Neural Processing Letters (2019) 50:149–164
https://doi.org/10.1007/s11063-018-9964-8

# Extraction of Product Evaluation Factors with a Convolutional Neural Network and Transfer Learning

**Younghoon Lee**[1,2] · **Minki Chung**[1] · **Sungzoon Cho**[1] · **Jinhae Choi**[2]

Published online: 2 January 2019


**Abstract**

Earlier studies have indicated that decision-making by a project development team can be improved throughout the design and development process by understanding the key factors that affect customers evaluations of a new product. Aspect extraction could thus be a useful tool for identifying important attributes when evaluating products or services. Aspect extraction based on deep convolutional neural networks has recently been suggested, demonstrating state-of-the-art performance when applied to a customer review of electronic devices. However, this approach is unsuited to the rapidly evolving smartphone industry, which involves a wide range of product lines. Whereas the previous approach required significant amounts of data labeling for each product, we propose a variant of that approach that includes transfer learning. We also propose a novel approach for transferring the architecture sequentially within the product group. The results indicate that the principal key feature of each product is extracted effectively by the proposed method without having to re-train the entire convolutional neural network model. Furthermore, the proposed method performs better than the previous method for each product line.

**Keywords** Product evaluation factor · Aspect extraction · Convolutional neural network · Transfer learning · Off-the-shelf features · Domain adaptation

✉ Sungzoon Cho
zoon@snu.ac.kr

Younghoon Lee
yhoon.lee@dm.snu.ac.kr

Minki Chung
minki.chung@dm.snu.ac.kr

Jinhae Choi
jin.choi@lge.com

1 Department of Industrial Engineering, Seoul National University, Seoul, Republic of Korea

2 Data Driven User Experience team, Mobile Communication Lab, LG Electronics, Seoul, Republic of Korea

## 1 Introduction

Consumers arguably make purchase decisions based on their evaluation of, and knowledge about, a product and its attributes [16,38]. Customer feedback can therefore play an important role in product development. By understanding the key factors that affect customers evaluation of a new product, a project team can improve decision making throughout the design and development process [46]. Thus, product developers or marketers are often interested in identifying the product attributes that are considered by consumers to be most important when evaluating a product [7].

For this purpose, aspect extraction (a subtopic of natural language processing or opinion mining) can be a useful means to determine important attributes for evaluating products or services. Aspect extraction aims to extract opinion targets from opinion texts [23]. For example, a sentence such as "I love the touchscreen on my phone but the battery life is too short" contains two "aspects or opinion targets, namely the *touchscreen* and the *battery life* [34]. Extracting these aspects from customer reviews or feedback can provide insight into the key factors affecting customers evaluation of a new product. The present study therefore focuses on proposing a method for extracting product evaluation factors from customer reviews, to determine the important attributes affecting purchase decisions by aspect extraction.

Aspect extraction involves two main approaches: *supervised* and *unsupervised*. The supervised approach [5,13,28] is mainly based on the conditional random field (CRF) and hidden Markov model (HMM) techniques, while the unsupervised approach [3,29,30] is mainly based on topic modeling. Topic modeling employed in the unsupervised approach often yields coarse topics rather than precise aspects (a "topic is not necessarily an "aspect). [23].

Recently, a supervised approach for aspect extraction based on a deep convolutional neural network was suggested, demonstrating state-of-art performance when applied to a customer review of electronic devices [32]. That approach utilized Google and Amazon embeddings to derive the feature vector of each word from sentences in the review and used a seven-layer-deep convolutional neural network to tag each word in opinion-laden sentences as either an "aspect or a" non-aspect word.

However, this approach required labeled data for each product, which makes it unsuitable for industries that involve a wide range of product lines or short production cycles. For example, the smartphone industry considered in this study is one of the most rapidly changing electronic-device industries. The factors used in product evaluation that affect users purchase decisions also evolve rapidly with the advancement of the products. It is therefore difficult to apply the earlier approach to the smartphone industry in its current form.

This study therefore proposes a variant of the previous convolutional neural network approach to extract product evaluation factors through transfer learning. Transfer learning is used when data becomes readily outdated. In the smartphone industry, labeled data obtained over one specific time period soon become outdated and may not follow the same distribution at a later time. This reality results from the very nature of this industry and makes transfer learning a suitable candidate for addressing these limitations.

The previous approach required that all the parameters of a convolutional neural network model be learnt for each product. In contrast, we herein preserve the global feature parameter of the convolutional neural network model and only re-learn the parameter used for the classification part of the model. This reduces the time needed for learning and the amount of labeled data required for each product. Moreover, the product evaluation factor is extracted better by the proposed approach than by the previous approach when there is only a limited amount of labeled data available. Additionally, we propose a novel idea, applying the concept of transfer learning while sequentially transferring the parameter within the product group.

The rest of this paper is structured as follows. Section 2 reviews the various studies done on aspect extraction and the previous approach based on convolutional neural networks. Section 3 describes the proposed method utilizing the convolutional neural network and the concept of transfer learning. Section 4 describes the data and experimental results to assess the performance of the proposed method, both qualitatively and quantitatively. Section 5 concludes with a discussion and suggestions for future work.

## 2 Literature Review

The following sections discuss related work regarding (1) aspect extraction as an approach for extracting evaluation factors for a given product, including the previous approach based on convolutional neural networks; and (2) transfer learning.

### 2.1 Aspect Extraction

The supervised and unsupervised approaches to aspect extraction are considered separately.

#### 2.1.1 Unsupervised Method

Unsupervised approaches are mostly based on topic modeling, especially LDA. Specific variants include the DF-LDA model [1]; sentic LDA [35], which improves LDA with semantic similarity; MaxEnt-LDA [55], which integrates a maximum-entropy approach into LDA; ILDA [29], which extends LDA by considering the interdependency between aspects and ratings; the two-step approach for detecting aspect-specific opinion words [3]; the aspect and sentiment unification model [18,19,22,47]; and domain knowledge incorporated into topic modeling [4,24,30].

In addition to the approach based on topic modeling, many rule-based approaches have utilized linguistic patterns [6,33], the syntactic-dependency method [50,54], the occurrence method [10,57], and the propagation method [37,56]. However, the topic modeling employed in the unsupervised approach yields course-grained topics rather than precise aspects. Indeed, a topical term does not necessarily constitute an aspect. Moreover, rule-based approaches require a knowledge of domain-based linguistics or custom-defined rules and features. The present study therefore focuses on supervised approaches.

#### 2.1.2 Supervised Method

Supervised learning methods have been widely used for aspect extraction. They are generally based on standard sequence-labeling approaches such as CRF [5,13,15,21,44,51] and HMM [17]. Further, only a few methods that require specific domain knowledge [14,49] have been studied within the context of supervised approaches.

These approaches usually consider datasets consisting of reviews on electronic devices in which aspect words are annotated. Using this information, a sequence-labeling algorithm is trained to identify aspect expressions.

Approaches based on convolutional neural networks [32] have recently been suggested, comparing state-of-the-art performance with earlier studies. In that study, the authors utilized

Amazon embeddings for word representation and constructed a seven-layer CNN architecture. They also used additional features, e.g., part-of-speech tags and additional linguistic rules. The present study utilizes variants of these CNN structures to address the limitations of the previous work within the particular context of the rapidly evolving smartphone industry.

### 2.2 Transfer Learning

Consider the situation where classification is required in one domain of interest, but the available training data was derived from another domain, lies in a different feature space, or follows a different data distribution. Then, a successful transfer learning would significantly improve the learning performance by avoiding expensive data-labeling efforts [31].

We present two approaches to transfer learning, which are utilized in the present study: using off-the-shelf features or domain adaptation. The former approach uses the outputs of one or more layers of a network trained on a different task to form a generic feature detector; a new shallow model is then trained using these features as target data [40,45]. Domain adaptation is based on the assumption that a model trained using data from some domain A also performs well on data from another domain B with a low level of tuning [15]. Although the domain-adaption approach has been utilized for various tasks [8], we focus here on the "off-the-shelf feature approach since it is better suited to our studies and target product.

Transfer learning is a useful concept in the field of natural language processing, particularly in situations where much training data is available in one domain but not in another [42]. One example is sentiment classification [2,9]. However, it has never yet been applied to aspect extraction. The present study is thus the first to present this application.

## 3 Method

As mentioned previously, even the state-of-art performance of the convolutional-neural-network-based approach suffers from limitations when applied to the extraction of product evaluation factors for smartphones. The previous approach requires a significant amount time to train the model for each product as well as labeled data for each model.

The previous approach is therefore not suitable for the rapidly evolving smartphone industry, where data becomes quickly outdated. In most cases, it is difficult to obtain large amounts of data for each smartphone. To overcome these limitations, we propose a variant of the previous approach by combining a convolutional neural network and transfer learning. The proposed framework is illustrated in Fig. 1 and explained below.

### 3.1 Training a Convolutional Neural Network

Before constructing the model for extracting product evaluation factors, we utilized the CBOW architecture of the word2vec model [27] to calculate word embeddings. The previous study had done this using the Amazon review dataset. However, that dataset did not contain all the words that appear in our smartphone review dataset. Thus, wWe therefore used both datasets. The Amazon dataset consists of 2,442,053 product reviews [25] and we kept the word embedding 300-dimensional.

After completing the word-embedding phase, we constructed and trained the convolutional neural network, as illustrated in Fig. 2. The features of an aspect term depend on its context words. Thus, we processed each word with a window size of $\pm 2$ as input. We formed local

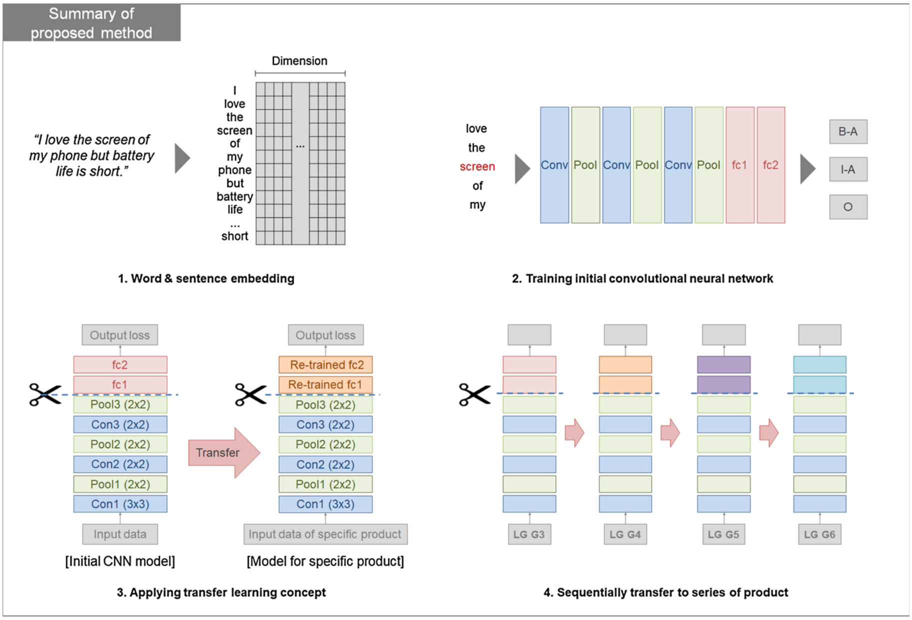



**Fig. 1** Overall framework of our proposed method

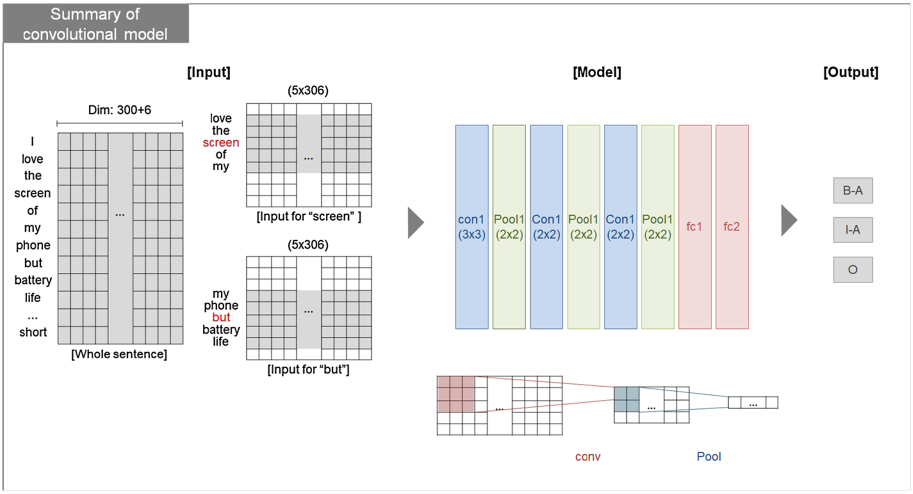



**Fig. 2** Proposed CNN architecture

features of that window and considered them to be features of the middle word. Then, the feature vector was fed into a convolutional neural network.

The network contained one input layer, two convolutional layers, two max pooling layers, and two fully connected layers with a softmax output. The first convolutional layer consisted of 100 feature maps with filter size $3 \times 3$. The second convolutional layer had 50 feature maps with filter size $2 \times 2 \times 100$. The third convolutional layer consisted of 25 feature maps with filter size $2 \times 2 \times 50$. The stride in each convolutional layer was 1, as we wanted to tag each

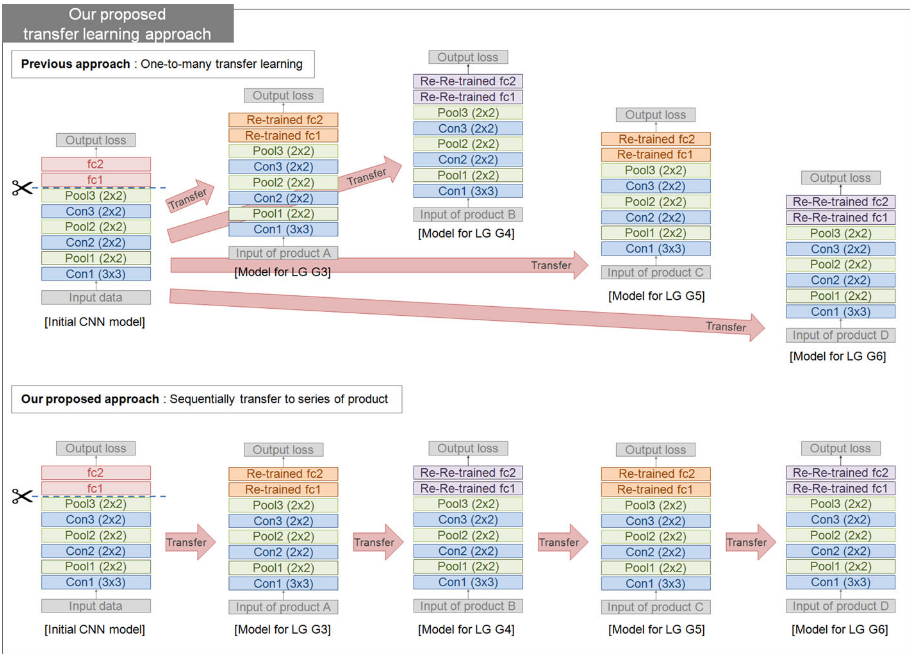



Fig. 3 Proposed transfer-learning approach

word. A max pooling layer followed each convolutional layer. The pool size used in the max pooling layer was $2 \times 2$. We used regularization with dropout on the penultimate layer with a constraint L2-norm of the weight vectors, with 30 epochs. The output of each convolutional layer was computed using a hyperbolic tangent, which yields a higher performance than other common activation functions [36].

Finally, we trained the convolutional neural network by using back-propagation after convolving all the tokens in the sentence. We stored the weight for each token after the convolution, and only back-propagated the summation of errors in order to correct them once all the tokens in the sentence had been processed.

## 3.2 Off-the-Shelf Features

After training the basic convolutional network by using the existing dataset of SemEval 2014 [39] and Qui et al. [37], we applied the "off-the-shelf features concept, as illustrated in Fig. 3.

This involves fixing the weights of the convolutional layer of the initial model and re-training only the fully connected layers with respect to each product. For instance, we stored the weights of the convolutional layer of the model for the initial product and set those stored weights as the initial weights of the model for a specific product. Then, we re-trained only the fully connected layers for the specific product to minimize the objective function of the aspect extraction model.

Previous studies have applied a one-to-many transfer-learning approach to train the aspect extraction model for each product. Based on the initial CNN model, it is transferred to each product separately. In this way, the dataset of the target product cannot be re-used. This is therefore not the most efficient approach for the group of products handled in our studies.

In our approach, the initial CNN model is sequentially transferred to another product group. For instance, the re-trained model for the LG G3 is transferred to the model for the LG G4, and that re-trained model is then transferred to the model for the LG G5 repeatedly. Our approach reuses all the dataset to train the aspect extraction model effectively.

Moreover, groups of products are usually released on the market sequentially, such that the latest product retains features of the previous product. This means that our transfer-learning approach effectively propagates the feature of the aspect extraction model of the previously released product.

### 3.3 Part-of-Speech Tagging

We also added linguistic features to further improve performance. Most product evaluation terms are either nouns or groups of nouns. We therefore utilized part-of-speech (POS) tags. Specifically, we used six basic POS features of the Stanford Tagger: Noun, Verb, Adjective, Adverb, Preposition, and Conjunction. It is encoded as a six-dimensional binary vector. Thus, for each word, the final feature vector is 306-dimensional.

## 4 Experiments

### 4.1 Data Description

As mentioned above, we utilized the dataset of SemEval and Qiu et al. to train the initial convolutional neural network, to serve as the basis for the proposed method. More specifically, we selected 3841 reviews on laptops from the SemEval dataset and 740 reviews of cellphones from the dataset of Qiu et al.

Further, we utilized our user review data for smartphones from LG Electronics to re-train the fully connected network for the transfer-learning approach. Our smartphone review dataset consisted of four major products of LG Electronics: the LG G3, G4, G5, and G6 phones. Each dataset consisted of 1000 reviews collected between April 23, 2014 and March 23, 2017, and were labeled by the domain experts of the Mobile Communication (MC) department in LG Electronics.

All the above-mentioned datasets were labeled using a widely used coding scheme for representing sequences. In this example, the first word of each aspect starts with a *B-A* tag. The *I-A* tag denotes the continuation of the aspect, and *O* is used to tag a word that is not an aspect:

*I/O love/O the/O touchscreen/B-A of/O my/O phone/O but/O the/O battery/B-A life/I-A is/O so/O short/O*

### 4.2 Experiment Setup

We performed qualitative and quantitative experiments to verify the performance of evaluation-factor extraction for the products. In the qualitative experiment, we verified that the proposed method extracted the factor specific to each product such as the LG G3, G4, G5, and G6 smartphones. The extraction performed for each specific product is very important for the smartphone industry, given its rapid evolution.

**Table 1** Key features of each product described on the official website

| Product | Key features |
|---|---|
| LG G3 | Quad HD IPS display, OIS+, Laser auto focus, Voice selfie, KnockCode, Smart notice, Smart Keyboard, Smart cleaning, Quick circle cover |
| LG G4 | Genuine leather cover, IPS Quantum display, F1.8 low-light lens, Manual mode camera, OIS 2.0, Quick shot, Smart notice 2.0 |
| LG G5 | Modular type, LG Friends, Wide angle lens, Magnifying zoom In and out, Auto selfie, Film effect, Pop-out pictures, Daylight mode, Always-on display, LG Backup, Smart doctor, Quick help |
| LG G6 | Full vision display, 18:9 screen ration, Hi-Fi quad DAC, 24 bit Hi-Fi recording, QHD+ Dolby vision, LG Pay, Wide angle front camera, Square camera mode, Water and dust resistant, Face unlock, Quick charge 3.0 |

Furthermore, the qualitative experiments compared the extraction performance and training time of the proposed and previous methods for each LG product G3, G4, G4, and G6. Performance was compared in terms of the following benchmarks: (1) a fully re-trained CNN without transfer learning [32]; (2) CRF-PR (a conditional random field with posterior regularization) [52]; (3) NBSVM (naive Bayes support vector machine) [48]; (4) SVMs (SVM with hand-coded rules) [41] and (5) MNB (multinomial naive Bayes) [48].

We performed additional experiments to assess the efficacy of sequential transfer learning, depth of transfer, window size, and adding POS features to word embedding. We trained each model over 50 epochs and tested the extraction performance with a 10-fold cross validation, as commonly used [26].

## 4.3 Experiment Results

### 4.3.1 Qualitative Result

First, we qualitatively verified the performance of the proposed method. Table 1 lists the specific features of each product, as described on the official LG Electronics website.

Table 2 summarizes the product evaluation factors extracted by the proposed method. A comparison with Fig. 1 suggests that the key features of each product were extracted effectively with the proposed method without re-training the entire convolutional neural network model. For instance, *Modular design* and *Friends* are related to LG "Friends (one of the key features), and *Always on* and *AoD* (which denotes "always on display, another key feature) were extracted effectively for the LG G5. In spite of having re-trained only a small portion of the entire network, the proposed method performs efficiently.

Nonetheless, some key features were not extracted by the proposed method. While being features of LG Electronics devices, they are not regarded as important evaluation factors by customers and do not appear frequently in customer reviews. The proposed method effectively extracts the product evaluation factors that are important from the point of view of the customer, and not of the manufacturer.

### 4.3.2 Quantitative Performance

Table 3 compares the extraction performance and training times of the proposed and previous methods applied to various products. Figure 4 also directly compares the earlier CNN

**Table 2** Key features of each product extracted by the proposed method

| Product | Key features |
|---|---|
| LG G3 | *QHD, Quad hd, Knock code, unlock, Quick circle, Quick Cover* |
| LG G4 | *Leather cover, F1.8 , Manual Mode, professional mode, Low light* |
| LG G5 | *Modular design, Friends, Wide angle, Always-on, AoD, Smart doctor* |
| LG G6 | *Full vision, 18:9, quad DAC, Hi-Fi recording, LG Pay, Wide angle front camera, Water- and dust-resistant* |

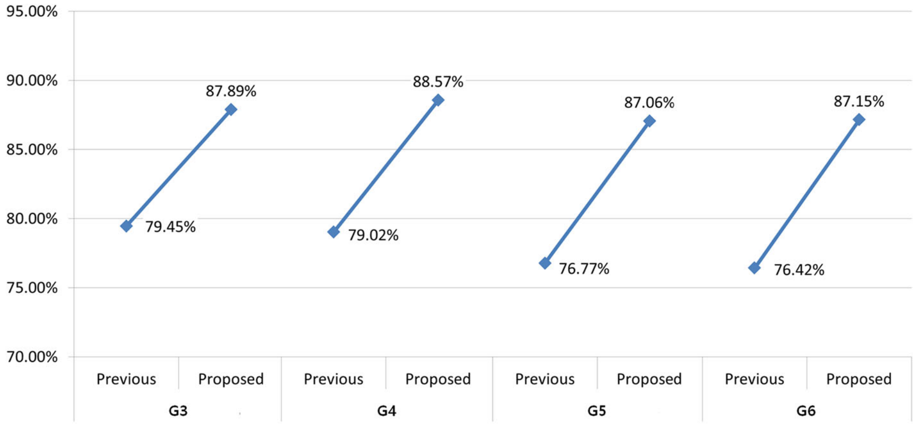
**Table 3** Extraction performance

| Product | Extraction model | Training time (min) | Precision (%) | Recall (%) | F-score (%) |
|---|---|---|---|---|---|
| LG G3 | Transfer learning* (CNN basis) | 12.28 | **90.48** | **85.45** | **87.89** |
| | Fully re-trained CNN | 175.35 | 81.59 | 77.41 | 79.45 |
| | CRF-PR | 16.92 | 72.42 | 70.98 | 71.69 |
| | NBSVM | 9.42 | 68.81 | 63.63 | 66.12 |
| | SVMs | 8.53 | 66.50 | 64.89 | 65.68 |
| | MNB | 8.83 | 62.03 | 61.50 | 61.76 |
| LG G4 | Transfer learning* (CNN basis) | 13.78 | **91.02** | **86.24** | **88.57** |
| | Fully re-trained CNN | 167.63 | 82.93 | 75.47 | 79.02 |
| | CRF-PR | 16.11 | 73.09 | 68.05 | 70.48 |
| | NBSVM | 9.16 | 68.64 | 64.61 | 66.56 |
| | SVMs | 8.61 | 65.01 | 64.96 | 64.98 |
| | MNB | 8.26 | 64.85 | 60.64 | 62.67 |
| LG G5 | Transfer learning* (CNN basis) | 13.22 | **90.12** | **84.21** | **87.06** |
| | Fully re-trained CNN | 170.05 | 80.17 | 73.64 | 76.77 |
| | CRF-PR | 16.77 | 72.27 | 70.33 | 71.29 |
| | NBSVM | 9.04 | 66.73 | 65.85 | 66.29 |
| | SVMs | 8.53 | 65.25 | 62.75 | 63.96 |
| | MNB | 8.11 | 64.57 | 60.82 | 62.64 |
| LG G6 | Transfer learning* (CNN basis) | 13.68 | **89.65** | **84.78** | **87.15** |
| | Fully re-trained CNN | 168.97 | 79.87 | 73.27 | 76.42 |
| | CRF-PR | 16.93 | 70.42 | 67.04 | 68.69 |
| | NBSVM | 9.39 | 66.52 | 64.58 | 65.54 |
| | SVMs | 8.81 | 66.86 | 64.85 | 65.84 |
| | MNB | 8.28 | 64.04 | 61.94 | 62.97 |

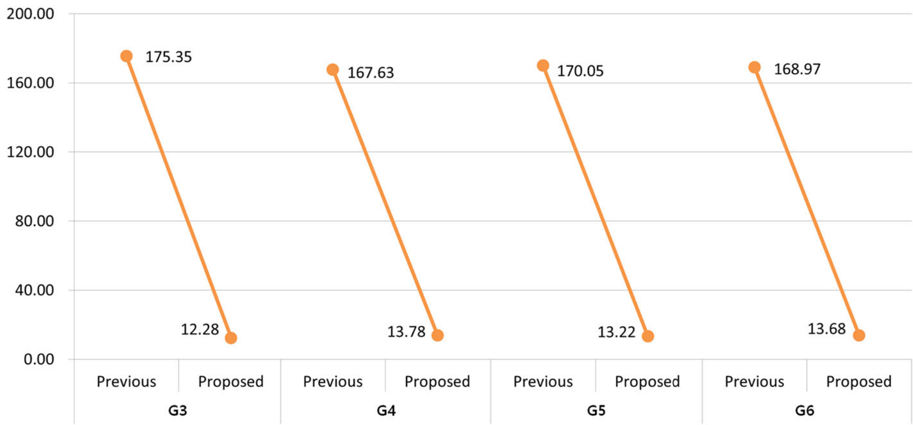
*Proposed method
Bold values indicate the highest performance score in each experiment variable

approach and our proposed method that includes transfer learning. This was investigated using an Intel Core-i5 2.6 GHz CPU and an NVIDIA GeForce GTX GPU system and progressed with Tensorflow.

The proposed method performed better than the previous methods for all the corresponding product lines. As is widely reported, convolutional neural networks cannot be trained well with a small number of datasets by over-fitting. For this reason, the previous method performed worse than the proposed method for various smartphones, although it worked well




**(a)** F-score




**(b)** Training time (min)

**Fig. 4** Extraction performance

with the fully trained initial convolution neural network model. Similarly, our experimental result indicates that other approaches, such as the CRF- or SVM-based approach, could not be trained effectively with only a few datasets.

Understandably, the proposed method requires significantly less time than the previous convolutional neural network based method, as the proposed method does not require re-training all the convolutional layers, unlike the previous method.

### 4.3.3 Effect of Sequential Transfer Learning

This section considers differences in the ways of conducting transfer learning. As mentioned above, we transferred the initial convolutional neural network model to other product group sequentially through a comparison with the earlier one-to-many approach.

**Table 4** Effectiveness of sequential transfer learning

| Product | Transfer-learning method | Recall (%) | Precision (%) | F-score (%) |
|---|---|---|---|---|
| LG G3 | One-to-many transfer learning | **90.48** | **85.45** | **87.89** |
| | Sequential transfer learning* | **90.48** | **85.45** | **87.89** |
| LG G4 | One-to-many transfer learning | 90.29 | 85.69 | 87.93 |
| | Sequential transfer learning* | **91.02** | **86.24** | **88.57** |
| LG G5 | One-to-many transfer learning | 89.45 | 81.95 | 85.54 |
| | Sequential transfer learning* | **90.12** | **84.21** | **87.06** |
| LG G6 | One-to-many transfer learning | 86.70 | 81.83 | 84.19 |
| | Sequential transfer learning* | **89.65** | **84.78** | **87.15** |

*Proposed method
Bold values indicate the highest performance score in each experiment variable

**Table 5** Effect of the transferred depth

| Product | Re-trained layer | Training time (min) | Precision (%) | Recall (%) | F-score (%) |
|---|---|---|---|---|---|
| LG G3 | FC (2-level)* | 12.28 | 90.48 | 85.45 | 87.89 |
| | FC (1-level) | 10.11 | 86.22 | 78.12 | 81.97 |
| | Conv/Pool + FC (2-level) | 24.93 | **90.61** | **85.68** | **88.08** |
| LG G4 | FC (2-level)* | 13.78 | **91.02** | **86.24** | **88.57** |
| | FC (1-level) | 11.23 | 84.73 | 80.91 | 82.78 |
| | Conv/Pool + FC (2-level) | 27.18 | 90.88 | 86.01 | 88.38 |
| LG G5 | FC (2-level)* | 13.22 | **90.12** | **84.21** | **87.06** |
| | FC (1-level) | 10.23 | 83.38 | 77.87 | 80.53 |
| | Conv/Pool + FC (2-level) | 89.92 | 84.39 | 86.96 | 62.64 |
| LG G6 | FC (2-level)* | 13.68 | 89.65 | **84.78** | **87.15** |
| | FC (1-level) | 10.97 | 84.07 | 80.06 | 82.02 |
| | Conv/Pool + FC (2-level) | 24.49 | **89.70** | 84.38 | 86.96 |

*Proposed method
Bold values indicate the highest performance score in each experiment variable

In the experiments, we transferred the convolutional neural network model in the order in which the data were released, namely G3->G4->G5->G5. Table 4 indicates that our proposed method performs better than the one-to-many approach and that the difference between the two methods increases with each new release. This means that aspect extraction performance increases as the product group is expanded since all the dataset of the previously released product is used.

### 4.3.4 Effect of Depth of Transfer

In order to check the effect of the depth to be transferred, we explored a few variations for transfer learning. In addition to our proposed architecture re-trained 2-level of fully connected network, we considered other options where: (1) only the last fully connected network layer is re-trained, or (2) the convolution and pooling layers are also re-trained. Table 5 reports the resulting performances, showing that a better performance is achieved when at least two levels of the fully connected layer are re-trained. And considering training time, it have no use for re-training the convolution and pooling layers.

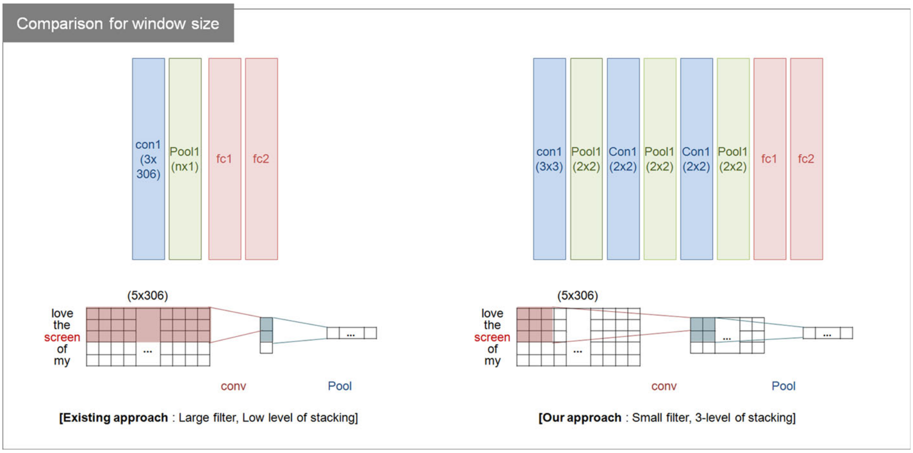



**Fig. 5** Significance of window size

**Table 6** Effect of window size

| Product | Filter size | Training time (min) | Precision (%) | Recall (%) | F-score (%) |
|---|---|---|---|---|---|
| LG G3 | Small-sized filter* | 12.28 | **90.48** | **85.45** | **87.89** |
| | Large-sized filter | 9.38 | 84.62 | 77.80 | 81.06 |
| LG G4 | Small-sized filter* | 13.78 | **91.02** | **86.24** | **88.57** |
| | Large-sized filter | 10.27 | 83.94 | 80.21 | 82.03 |
| LG G5 | Small-sized filter* | 13.22 | **90.12** | **84.21** | **87.06** |
| | Large-sized filter | 9.88 | 84.66 | 77.45 | 80.90 |
| LG G6 | Small-sized filter* | 13.68 | **89.65** | **84.78** | **87.15** |
| | Large-sized filter | 10.94 | 83.45 | 78.78 | 81.05 |

*Proposed method
Bold values indicate the highest performance score in each experiment variable

### 4.3.5 Significance of the Window Size

Our convolutional network utilized a small convolution filter and the three levels of convolution and pooling layers were stacked, as is commonly done in the image-processing field [53]. In contrast with our approach, existing text-mining studies have tended to utilize large filters with dimension of word embedding and to not stack the convolution/pooling layers [20,32]. This is illustrated in Fig. 5.

Table 6 indicates that a small window filter yields better performance than a large one with low-level stacking. Even a shorter training time for a large filter would strongly affect aspect extraction performance.

### 4.3.6 Effect of POS Tagging

POS tagging information yields a further small increase in extraction performance, as shown in Table 7. However, it does not affect the result when critically compared with other features.

**Table 7** Effectiveness of POS tagging

| Product | POS feature | Recall (%) | Precision (%) | F-score (%) |
|---|---|---|---|---|
| LG G3 | W/O POS | 89.72 | 84.13 | 86.84 |
| | W/ POS* | **90.48** | **85.45** | **87.89** |
| LG G4 | W/O POS | 88.79 | 86.05 | 87.40 |
| | W/ POS* | **91.02** | **86.24** | **88.57** |
| LG G5 | W/O POS | 89.16 | 84.05 | 86.53 |
| | W/ POS* | **90.12** | **84.21** | **87.06** |
| LG G6 | W/O POS | 89.13 | 82.79 | 85.84 |
| | W/ POS* | **89.65** | **84.78** | **87.15** |

*Proposed method
Bold values indicate the highest performance score in each experiment variable

## 5 Conclusion

The present study proposed an advanced method for extracting product evaluation factors, for application to the rapidly evolving smartphone industry. Understanding the key factors that affect customers evaluations of a new product can improve decision making throughout the design and development process. Product developers and marketers often seek to identify the product attributes that consumers consider to be most important when they evaluate a product for purchase.

In this respect, aspect extraction (as a subtopic of natural language processing or opinion mining) could be a useful tool for determining these important attributes for the purpose of evaluating products or services. Recently, aspect extraction based on deep convolution neural networks has been suggested and shown to achieve state-of-the-art performance when applied to a customer review of electronic devices.

However, that approach required a significant amount of labeled data and a long time to train the convolutional neural network for each product. This makes it inappropriate for application to the smartphone industry, which has a wide range of product lines.

We therefore proposed an improvement over the previous convolutional-neural-network-based approach, whereby product evaluation factors are extracted through transfer learning. Unlike the previous approach, we maintained the global feature parameters of the convolutional neural network model and only re-trained the parameter for the classification part of the model.

As demonstrated in the experiments, key features of each product were extracted effectively by the proposed method without the need to re-train the entire convolutional neural network model. The proposed method also showed better extraction performance compared with the previous method for all the corresponding product lines.

On this basis, future studies could seek to construct an entire set of product evaluation factors by analyzing customers spoken words. This study applied transfer learning to products within the same general domain. Future studies could thus consider extending the analysis to span a broader range of domains.

Our application of the convolutional-neural-network-based approach was relatively focused. Future studies could extend the range of utilized architectures. For instance, we intend to perform aspect extraction using an auto-encoder that is utilized for various tasks [11,12,43].

Furthermore, our proposed method should be applicable to the marketing or development of products in real business environments after in-depth analysis.

**Acknowledgements** We are very grateful to LG Electronics for having provided us with the customer-review dataset used in our comparison experiments.

## References

1. Andrzejewski D, Zhu X, Craven M (2009) Incorporating domain knowledge into topic modeling via dirichlet forest priors. In: Proceedings of the 26th annual international conference on machine learning. ACM, pp 25–32
2. Blitzer J, Dredze M, Pereira F et al (2007) Biographies, bollywood, boom-boxes and blenders: domain adaptation for sentiment classification. In: ACL, vol 7. pp 440–447
3. Brody S, Elhadad N (2010) An unsupervised aspect-sentiment model for online reviews. In: Human language technologies: the 2010 annual conference of the North American chapter of the association for computational linguistics. Association for Computational Linguistics, pp 804–812
4. Chen Z, Mukherjee A, Liu B, Hsu M, Castellanos M, Ghosh R (2013) Exploiting domain knowledge in aspect extraction. In: Emnlp, pp 1655–1667
5. Choi Y, Cardie C (2010) Hierarchical sequential learning for extracting opinions and their attributes. In: Proceedings of the ACL 2010 conference short papers. Association for Computational Linguistics, pp 269–274
6. Federici M, Dragoni M (2016) Towards unsupervised approaches for aspects extraction. In: EMSA-RMed@ ESWC
7. Fuchs C, Prandelli E, Schreier M (2010) The psychological effects of empowerment strategies on consumers' product demand. J Mark 74(1):65–79
8. Ghifary M, Kleijn WB, Zhang M (2014) Domain adaptive neural networks for object recognition. In: Pacific rim international conference on artificial intelligence, Springer, pp 898–904
9. Glorot X, Bordes A, Bengio Y (2011) Domain adaptation for large-scale sentiment classification: a deep learning approach. In: Proceedings of the 28th international conference on machine learning (ICML-11), pp 513–520
10. Hai Z, Chang K, Kim Jj (2011) Implicit feature identification via co-occurrence association rule mining. In: Computational linguistics and intelligent text processing, pp 393–404
11. Hong C, Chen X, Wang X, Tang C (2016) Hypergraph regularized autoencoder for image-based 3d human pose recovery. Signal Process 124:132–140
12. Hong C, Yu J, Wan J, Tao D, Wang M (2015) Multimodal deep autoencoder for human pose recovery. IEEE Trans Image Process 24(12):5659–5670
13. Huang S, Liu X, Peng X, Niu Z (2012) Fine-grained product features extraction and categorization in reviews opinion mining. In: 2012 IEEE 12th international conference on data mining workshops (ICDMW). IEEE, pp 680–686
14. Jagarlamudi J, Daumé III H, Udupa R (2012) Incorporating lexical priors into topic models. In: Proceedings of the 13th conference of the European chapter of the association for computational linguistics. Association for Computational Linguistics, pp 204–213
15. Jakob N, Gurevych I (2010) Extracting opinion targets in a single-and cross-domain setting with conditional random fields. In: Proceedings of the 2010 conference on empirical methods in natural language processing. Association for Computational Linguistics, pp 1035–1045
16. Jamal A, Goode M (2001) Consumers' product evaluation: a study of the primary evaluative criteria in the precious jewellery market in the UK. J Consum Behav 1(2):140–155
17. Jin W, Ho HH, Srihari RK (2009) A novel lexicalized hmm-based learning framework for web opinion mining. In: Proceedings of the 26th annual international conference on machine learning, pp. 465–472
18. Jo Y, Oh AH (2011) Aspect and sentiment unification model for online review analysis. In: Proceedings of the fourth ACM international conference on Web search and data mining. ACM, pp 815–824
19. Kim S, Zhang J, Chen Z, Oh AH, Liu S (2013) A hierarchical aspect-sentiment model for online reviews. In: AAAI
20. Kim Y (2014) Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882
21. Li F, Han C, Huang M, Zhu X, Xia YJ, Zhang S, Yu H (2010) Structure-aware review mining and summarization. In: Proceedings of the 23rd international conference on computational linguistics. Association for Computational Linguistics, pp 653–661

22. Lin C, He Y (2009) Joint sentiment/topic model for sentiment analysis. In: Proceedings of the 18th ACM conference on Information and knowledge management. ACM, pp 375–384
23. Liu Q, Liu B, Zhang Y, Kim DS, Gao Z (2016) Improving opinion aspect extraction using semantic similarity and aspect associations. In: AAAI, pp 2986–2992
24. Liu B, Zhiyuan C, Arjun M (2014) Aspect extraction with automated prior knowledge learning. In: Proceedings of the 52nd annual meeting of the Association for Computational Linguistics, pp 347–358
25. McAuley J, Leskovec J (2013) Hidden factors and hidden topics: understanding rating dimensions with review text. In: Proceedings of the 7th ACM conference on Recommender systems. ACM, pp 165–172
26. McLachlan G, Do KA, Ambroise C (2005) Analyzing microarray gene expression data, vol 422. Wiley, New York
27. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781
28. Mitchell M, Aguilar J, Wilson T, Van Durme B (2013) Open domain targeted sentiment
29. Moghaddam S, Ester M (2011) Ilda: interdependent lda model for learning latent aspects and their ratings from online product reviews. In: Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval. ACM, pp 665–674
30. Mukherjee A, Liu B (2012) Aspect extraction through semi-supervised modeling. In: Proceedings of the 50th annual meeting of the association for computational linguistics: long papers, vol 1. Association for Computational Linguistics, pp 339–348
31. Pan SJ, Yang Q (2010) A survey on transfer learning. IEEE Trans Knowl Data Eng 22(10):1345–1359
32. Poria S, Cambria E, Gelbukh A (2016) Aspect extraction for opinion mining with a deep convolutional neural network. Knowl Based Syst 108:42–49
33. Poria S, Cambria E, Gelbukh A, Bisio F, Hussain A (2015) Sentiment data flow analysis by means of dynamic linguistic patterns. IEEE Comput Intell Mag 10(4):26–36
34. Poria S, Cambria E, Ku L W , Gui C, Gelbukh A (2014) A rule-based approach to aspect extraction from product reviews. In: Proceedings of the second workshop on natural language processing for social media (SocialNLP), pp 28–37
35. Poria S, Chaturvedi I, Cambria E, Bisio F (2016) Sentic lda: improving on lda with semantic similarity for aspect-based sentiment analysis. In: 2016 international joint conference on neural networks (IJCNN). IEEE, pp 4465–4473
36. Qian S, Liu H, Liu C, Wu S, San Wong H (2018) Adaptive activation functions in convolutional neural networks. Neurocomputing 272:204–212
37. Qiu G, Liu B, Bu J, Chen C (2011) Opinion word expansion and target extraction through double propagation. Comput Linguist 37(1):9–27
38. Samiee S (1994) Customer evaluation of products in a global market. J Int Bus Stud 25(3):579–604
39. Semeval-2014 dataset (2014) In: International workshop on semantic evaluation. https://alt.qcri.org/semeval2014
40. Sharif Razavian A, Azizpour H, Sullivan J, Carlsson S (2014) Cnn features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 806–813
41. Silva J, Coheur L, Mendes AC, Wichert A (2011) From symbolic to sub-symbolic information in question classification. Artif Intell Rev 35(2):137–154
42. Socher R, Ganjoo M, Manning CD, Ng A (2013) Zero-shot learning through cross-modal transfer. In: Advances in neural information processing systems, pp 935–943
43. Socher R, Pennington J, Huang EH, Ng AY, Manning CD (2011) Semi-supervised recursive autoencoders for predicting sentiment distributions. In: Proceedings of the conference on empirical methods in natural language processing. Association for Computational Linguistics, pp 151–161
44. Toh Z, Wang W (2014) Dlirec: Aspect term extraction and term polarity classification system. In: Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014), pp 235–240
45. van Ginneken B, Setio AA, Jacobs C, Ciompi F (2015) Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. In: 2015 IEEE 12th international symposium on biomedical imaging (ISBI). IEEE, pp 286–289
46. Veryzer RW (1998) Key factors affecting customer evaluation of discontinuous new products. J Prod Innov Manag 15(2):136–150
47. Wang H (2015) Sentiment-aligned topic models for product aspect rating prediction. Ph.D. thesis, Applied Sciences, School of Computing Science
48. Wang S, Manning CD (2012) Baselines and bigrams: Simple, good sentiment and topic classification. In: Proceedings of the 50th annual meeting of the association for computational linguistics: short papers, vol 2. Association for Computational Linguistics, pp 90–94

49. Wang T, Cai Y, Leung Hf, Lau RY, Li Q, Min H (2014) Product aspect extraction supervised with online domain knowledge. Knowl Based Syst 71:86–100
50. Wu Y, Zhang Q, Huang X, Wu L (2009) Phrase dependency parsing for opinion mining. In: Proceedings of the 2009 conference on empirical methods in natural language processing, vol 3. Association for Computational Linguistics, pp 1533–1541
51. Yang B, Cardie C (2013) Joint inference for fine-grained opinion extraction. In: ACL, vol 1, pp 1640–1649
52. Yang B, Cardie C (2014) Context-aware learning for sentence-level sentiment analysis with posterior regularization. In: Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 1: long papers), vol 1, pp 325–335
53. Yang M, Liu Y, You Z (2017) The euclidean embedding learning based on convolutional neural network for stereo matching. Neurocomputing 267:195–200
54. Zhang L, Liu B, Lim S H , O'Brien-Strain E (2010) Extracting and ranking product features in opinion documents. In: Proceedings of the 23rd international conference on computational linguistics: posters. Association for Computational Linguistics, pp 1462–1470
55. Zhao WX, Jiang J, Yan H, Li X (2010) Jointly modeling aspects and opinions with a maxent-lda hybrid. In: Proceedings of the 2010 conference on empirical methods in natural language processing. Association for Computational Linguistics, pp 56–65
56. Zhou X, Wan X, Xiao J (2013) Collective opinion target extraction in chinese microblogs. In: EMNLP, vol 13, pp 1840–1850
57. Zhu J, Wang H, Tsou B K, Zhu M. (2009). Multi-aspect opinion polling from textual reviews. In: Proceedings of the 18th ACM conference on information and knowledge management. ACM, pp 1799–1802

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.