Application of Text Classification Technology in Intelligent Customer Service Systems for the Newspaper Industry (Postprint)

Authors: Zheng Chuangwei, Xie Zhicheng, Gutao Xing, Chen Shaobin, Chen Yifei

Date: 2023-10-08T00:00:00+00:00

Abstract

The newspaper industry, as a mainstream media sector, handles a massive volume of customer information daily, encompassing advertising services, newspaper printing, subscription management, and news tip-offs. In recent years, with continuously rising labor costs, the media industry' s demand for intelligent customer service systems that can substantially reduce the workload of human customer service has become increasingly urgent. This paper presents an in-depth study of question classification technology aimed at improving the accuracy and efficiency of AI customer service systems in the newspaper industry, primarily employing the fastText algorithm for question classification. Experimental results demonstrate that, compared with traditional SVM+TF and BERT+SVM approaches, this algorithm effectively balances system query accuracy and time overhead to satisfy user requirements for both query accuracy and response speed.

Full Text

Preamble

Application of Text Classification Technology in Newspaper Industry Intelligent Customer Service Systems

Zheng Chuangwei, Xie Zhicheng, Xing Gutao, Chen Shaobin, Chen Yifei

(Shenzhen Creative Smart Harbor Technology Co., Ltd., Shenzhen, Guangdong 518034)

Abstract: As a mainstream media outlet, the newspaper industry processes massive volumes of customer information daily, ranging from advertising services and print subscriptions to news tips and circulation inquiries. In recent years, rising labor costs have created urgent demand within the media sector for intelligent customer service systems that can substantially reduce manual workload. This paper investigates question classification techniques to improve the accuracy and efficiency of AI-powered customer service systems in the newspaper industry, primarily employing the fastText algorithm for question categorization. Experiments demonstrate that compared with traditional SVM+TF and BERT+SVM approaches, this algorithm effectively balances query accuracy and computational overhead to meet user demands for both precision and speed.

Keywords: Intelligent customer service system; Question classification technology; fastText; Hierarchical softmax

CLC Number: G211
Document Code: A
Article ID: 1671-0134(2021)10-149-03
DOI: 10.19483/j.cnki.11-4653/n.2021.10.045

Citation Format: Zheng Chuangwei, Xie Zhicheng, Xing Gutao, Chen Shaobin, Chen Yifei. Application of Text Classification Technology in Newspaper Industry Intelligent Customer Service Systems[J]. China Media Technology, 2021(10): 149-151.

---

1. Concept of Intelligent Customer Service Systems

The intelligent customer service system represents a sophisticated integration of multiple advanced technologies, including large-scale knowledge processing, natural language understanding, knowledge management, automatic question-answering systems, and inference engines. These systems not only provide enterprises with fine-grained knowledge management capabilities but also establish an efficient technical framework for communication between businesses and massive user bases through natural language interfaces. Additionally, they deliver statistical analysis information essential for refined user management.

The advent of the big data era has accelerated the development of deep learning classification algorithms. For instance, TextCNN, proposed by Yoon Kim in 2014, exhibits a structure remarkably similar to image processing pipelines, comprising a convolutional layer followed by a max pooling layer. The input takes the form of a word vector matrix, where convolutional kernels of varying widths slide across the entire sentence length to generate n activation values. These are subsequently processed through a max pooling layer to produce a feature map composed of m feature values, which serves as the basis for classification by downstream classifiers [2].

2. Development Background

As multimedia technologies mature, information access channels have proliferated from traditional branches, telephone, websites, and email to instant messaging, microblogging, and WeChat. Consequently, online information has become characterized by fragmentation, mobility, real-time delivery, personalization, multimedia integration, and massive data volumes. These transformations pose novel challenges to information management and service delivery, rendering traditional newspaper customer service systems inadequate for public service demands.

Simultaneously, intelligent robot technology in artificial intelligence has advanced rapidly, revolutionizing production and lifestyle patterns. For the newspaper industry, intelligent robot technology can significantly reduce manual customer service workloads and improve information processing efficiency amid the current explosion of internet information. Therefore, applying intelligent robots to newspaper customer service has become an inevitable trend.

In scenarios with particularly stringent speed requirements, the fastText algorithm presents an excellent choice. It not only ensures high-speed model training but also maintains model precision [3]. This study employs fastText as a domain classifier within intelligent customer service systems and compares it against traditional classification algorithms.

3. Related Work

In recent years, text classification technology has been widely applied in intelligent customer service systems. Traditional classification algorithms include Multilayer Perceptrons, Naive Bayes, and Support Vector Machines (SVM). SVM represents a generalized linear classifier for supervised binary classification that offers robustness by considering both empirical and structural risk minimization in its optimization objective. The hinge loss function' s properties confer sparsity advantages. However, a significant drawback emerges when data volume exceeds a certain threshold: training time grows exponentially. Consequently, with large-scale datasets, traditional classification algorithms lose their competitive edge and become the primary factor limiting computational efficiency [1].

4. FastText Algorithm

This classification algorithm establishes a simple yet effective baseline by representing sentences as bag-of-words and feeding words from news reports into a linear classifier for training. However, linear classifiers cannot share parameters between features and classes, potentially limiting generalization. Common solutions involve decomposing the linear classifier into low-rank matrices or employing multi-layer neural networks. Neural networks enable information sharing through hidden layers. The fastText methodology comprises three components: model architecture, hierarchical softmax, and N-gram features. It processes word sequences as input and predicts probability distributions over word categories using a softmax function, with model training performed via stochastic gradient descent and backpropagation.

4.1 Model Architecture

The fastText model architecture resembles the CBOW model in word2vec [4]. While CBOW predicts target words from context, fastText predicts text categories from context. Both 本质上 constitute three-layer networks (input layer, single hidden layer, output layer), with fastText employing supervised learning as opposed to word2vec' s unsupervised approach. The specific model structure is illustrated below.

[Figure 1: see original paper] fastText Model Structure

In the diagram, *ix* represents the feature vector of the *i*-th word in the text. The model takes a word sequence as input and generates a probability distribution over predefined classes, training asynchronously across multiple CPUs for exceptional speed. Probabilities are computed through a softmax equation built upon a Huffman coding tree, where each node associates with the path probability from the root to that node, expressed as:

$$P(n) = \prod_{i=1}^{L(n)} P(d_i|n_i)$$

This formulation shows that a node's probability is smaller than its parent node's probability. Depth-first search through the entire Huffman coding tree identifies maximum probability paths between nodes, enabling complexity reduction by pruning low-probability branches.

4.2 Hierarchical Softmax

For datasets with numerous categories, linear classifier computational load becomes substantial, with complexity scaling as *O(Nd)*, where N represents the number of categories and d denotes hidden layer dimensionality. To reduce computation, the model employs hierarchical softmax based on Huffman coding trees instead of flat architectures. This approach generalizes logistic regression for multi-class tasks, integrating different categories into a tree structure that accounts for class imbalance (some categories appear more frequently) [5]. As shown in [Figure 2: see original paper], leaf nodes comprise K distinct class labels at the bottom, with *K-1* internal nodes serving as internal parameters, formulated as:

$$P(c|w) = \prod_{j=1}^{L(c)} \sigma(\mathbf{v}_{n_j}^\top \mathbf{h})$$

The tree structure ensures that frequently occurring categories have shallower depths than infrequent ones, further improving computational efficiency. During training, complexity reduces to *O(d log N)*.

[Figure 2: see original paper] Hierarchical Softmax Example Diagram

4.3 N-gram Features

Both text and sentence classification commonly employ bag-of-words models. However, such models ignore word order relationships. To address this limitation, the model incorporates N-gram features. The fundamental approach involves sliding a window of size N over customer consultation text to generate multiple byte fragment sequences of length N. For efficiency, low-frequency N-grams are filtered according to a predetermined threshold to form a key list [6]. N-grams capture local word order information to evaluate the reasonableness of customer queries, achieving a balanced acquisition of word order information and computational resources.

In practical newspaper intelligent customer service applications, customer consultations generate massive text datasets. N-gram processing produces numerous redundant terms during classification, necessitating term reconstruction to identify customers' actual consultation intents. This process also learns additional contextual relationships between words. During reconstruction, grammatical and sequential relationships are disregarded—each word is treated as independent—before removing terms present in N-gram processing but absent from the text bag-of-words.

5. Experimental Results and Analysis

This section compares fastText against traditional SVM+TF and SVM+BERT (using BERT [6] for word vector extraction) to demonstrate fastText' s advantages with sufficient data volume and numerous categories.

5.1 Binary Classification

Using online intelligent customer service system data for training and testing, the dataset comprised two categories: casual chat and service consultation. Preprocessing removed non-compliant data, yielding 50,000 samples per category. Experiments selected 10,000 test samples, evaluating models based on test accuracy, training time, and ROC curves (ROC and AUC being standard binary classification metrics).

Algorithm Performance Comparison 1

Method	Accuracy	Train Time (s)
SVM+TF	99.81%	
SVM+BERT	99.47%	

Method	Accuracy	Train Time (s)
fastText	99.60%	

Table 1 shows SVM+TF achieves slightly higher accuracy than the other algorithms with lower training time. SVM+BERT' s excessive training time renders it unsuitable for the current application scenario, so subsequent experiments focus on SVM+TF and fastText.

Building on the previous experiment, the dataset was expanded from 100,000 to 1.65 million samples, split 9:1 for training and testing. Results appear in .

Algorithm Performance Comparison 2

Method	Accuracy	Train Time (s)	F1-score
SVM+TF	99.74%	98.08%	99.93%
fastText	99.62%	99.46%	99.76%

Overall accuracy improved with the larger dataset. While fastText' s accuracy is marginally lower than SVM+TF, its training time is substantially lower, demonstrating a clear advantage. [Figure 3: see original paper] presents ROC curves for both algorithms.

[Figure 3: see original paper] ROC Curves for SVM+TF and fastText Algorithms

The AUC value for fastText exceeds that of SVM. Combined with , these results indicate that fastText delivers classification performance comparable to SVM under large-scale data conditions.

5.2 Multi-classification (Three-class)

SVM functions as a linear classifier particularly convenient for binary classification but less so for multi-class problems. This experiment implements three-class SVM using an indirect approach with two binary classifiers. Data were divided into three categories: casual chat, geographic services, and general consultation. Results appear in .

Algorithm Performance Comparison 3

Method	Accuracy	Train Time (s)	F1-score	Route
SVM+TF	99.10%	98.04%	99.88%	99.97%
fastText	99.71%	99.69%	99.38%	99.72%

fastText demonstrates superior performance in both accuracy and training time compared to SVM.

6. Conclusion

In summary, for intelligent customer service applications with low customer consultation volumes, small datasets, and low problem complexity (few categories), SVM classification outperforms fastText. However, as consultation volume increases or classification categories multiply, fastText demonstrates clear advantages—delivering excellent classification performance with exceptional speed, making it highly suitable for intelligent customer service applications. As user scales continue growing, newspaper operations face exponentially increasing customer inquiries and data volumes. Therefore, fastText represents a promising modeling approach for future newspaper intelligent customer service systems.

References

[1] Xie Jichuan, Zong Zhenguo, Liu Hongguo, Zhang Chunqiu, Tian Xiao. Text Mining of 95598 Work Orders Based on Word Vector Models[J]. Electronic World, 2017(23): 176+178.

[2] Feng Mengying, Li Hong. Application of Text Convolutional Neural Network Models in Short Text Multi-classification[J]. Financial Technology Times, 2020(1): 38-42.

[3] Joulin A, Grave E, Bojanowski P, et al. Bag of Tricks for Efficient Text Classification[J]. 2016.

[4] Xiong Fulin, Deng Yihao, Tang Xiaosheng. Core Architecture and Applications of Word2vec[J]. Journal of Nanjing Normal University (Engineering and Technology Edition), 2015(1).

[5] Li Hao, Zhang Jihao. Discussion on Automatic Classification of Customer Service Complaint Work Orders Based on Text Mining Technology[J]. Mobile Communications, 2017(23): 66-72.

[6] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[J]. 2018.

Author Biographies:

Zheng Chuangwei (1978-), male, from Shantou, Guangdong, Senior Engineer. Research interests: Big data, Artificial intelligence.

Xie Zhicheng (1980-), male, from Shantou, Guangdong, Intermediate Professional Title. Research interests: Big data, Cloud computing.

Xing Gutao (1984-), male, from Wenchang, Hainan, Intermediate Professional Title. Research interests: Cloud computing.

Chen Shaobin (1973-), male, from Jieyang, Guangdong, Intermediate Professional Title. Research interests: Big data.

Chen Yifei (1981-), male, from Zhanjiang, Guangdong, Intermediate Professional Title. Research interests: Big data.

(Responsible Editor: Chen Xuguan)

Note: Figure translations are in progress. See original paper for figures.