Title	Image Preference Estimation with Word Embedding Model and Convolutional Neural Network
Author(s)	万, 樺
Citation	
Issue Date	2021-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/17149
Rights	
Description	Supervisor: 長谷川 忍, 先端科学技術研究科, 修士(情報科学)

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

Image Preference Estimation with Word Embedding Model and Convolutional Neural Network

1810443   Wan Hua

To improve the traditional image recommendation system and classification methods, we propose measuring a distance between two vectorized representations: **User Preference Vector (UPV)** and **Image Classification Vector (ICV)**. Both vectors are obtained by our designed procedures together with the calculation formula. For the purpose of individual preferences toward images into consideration, instead of a conventional approach which focus on explore features and attributes only appear within particular contents, along with whole processing procedure lack of combination with contemporary techniques, such as Natural Language Processing (NLP) for embed user keywords, as well as probabilities draw from the deep neural network concerning each label also fulling into deliberation.

Regarding our experiment procedures, the creation of **UPV** can be divide into several steps:

i.   Construct an original website based on Python programming language and FLASK framework, combining with an original database to collect and save user inputs as the plain natural language recourses.

ii. Derive the data from the website and pass them into the data cleaning section, which is the text preprocessing for raw text data, such as lower the capital letters, removing punctuations and whitespaces, tokenization, etc.

iii. Pass the data being cleaned into the word embedding model to output each vectorized representation for user preferences.

iv. Calculate the converted vectors based on assigned weights as well as designed formula to procure the **UPV**.

On the other hand, Convolutional Neural Network (CNN) takes place in the creation of **ICV**:

i.   Apply the Cifar-10 image dataset (a total of 60000 images and 10 labels) to train the neural network, which is constructed by the Python language and Keras framework in a cloud environment.

ii.   Feed the model by an original image dataset (Kawaii dataset, a total of 30 images, divided into 3 branches), records the probabilities of each image come by the output layer (output of SoftMax function), which is an array telling us the probabilities concerning each image being categorized into all labels.

iii.   Pass the probabilities into our secondary designed formula to procure the **ICV**.

After obtaining both **UPV** and **ICV**, we calculate the distance between those vectors to digitize individual preference toward images. Based on our hypothesis, a small distance as the output draw from our proposed model represent a higher priority for our respondents. Contemplation of justification as well as evaluation for the purposed model, we conduct the Single User Estimation together with Multiuser Estimation as our preliminary experiment at the beginning of the evaluation section. In Single User Estimation, our proposed model works well and distinguished image priority toward User ID.01. However, in the case of Multiuser Estimation, an ambiguous prediction for User ID.03 appeared.

Conducive to evaluate our proposed model properly, we determined a further investigation in the sense of both scientific viewpoints together with user-oriented viewpoint. Specifically, we designed an inquiry form with the same content (10 image sets, each including 3 images pick up from the Kawaii dataset, along with a table of keywords) for all respondents. Later, we introduced the statistic approach toward User ID.04 and understood that the assigned weights for images and system output (distance) were statistically insignificant (R-value equals 0.24 but P-value ≈ 0.19 which is large than 0.05). Besides, we also apply other correlation analysis methods to calculate the results and draw the regression line for the User ID.04.

Subsequently, we continue to distribute the inquire form and collect information until User ID.07 and apply several data analysis methods to the output data raised by our experiment, such as correlation (statistic approach notifies that the assigned weight and output distance is statistically insignificant, P-value equals to $0.27 > 0.05$), logistic regression (75% accuracy), draw confusion matrix, F-measure, etc. As the results of this research, a total of 43 samples being tested (3 samples draw from the preliminary experiment and 40 samples draw form inquire from). 8 of the samples are perfectly hit (18.6% accuracy), and 21 of the sample identified that our proposed model affected somehow (Partially correct, floating from 18.6% ~ 67.4%). Nevertheless, we also have a total of 14 samples of image sets that could not properly distinguish by our proposed

model. Overall, after we checked the assigned weights from our respondents, comparing with the calculation results (distance), we understand that the user performance in the sense of image priority assignment gives a significant influence on the results.

Furthermore, we also understand several weaknesses regarding our proposed model. Such as we should update the Kawaii dataset closer to the training dataset instead of 2/3 branches are not affected, in the sense of images can be learned and distinguished by the convolutional neural network. We should also update our policy regarding the inquiry form, which notifies our respondent keywords and image priorities should be related. Lastly, scientific approaches such as correlation and linear/logistic regression could not make much sense for experiment results. We should also consider other approaches.

At last, this research investigated a digitalized representation between **UPV** and **ICV.** Which gives inspiration to whose works in image recommendation and classification problems. We believe that all procedures combined in this experiment, such as website and database for data collection. Word embedding and data cleaning for text processing, neural network, distance calculation following our design formula, etc. The flow that appears in this experiment will be a ignite.

Keywords: **User Preference Vector (UPV)**, **Image Classification Vector (ICV)**, Distance Calculation, Word Embedding, Convolutional Neural Network, Kawaii Dataset.