Bus Inf Syst Eng 64(4):407–419 (2022)
https://doi.org/10.1007/s12599-021-00737-5

RESEARCH PAPER

Identifying Competitive Attributes Based on an Ensemble of Explainable Artificial Intelligence

Younghoon Lee

Received: 23 August 2020 / Accepted: 28 October 2021 / Published online: 14 January 2022

Abstract Competitor analysis is a fundamental requirement in both strategic and operational management, and the competitive attributes of reviewer comments are a crucial determinant of competitor analysis approaches. Most studies have focused on identifying competitors or detecting comparative sentences, not competitive attributes. Thus, the authors propose a method based on explainable artificial intelligence (XAI) that can detect competitive attributes from consumers' perspectives. They construct a model to classify the reviewer comments for each competitive product and calculate the importance of each keyword in the reviewer comments during the classification process. This is based on the assumption that keywords significantly influence product classification. The authors also propose an additional novel methodology that combines various XAI techniques such as local interpretable model-agnostic explanations, Shapley additive explanations, logistic regression, gradient-based class activation map, and layer-wise relevance propagation to build a robust model for calculating the importance of competitive attributes for various data sources.

Keywords XAI · Ensemble · Competitor analysis · Competitive factors · Home appliance

Accepted after three revisions by Stefan Lessmann.

Y. Lee (✉)
Department of Industrial Engineering, Seoul National University of Science and Technology, 232 Gongneung-ro, Nowon-gu, Seoul 01811, Korea
e-mail: yhoon.lee@seoultech.ac.kr

1 Introduction

Competitor analysis is the identification of the strengths and weaknesses of competitors' products and services (Davcik and Sharma 2016). Thus, it is closely affiliated with strategic decision-making. A company should be aware of the current strategy and future goals of its competitors. Further, it should be aware of the assumptions about capabilities and priorities to understand how a competitor is likely to respond (Chakraborti and Dey 2019). Therefore, competitor analysis is a fundamental requirement in both strategic and operational management (Fan et al. 2015).

With the rapid development of mobile and web technologies, competitors often use the online textual reviews of their products (Archak et al. 2011). Online reviews furnish rich information on customers' concerns and allow designers to improve products by providing them with a general idea of their competitors (Mudambi and Schuff 2010). In addition to customer reviews, which are becoming an essential source in competitor analysis, the competitive attributes in reviewer comments are a key determinant of competitor analysis approaches (Raut et al. 2018).

Many studies based on customer reviews have been conducted. However, most have focused on the identification of competitors or detection of comparative sentences, not competitive attributes (Kim and Kang 2018; Bi et al. 2019; Jin et al. 2016; Lee and Lee 2017; Raut et al. 2018; Gao et al. 2018; Guo et al. 2017; Jin et al. 2019). Although existing studies shed light on competitors and find comparative sentences (e.g., "LG TVs perform better than Samsung TVs"), they fail to detect which competitive factor is important (e.g., design, operating system). To extract the competitive attribute from the comparative sentence, time-consuming heuristic work is required, and this may offer a subjective judgment depending on the person. Recently, competitive attributes were detected in a small number of studies (Lee 2021; Kim and Kang 2018; Han and Lee 2021). However, these studies have limitations in performance and robustness because existing methods such as neural networks are applied to a single data source.

Based on the foregoing, we propose a method based on explainable artificial intelligence (XAI) to detect competitive attributes from consumers' perspective. Importantly, we construct a model to classify the reviewer comments for each competitive product and calculate each keyword's importance in the reviewer comments during the classification process. We then extract and prioritize the keywords and determine their competitiveness based on importance. We assume that keywords significantly influence product classification and are considered to be meaningful points of differentiation by customers.

Additionally, we collect customer review data from multiple sources such as blogs, communities, and ecommerce sites. Our experimental results show that the method of detecting competitive attributes based on individual XAI algorithms performs significantly differently depending on the data source. Therefore, we propose an additional novel methodology that combines various XAI techniques such as LIME, SHAP, logistic regression, Grad-CAM, and LRP to build a robust model for calculating the importance of competitive attributes. Since each XAI model has a different method of calculating the importance of competitive attributes, a detailed methodology for normalizing each model's importance score for the ensemble is also proposed.

We verify the performance of our proposed methodology, both qualitatively and quantitatively. We then review the extracted competitive factors qualitatively and compare them with the product attributes that customers consider to be the most important, as found through a survey. We further verify how the competitive factors extracted using our proposed method influence each product's overall customer evaluation.

The rest of this paper is structured as follows. Section 2 discusses various studies of competitor analysis and the related architecture. Section 3 describes our proposed method used to extract the competitive factors perceived by consumers using the customer's ensemble. Section 4 describes the data applied and presents the experimental results to demonstrate our proposed method's performance. Finally, Sect. 5 offers concluding remarks and future research directions.

2 Literature Review

2.1 Studies of Competitor Analysis

The first literature stream of competitor analysis identifies the competitor or firm's position. Guo et al. built an automated competitor analysis system using big data analytics, focusing on monitoring a firm's market position and competitors (Guo et al. 2017). Lee et al. analyzed companies' position in complex markets using competitor intelligence (Lee and Lee 2017). Gao et al. proposed a novel method of identifying competitors and the market environment by mining customers' opinions (Gao et al. 2018). Raut et al. proposed a framework to determine the top-k competitors using large unstructured textual datasets (Raut et al. 2018). Boniface et al. enriched existing theories by suggesting a customer–product–competitor analysis model determining whether to reconfigure and modify products to create new value, thereby contributing to a firm's market repositioning, continuity, and sustainability (Boniface 2017). Additionally, Gur et al. and Hatzijordanou et al. summarized competitor identification studies (Gur and Greckhamer 2019; Hatzijordanou et al. 2019).

Competitor analysis also occurs in other literature streams focusing on detecting comparative sentences. Wang et al. categorized the comparative opinions of Chinese online reviews and proposed a combined method for extracting comparative elements (Wang et al. 2017). Yan et al. outlined a framework for competitor analysis by extracting customer concerns from reviews of a series of products (Yan et al. 2017). Alharbi focused on identifying comparative sentences from social comments using a sequential pattern mining approach (Alharbi and Khan 2019). Simultaneously, Jin et al. proposed a framework representing shared customer feedback extracted from reviews comparing a series of products (Jin et al. 2019). Jin et al. also proposed a framework to select pairs of representative but comparative sentences related to specific competitive product features (Jin et al. 2016). Paul et al. proposed a two-stage approach to summarize consumers' contrasting opinions and uncover their different concerns (Paul et al. 2010). Moreover, Varathan et al. reviewed studies detecting comparative sentences (Varathan et al. 2017).

However, as mentioned in the Introduction, only a few studies focus on extracting competitive attributes, which are a critical determinant of competitor analysis approaches. Lee et al. and Kim et al. proposed methodologies for extracting the distinct attributes of competing products using a neural network-based algorithm and a text mining approach (Lee 2021; Kim and Kang 2018; Han and Lee 2021). However, these studies have limitations in performance and robustness because existing methods such as neural networks are applied to a single data source. Thus, we propose a novel methodology that combines various XAI techniques such as LIME, SHAP, logistic regression, Grad-CAM, and LRP to build a robust model for calculating the importance of competitive attributes.

2.2 Related Architecture

2.2.1 *XAI Algorithms*

XAI is an artificial intelligence programmed to describe its purpose, rationale, and decision-making process in a way that the average person can understand. This study introduces an ensemble of variations of XAI algorithms such as logistic regression, LIME, SHAP, Grad-CAM, and LRP.

The most common way of understanding a linear model such as logistic regression is to examine the coefficients learned for each feature. These coefficients explain the extent to which the model output changes when we change each of the input features. The LIME algorithm can accurately explain the predictions of any classifier or regressor by approximating it locally using an interpretable model (Ribeiro et al. 2016). It modifies every data sample by tweaking the feature values and observing the resulting impact on the output. LIME explains each data sample's predictions for each feature in the form of local interpretability.

The core idea behind SHAP-based explanations of machine learning models is using a fair allocation from cooperative game theory to allocate credit for a model's output $f(x)$ among its input features. To connect game theory with machine learning models, it is necessary to match both a model's input features with the players in a game and the model function with the rules of the game (Lundberg and Lee 2017).

Grad-CAM is a popular technique for visualizing a convolutional neural network (CNN) model. It is based on the belief that image pixel attributions can be better visualized by calculating the gradient from the output to a given deeper layer (as opposed to calculating the gradient up to the input layer of the model). Grad-CAM reconstructs maps as a weighted combination of the forward neuron activation; the weights are based on the global average pooling and backpropagation of the outputs to a target layer (Zhou et al. 2016).

The core idea of an LRP algorithm attributing relevance to individual input nodes is to trace back the contributions to the final output node layer by layer. The LRP algorithm has several versions, but they all share the same principle: total relevance. For example, the activation strength of an output node for a particular class is conserved by layer; that is, each node in layer l contributing to the activation of node j in the subsequent layer $l+1$ is attributed a certain share of the relevance R^j_{l+1} of that node. Overall, the relevance of all nodes i contributing to neuron j in layer l must sum to R^j_{l+1}, thus conserving total relevance by layer.

Furthermore, we use other algorithms such as attention mechanisms as baselines for the comparison with our proposed method. These algorithms include a sequence model based on an attention mechanism (Wang et al. 2016). In a typical sequence model, the encoder LSTM is used to process the entire input sentence and encode it into a context vector, the last hidden LSTM/RNN state. The decoder LSTM or RNN units generate words one after another to form a sentence. After that, it tends to become forgetful in specific cases. Moreover, some of the input words cannot be given more importance than others when translating sentences. Therefore, when the proposed model generates a sentence, it searches for a set of positions in the encoder-hidden states containing the most relevant information. This idea is called "attention."

However, the performance of the XAI techniques mentioned above differs significantly depending on the characteristics of the data source since they adopt different approaches. Thus, in this study, competitive attributes are extracted with an ensemble of XAI techniques because several studies have demonstrated that individual ensemble algorithms in supervised learning show robust performance (Hu et al. 2012; Alobaidi et al. 2018; Zameer et al. 2017; Lee and Chung 2019; Wei et al. 2018).

2.2.2 *Aspect Extraction Method*

As explained in the next section, we apply an aspect extraction method to improve performance by extracting only essential words called "aspect" before placing the entire reviewer comment into the model. For example, a sentence such as "I love the touchscreen of my phone, but the battery life is too short" contains two aspects, or opinion targets, namely, *touchscreen* and *battery life* (Poria et al. 2014). We train the aspect extraction model to extract only the word that can be considered to be a key factor and not relatively meaningless words such as "I" and "my."

We use a state-of-the-art supervised CNN approach following Poria et al. (2016). This is because an unsupervised approach (e.g., topic modeling) usually provides rough topics rather than precise aspects. A topical term does not necessarily have to be an aspect.

The network includes one input layer, three convolutional layers, three max-pooling layers, and two layers fully connected with a softmax output. They construct the convolutional layers described in Table 1. Each convolutional layer's stride is one because we need to tag each word.

Table 1 Structure of the CNN for aspect extraction

Layer	Number of feature maps	Filter size
First layer	100	3×3
Second layer	50	$2 \times 2 \times 100$
Third layer	25	$2 \times 2 \times 50$

The pool size used in the max-pooling layer is 2×2. This computes the output of each convolutional layer using a hyperbolic tangent. Additionally, we use other algorithms as baselines for the comparison with our proposed method, namely, the hierarchical conditional random field (CRF) (Huang et al. 2012) and Dlirec (Toh and Wang 2014) approaches. A set of increasingly powerful CRF-based models is proposed in hierarchical CRF-based aspect extraction. This includes a hierarchical multi-label CRF scheme that jointly models the overall opinion expressed in the review and set of aspect-specific opinions expressed in each of its sentences. Further, Dlirec models aspect extraction as a sequential labeling task and extracts the features to be used for CRF training. Besides the common features used in traditional Named Entity Recognition systems, Dlirec also uses extensive external resources to build various name lists and word clusters.

3 Method

3.1 Extraction of Product Aspects

Before training the classification model, we must train the product aspect extraction model to choose only relevant factors from entire sentences. These aspects provide the essential attributes for evaluating products and services. First, we embed all the customer reviews in a 300-dimensional word2vec representation (Mikolov et al. 2013). We use Amazon datasets for the word2vec embedding task. We also add linguistic features to improve the performance of our proposed method. Most product evaluation terms are either nouns or groups of nouns. Therefore, we use parts-of-speech tags. Specifically, we use six basic Stanford tagger parts-of-speech words, namely, noun, verb, adjective, adverb, preposition, and conjunction, encoding them as a six-dimensional binary vector.

Furthermore, we construct and train the CNN after the word embedding tasks using existing datasets, that is, SemEval 2014 (International Workshop on Semantic Evaluation 2014) and the dataset developed by Qiu et al. (2011). We input each word with a window size of five into the CNN because the aspect terms' features depend on the context words. The other CNN parameters are based on the previous studies described in the Sect. 2. Additionally, we use a regularization with dropout on the penultimate layer, where the constraint L2-norms of the weight vectors have 50 epochs.

We label all the datasets mentioned above using a coding scheme widely employed to represent sequences. In this example, each aspect's first word starts with a *B-A* tag. An *I-A* tag denotes the continuation of this aspect, whereas *O* tags a word that is not an aspect.

3.2 Extraction and Prioritization of Competitive Attributes

As previously mentioned, a keyword that significantly influences the classification decision is considered to be a meaningful point of differentiation by customers. Furthermore, we combine various XAI methodologies to ensure robustness by considering data collected from multiple sources. As shown in the Sect. 4, individual XAI algorithms show different results depending on the data source. Conversely, if various XAI methods are ensembled, performance is high regardless of the data source. In this study, we experiment with various combinations of XAI ensembles and select the ensemble showing the best performance. The overall structure is illustrated in Fig. 1, and we describe the XAI we use in the next section. The parameter presented in each XAI methodology is that performing the best in the empirical experiments.

3.2.1 Logistic Regression

Logistic regression describes data and explains the relationship between one dependent variable and one or more independent variables. Mathematically, it estimates a multiple linear regression function defined as

$$\log\left(\frac{p(y=1)}{1-p(p=1)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots.$$

In logistic regression analysis, the value of each word's regression coefficient is regarded as the importance score of the word. As each reviewer comment is processed, each word's importance score is accumulated; however, if it is less than a specific coefficient λ, it is excluded from the accumulation. In this study, lambda is set to 0.1. Additionally, L2 regularization is applied and the Newton-cg technique is used as a solver.

3.2.2 LIME and SHAP

LIME is a novel technique that explains any classifier's prediction in an interpretable and faithful manner by learning an interpretable model locally around the prediction. To find a model that locally approximates a black-box

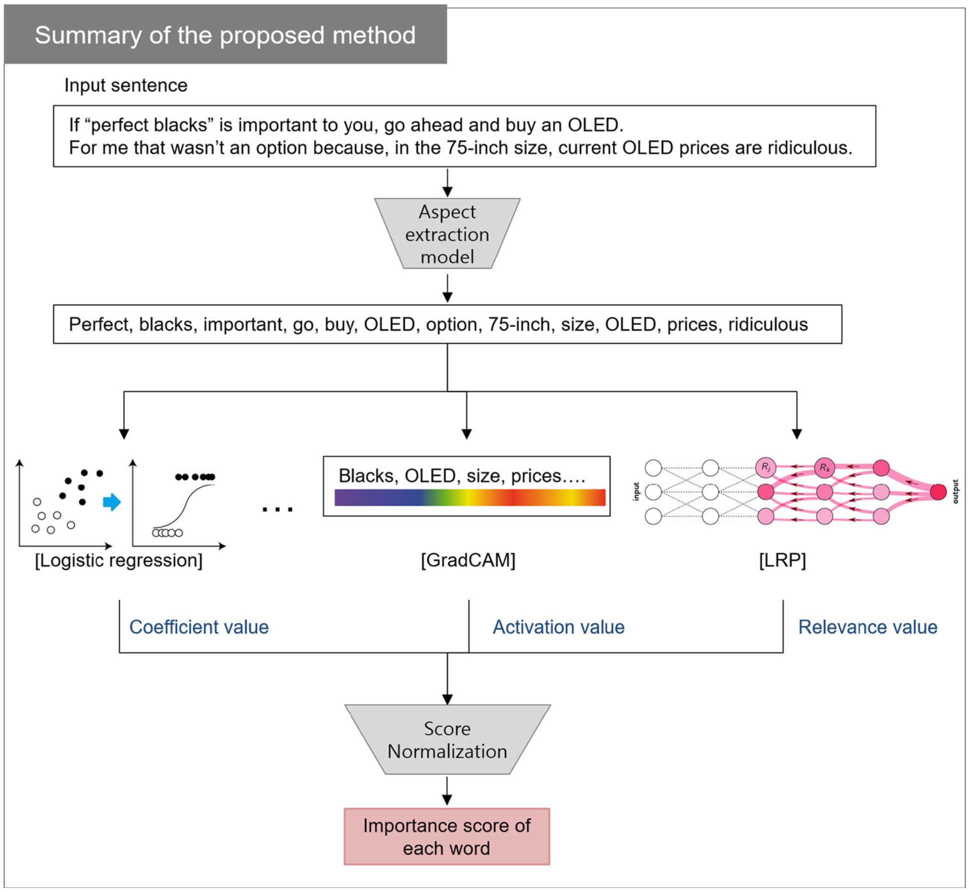

Fig. 1 Summary of our proposed method

model $f(x)$ around the instance of interest, we minimize the following equation:

$$\hat{g} = \text{argmin}_{g \in G} L(f, g, \pi_x) + \Omega(g)$$

where f is an original predictor and x is an original feature. g is an explanatory model that could be linear, decision tree, or falling rule lists. π is a proximity measure between an instance of z and x to define the locality around x.

Using the above equation, the importance of a word is calculated based on the extent to which the model's output value changes when that word is removed. Owing to the nature of LIME under which a considerably high importance score is assigned to most words, only words greater than 0.3 are used for the importance accumulation.

SHAP belongs to the class of models called additive feature attribution methods, where the explanation is expressed as a linear function of the features. Under linear regression, an increase in the independent variable of 1, the dependent variable by a coefficient. SHAP tries to build such a model for each data point. Instead of the original feature, SHAP replaces each feature (x_i) with the binary variable (zi) that represents whether x_i is present:

$$g(z) = \phi_0 + \sum_{i=1}^{M} \phi_i z_i = bias + \sum contribution\ of\ each\ word$$

where $g(z)$ is a local proxy model of the original model $f(x)$ and ϕ_i represents the extent to which the presence of a feature i contributes to the final output, which helps us interpret the original model.

The average of the marginal contributions of all possible combinations is regarded as the importance score for each word using this equation. Similar to LIME, only words greater than 0.3 are used for the importance accumulation because a reasonably high importance score is assigned to most words.

Table 2 summarizes the parameters used for LIME and SHAP.

3.2.3 Grad-CAM and LRP

Grad-CAM considers the gradient value in a convolution layer, which is calculated using the backpropagation of the

Table 2 Parameter settings for LIME and SHAP

Model	Initial model	Explainer	Feature selection	Feature independence
LIME	Logistic regression	LimeTextExplainer	Lasso path	–
SHAP	Logistic regression	LinearExplainer	–	Independent

CNN classification model, to be the importance score of each word. In detail, our proposed architecture is based on the fundamental assumption that the relative importance weight Y^c of a particular class c can be written as a linear combination of its last global average-pooled convolutional layer feature maps A^k, as in the following equation:

$$Y^c = \sum_k w_c^k \sum_i A_i^k$$

Moreover, we assume that class c is equal to $w_k^c = \partial Y^c / \partial A_i^k$ when i is the sequential location of a word in a sentence. However, this formulation makes the weights w_c^k independent of the positions i of a particular activation map A^k. We overcome this limitation by taking the global average pool of the partial derivatives, as in the following equation, which is the same as in the original Grad-CAM approach:

$$w_k^c = \sum_i \frac{\partial Y^c}{\partial A_i^k}$$

In this study, given the considerable noise in Grad-CAM, only the case in which the model output value is 0.7 or more, not the entire sentence, is reflected in the importance accumulation.

Under the LRP methodology, the contribution value of each node is calculated when the LSTM classification model is backpropagated, and this value is regarded as the importance of each word.

Assume that $g(x)$ is the model's prediction. We redistribute this prediction to each input aspect word, assigning the relevance score R_i to each input word i. The central idea of this relevance propagation is relevance conservation: $\sum_i R_i^{(1)} = \cdots = \sum_i R_j^{(i)} = \cdots = g(x)$, where l denotes a generic network layer. This implies that total relevance is conserved at each layer.

In essence, at each layer of the network, total relevance, which equals prediction $g(x)$, is conserved. The relevance score of each input variable determines that variable's contribution to the prediction. Consider a neuron in our artificial neural network. This maps a set of inputs, x_i, to an output, x_j, through a combination of weights, w_{ij}, and an activation function. Let us call it $h(.)$. Now, $x_j = h(\sum_i w_{ij} x_i)$.

The relevance assignment mechanism works by computing relevance R_i for neuron x_i (input) given all the relevances R_j of outputs x_j. Of the various formulas for this propagation, we use the following:

$$R_i^{(l)} = \sum_j \frac{x_i^{(l)} w_{ij}^{(l,l+1)}}{\sum_k x_k^{(l)} w_{kj}^{(l,l+1)} + \epsilon \times sign(\sum_k x_k^{(l)} w_{kj}^{(l,l+1)})} R_j^{(l+1)}$$

The layer-wise relevance propagation output is essentially the importance of the input words. Like Grad-CAM, only the case in which the model output value is 0.7 or more, not the entire sentence, is reflected in the importance accumulation.

Figure 2 summarizes the backbone networks of Grad-CAM and LRP.

Table 3 summarizes the parameters and structures of the CNN used for Grad-CAM and LSTM used for LRP.

3.2.4 Normalization and Refinement

We combine the various XAI methodologies to build a robust model for calculating the importance of competitive attributes. Since each XAI model has a different method of calculating the importance of competitive attributes, we normalize the importance score derived from each method in each review text by dividing by total importance so that the value is between 0 and 1. Further, the importance score derived from each XAI method is averaged and considered to be each word's final importance score. We then sort the attributes by order of importance score to find their relative importance. We can then easily select the important competitive attributes by sorting them.

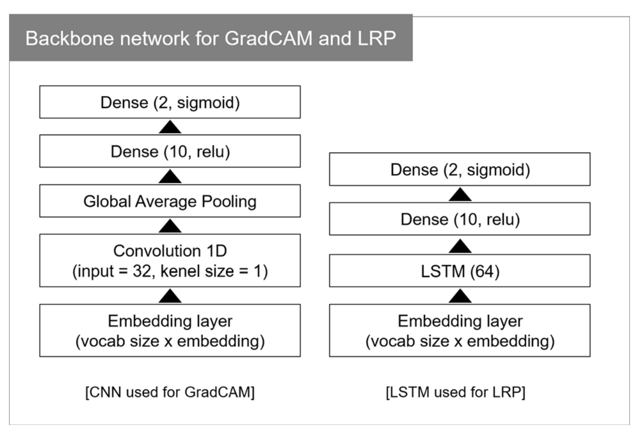

Fig. 2 Backbone networks for Grad-CAM and LRP

Table 3 Parameter settings for Grad-CAM and LRP

Model	Optimizer	Loss	Learning rate	Epoch	Batch size
CNN (used for Grad-CAM)	Adam	Binary cross-entropy	0.001	20	16
LSTM (used for LRP)	Adam	Binary cross-entropy	0.001	20	16

Furthermore, to improve performance, we make minor refinements. Our observations of the extracted attribution factors show different words with the same meaning, which occurs because customer review data comprise extremely unstructured text. Thus, we assign synonymous words representing an extracted attribution factor to the same cluster using a clustering technique.

To obtain the lowest silhouette index, we cluster the words based on the extracted factors' embedding vector calculated in the first step using the spherical *k*-means method (Zhong 2005). The cosine dissimilarity $1 - cos(x, y)$ is a distance measure used in the spherical *k*-means method. The proposed method is affected by the number of clusters. Aspects such as *operating system* and *Android* are assigned to the same cluster.

4 Experiments

4.1 Experimental Setup

To verify our proposed method's performance, we carry out two quantitative experiments, adopting a qualitative approach. The first experiment reviews the competitive analysis factor results using our approach and proves our proposed method's effectiveness qualitatively. Additionally, it compares the attributes of our extracted competitive factors with the results of a survey conducted by LG Electronics to identify real-world customers' product attributes as the most critical competitive and differentiation factors.

The survey results, used as an answer set, consist of product attributes ordered by the importance of the competitive and differentiation factors that customers consider to be the most significant. To compare the order of the product attributes in the answer set with that from the results of our proposed method, we measure the results using normalized discounted cumulative gain (NDCG), one of the best-known evaluation measures for ranking systems in information retrieval (Järvelin and Kekäläinen 2002, 2017). While most traditional ranking measures allow for only binary relevance, NDCG allows each retrieved result to achieve a graded relevance. Moreover, many other measures uniformly weigh all positions but associate a discount function with a rank (Wang et al. 2013).

Specifically, at a particular rank position *p*, the naive cumulative gain, which is the predecessor of the discounted cumulative gain and does not include the position of the result because of the usefulness of the result set, is defined as $CG_p = \sum_{i=1}^{p} rel_i$. Moreover, the discounted cumulative gain at a particular rank position *p* is defined as $DCG_p = rel_1 + \sum_{i=2}^{p} rel_i / log_2(i+1)$, indicating that the highly relevant documents appearing lower in a search result list should be penalized because the graded relevance value is logarithmically reduced in proportion to the position of the result.

The performance of a search engine cannot be consistently compared from one query to the next using the discounted cumulative gain alone; therefore, the cumulative gain for a chosen value of *p* should be normalized across the queries at each position by sorting all the relevant documents in the corpus by their relative relevance. Thus, we compute NDCG as $NDCG_p = DCG_p / IDCG_p$, where $IDCG_p = \sum_{i=1}^{|REL_p|} (2^{rel_i} - 1)/log_2(i+1)$ and REL_p represents the list of relevant documents in the text until position *p*.

We measure the NDCG value of the top-30 extracted attributes and compare the results with those of the other baselines, as presented in Table 4. Each method's details have already been described in the Sect. 2. Highly cited studies that present each methodology are selected as benchmarks.

We then assign a relevance weight on a scale from 1 to 10 to each of the three attributes based on a discussion with the domain experts at LG Electronics and reduce the weights from 1 to nearly 0 using a logarithm function. For instance, we assign the attributes in the answer set values of $[10, 10, 10, 9, 9, 9, 8, 8, \ldots, 2, 1, 1, 1]$, and reduce the weights to $[1.0, 1.0, 1.0, 0.6309, 0.6309, 0.6309, 0.5, 0.5, 0.5, 0.4307, \ldots]$. Notably, the NDCG value increases when the largest number is listed first.

In the second experiment, we compare the effectiveness of our proposed method with that of the other benchmarks. The experiment consists of two five-point Likert scale-based customer surveys of 30 participants on the following points: each attribute's influence in the attribute set on the participant's satisfaction with each product and their

Table 4 Baselines used for the first experiment

No.	Extraction method	Prioritization method	No.	Extraction method	Prioritization method
1	CNN-based	LRP	6	CNN-based	Logistic regression
2	CNN-based	Grad-CAM	7	Dlirec	LRP
3	CNN-based	LSTM attention	8	Dlirec	Grad-CAM
4	CNN-based	LIME	9	Hierarchical CRF	LRP
5	CNN-based	SHAP	10	Hierarchical CRF	Grad-CAM

satisfaction with the overall product. We then construct multiple regression satisfaction models with each attribute's comprehensive product and compare the coefficients of determination (R^2). The regression model demonstrates that the composition of the competitive factors is complete. To construct the regression model in the experiment, we set attribute numbers such as 10.

In particular, $\bar{y}$ provides the mean of the observed data: $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. The variability of the dataset can be measured with three sums-of-squares formulas: the total sum of squares (in proportion to the variance in the data), $SS_{tot} = \sum_i (y_i - \bar{y})^2$; regression sum of squares, also called the explained sum of squares, $SS_{reg} = \sum_i (f_i - \bar{y})^2$; and sum of squares of residuals, also called the residual sum of squares, $SS_{res} = \sum_i (y_i - f_i)^2$, where f represents a fitted value. Here, R^2 is calculated as $1 - SS_{res}/SS_{tot}$.

Additionally, we conduct ablation experiments on the various ensembles to check the performance difference between them. Threefold and fourfold ensembles are randomly selected. The experiment is then conducted and the results are compared. Finally, the various data sources are compared to verify the robust performance of the ensemble.

4.2 Data Description

In the experiments, we reviewed data from LG and Samsung Electronics, which are representative competitors in the home appliances industry. We collected data on three types of home appliances: refrigerators, laundry appliances, (e.g., washing machines and tumble dryers), and air-care appliances (e.g., air conditioners, air purifiers, and vacuum cleaners).

Data were collected using a crawler module developed by LG Electronics, and five data sources were used: BestBuy, Amazon, CNET, The Verge, and Engadget. We also collected review data from YouTube and social networks searched using the model name of each product.

The review dataset consisted of 80,000 reviews (refrigerators: 40,000, laundry appliances: 25,000, air-care appliances: 15,000) from July 1, 2019 to October 31, 2020, with the aspect keywords of the dataset labeled by domain experts from LG Electronics. The survey data consisted of product attributes ordered by importance; these are considered to be the most differentiated competitive factors by customers. LG Electronics surveys each type of product periodically.

4.3 Experimental Results

4.3.1 *Qualitative Review of the Extracted Competitive Factors*

Table 5 shows the top-10 competitive attributes of LG and Samsung for each product derived using our proposed method. Customers often perceive the brand name emphasized by the manufacturer (e.g., Twinwash, ThinkQ) as an essential competitive factor. However, in some cases, the factors emphasized by the manufacturer are not recognized as vital competitive factors by customers (e.g., AutoFill in a Samsung refrigerator, Insta-view in an LG refrigerator). This highlights one difference between the core competitive factors perceived by manufacturers and customers.

Furthermore, although there were differences by product, LG home appliances had strengths in performance and Samsung home appliances were strong in design and Internet of things (IoT). This result proves our proposed method's effectiveness because it agrees with the heuristic analysis results judged by the domain experts.

In the case of Samsung refrigerators, the key competitive and differentiation factors extracted were "Bespoke" and keywords related to "Interior." We can infer that customers recognize the harmony of a bespoke design and interior as the greatest strength of Samsung refrigerators. Other key differentiation factors were "3-door," "Flex," and "Optimal." This finding indicates that customers also recognize space efficiency as a competitive factor of Samsung refrigerators.

For LG refrigerators, customers recognized "Door-in-door" and "Transparency" as competitive factors. The competitiveness of an LG refrigerator is therefore considered to be the design concept with a transparent inside. Another group of competitive factors was "ThinkQ" and "IoT." This indicates that IoT connection, generally recognized as an advantage of Samsung home appliances, is

Table 5 Top-10 competitive attributes based on our proposed method

Product	Extracted competitive attribute
Samsung refrigerators	Bespoke, harmony, interior, design, 3-door, French door, water, flex, capacity, optimal
LG refrigerators	Door-in-door, transparency, smart, Wi-Fi, ThinkQ, wine, performance, IoT, connection, energy
Samsung laundry appliances	AddWash, smart, super, speed, sensor, AI, IoT, control, vibration, sanitize
LG laundry appliances	Twinwash, combo, SideKick, design, look, compact, steam, turbo, 6Motion, technology
Samsung air-care appliances	Windfree, comfortable, fast, quiet, easy, cyclone, IoT, control, connection, application
LG air-care appliances	ThinkQ, Puri-care, fresh, inverter, dual, powerful, energy, efficiency, performance, maximum

Table 6 Performance of the competitive factor extraction (NDCG)

Method	Samsung refrigerators	LG refrigerators	Samsung laundry appliances	LG laundry appliances	Samsung air-care appliances	LG air-care appliances
Proposed method	**0.9403** (±.0135)	**0.9435** (±.0148)	**0.9486** (±.0184)	**0.9452** (±.0157)	**0.9545** (±.0141)	**0.9411** (±.0157)
CNN & LRP	0.9039	0.9103	0.9138	0.9052	0.9042	0.9112
CNN & Grad-CAM	9035	0.9052	0.9005	0.8967	0.9005	0.8983
CNN & LSTM attention	0.8834	0.8947	0.8903	0.8836	0.8920	0.8868
CNN & LIME	0.8516	0.8553	0.8604	0.8512	0.8530	0.8616
CNN & SHAP	0.8439	0.8319	0.8409	0.8303	0.8401	0.8448
CNN & logistic regression	0.7809	0.7858	0.7906	0.7949	0.7863	0.7892
Dlirec & LRP	0.8430	0.8363	0.8318	0.8438	0.8322	0.8365
Dlirec & Grad-CAM	0.8420	0.8398	0.8090	0.8017	0.8049	0.7948
Hierarchical CRF & LRP	0.7908	0.7955	0.8029	0.8009	0.7948	0.8046
Hierarchical CRF & Grad-CAM	0.7925	0.7916	0.7988	0.8049	0.7945	0.8037

Bold font indicates the highest accuracy

recognized as a competitive attribute only for LG refrigerators.

In the case of Samsung laundry appliances, the vital competitive attributes extracted were “AddWash” and related keywords. The ability to add extra clothes while washing is therefore the most significant competitive factor for Samsung laundry appliances. Meanwhile, for LG laundry appliances, the key differentiation factors extracted were “Twinwash” and “Combo.” This shows that the LG washer/dryer combination is receiving considerable attention from customers. Further, “6motion” and “Turbo” related to performance were also important competitive factors perceived by customers.

For Samsung air-care appliances, the competitive factors were “Windfree” and related keywords. The windless air conditioner of Samsung was also evaluated as a crucial competitive factor by customers. By contrast, keywords related to efficiency and technology were recognized as the critical competitive factors of LG air-care appliances.

4.3.2 *Comparison of the Competitive Factors Extracted Using Our Method with Those Perceived by Customers*

Table 6 lists the NDCG results of comparing the extracted product attributes with the customer responses in the survey by LG Electronics. As mentioned previously, we test our proposed method’s performance and that of a baseline approaches using LG and Samsung home appliances. All the results are the average values of the fivefold validation.

Although the NDCG results for each product differ, our proposed method outperforms all the other methods. As for the prioritization method applied, our proposed ensemble of XAI methods yields better results than those obtained using other XAI approaches such as LRP, Grad-CAM, and

Table 7 Results of the effectiveness comparison (Samsung)

Method	Average influence score			R^2		
	Refrigerators	Laundry appliances	Air-care appliances	Refrigerators	Laundry appliances	Air-care appliances
Proposed method	**4.45**	**4.37**	**4.41**	**0.5516**	**0.5475**	**0.5528**
CNN & LRP	4.23	4.18	4.19	0.5149	0.5204	0.5137
CNN & Grad-CAM	4.16	4.18	4.13	0.5057	0.5142	0.5084
CNN & LSTM attention	3.98	3.78	3.85	0.4945	0.4848	0.4834
CNN & LIME	3.28	3.34	3.30	0.4664	0.4627	0.4544
CNN & SHAP	3.19	3.24	3.29	0.4579	0.4476	0.4518
CNN & logistic regression	3.04	3.12	3.08	0.4113	0.4073	0.4136
Dlirec & LRP	3.46	3.47	3.34	0.4418	0.4516	0.4447
Dlirec & Grad-CAM	3.31	3.41	3.35	0.4332	0.4410	0.4464
Hierarchical CRF & LRP	3.23	3.19	3.15	0.4274	0.4233	0.4212
Hierarchical CRF & Grad-CAM	3.16	3.08	3.12	0.4015	0.4167	0.4074

Bold font indicates the highest accuracy

Table 8 Results of the effectiveness comparison (LG)

Method	Average influence score			R^2		
	Refrigerators	Laundry appliances	Air-care appliances	Refrigerators	Laundry appliances	Air-care appliances
Proposed method	**4.51**	**4.44**	**4.49**	**0.5627**	**0.5557**	**0.5603**
CNN & LRP	4.29	4.21	4.23	0.5224	0.5298	0.5207
CNN & Grad-CAM	4.16	4.18	4.13	0.5114	0.5167	0.5203
CNN & LSTM attention	4.05	3.84	3.92	0.5004	0.4945	0.4881
CNN & LIME	3.31	3.39	3.33	0.4681	0.4655	0.4593
CNN & SHAP	3.22	3.25	3.24	0.4587	0.4551	0.4571
CNN & logistic regression	3.11	3.15	3.09	0.4204	0.4114	0.4207
Dlirec & LRP	3.42	3.42	3.39	0.4472	0.4507	0.4498
Dlirec & Grad-CAM	3.32	3.39	3.31	0.4400	0.4375	0.4415
Hierarchical CRF & LRP	3.27	3.21	3.20	0.4217	0.4245	0.4205
Hierarchical CRF & Grad-CAM	3.12	3.11	3.17	0.4077	0.4098	0.4115

Bold font indicates the highest accuracy

LIME. This proves that our proposed method can derive a similar result to the primary competitive factors frequently derived by field experts using heuristic techniques.

An analysis of the detailed results shows that the LRP and Grad-CAM-based method is inferior than our proposed method but performs much better than our methods. Thus, we can conclude that LRP and Grad-CAM are considerably more effective XAI techniques than the other algorithms.

Moreover, the LSTM attention approach shows inconsistent weight calculations based on each textual review's length, leading to a bias in the overall weight calculation. Moreover, the LIME and logistic regression approaches are more appropriate for the binary decision-making of each aspect considered than for calculating the importance weight. Furthermore, the naive CNN-based approach outperforms the other methods (e.g., the CRF and Dlirec approaches) in terms of extraction, as verified by Poria et al. (2016).

4.3.3 *Influence of Competitive Factors on Customer Satisfaction Using Our Method*

Tables 7 and 8 show that our proposed method has a higher importance score and larger coefficients of determination

Table 9 Ablation results of the ensembles (NDCG)

Ensemble	Samsung refrigerators	LG refrigerators	Samsung laundry appliances	LG laundry appliances	Samsung air-care appliances	LG air-care appliances
Grad-CAM + SHAP + LRP + LIME + logistic regression	**0.9403**	**0.9435**	**0.9486**	**0.9452**	**0.9545**	**0.9411**
Grad-CAM + SHAP + LRP + LIME	0.9389	0.9394	0.9361	0.9345	0.9402	0.9364
Grad-CAM + LRP + LIME + logistic regression	9376	0.9407	0.9391	0.9401	0.9374	0.9371
Grad-CAM + SHAP + LIME + logistic regression	0.9344	0.9367	0.9375	0.9378	0.9341	0.9368
SHAP + LRP + LIME + logistic regression	0.9334	0.9327	0.9302	0.9328	0.9304	0.9316
Grad-CAM + SHAP + logistic regression	0.9214	0.9207	0.9227	0.9206	0.9211	0.9217
SHAP + LRP + LIME	0.9203	0.9201	0.9196	0.9189	0.9203	0.9208
Grad-CAM + LRP + LIME	0.9188	0.9175	0.9168	0.9185	0.9173	0.9166
LRP + LIME + logistic regression	0.9128	0.9144	0.9137	0.9114	0.9146	0.9135

Bold font indicates the highest accuracy

Table 10 Performance differences by data source (Samsung refrigerators)

Data source	Amazon	Engadget	YouTube	Social networks
Proposed method	**0.9411**	**0.9408**	**0.9397**	**0.9345**
CNN & LRP	0.9101	0.9114	0.8916	0.8883
CNN & Grad-CAM	9114	0.9084	0.8904	0.8847
CNN & LSTM attention	0.8964	0.8912	0.9014	0.9019
CNN & LIME	0.8427	0.8433	0.8278	0.8116
CNN & SHAP	0.8401	0.8356	0.8337	0.8278
CNN & logistic regression	0.7759	0.7679	0.7437	0.7416

Bold font indicates the highest accuracy

than the values for each of the corresponding numbers of attributes. Thus, it effectively reflects customers' interests and identifies the essential elements influencing their purchase intention.

4.3.4 Ablation Experiments

Table 9 presents the results of the ablation experiments for the various ensembles. Overall, although the fivefold combination performs the best, there are no significant differences between the fourfold combinations. Moreover, the threefold combinations show worse overall performance than the fourfold combinations. In terms of the individual algorithms, the overall performance of the combinations including LRP and Grad-CAM is strong.

4.3.5 Results by Data Source

Tables 10 and 11 describe the differences in the results according to the data source used. As shown from the experimental results, our proposed ensemble methodology performs the best for all the data sources. Conversely, LRP and Grad-CAM perform worse than the benchmark LSTM attention model for the data collected from YouTube and social networks. Additionally, the performance of LIME and SHAP changes depending on the data source. The experimental results highlight that our proposed methodology performs better than those approaches based on an individual XAI method and shows robust performance regardless of the data source.

5 Conclusion

This study proposes an advanced methodology for extracting competitive factors using an XAI ensemble. Based on the assumption that keywords, which significantly influence the classification decision, are considered to be a meaningful point of differentiation by customers, we construct a model to classify the reviewer comments for

Table 11 Performance differences by data source (LG laundry appliances)

Data source	Amazon	Engadget	YouTube	Social networks
Proposed method	**0.9446**	**0.9477**	**0.9401**	**0.9386**
CNN & LRP	0.9287	0.9267	0.9012	0.9004
CNN & Grad-CAM	9124	0.9143	0.8824	0.8967
CNN & LSTM attention	0.8924	0.8907	0.9037	0.9054
CNN & LIME	0.8473	0.8441	0.8279	0.8254
CNN & SHAP	0.8454	0.8384	0.8186	0.8204
CNN & logistic regression	0.7746	0.7824	0.7615	0.7678

Bold font indicates the highest accuracy

each competitive product and classify each keyword's importance in the reviewer comments. Further, we propose an additional ensemble methodology to maintain robust performance despite differences in data sources.

The qualitative experimental results found in this study demonstrate that our proposed method can extract the competitive factors of a product effectively. The method also achieved higher NDCG values and higher influence scores than the other methods in the quantitative experiments. In particular, the proposed ensemble technique showed the highest and most robust performance for all the data sources. This proves that it can effectively and robustly extract competitive attributes from customer review data quantitatively.

Future studies could extend the scope of XAI to extract competitive factors from customer reviews. Such studies may involve sentiment analysis because the competitive factors would be advantageous to certain products over others. Studies could also extract the competitive factors of various products and services in highly competitive markets. Such approaches might then increase the application of our proposed method to various tasks in the real-world business environment.

Acknowledgements This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2020R1F1A1067914).

References

Alharbi FR, Khan MB (2019) Identifying comparative opinions in Arabic text in social media using machine learning techniques. SN Appl Sci 1(3):213

Alobaidi MH, Chebana F, Meguid MA (2018) Robust ensemble learning framework for day-ahead forecasting of household based energy consumption. Appl Engery 212:997–1012

Archak N, Ghose A, Ipeirotis PG (2011) Deriving the pricing power of product features by mining consumer reviews. Manag Sci 57(8):1485–1509

Bi JW, Liu Y, Fan ZP, Zhang J (2019) Wisdom of crowds: conducting importance-performance analysis (IPA) through online reviews. Tour Manag 70:460–478

Boniface O (2017) Using customer–product competitor analysis as drivers for a business reconfiguration and market re-positioning. J Contemp Manag 14(1):385–415

Chakraborti S, Dey S (2019) Analysis of competitor intelligence in the era of big data: an integrated system using text summarization based on global optimization. Bus Inf Syst Eng 61(3):345–355

Davcik NS, Sharma P (2016) Marketing resources, performance, and competitive advantage: a review and future research directions. J Bus Res 69(12):5547–5552

Fan S, Lau RY, Zhao JL (2015) Demystifying big data analytics for business intelligence through the lens of marketing mix. Big Data Res 2(1):28–32

Gao S, Tang O, Wang H, Yin P (2018) Identifying competitors through comparative relation mining of online reviews in the restaurant industry. Int J Hosp Manag 71:19–32

Guo L, Sharma R, Yin L, Lu R, Rong K (2017) Automated competitor analysis using big data analytics: evidence from the fitness mobile app business. Bus Process Manag J 23(3):735–762

Gur FA, Greckhamer T (2019) Know thy enemy: a review and agenda for research on competitor identification. J Manag 45(5):2072–2100

Han J, Lee Y (2021) Explainable artificial intelligence-based competitive factor identification. ACM Trans Knowl Discov Data 16(1):1–11

Hatzijordanou N, Bohn N, Terzidis O (2019) A systematic literature review on competitor analysis: status quo and start-up specifics. Manag Rev Q 69(4):415–458

Hu C, Youn BD, Wang P, Yoon JT (2012) Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. Reliab Eng Syst Saf 103:120–135

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

International Workshop on Semantic Evaluation (2014) Semeval-2014 dataset. https://alt.qcri.org/semeval2014. Accessed 23 Aug 2014

Järvelin K, Kekäläinen J (2002) Cumulated gain-based evaluation of IR techniques. ACM Trans Inf Syst (TOIS) 20(4):422–446

Järvelin K, Kekäläinen J (2017) IR evaluation methods for retrieving highly relevant documents, vol 51. ACM SIGIR Forum, ACM, New York, pp 243–250

Jin J, Ji P, Gu R (2016) Identifying comparative customer requirements from product online reviews for competitor analysis. Eng Appl Artif Intell 49:61–73

Jin J, Ji P, Yan S (2019) Comparison of series products from customer online concerns for competitive intelligence. J Ambient Intell Hum Comput 10(3):937–952

Kim SG, Kang J (2018) Analyzing the discriminative attributes of products using text mining focused on cosmetic reviews. Inf Process Manag 54(6):938–957

Lee Y (2021) Extraction of competitive factors in a competitor analysis using an explainable neural network. Neural Process Lett 53:1979–1994

Lee J, Chung SY (2019) Robust training with ensemble consensus. arXiv preprint arXiv:191009792

Lee M, Lee S (2017) Identifying new business opportunities from competitor intelligence: an integrated use of patent and trademark databases. Technol Forecast Soc Change 119:170–183

Lundberg S, Lee SI (2017) A unified approach to interpreting model predictions. arXiv preprint arXiv:170507874

Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. arXiv preprint arXiv: 13013781

Mudambi SM, Schuff D (2010) Research note: what makes a helpful online review? a study of customer reviews on Amazon. com. MIS Q 34(1):185–200

Paul MJ, Zhai C, Girju R (2010) Summarizing contrastive viewpoints in opinionated text. In: Proceedings of the 2010 conference on empirical methods in natural language processing. Association for Computational Linguistics, pp 66–76

Poria S, Cambria E, Ku LW, 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

Poria S, Cambria E, Gelbukh A (2016) Aspect extraction for opinion mining with a deep convolutional neural network. Knowl Based Syst 108:42–49

Qiu G, Liu B, Bu J, Chen C (2011) Opinion word expansion and target extraction through double propagation. Comput Linguist 37(1):9–27

Raut RR, Mante RV, Chatur P (2018) Competitor mining and identification from large unstructured datasets. In: 2018 3rd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT). IEEE, pp 713–717

Ribeiro MT, Singh S, Guestrin C (2016) Why should I trust you?: Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 1135–1144

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

Varathan KD, Giachanou A, Crestani F (2017) Comparative opinion mining: a review. J Assoc Inf Sci Technol 68(4):811–829

Wang Y, Wang L, Li Y, He D, Liu TY (2013) A theoretical analysis of NDCG type ranking measures. In: Conference on learning theory, pp 25–54

Wang Y, Huang M, Zhao L et al (2016) Attention-based LSTM for aspect-level sentiment classification. In: Proceedings of the 2016 conference on empirical methods in natural language processing, pp 606–615

Wang H, Gao S, Yin P, Liu JNK (2017) Competitiveness analysis through comparative relation mining: evidence from restaurants' online reviews. Ind Manag Data Syst 117(4):672–687

Wei P, Ball JE, Anderson DT (2018) Fusion of an ensemble of augmented image detectors for robust object detection. Sensors 18(3):894

Yan S, Jin J, Ji P, Geng Z (2017) Extracting customer concerns from online reviews of series products for competitor analysis. In: iConference 2017 proceedings

Zameer A, Arshad J, Khan A, Raja MAZ (2017) Intelligent and robust prediction of short term wind power using genetic programming based ensemble of neural networks. Energy Convers Manag 134:361–372

Zhong S (2005) Efficient online spherical k-means clustering. In: Proceedings. 2005 IEEE international joint conference on neural networks, 2005, vol 5. IEEE, pp 3180–3185

Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A (2016) Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2921–2929