm{y}_\mathrm{n}\}$ through the input layer and the coding layer is.

$$P(\boldsymbol{y}|\boldsymbol{x}) = \frac{\exp(\sum_{i=1}^{n} W_{crf}^{y_i} h_i^{(output)} + b_{crf}^{(y_{i-1}, y_i)})}{\sum_{L^{'} \in C} \exp(\sum_{i=1}^{n} W_{crf}^{y_i^{'}} h_i^{(output)} + b_{crf}^{(y_{i-1}^{'}, y_i^{'})})} \tag{15}$$

where $h_i^{(output)}$ is the encoded user feature vector representation, C is the set of all possible tag sequences, $W_{crf}^{y_i} \in R^{out}$ is the scoring vector corresponding to the tag y_i, which is multiplied by the representation of each word. The probability of getting the next character's label y_i from the previous character's label y_{i-1} is indicated by $b_{crf}^{(y_{i-1}, y_i)}$, and $W_{crf}^{y_i}$ and $b_{crf}^{(y_{i-1}, y_i)}$ are both learnable parameters.

The Viterbi algorithm is adopted to calculate the optimal sequence in the observed label sequence and reduce the complexity of the decoding process. The Viterbi algorithm is a dynamic planning algorithm, which is used to find the Viterbi path that is most likely to produce the sequence of observed events, i.e., to find the optimal sequence among all observed sequences. The Viterbi algorithm records the optimal sequence of each observation label at each moment, assuming that the optimal path from moment 0 to moment t has been saved at moment t, so that the optimal sequence from t to t + 1 only needs to be computed at moment t + 1. The optimal value from the last moment is backtracked to the start position, and after the backtracking is completed, this path from the start to the end is optimal.

3.2.4. Optimization objective

The model is trained by minimizing a loss function, which is defined as.

$$Loss = -\mathrm{E}[\log P(\boldsymbol{y}|\boldsymbol{x})] \tag{16}$$

where × and y represent the positive example of the model, while $P(\boldsymbol{y}|\boldsymbol{x})$ represents the decoding probability of positive examples. The loss function is a negative log-likelihood function, which is commonly applied for maximum likelihood estimation and related fields.

4. Comparison and ablation experiment

To verify the effectiveness and accuracy of our new fine-grained sentiment analysis approach, the paper uses the "integrated stove" as an example product. A host of consumer reviews on integrated stoves were crawled from Taobao, Tmall.com and JD.com, major Chinese e-commerce websites. Data cleaning technology was then used to filter out texts that were too short or too long, whilst only texts with actual evaluations of the product were retained. Afterwards, the filtered comments were manually labeled, and each comment where product features were mentioned together with its corresponding sentiment polarity marked. This provides a dataset that can be employed for model training and evaluation. Basic statistics of the dataset are summarized in Table 2. The data set is divided into a training set, a development set, and a test set at a ratio of 8:1:1.

4.1. Comparison experiment

4.1.1. Experimental design

To verify the effectiveness of our new approach its performance is first compared against six alternative approaches, which are as follows.

- **CNN:** The CNN (Convolutional Neural Network) is used as the coding layer to model semantics of the text. Multiple parallel convolution kernels of different scales are applied, and then the results spliced to gain the final character representation. After obtaining the vector representation of each character, the softmax layer is used as the decoding layer to decode the predicted sentiment label sequence. Specifically, each vector passes through a linear layer to obtain the score of each character for different labels. Then the softmax function is applied to normalize and gain the probability values of different labels.
- **CNN + CRF:** The coding layer of CNN + CRF (Convolutional Neural Network + Conditional Random Field) adopts the same convolution structure as CNN to calculate the vector representation of words, but the decoding layer adopts a different decoding algorithm. The softmax decoding layer is replaced with the CRF decoding layer. This means it takes the dependency between labels into account during the decoding process.
- **BiLSTM:** The BiLSTM (Bidirectional Long- Short-Term Memory Recurrent Neural Network) first forwards the text to an LSTM network. Then, the character order of the text is reversed and sent to another LSTM network. Finally, the output of the two LSTM networks is spliced as the vector representation of the character. The softmax layer is applied for decoding.
- **BiLSTM + CRF:** The BilSTM + CRF draws on the same LSTM structure as BiLSTM to encode text semantics to obtain a vector representation of the words, but the decoding process occurs in the CRF layer.
- **BERT:** This model only makes use of the pre-trained language model BERT (Bidirectional Encoder Representations from Transformers) to encode semantics. However, unlike the model proposed in this article, BERT only takes advantage of word vectors and position vectors without incorporating linguistic knowledge (POS information and word segmentation information). Further, the final output does not additionally pass a BiLSTM layer to perform information fusion, but rather only draws on the softmax layer.
- **BERT + CRF:** BERT + CRF employs the same pre-trained language model BERT to encode semantics, but the decoding layer uses CRF instead of the softmax layer to decode the output.

All models are trained on the constructed training set, parameters adjusted on the development set, and the sentiment analysis results obtained for the test set. All the models use the same hyperparameters to control the experimental error and mitigate the influence of random

Table 2
Statistics of the Manually Labeled Data Set.

	Value
Number of comments	2113
Average length of comments	23
The total number of positive product features	3307
The total number of negative product features	4812
The total number of neutral product features	2003

factors. The optimization algorithm uniformly adopts the stochastic gradient descent (SGD) method, and the learning rate is set to 2e-5. The model parameters are initialized by the same initialization seed, and Dropout technology is used to prevent overfitting. The specific settings of the model are summarized in Table 3.

4.1.2. Experimental process

When training the model, its parameters are continuously updated by SGD based on the calculated loss function. Meanwhile, early stop technology is employed to avoid over-fitting. This means, every time when parameters are updated, the value of the loss function of the model is recorded on the training set and the development set. When the loss value on the development set no longer drops, or even starts to rise, it indicates that the model begins to overfit on the training set. The current model is saved, and parameters are considered as the final training result. The loss function value during training is visualized in Fig. 2.

As can be seen from Fig. 2, the loss function value of the model on the training set decreases with the increase of the number of training rounds, eventually approaching a value of 0. However, the loss function value of the model on the development set first continues to decrease, reaching the lowest point in the 12th round, and then begins to gradually increase, demonstrating that the model has overfitted the training set at this time and under these circumstances.

To further verify the rationality and effectiveness of the early stopping technology, the F1 values on the training set and development set are recorded after each update iteration of the model during training. Results are given in Fig. 3.

Fig. 3 shows that the F1 value on the training set continuously increases with the number of iterations, approaching 100% accuracy. Nevertheless, the F1 value on the development set gradually increases at first until it reaches the highest point of 92.71 in the 12th round, and then it declines. The 12th round model is therefore selected as the model to test the final performance on the test set.

4.1.3. Comparison of experimental results

After training the approaches, all seven were applied to predict the fine-grained sentiment polarity of the test set. Results where compared with the manually obtained labels. Three evaluation indicators, the precision rate P, the recall rate R and the F1 score, are adopted as follows:

$$P = \frac{TP}{TP + FP} \tag{17}$$

$$R = \frac{TP}{TP + FN} \tag{18}$$

$$F1 = \frac{2^{*}P^{*}R}{P + R} \tag{19}$$

where TP refers to the number of true positive cases, TF to the number of true negative cases, FP to the number of false positive cases, and FN to the number of false negative cases. Since the fine-grained sentiment analysis involves multiple tags, macro-averaging is employed to synthesize various precision rates, recall rates and F1 scores

Table 3
Settings of Model Hyperparameters.

Hyperparameters	Value
Dimensions of character vector	512
Dimensions of position vector	5
Dimensions of POS vector	100
Dimensions of word segmentation vector	100
Dimensions of hidden(implicit) vector	768
Optimization algorithm	SGD
Learning rate	2e-5
Number of training rounds	200
Dropout rate	0.2

into the final performance indicators. Results are given in Table 4.

Based on the results in Table 4 the following conclusions can be obtained:

- Our new approach outperforms existing approaches. More specifically, and compared to the best alternative approach, its accuracy is 1.57% higher, the recall rate is 3.40% higher, and the F1 value is 2.47% higher.
- The application of a pre-trained language model BERT as the coding layer can significantly enhance performance.
- The use of a CRF model in the decoding layer can effectively improve the decoding accuracy.

To further provide a fine-grained evaluation of the three kinds of sentiment labels encompassing positive, negative, and neutral attitudes, the F1 values are calculated for the three sentimental polarities. The specific comparison results for our seven approaches are given in Fig. 4.

Fig. 4 shows that the F1 value of our approach is higher than that of the alternatives considering all three kinds of sentiment polarity. Meanwhile, among the three kinds of sentiment polarity, all seven approaches have the highest recognition rate for negative emotions, followed by positive emotions. This phenomenon can be explained by users being more inclined to comment on product features that they like or dislike. Neutral features are seldom mentioned, and neutral sentiment labels are the fewest in the training data. This in turn affects the model's recognition capability of neutral labels.

4.2. Ablation experiment

To verify the role of each constituting component of our new approach, an ablation experiment is designed. This experiment removes the different components of the approach, and then conducts performance evaluation on the development set. Results are summarized in Table 5.

The following can be observed from Table 5:

- The precision rate, the recall rate and the F1 score decline by 2.05%, 1.69% and 1.87% respectively, if POS knowledge is not used and POS tags are absent in the coding layer. This further demonstrates that the inclusion of POS knowledge strengthens the accuracy of the recognition of text sentiment.
- The accuracy rate, the recall rate and the F1 score are reduced by 1.36%, 1.16% and 1.26%, respectively, when suppressing word segmentation and not splicing the word segmentation vector into the model input.
- Suppressing the POS information, word segmentation information and the BILSTM layer, the reduction of the accuracy rate, the recall rate and the F1 value of the model is 3.16%, 2.87% and 3.02%, respectively. This highlights the significant role of linguistic knowledge in fine-grained sentiment analysis.
- The precision rate, the recall rate and the F1 value decline by 3.61%, 3.72% and 3.66%, respectively, by replacing the CRF layer with the softmax function to decode. This supports the argument that the use of a probabilistic graphical model to describe the labels' relationship can improve performance.

5. Use case

Our new approach was applied to analyze customer data for a large household appliance manufacturer in China. We next provide an example on how it can be applied to obtain the frequency of positive, negative, and neutral evaluations of integrated stoves and their product features. Results that can be used to benchmark products, and to extract what the customer values for future product development and process improvement. Different brands of integrated stoves, namely Midea, Changhong, SUPOR, FOTILE and ROBAM, are analyzed. These brands

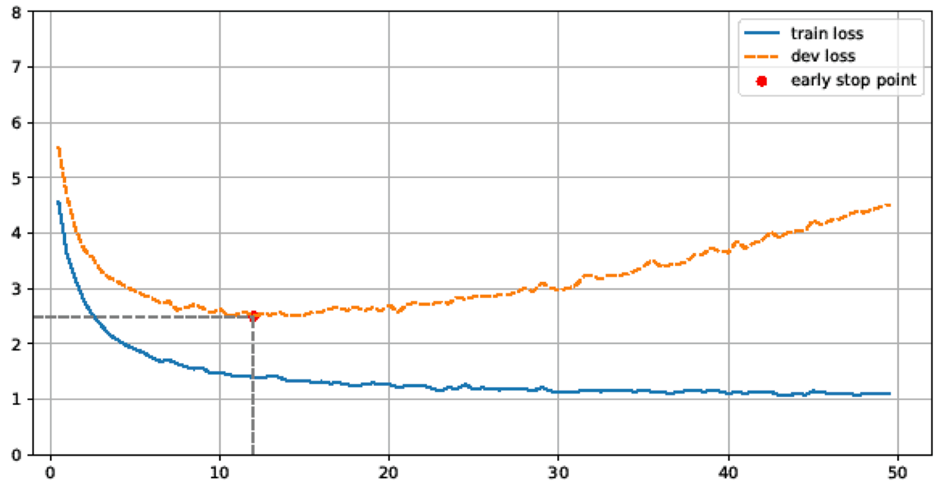

Fig. 2. Loss Function Value during Training.

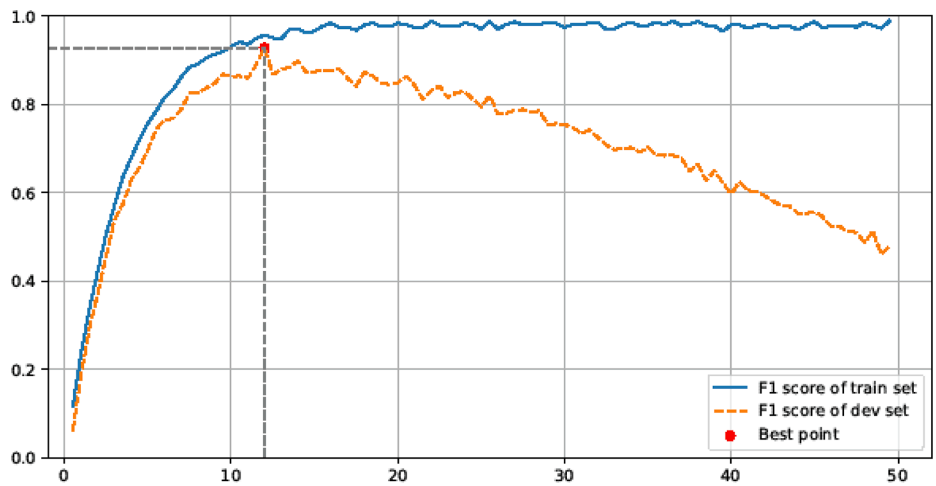

Fig. 3. F1 Value of Training and Development Set.

Table 4
Summary of Results.

Models	Accuracy/ P	Recall rate /R	F1 score
CNN	69.98	70.44	70.21
CNN + CRF	72.48	74.21	73.33
BiLSTM	82.77	83.76	83.26
BiLSMT + CRF	85.86	86.55	86.20
BERT	89.12	89.79	89.45
BERT + CRF	90.88	91.91	91.39
Our model	92.45	95.31	93.86

are typical representatives of the industry in China. After training the model, web crawler technology is employed to crawl large-scale unlabeled data of customer reviews on integrated stoves from Taobao, Tmall.com, and JD.com. The statistic of this date is shown in Table 6.

Our approach is first used to extract different product features. Fig. 5 visualize the frequency of product features referenced by users of integrated stoves. Note that results are in Chinese given our approach was developed for a Chinese context. The four product features that users care most about are appearance, installation service, absorptivity of smoke, and firepower.

Using these four product characteristics, the second step obtains the emotional polarity. This provides statistics on positive, negative, and neutral evaluations for five types of integrated stoves and four types of product characteristics, as shown in Table 7. Table 7 clearly identifies potential benchmarks (e.g. FOTILEX and ROBAM), and a ranking of features that need improvement, e.g. improving the capability to absorb smoke should be a higher priority for Midea than improving the appearance.

To further guide improvement efforts, the characteristics with the highest frequency of receiving negative reviews can be listed. An example is given in Table 8, which provides the three most mentioned negative characteristics of the five different products.

The characteristics listed in Table 8 provide very detailed input for product and process improvement. For example, the loud noise, about which more than 40% of the customers complain for the Midea, can be due to poor design or poor production processes. The data obtained through our approach can be compared to data obtained from process analysis. For example, Fig. 6 gives an analysis of the quality defects in the manufacturing process producing the Midea. As can be seen, the noise is a main issue occurring during the production process. It should be the highest priority when deciding on improvement objectives in the context of Lean production.

6. Conclusions

This paper presented a new approach to extract what the customer wants from large amounts of Chinese consumer data. This provides a faster and cheaper alternative to large-scale surveys typically applied in

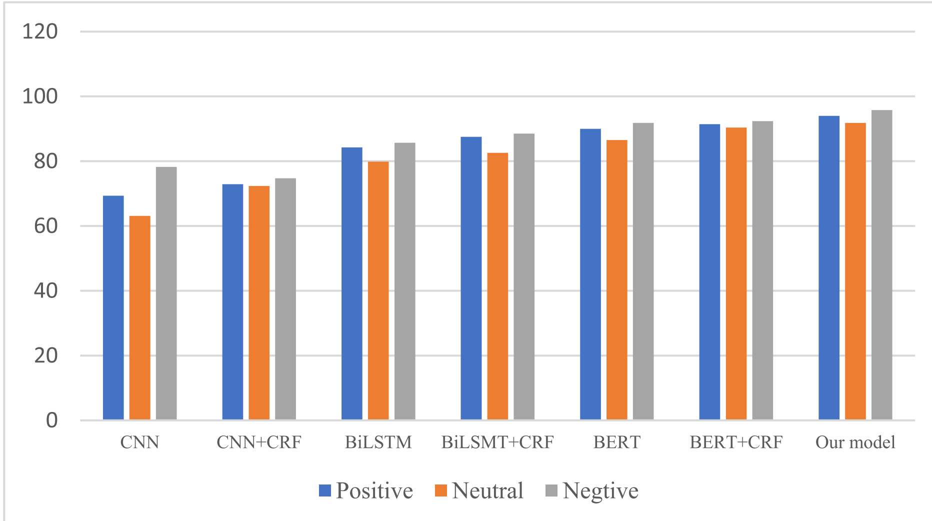

Fig. 4. Comparative Result of the F1 Score for Fine-grained Emotion Polarity.

Table 5
Result of the Ablation Experiment.

Models	Accuracy	Recall Rate	F1 Score
Model proposed	92.38	93.04	92.71
POS knowledge suppressed	90.33	91.25	90.84
Word segmentation knowledge suppressed	91.02	91.88	91.44
BiLSTM suppressed	89.22	90.17	89.69
CRF suppressed	88.77	89.32	89.04

Table 6
Statistic of the Unlabeled Data Set.

E-Commerce Platform	Number of data
Taobao	43,551
Tmall.com	36,221
JD	51,378

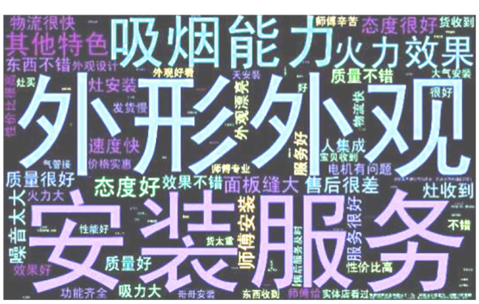

Fig. 5. Word Cloud Image obtained from the Data Set.

Lean to identify customer value. Sentiment analysis is one of the most ubiquitous research fields in natural language processing. However, classical fine-grained sentiment analysis methods typically use statistical machine learning methods that rely on manually constructed characteristic. The representation and learning capabilities of deep neural networks provided major improvements to fine-grained sentiment analysis in recent years. But these neural network models typically employ the coding abilities of network structures itself to extract the characteristics of the text. They neglect the linguistic knowledge contained in the language, for instance, POS knowledge and word segmentation knowledge, which may result in incomplete feature extraction. Furthermore, the existing neural network models often use

Table 7
Summary of Four Main Characteristics of Five Brands of Integrated Stoves.

Model	Feature	Positive	Neutral	Negative
Midea WD26	appearance	78.33	0	21.67
	installation service	97.81	1.03	1.16
	smoke absorbance	62.15	6.34	31.51
	firepower	92.07	0	7.93
Changhong JJZT-H3K	appearance	88.65	8.23	3.12
	installation service	90.69	0	9.31
	smoke absorbance	84.24	0	15.76
	firepower	79.28	4.81	15.91
SUPOR UX22	appearance	85.92	1.97	12.11
	installation service	87.66	0	12.34
	smoke absorbance	91.06	0	8.94
	firepower	79.98	0	20.02
FOTILEX2A + X2.iA	appearance	98.23	1.17	0.6
	installation service	97.10	2.79	0.11
	smoke absorbance	99.21	0.64	0.15
	firepower	98.39	1.61	0
ROBAM 901FZZ	appearance	96.49	1.61	1.90
	installation service	97.48	1.41	0.74
	smoke absorbance	97.95	0.20	1.85
	firepower	97.87	1.84	0.29

Table 8
Summary of the 3 Characteristics with Highest Frequency of Negative Reviews.

Model	Quality index	Feature	Negative evaluation rate
Midea WD26	1	loud noise	42.57
	2	too heavy	38.32
	3	poor absorptivity of smoke	31.51
Changhong JJZT-H3K	1	large panel gap	41.93
	2	bad after-sales service	33.82
	3	loud noise	29.52
SUPOR UX22	1	bad after-sales service	37.69
	2	something wrong with the motor	21.14
	3	poor firepower	20.02
FOTILEX2A + X2. iA	1	too expensive	67.44
	2	high energy consumption	49.94
	3	bad after-sales service	35.71
ROBAM 901FZZ	1	too expensive	47.83
	2	take up much space	30.79
	3	the edge cut the hand	21.55

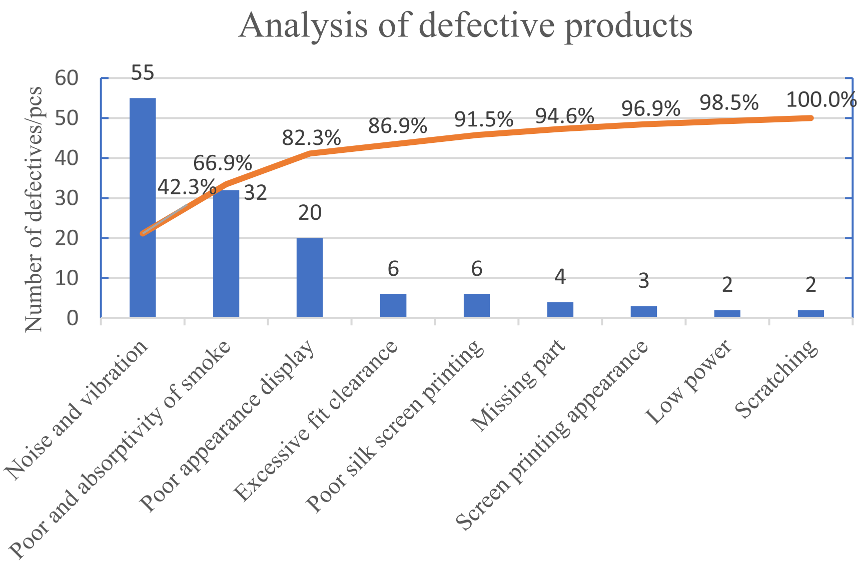

Fig. 6. Analysis of Quality Defects in the Manufacturing Process.

conventional sequence models for encoding, such as RNN, or LSTM. This study started with the proposition to design a new mechanism for fine grained sentiment analysis of Chinese consumer data that by focusing on product features provides information on value (positive) and process improvement possibilities (negative). The new model significantly extends existing literature. It uses a pre-trained language model BERT to encipher the text, and CRF to decipher the predicted label sequence. Linguistic knowledge in the form of POS knowledge and word segmentation knowledge is embedded into the textual encoding. A comparison experiments showed that our approach has the potential to outperform alternative approaches. Meanwhile, that all components integrated into our approach are necessary was shown via an ablation experiment. This has important implication for future design of fine-grained sentiment analysis. Meanwhile, the practical usefulness of our approach was proven in a use case.

A main limitation of our study is its focus on Chinese, Chinese is characterized by each word having its own semantic information, which enhances the difficulty of sentiment analysis to a certain extent. The solution of this paper is to label individual words as a unit. Since product features usually consist of multiple Chinese characters, this paper employs a BIO template to mark the beginning and end of words in Chinese, and combines them with three sentiment labels (i.e., positive, neutral and negative). We consider that our focus on Chinese is justified by the lack of research on fine-grained sentiment analysis of Chinese texts, and the increasing importance of the Chinese market. Future research could however explore how our approach can be adapted to other languages of the CJK (Chinese, Japanese, Korean) group. While all three belong to different language groups they share to a certain extend a common writing system. Future research should thereby not forget that the use of a specific writing system itself has semantic meaning. Meanwhile, our use case demonstrated how our approach can identify specific problems related to specific product characteristics. Comparing this data with data on defects obtained from an analysis of the production process, a clear link could be established. This allows to link customer value to process improvement. Future research could explore how additional layers could be integrated into our approach to automatically transform customer value, obtained through our current approach, in input for process improvement.

Uncited references

CRediT authorship contribution statement

Yan Xiao: Conceptualization, Writing – original draft. **Congdong Li:** Conceptualization, Methodology, Supervision, Funding acquisition. **Matthias Thürer:** Writing – review & editing. **Yide Liu:** Methodology, Software. **Ting Qu:** Validation.

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

1. National Natural Science Foundation of China, No. 71672074.
2. National Natural Science Foundation of China, No. 72072072.
3. Natural Science Foundation of Guangdong Province of China, No. 2019A1515010045.
4. 2018 Guangzhou Leading Innovation Team Program (China), No. 201909010006

References

Abirami, A. M., & Askarunisa, A. (2016). Feature Based Sentiment Analysis for Service Reviews. *Journal of Universal Computer Science, 22*(5), 650–670.
Agarwal, N., Sikka, G., & Awasthi, L. K. (2020). Enhancing web service clustering using Length Feature Weight Method for service description document vector space representation. *Expert Systems with Applications, 161*.
Bamford, D., Forrester, P., Dehe, B., & Leese, R. G. (2015). Partial and iterative Lean implementation: Two case studies. *International Journal of Operations & Production Management, 35*(5), 702–727.
Bicen, P., & Johnson, W. H. A. (2015). Radical Innovation with Limited Resources in High-Turbulent Markets: The Role of Lean Innovation Capability. *Creativity and Innovation Management, 24*(2), 278–299.
Bittencourt, V. L., Alves, A. C., & Leão, C. P. (2020). Industry 4.0 triggered by Lean Thinking: Insights from a systematic literature review. *International Journal of Production Research*, 1–15.
Bortolini, R.F., Nogueira Cortimiglia, M., Danilevicz, A.d.M.F. & Ghezzi, A. (2018). Lean Startup: a comprehensive historical review. Management Decision, ahead-of-print (ahead-of-print).
Cambria, E. (2016). Affective Computing and Sentiment Analysis. *Ieee Intelligent Systems, 31*(2), 102–107.

Chae, B. (2015). Insights from hashtag #supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics, 165*, 247–259.
Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K. & Kuksa, P.J.J.o.M.L.R. (2011). Natural Language Processing (almost) from Scratch. 12(1), 2493-2537.
Denyer, D., Tranfield, D., & van Aken, J. E. (2008). Developing design propositions through research synthesis. *Organization Studies, 29*(3), 393–413.
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.
Gautam, N., & Singh, N. (2008). Lean product development: Maximizing the customer perceived value through design change (redesign). *International Journal of Production Economics, 114*(1), 313–332.
Ghezzi, A. (2019). Digital startups and the adoption and implementation of Lean Startup Approaches: Effectuation, Bricolage and Opportunity Creation in practice. *Technological Forecasting and Social Change, 146*, 945–960.
Goodfellow, I., Bengio, Y., Courville, A., Goodfellow, I., Bengio, Y. & Courville, A. (2016). Deep Learning Introduction.
Hino, S. (2005). *Inside the Mind of Toyota: Management Principles for Enduring Growth*. New York: NY: Productivity Press.
Jasti, N. V. K., & Kodali, R. (2014). A literature review of empirical research methodology in lean manufacturing. *International Journal of Operations & Production Management, 34*(8), 1080–1122.
Jasti, N. V. K., & Kodali, R. (2015). Lean production: Literature review and trends. *International Journal of Production Research, 53*(3), 867–885.
Jiang, L., Yu, M., Zhou, M., Liu, X., & Zhao, T. (2011). In *Target-dependent Twitter sentiment classification* (pp. 151–160). Portland, Oregon: Association for Computational Linguistics.
Jin, J., Liu, Y., Ji, P., & Liu, H. G. (2016). Understanding big consumer opinion data for market-driven product design. *International Journal of Production Research, 54*(10), 3019–3041.
Jin, W., & Ho, H. H. (2009). In *A novel lexicalized HMM-based learning framework for web opinion mining* (pp. 465–472). Montreal, Quebec, Canada: Association for Computing Machinery.
Khattak, A., Paracha, W. T., Asghar, M. Z., Jillani, N., Younis, U., Saddozai, F. K., & Hameed, I. A. (2020). Fine-Grained Sentiment Analysis for Measuring Customer Satisfaction Using an Extended Set of Fuzzy Linguistic Hedges. *International Journal of Computational Intelligence Systems, 13*(1), 744–756.
Khurum, M., Petersen, K., & Gorschek, T. (2014). Extending value stream mapping through waste definition beyond customer perspective. *Journal of Software-Evolution and Process, 26*(12), 1074–1105.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.
Li, F., Han, C., Huang, M., Zhu, X., & Yu, H. (2010). Structure-Aware Review Mining and Summarization. *International Conference on Computational Linguistics*.
Li, X., Bing, L., Zhang, W. & Lam, W. (2019). Exploiting BERT for End-to-End Aspect-based Sentiment Analysis. In: W-NUT@EMNLP.
Li, Y., Jin, Q., Zuo, M., Li, H., & Liu, X. (2020). Multi-neural network-based sentiment analysis of food reviews based on character and word embeddings. *J International Journal of Electrical Engineering Education.*
Liu, S. Y., Xiao, J., & Xu, X. K. (2020). Sign prediction by motif naive Bayes model in social networks. *Information Sciences, 541*, 316–331.
Long, D., Cohn, T., Bird, S., & Cook, P. (2015). Low Resource Dependency Parsing: Cross-lingual Parameter Sharing in a Neural Network Parser. *In: Meeting of the Association for Computational Linguistics & International Joint Conference on Natural Language Processing*.
Ma, X.Z. & Hovy, E. (2016). End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF.
Maheshwari, S., Gautam, P., & Jaggi, C. K. (2020). Role of Big Data Analytics in supply chain management: Current trends and future perspectives. *International Journal of Production Research*.
Mazur, J., Drabek, R., Goldman, A.J.F.Q. & Preference. (2017). Hedonic contrast effects in multi-product food evaluations differing in complexity. S0950329317301532.
Mcdonald, R. T., Hannan, K., Neylon, T., Wells, M., & Reynar, J. C. (2007). Structured Models for Fine-to-Coarse Sentiment Analysis. *In: Meeting of the Association for Computational Linguistics*.
Miao, Q. L., Li, Q. D., & Zeng, D. (2010). Fine-Grained Opinion Mining by Integrating Multiple Review Sources. *Journal of the American Society for Information Science and Technology, 61*(11), 2288–2299.
Mishra, R. K., Urolagin, S., & Angel, A. J. J. (2019). A Sentiment analysis-based hotel recommendation using TF-IDF Approach. In *In: International Conference on Computational Intelligence and Knowledge Economy* (pp. 811–815).
Ohno, T. (1988). *Toyota Production System: Beyond Large-Scale Production*. New York: Productivity Press.
Pang, B., Lee, L. & Vaithyanathan, S. (2002). Thumbs up? sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10 (pp. 79–86): Association for Computational Linguistics.
Peng, H. Y., Cambria, E., & Hussain, A. (2017). A Review of Sentiment Analysis Research in Chinese Language. *Cognitive Computation, 9*(4), 423–435.
Peralta, C. B. D., Echeveste, M. E., Lermen, F. H., Marcon, A., & Tortorella, G. (2020). A framework proposition to identify customer value through lean practices. *Journal of Manufacturing Technology Management, 31*(4), 725–747.
Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., & Eryiğit, G. (2018). SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In *In: International Workshop on Semantic Evaluation* (pp. 19–30).
Quan, C. Q., & Ren, F. J. (2014). Unsupervised product feature extraction for feature-oriented opinion determination. *Information Sciences, 272*, 16–28.
Rezaeian, A., Rezaeian, M., Khatami, S. F., Khorashadizadeh, F., & Moghaddam, F. P. (2020). Prediction of mortality of premature neonates using neural network and logistic regression. Journal of Ambient Intelligence and Humanized. *Computing.*
Schouten, K., & Frasincar, F. (2016). Survey on Aspect-Level Sentiment Analysis. *Ieee Transactions on Knowledge and Data Engineering, 28*(3), 813–830.
Sheng, J., Amankwah-Amoah, J., & Wang, X. J. (2017). A multidisciplinary perspective of big data in management research. *International Journal of Production Economics, 191*, 97–112.
Song, Y.W., Wang, J.H., Jiang, T., Liu, Z.Y. & Rao, Y.H. (2019). Targeted Sentiment Classification with Attentional Encoder Network. In: I.V. Tetko, V. Kurkova, P. Karpov & F. Theis, Artificial Neural Networks and Machine Learning - Icann 2019: Text and Time Series, Pt Iv (Vol. 11730, pp. 93-103).
Tang, D., Qin, B., Feng, X. & Liu, T.J.C.e. (2015). Effective LSTMs for Target-Dependent Sentiment Classification.
Tang, D., Qin, B. & Liu, T. (2016). Aspect Level Sentiment Classification with Deep Memory Network.
Thomas, P. S. (1989). Toyota production system - beyond large-scale production - OHNO. *T. Interfaces, 19*(5), 83–84.
Tontini, G., Gomes, G., Daros, S., & Feldmann, M. P. (2012). Scientific production based on the Kano model of attractive and must-be quality: A research in the databases of emerald, gale, science direct and wiley. *Revista Gestao Organizacional, 5*(2), 180–191.
Tortorella, G. L., Narayanamurthy, G., & Thurer, M. (2020). Identifying pathways to a high-performing Lean Automation implementation: An empirical study in the manufacturing industry. *International Journal of Production Economics*.
Wagner, J., Arora, P., Cortes, S., Barman, U., & Tounsi, L. (2014). DCU: Aspect-based polarity classification for SemEval task 4. *In: International Workshop on Semantic Evaluation*.
Womack, J. P., Jones, D. T., & Roos, D. T. (1990). *The Machine That Changed the World*. USA: Simon & Schuster Inc.
Xiao, Y., Li, C. D., Song, L. J., Yang, J., & Su, J. F. (2021). A Multidimensional Information Fusion-Based Matching Decision Method for Manufacturing Service Resource. *Ieee Access, 9*, 39839–39851.
Xu, D., Tian, Z., Lai, R., Kong, X., & Shi, W. J. I. F. (2020). *Deep Learning Based Emotion Analysis of Microblog Texts., 64*.
Yan, Z. J., Xing, M. M., Zhang, D. S., & Ma, B. Z. (2015). EXPRS: An extended pagerank method for product feature extraction from online consumer reviews. *Information & Management, 52*(7), 850–858.
Yazdani, B. (1994). Toyota production system: an integrated approach to Just-In-Time (2nd ed) J. Computer Integrated Manufacturing Systems, 423.
Yu, H., & Hatzivassiloglou, V. (2003). Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences. *Conference on Empirical Methods in Natural Language Processing*.
Zhang, C. L., Zeng, D., Li, J. X., Wang, F. Y., & Zuo, W. L. (2009). Sentiment Analysis of Chinese Documents: From Sentence to Document Level. *Journal of the American Society for Information Science and Technology, 60*(12), 2474–2487.
Zhang, W. H., Xu, H., & Wan, W. (2012). Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis. *Expert Systems with Applications, 39*(11), 10283–10291.
Zhang, Y. B., Zhang, Z. F., Miao, D. Q., & Wang, J. Q. (2019). Three-way enhanced convolutional neural networks for sentence-level sentiment classification. *Information Sciences, 477*, 55–64.