Packaging Domain-Based Named Entity Recognition with Multi-Layer Neural Networks

Changyun Li[1,2], Yuezhong Wu[1,2]*, Fanghuai Hu[3], Changsheng Liu[4]

ABSTRACT

Artificial neural networks (ANNs) are the greatest success story that inspired by biological neural networks and neuroscience; ANNs model realistic problems by a network of neurons which are designed by simulating biological neurons. This paper attempts to design a multi-layer neural network to recognize the named entities in packaging domain. For this purpose, a neural network language model was designed to automatically learn the distributed word features and partial speech features. Based on these distributed features, a multi-layer deep neural network model was constructed for the NER of packaging products. The experiments prove that our model can automatically extract more and better advanced features than traditional methods, thus minimizing the workloads of manual feature selection. The results show that the model outperformed the traditional sequence labelled CRF model by 10% in precision and 6% in recall, and that the four-layer neural network with two hidden layers boasted the best NER of packaging products.

Key Words: : Packaging, Named Entity Recognition (NER), Neural Network, Computational Neuroscience

DOI Number: 10.14704/nq.2018.16.6.1555 **NeuroQuantology 2018; 16(6):564-569**

Introduction

The human brain is a staggeringly complex computational system, consisting of some 100 billion neurons, connected by an estimated 100 trillion synapses; it allows us to make sense of a complex and ever-changing sensory world, to plan complex actions, to navigate our social environment and intuit the minds of others, and to learn and remember across our entire lifespans (David, 2014). In order to improve the capacity of computers, scientists have made long-term effort in simulating the human brains, instantiations including, "the artificial neuron" (see Figure 1 a), the "perceptron", the "artificial neural network", the "convolutional neural network" (see Figure 1 b), and so on. Neuroscience has provided guiding force for the development of artificial neural networks, providing inspiration for architectural features of neural networks (Gonçalves, 2017).

With the continued development of information technology, intelligence has become an important direction for the development of packaging industry (Ge, 2015). In the highly intelligent environment, a large number of files, including but not limited to product descriptions and user manuals, are presented in a digital format. However, it is still more difficult to recognize product entities in the industry than in other fields, as the complex structure of product names has made ordinary entity recognition virtually useless. To overcome the difficulty, the named entity recognition (NER) should be applied to the packaging industry, seeking to construct knowledge graph, intelligent Q&A engine and other basic applications based on information in

Corresponding author: Yuezhong Wu
Address: [1]School of Artificial Intelligence, Hunan University of Technology, Zhuzhou 412007, China; [2]Intelligent Information Perception and Processing Technology Hunan Province Key Laboratory, Zhuzhou 412007, China; [3]Department of Computer Science and Engineering, East China University of Science and Technology, Shanghai 200237 China; [4]Changsha Aeronautical Vocational and Technical College, Changsha 410124 China
e-mail ✉ yuezhong.wu@163.com
Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
Received: 2 March 2018; **Accepted:** 1 May 2018

packaging domain.

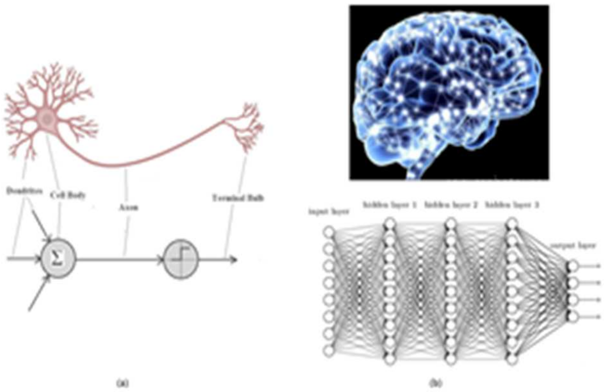

Figure 1. A biological neuron and an artificial neuron

Literature Review

First proposed at the Sixth Message Understanding Conference (MUC-6), the concept of NER (Chen, 2016) refers to a subtask of information extraction that seeks to locate and classify named entities in text into pre-defined categories, such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc. In the packaging industry, the NER has been mainly applied to the extraction of enterprise files, experts, products and others.

The commonly used NER methods mainly fall into two categories: the rule-based methods and the statistical methods. The first category is old, simple and convenient, but features a high demand of artificial prior knowledge and poor portability. In the second category, the NER is either treated as a classification problem using support vector machine (SVM) (Hu, *et al.*, 2017; Lakshmi *et al.*, 2017), Bayesian method (Mahalakshmi *et al.*, 2016) and other classification models, or treated as a sequence labelling problem using the hidden Markov model, the maximum entropy Markov model, the conditional random field (CRF) and other sequence annotation models in machine learning (Zheng and Liu, 2016; Huai *et al.*, 2014; Zhong *et al.*, 2014). The statistical methods require the manual design of various features by intuition or learning of training corpus. The performance mainly depends on the accuracy of the features. In most cases, the two kinds of methods are combined to achieve the desirable outcome of NER.

Owing to the changing structure and blurred boundaries of product entities, the traditional NER methods cannot identify entities from the corpus of the packaging domain in a correct and efficient manner. It is generally believed that a product naming entity covers five aspects: brand, series, model, type and attribute. Of course, not all of these aspects are indispensable. Sometimes, a product can have a brand name without a series name, and vice versa. In this research, the NER of packaging products mainly focuses on the product name, such as toilet paper rolling machine.

Much research has been done on product NER at home and abroad. For example, Liu *et al.* (2016) proposed a product NER method based on hierarchical hidden Markov model (HHMM), and obtained the F-statistic value of about 80% for the recognition of product, model and brand. However, the HHMM was not compared with other models, and the labelling strategy was relatively simple. Mei *et al.* (2010) integrated the maximum entropy model and the knowledge base into the NER of network text products in China. Applying the CRF to English product NER, Zhang *et al.* (2010) selected the contexts and English product names with excellent instructions, and recognized the product entities based on a hand-built brand library. The method achieved the precision of 93.6% and the recall rate of 92.4%, but failed to consider or validate the semantic features. Focusing on product entities in English e-commerce sites, Putthividhya *et al.*, (2011) expanded the list of seeds for product recognition through the integration of supervised learning and bootstrap methods. Wu *et al.*, (2013) implemented the matching model, the rule-based model, and the CRF model in the product NER of English short texts. In view of the peculiarity of Chinese microblogs, Wang *et al.*, (2017) developed a method for selecting feature words vectors that combines global context information. The method relies on the subject model and the neural network to obtain deep semantic information, e.g. the NER of microblog products for airports.

To sum up, the above studies, with relatively heavy workloads of artificial annotation, often cannot dig out in-depth semantic relationships. Fortunately, the recent success in image and speech recognition proves that deep learning is possible without human intervention when the original feature set has been extracted. Therefore, this paper attempts to use the deep learning technique to extract the unsupervised word features from the corpus of various unlabelled packaging products. Based on the previous research, the neural network language model was introduced to the construction of distributed word features, adding to the feasibility of the packaging product NER.

Algorithm Design

This paper designs a deep neural network architecture for the NER. In essence, the architecture is a multi-layer neural network for learning useful features and improving the recognition performance. The architecture was constructed in the following steps: replace the traditional sparse features by dense distributed features; combine the structural features of the named entity domain; compare the traditional sparse features with the commonly used models in natural language processing tasks such as the CRF, the SVM, and rich distributed feature representation; apply the deep learning structure in more advanced features.

Neural network of NER

The neural network was designed with at least three layers: an input layer, a hidden layer and an output layer. The input layer was a distributed expression of the word of the packaging domain. The model parameters were trained and optimized at the presence of the input word vector. The hidden layer could contain multiple layers. For better training speed, the single-layer structure was adopted for the hidden layer. The output layer relied on a logical classifier with a loss function of binary cross entropy.

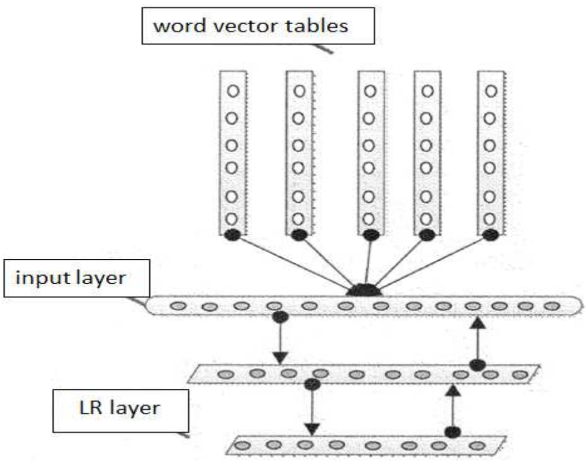

Figure 2. Network Structure of NER

Taking the NER of packaging products as a classification problem, the author regarded its input as a word vector that combines the vector expression of the domain structure with the context vocabulary. In the neural network, the word vector was converted into another vector in the hidden layer; then, the entity name was obtained based on the probability of each word through the classification on the logical regression layer (Figure 2).

Distributed representation

The traditional word and partial speech features ignore the connection between two words or two parts of word tag. Here, the word feature and partial speech feature are distributed, that is, each word or each partial speech feature is expressed as a low-dimensional real vector, so as to shorten the Euclidean distance of any two words or two partial speech features. The distributed representation of lexical features can solve the curse of dimensionality and the local optimum trap in machine learning. Compared with the traditional feature representation, the distributed representation helps to explore the connection between input data and capture its internal grammar and semantic similarity.

Feedforward neural network

To recognize the exact entity of a word, it is necessary to consider the context of the word. Therefore, the traditional single word vector input was replaced by the word vectors of the left and right windows as the neural network input. For the defined word, let the size of the left window as LC and that of the right window as the RC. When LC=0 and RC=0, the input is a word vector, and the input of the hidden layer is (LC+1+RC)×M, where LC+1+RC is the current window size and M is the dimension of the word vector. The output of the hidden layer acts as a feature of the logical regression layer. In the logical regression layer, the window of the central word is calculated for the probability of each category. Therefore, the feedforward neural network functions of the network architecture are as follows: $z=W(x^1, x_2)+b^1(1)$, $a=f(z)$, $h=g(U^Ta+b^2)$.

Where (x^1, x_2) is the input window word vector; a is the output value of the hidden layer; z is the eigenvector after the linear transformation of the input word window vector; W and U are the model parameters; f and g are the model functions; g and b are the scalars of linear transformation; h is the probability of each category; f is the activation function (either sigmoid function or hyperbolic tangent function). The hyperbolic tangent function was chosen for this research, because the derivative of the hyperbolic tangent function can also be expressed as a hyperbolic tangent function.

$d/d_x tanx=1-tanh^2x.$

To return a probability value, the sigmoid function was adopted as the g function.

$g(z)=sigmod(z)=1/(1+e^{-z}).$

Based on neural networks, the entities are named in the following manner: The input training samples contain a series of input pairs (x^i, y^i), where i ranges from 1 to the total length of training corpus, $x_i=[x_{i-1}, x_i, x_{i+1}]$ is the window word vector, and $y^i \in [0, 1]$ is the entity probability. The parameters of θ were defined as the whole network $\theta=(W,b^{(1)}, b^{(2)})$. By this definition, the whole neural network function can be expressed as: $h_\theta(x^i)=g(U^T f(W(x^1, x_2)+b^1)+b^2)$.

Randomized parameters

During network training, the parameters should be assigned with random values. The most efficient way to randomize parameters is to assign the parameter W with a random value from the range domain [-∈, ∈]. Here, ∈ is calculated by the number of neurons per neural network and the number of output neurons: $\in=\sqrt{6}/\sqrt{(in+out)}$.

Where in is the number of input neurons in the neural network of this layer; out is the number of output neurons by this layer of neural network.

Parameters training

In this paper, the training of deep network structure is essentially the calculation of unknown parameters in the training corpus. The unknown parameters mainly include several parameters of the hidden layer, and the change matrix and the offset matrix of the logistic regression layer. The training mainly uses the back-propagation algorithm and the stochastic gradient descent (SGD) algorithm. The specific steps of the training process include:

Step 1. Randomly initialize all the parameters of the network, including those on the hidden layer and the logical regression layer.

Step 2. Randomly select a training sample pairs (x^i, y^i), perform the first forward communication from the hidden layer of the output information to the logical regression layer, and extract the most advanced features mapped to the corresponding tag information. Meanwhile, supervise the model by the tag value of the data, and adjust the connection weight continuously to lower the probability error between the target prediction feature and the actual feature.

Step 3. Reverse propagation calculation: calculate the conceptual error between the target prediction feature and the actual feature in the forward propagation process, propagate the error from the logical regression layer to the hidden layer, and do not adjust the hidden layer parameter $\theta=(W,b^{(1)})$.

Implementation Process

Targeted at the text information of packaging products, this paper mainly considers the NER of packaging product names. As shown in Figure 3, the entity recognition process includes data acquisition, data pre-processing, distributed word feature training, neural network model training, entity recognition and recognition result extraction. First, the texts related to packaging products were collected from the China Packaging Network (http://www.pack.cn/); then, the webpage text data were pre-processed, including filtering of special symbols, word segmentation, and stop word processing; following the Pareto principle, the data were divided into the training set (80%) and the test set (20%); next, the training set was inputted to the model for training; finally, the trained parameters were used to verify the effect of the test model.

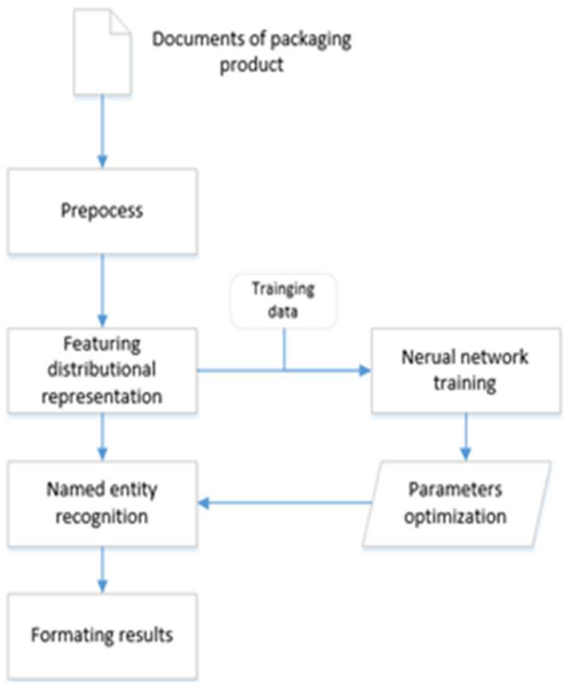

Figure 3. Workflow of packaging named entity recognition

Table 1. Result matrix

Experiment	Features	Precision	Recall	F values
Exp 1	POS	72.34%	71.56%	71.95%
	Syntax	74.23%	76.87%	76.03%
	Intonation	60.50%	45.76%	52.11%
	POS + Syntax	79.45%	80.35%	79.95%
	POS + Syntax + Intonation	78.34%	80.56%	79.44%
Exp 2	Word distribution	86.48%	81.90%	84.13%
	POS distribution	83.46%	79.45%	81.41%
	Word Distribution + POS distribution	90.13%	86.54%	88.30%

Methods

Experiment design

The relevant codes and the NER model were implemented and trained on a CentOS system with Python. Google's open-source version of Word2vec was adopted to build a neural network language model. Through the continuous word bag model (CBOW), Word2vec, a simplified nonprofit leadership and management (MNLM) model, removes the most time-consuming nonlinear hidden layer, shares all words in the hidden layer, and trains the word features and partial distributed features in an unsupervised manner. Two comparative tests were set up to verify the effect of our algorithm for text design.

In test 1, the CRF model was trained on the training corpus with the feature set, and the trained CRF model was used to test the data in the test set. Then, the NER results of packaging products were evaluated manually.

In test 2, the distributed expression of words and partial speech features were unsupervised in the training corpus; then, the three-layer neural network model was constructed and trained using these distributed features; next, the trained neural network model was used to carry out the NER on the test corpus; finally, the recognition results were evaluated manually.

Results

The NER results of packaging products in tests 1 and 2 were evaluated against such three indices as precision, recall, and F-statistic:

Precision=Correctly Recognized Entities/Total Entities Recognized;

Recall=Correctly Recognized Entities/Total Entities Existed;

F-statistic=2∗precision∗Recall/(Precision+Recall).

According to the test results in Table 1, there are no many results on the intonation feature for the NER of the CRF sequence. Therefore, the lexical feature + grammar rule was taken as the feature input of the CRF model. The precision, recall and F-statistic were all about 80%. For the deep neural network model, the results were evaluated against word distribution feature, partial distributed speech feature and joint feature, respectively. It can be seen from Table 1 that the joint feature had a great effect on the deep neural network model in the NER of packaging products, with the recall and the F-statistics at about 90%. Hence, the proposed deep neural network model outperformed sequence labeled CRF model in the NER of packaging products.

Figure 4 compares the recognition results of the neural network models with different layers. With the continuous increase in the number of layers, the number of training samples was insufficient to complete the parameter training of the model, resulting in the decrease of the test results and the recall. Overall, the four-layer neural network with two hidden layers boasted the best NER of packaging products.

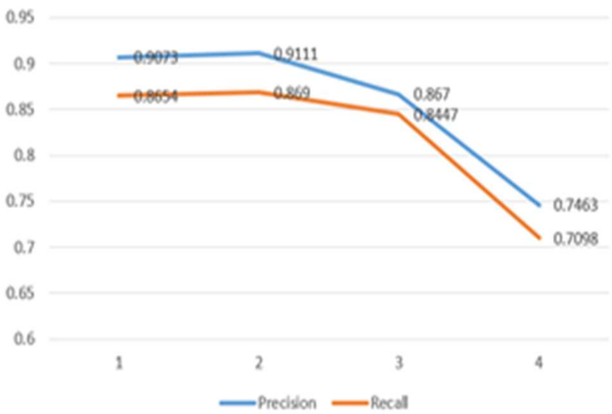

Figure 4. Impacts of hidden layer scales on the P&R value

Conclusions

This paper designs a neural network language model to automatically learn the distributed word features and partial speech features. Based on these distributed features, a multi-layer deep neural network model was constructed for the NER of packaging products. The experiments prove that our model can automatically extract more and better advanced features than traditional methods, thus minimizing the workloads of manual feature selection. The results show that the model outperformed the traditional sequence labelled CRF model by 10% in precision and 6% in recall, and that the four-layer neural network with two hidden layers boasted the best NER of packaging products.

Acknowledgments

This paper is funded by the Hunan Provincial Natural Science Foundation of China under grant numbers 2016JJ5035; the Hunan Provincial Science and Technology Project (key project) under grant numbers 2015TP1003; the Scientific Research Project of Hunan Province Department of Education under grant numbers 16C0481; the Project of China packaging federation under Funding Support Number 2014GSZJWT001KT001, 2014GSZJWT001KT010, 17ZBLWT001KT010; the Scientific Research Project of Zhuzhou City Department of science and technology bureau under grant numbers 2016-68-12; the Project of National Packaging Advertising Research Base and Hunan Packaging Advertising Creative Base under Funding Support Number 17JDXMA03.

References

Cannistraro M, Lorenzini E. The applications of the new technologies “e-sensing” in hospitals. International Journal of Heat and Technology 2016; 34(4): 551-57.

Chen J. Survey of named entity recognition. Modern Computer 2016; 3: 24-26.

David DC, Thomas D. Neural networks and neuroscience-inspired computer vision. Current Biology 2014; 24(18): 921-29.

Ge WM. Internet artificial intelligence: seize the opportunity to seize the top of the industrial revolution. World Telecommunications 2015; 8: 2-74.

Gonçalves CP. Quantum neural machine learning: backpropagation and dynamics. NeuroQuantology 2017; 15(1): 22-41.

Hu C, Ding C, Dai L. Modeling and development of medical information system based on support vector machine in web network. International Journal Bioautomation 2017; 21(S4): 283-292

Huai BX, Bao TF, Zhu HS. A named entity linking method based on probability topic model. Journal of Software 2014; 9: 2076-87.

Lakshmi G, Panicker JR, Meera M. Named entity recognition in Malayalam using fuzzy support vector machine. International Conference on Information Science 2017: 201-06.

Liu FF, Zhao J, Lv BB. Research on product named entity recognition for business information extraction. Journal of Chinese Information Processing 2006; 20(1): 7-13.

Mahalakshmi GS, Betina AJ, Bagawathi RS. domain based named entity recognition using naive bayes classification. Social Science Electronic Publishing 2016; 10(102): 234-39.

Mei F, Sun CJ, Sun K. Chinese product named entity recognition based on network text. Journal of Zhengzhou University: Natural Science Ed 2010; 42(1): 62-66.

Putthividhya D, Hu J. Bootstrapped named entity recognition for product attribute extraction. Conference on Empirical Methods in Natural Language Processing 2011; 1557-67.

Wang HL. Named entity recognition of Chinese microblog product based on word-vector clustering. Journal of Lanzhou University of Technology 2017; 43(1): 104-10.

Wu S, Fang Z, Tang J. Accurate Product Name Recognition from user generated content. International Conference on Data Mining Workshops 2013: 874-77.

Zhang CS, Guo JY, Yan XT. Named entity recognition of English products based on conditional random field. Computer Engineering and Science 2010, 32(6): 115-17.

Zheng QS, Liu SX. Research of web text named entity recognition based on CRF. Journal of Zhongyuan University of Technology 2016; 27(1): 70-73, 95.

Zhong ZN, Liu FC, Wu Y. Active learning and self-learning Chinese named entity recognition. Journal of National University of Defense Technology 2014; 4: 82-88.