International Journal of Engineering & Technology, 7 (3.33) (2018) 179-182

SPC

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



Deep Neural Network for Product Classification System with Korean Character Image

Jae-Kyung Sung[1], Sang-Min Park[2], Sang-Yun Sin[3], Yung Bok Kim[4], Yong-Guk Kim[5]*

[1]*Research Professor. Dept. of Computer Engineering, Sejong University*
[2, 3] *Graduate Student. Dept. of Computer Engineering, Sejong University*
[4, 5]*Professor. Dept. of Computer Engineering, Sejong University,*
**Corresponding author E-mail: ykim@sejong.ac.kr*

Abstract

This paper proposes a product classification system based on deep learning using Korean character images (Hangul) to search for products in the shopping mall. Generally, an online shopping mall customer searches through a category classification or a product name to purchase a product. When the exact product name or category is not clear, the user has to search its name. However, the product image classification is degraded because the product logos and characters in the package often interfere. To solve such problems, we propose a classification system based on Deep Learning using Korean character images. The learning data of this system uses Korean character images of PHD08, a Hangul (Korean-language) database. The experimental is carried out using product names collected on the web. For the performance experiment, 10 categories of online shopping mall are selected and the classification accuracy is measured and compared with the previous systems.

Keywords: *Deep Learning, Product Image, Korean Character image, Shopping Mall, Product Classification*

1. Introduction

The online shoppers search their product using either category or product name. However, if he does not know the exact category of the product, it is difficult to search the product. If product category classification system in shopping mall and purchasing user classification system are different, product search can be also difficult. When the category search is difficult in this way, the character search method is generally used centering on the product name. However, character search is difficult to search if you do not know the exact name of the product. If the purchasing products come from overseas shopping malls, users who lack the language ability have difficulty in category classification. Therefore, when the search is difficult in this way, the search using the product image is possible. For example, if the product information can be retrieved by an image taken on a smart phone, it provides users with speed and convenience [1]. This image search has been provided by many companies through the recent AI (Artificial Intelligence) technology. The product classification service is based on product image search technology. However, category classification through product image is an obstacle to the logos or text included in the product packaging image. In fact, most of the images of products appearing in online shopping are not unique products, but rather images that are packaging products. The packaging image consists of various text messages describing the product and images related to the product. As a result, there is a performance degradation in product classification based on AI technology, Deep Learning. However, online product classification should be categorized based on the design image of the product. The product packaging image generally includes images related to the product, logos and letters of the company that manufactured the product, and the like. Among them, the product text is composed of a character image which is made in-to a graphic and a product explaining the contents of the product.

In this paper, we propose a classification technique based on deep learning by selecting 10 product categories among the words describing the product name and contents in the packaged image. The 10 product categories are tested for classification performance using Korean character image data such as coffee, ramen, milk, soy milk, cookie, toothpaste, soap, soup, apple, grape, among other products sold in online shopping malls.

2. Related research

This chapter introduces the status of image-based product search service and image search technology that companies are servicing.

2.1. Image based product search service

Image search can be easily combined with smart phone camera to find the information you want. So by adding machine learning technology to AI, many companies are providing various image search services. Image-based automatic retrieval is providing services for shopping-related image search in global companies such as Google, Amazon and Taobao (Alibaba), and is also in service in Naver, 11th Street and KakaoTalk in Korea. Google's AI image search 'Google Lens' provides information on menus, food quality, etc., when it is seen in a restaurant, and can be settled, translated, and linked with Google Photos. Amazon has been providing image search services since 2014, and it offers a service called Spark, which can be called an image-based shopping SNS. It is a service that works with the products that are sold in Amazon so that you can purchase the goods in the images uploaded to SNS immediate-

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ly. Naver also officially released 'Shopping Lens' through mobile app updates. It is currently providing 'shopping lenses' by adding Deep Learning technology to its existing image search service 'Smart Lenses'. The 'shopping lens' is a photograph of a product, and the related image is retrieved and the search result is provided. It applies to all product categories and works with existing services such as Naver Shopping and Naver Pay. SK Planet, which operates 11th Street, launched the image search shopping service using deep-learning technology at the end of 2015 and introduced first in Korea. The 11th Street offers image search shopping service on clothing, shoes, and bags with a high proportion of various categories [2]. The Kakao started AI image search service as 'flower search' which searches flowers after mobile. In addition, long tap image search function that can be used immediately in KakaoTalk is added. If you press and hold the image received from KakaoTalk, it shows the similar product image. The Kakao plans to develop it as a function to find similar products in connection with shopping sites in the future [3].

2.2. Image Retrieval Technology

Deep Learning drew attention when the CNN model, which mimics the human brain, won the 2012 ImageNet Challenge, which is one of the international image recognition competitions. Since the competition, CNN Deep Learning models have changed the flow of image related research in various image recognition areas. This deep learning based image recognition study also contributed to the object recognition research. R-CNN (Regions with CNN features) implements the object recognition model using the deep learning method by adding the location detection function to the existing deep learning based image recognition model [4]. Before 2013, when deep study research becomes active, image based product search function using CRBM (Convolutional Restricted Boltzmann Machines) and SIFT (Scale Invariant Feature Transform) technology were developed [5]. The SIFT algorithm [6, 7] extracts feature points and uses image matching to search for images by user interaction [8].

Image-based search can be used to search for similar items in fashion product search, and is used actively in related products such as socks and accessories to suit the style when the user selects clothing or shoes [9].

3. Korean character images classification system

This paper proposes a deep learning based classification system using product images. The product image data is based on the product image of Korean character images and product image of 10 product categories in the online shopping mall mentioned above. In general, Korean character images are text files in product descriptions, but there are many graphic characters that are designed in product titles. Such graphical characters cause performance degradation in various design shapes during classification experiment by the deep learning technique. However, since the shape of the same product name character is basically similar, accuracy can be improved by learning a lot of data. The product image uses the image data provided in the online shopping mall catalogue. This chapter examines the process of generating experimental data and how to implement the Korean character image classification system.

3.1. Experimental data

The Korean character images convert the PHD08 data, which is a Korean data-base, into a binary image and uses it as learning data. PHD08 is a database in which Korean characters are generated under various conditions and scanned images are stored in binary form after printing. There are 2,187 samples per character for 2,350 Korean characters, each of which has a different font, character size, scan resolution, noise level, rotation, and angle [06]. Currently, the databases for Hangul (Korean language) recognition are divided into handwritten and printed ones according to recognition purpose. The conventional PHD08 input image with various sizes is changed to the same size of 64x128 and used for the experiment. The 10 classification categories (coffee, ramen, milk, soy milk, cookie, tooth-paste, soap, soup, apple, grape) used in the experiment consist of two Korean characters. An example of the Korean input learning data used in the experiment is shown in [Fig. 1].



Fig. 1: Example of Input training data transformed from PHD08.

In this paper, the number of learning data used constitutes 2,187 learning data sets for each 10 product categories. The experimental data consists of 10 categories of letters, each of which consists of 100 data. And each category experiment data consists of product name collected from the web and font data not included in the PHD08 database. [Fig. 2]



Fig. 2: The 10 categories Korean character experiment data (left: font data not in PHD08, right: product image of online shopping mall)

It is Increased the quantity of learning data through augmentation to improve learning performance. In the type of augmentation used, Flip (top, bottom, left and right), Rotation, and Crop are not used because of the distortion of character behavior and cause accuracy degradation. In this experiment, about 218,700 pieces of learning data are used, which is 10 times more than the original data 21,870 pieces due to blur, sharpening, noise addition and change of letters and background color. The 10,000 experimental images are used in the same way. The product image learning data extracts 14,936 images related to food from shopping mall. Among them, 6,000 product images are extracted in the category of 13 relatively good quality data (Fruits, Vegetables, Fresh Meat, Fish, Noodles, Drink, Tea, Grain, Eggs, Cigarettes, Milk, Ice Cream and Sauce). As in the case of Korean character images, there is not enough learning data for each category, so the number of learning data is increased through augmentation. The types of

augmentation were randomly processed, including Flip, Rotation, Crop, Blur, Sharpening and Illumination changes. Then, it is experimented with about 90,000 learning data, which is 15 times more than the original data, about 6,000.

3.2. Implementation of Korean character images classification system

Among the deep learning techniques of this system, the structure is simplified so that it can be operated as fast as possible by using CNN (Convolutional Neural Network). CNN is the most popular network model for deep learning for image recognition [10]. Currently, domestic and foreign IT companies, such as Google, Facebook, Amazon, Naver, etc., provide automatic classification and content-based image search. CNN is located at the front end of the network structure and consists of a part extracting the characteristics of the input data image and a part classifying the image into the target class. CNN computes each pixel association of image data. After modelling only a part of the pattern using a small amount of computationally small filters, the number of pixels of the data is reduced and the entire region is modelled in common. And the convolution filters apply to different locations with the same weight value. To extract image features, the images use local features to gather the whole pattern [11, 12]. Feature extraction gradually reduces the size of the image by selecting the input values of the pixels as the maximum value or the average value. Such an arithmetic operation results in that a convolution filter of the same size processes a relatively large area as it goes up to the upper layer, and the contour feature of the lower layer is learned by the shape reflected from the upper layer. As a result, as the hierarchy rises, it gradually recognizes shapes and objects with higher abstraction levels [11, 12]. The proposed system is a simplified network of AlexNet [13], which is the most widely used network in CNN. It consists of 3 Convolutional layers, 2 fully connected layers, and 10 Output layers to distinguish categories. The input image was 64x128 in size and was smaller than the 227x227 used in the previous ImageNet. The reason for reducing the structure of the network and reducing the input image is as follows. The proposed system is a method for enabling service on a GPU based general performance PC for deep learning. All Layer output except Output and Pooling Layer used ReLU Activation and the output of two Fully Connected layers was prevented from being Overfitting by using 0.3 ratio Drop Out. Classification must be performed using a small number of layers and kernels, so Overfitting occurs frequently in fast Epoch.

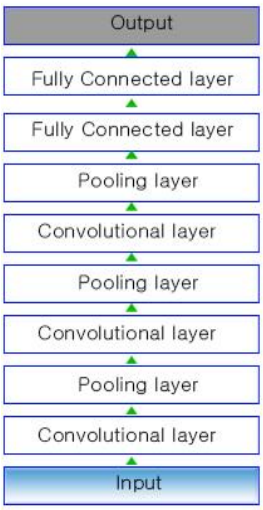

Fig. 3: Network structure used in category decision system

This paper corrects Drop Out, Learning parameter and hyper parameter to prevent Overfitting. Fig. 3 is the network used by the system. CNN's research has increased the layer of the network exponentially and improved performance. Recently, ResNet has developed existing performance using 152 layer. However, the proposed system is a version that can be used in a low computing environment, the learning is done using the GPU, but the execution is configured as shown in Fig. 3.

A total of 150 epochs were studied. Fig. 4 is the train loss, validation loss, and validation accuracy up to 50 epochs during the learning process. As shown in Fig. 4, the Hyper Parameter was modified to increase performance very quickly and prevent overfitting.

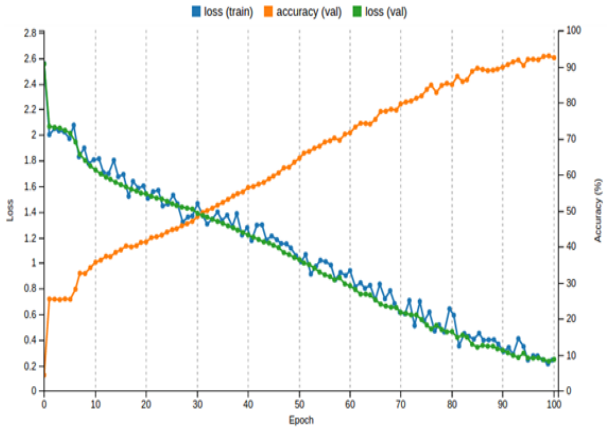

Fig. 4: The Training Result of Train Loss, Validation Loss and Validation accuracy

4. Experiments and Results

In this chapter, we briefly examine the experimental environment and method, and compare the results of Korean character images and product image classification experiment through image based classification system.

4.1. Experimental environment and method

The experimental environment for this system implementation is learning using Intel Core i7 CPU 4.40GHz, 16.0GB RAM and NVIDIA GeForce GTX 1080 video card.

This experiment quantifies the results through two experiments using Korean character image and product image data through the same classification system. Korean character images classification test data uses about 200,000 learning data in 10 categories (coffee, ramen, milk, soy milk, cookie, toothpaste, soap, soup, apple, grape). And the accuracy is quantified through experimental data that is about 5% of the learning data.

The product image classification test data is divided into 13 pieces (Fruits, Vegetables, Fresh Meat, Fish, Noodles, Drink, Tea, Grain, Eggs, Cigarettes, Milk, Ice Cream and Sauce) and the accuracy is quantified through experimental data equivalent to 10% of the learning data. And the accuracy is quantified through the experimental data corresponding to 10% of the learning data [Fig. 5].

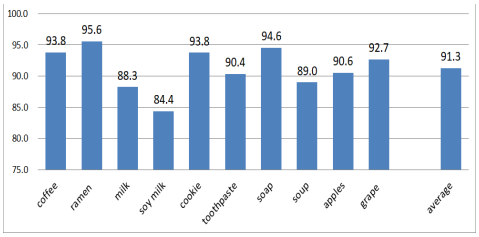

Fig. 5: The accuracy of product classification using Korean character image

The reason why the category classification and number of data used in Korean character image and product image experiment are

different is the result considering convenience and suitability of data collection for each category.

4.2. Experiment result

The accuracy test was conducted on 10 categories of Korean characters using experimental data [Fig. 6].

The average accuracy was 91.3%, of which more than 90% were tested in seven groups: coffee, instant noodles, cookie, toothpaste, soap, soup, apple, grape. And the group below the overall average was milk and soy milk. Overall, milk and soy milk performed relatively poorly, with more than 90 percent accuracy. This may be the reason why milk and soy milk are similar to the Korean character image, but there is no big difference.

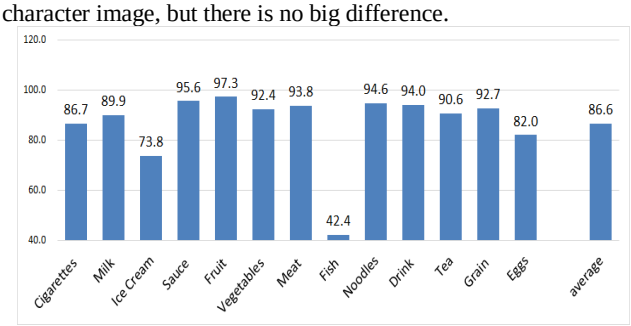

Fig. 6: The accuracy of product classification using product image

Among them, more than 90% of the group tested seven out of 13 groups, including sauces, fruits, vegetables, meat, noodles, beverages, teas and grains. And the group below the overall average was ice cream, fish and eggs. Overall, it showed 85% accuracy but very low performance in fish. Through these two experiments, the experiment using the Korean character image was about 5% higher than the experiment using the product image. Of course, it is problematic to evaluate the performance with only the accuracy average value in the state where the number of categories and data of the two experiments is different. But the ultimate goal of this experiment is to improve the category classification performance of the products. It is confirmed that the character image data can be used for higher performance. Therefore, if Korean character image classification technology is added to the classification system using the existing product image, the performance can be improved. In order to commercialize this system, Korean character image data of a variety of fonts is required. And it is necessary to study the category definition for systematic classification of products.

However, the definition of product categories is not an easy problem, and obtaining large amounts of image data for each category is a research task in the future because it takes time.

5. Conclusion

In this paper, we compare the accuracy of Korean character image and product image classification test through the same classification system. The Korean character image data used in this system can be applied to characters of other languages. Although practicality is confirmed through experimental results, many character and product image data are needed for commercialization of individual product search. Deep learning based product automatic search service using images can be utilized variously in a mobile environment if quantitative enough individual product image data is secured.

In the future, we will improve the performance of individual product image retrieval system and propose various classification systems through online and offline links.

Acknowledgement

This research was supported by the Industrial Core Technology Development Project (Knowledge Service) from KEIT[2016-Knowledge Service-General- Designa-tion-13, Project Number: 10067784]

References

[1] Sung-Ho Cho, Youn-Joon Lee(2016), A Comparative Study on Design of User Interface for O2O (Online to Offline) Coupon Application Service, Master's Thesis, p.15-80, Hongik University
[2] http://www.elec4.co.kr/article/articleView.asp?idx=18816
[3] http://www.itworld.co.kr/news/87229#csidx7dd4d034afb87a3a735ca646a16946d
[4] Kitae Kim, Wonseok Oh, Geunwon Lim, Eunwoo Cha, Minyoung Shin, Jongwoo Kim(2018), The way to make training data for deep learning model to recognize keywords in product catalog image at E-commerce. Bibliographic info: J Intell Inform Syst 2018March: 24(1): 1~23
[5] Chung, S. H, Goswami. A, Lee. H, & Hu, J(2012), The impact of images on user clicks in product search, In Proceedings of the Twelfth International Workshop on Multimedia Data Mining, pp. 25-33, ACM, Aug. 2012.
[6] D. Lowe(1999), Object Recognition from Local Scale Invariant Features, In International Conference on Computer Vision, pp. 1150-1157, 1999.
[7] D. Lowe(2004), Distinctive Image Features from Scale Invariant Keypoints, International Journal of Computer Vision, vol.2, no.60, pp. 91-110, 2004
[8] Gi-Ryong Choi, Hye-Wuk Jung and Jee-Hyoung Lee(2012), Contents-based Image Retrieval System Design of Shopping, Proceedings of KIIS Spring Conference, Vol. 22, No. 1, 2012.
[9] Veit. A, Kovacs. B, Bell. S, McAuley. J, Bala. K, & Belongie. S(2015), Learning visual clothing style with heterogeneous dyadic co-occurrences, In Proceedings of the IEEE International Conference on Computer Vision, pp. 4642-4650, 2015.
[10] DOI: Yeon-gyu Kim, Eui-young Cha(2016), Streamlined GoogLeNet Algorithm Based on CNN for Korean Character Recognition, J. Korea Inst. Inf. Commun. Eng., Vol. 20, No. 9 : 1657~1665, Sep. 2016. http://dx.doi.org/10.6109/jkiice.2016.20.9.1657
[11] DOI: Seung-Cheol Baek(2016), Fast and All-Purpose Area-Based Imagery Registration Using ConvNets, Journal of KIISE,, Vol. 43, No. 9, pp. 1034-1042, 2016. http://dx.doi.org/10.5626/JOK.2016.43.9.1034
[12] Andreas Veit, Balazs Kovacs, Sean Bell, Julian McAuley, Kavita Bala, Serge Belongie(2015), Deep Learning of Binary Hash Codes for Fast Image Retrieval, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 27-35, 2015.
[13] A. Krizhevsky, I. Sutskever, G. E. Hinton(2012), Imagenet classification with deep convolutional neural networks, Advances in Neural Information Processing Systems 25, pp. 1097–1105, 2012.