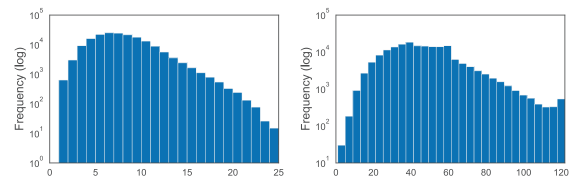
The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

Vertical Domain Text Classification: Towards Understanding IT Tickets Using Deep Neural Networks

Jianglei Han, Mohammad Akbari
SAP Innovation Center, Singapore
{ray.han, mohammad.akbari}@sap.com

Abstract

It is challenging to directly apply text classification models without much feature engineering on domain-specific use cases, and expect the state of art performance. Much more so when the number of classes is large. Convolutional Neural Network (CNN or ConvNet) has attracted much in text mining due to its effectiveness in automatic feature extraction from text. In this paper, we compare traditional and deep learning approaches for automatic categorization of IT tickets in a real world production ticketing system. Experimental results demonstrate the good potential of CNN models in our task.



(a) By words (b) By characters

Figure 1: Title length distribution

Introduction

In Information Technology (IT) support environments, customers' incidents, problems, and change requests are commonly managed in the form of *tickets*. A ticket is created by a customer or the system at fault, before being sent to a centralized ticketing system. Subsequently, it is dispatched to an appropriate domain expert to investigate the issue. Dispatching decisions are made based on the ticket type, which is usually selected by the customers. In an extensive support system, the number of categories can easily expand to over a few thousand, making manual selection costly in terms of both time and labor. Further, choosing the correct ticket category requires domain expertises of which grassroots users may not be aware, introducing error in choosing a proper category. For instance, in the system we studied, categories for nearly 40% of the tickets need to be modified after initial assignment. Therefore, in this paper, we leverage machine learning techniques to automatically classify IT tickets.

Automatic categorization of IT tickets can reduce the overall ticket turnaround time, benefit both service providers and customers. Prior studies have proposed various approaches. For instance, Diao, Jamjoon, and Loewenstern (2009) designed a set of rules to perform multi-class classification on IT tickets. However, rule-based system is hard to extend and to cope with a large number of classes. SmartDispatch (Agarwal, Sindhgatta, and Sengupta 2012) is proposed to classify tickets to their perspective processing groups in which an SVM-based classifier is used in a semi-automated dispatching framework. Since each group is responsible for a few problem types, the task is similar to ticket classification problem that we concern ourselves with, but with much fewer categories (less than 80 processing groups). Although some research has been devoted to ticket classification, they mostly focus on specific cases and datasets and fail to generalize to other scenarios and so their effectiveness is hardly comparable.

Recent years witnessed an increasing trend of applying deep neural network in a variety of machine learning problems in text processing. Some promising results are shown when applied on text categorization (Zhang, Zhao, and LeCun 2015; Kim 2014). A deep neural network learns latent features from data without explicit feature engineering. It is also known to scale well with lots of training data. In this paper, we present the results of conventional neural network approaches for categorization of IT tickets as part of a real-life industrial project.

Dataset Description

We study data exported from a real production ticketing system from a software enterprise. We randomly sampled $177,851$ tickets among tickets which were created in 2016 and solutions had been accepted by customers. For the this investigation, we only consider ticket titles. Distributions of the data are illustrated in Figure 1a and 1b, respectively. Each ticket has a label verified by experts indicating the functional area the problem originated from. We considered these labels as ground truths. Figure 2 shows the cumulative frequency of tickets from 559 labels in the dataset.

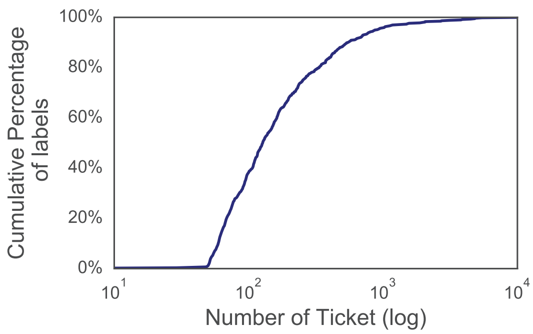


Figure 2: Label distribution

Classification approaches

Rule-based Inspired by (Diao, Jamjoom, and Loewenstern 2009), we leveraged $1,471,093$ manually crafted lexical rules to classify IT tickets.

IR-based Based on (Sun 2012), we constructed an index for the ticket corpus and post-processed the retrieval results to identify a ticket category.

SVM We employed linear SVM classifier using Tf-Idf feature extractions for ticket classification (Agarwal, Sindhgatta, and Sengupta 2012). It was reported to have the state of art performance for several text classification (Aggarwal and Zhai 2012).

CNN-Rand Inspired by recent research efforts in Convolutional Neural Network (CNN) (Zhang, Zhao, and LeCun 2015), we leveraged deep networks for learning the representation of IT tickets in latent space, where word vectors are randomly initialized before training.

CNN-Pre Similar to **CNN-Rand** where we utilized pre-trained word vectors extracted from our dataset using (Bojanowski et al. 2016).

Experiment and Result

Experimental Setting

From the data extracted, $142,280$ tickets are used for training and $35,571$ for evaluation. The rule-based method maps 1,471,093 keywords to labels, before using the aggregation of all matched keywords in a ticket as the prediction. The IR-based method follows the setup used by Sun (Sun 2012) in which the query words are selected by the product of term frequency and a clarity score. Due to a large number of labels, we use the voting scores from top-200 hits for each prediction in the test set. SVM model uses 50,686 terms as term features with TF-IDF weights. No preprocessing step other than removing non-ascii characters is performed. We evaluate prediction accuracy@k for $k = [1, 3, 5, 10]$.

Results

Without much manual feature engineering, we experiment the methods in vertical domain text classification. As shown in Table 1, rule-based classifier performs poorly, due to the limitation of keyword matching and the ambiguous linkages exist between keywords and labels. CNN achieved the best performance due to its ability to model local semantic context using convolution as well as considering composite features (N-Grams) in the text. Nearly 15% improvement at Accuracy@1 is observed, by replacing the random word embedding to pre-trained vectors. Interestingly, the retrieval-based method exhibits better results compared to other shallow models. Despite of the sheer number of classes, aggregating search engine returned results tend to be a scalable and promising approach.

Table 1: Experiment results. The best results are in bold face and second best are underlined.

	Accuracy@1	@3	@5	@10
Rule-based	0.037	0.072	0.098	0.140
IR-based	0.370	<u>0.545</u>	<u>0.615</u>	<u>0.706</u>
SVM	0.232	0.341	0.386	0.460
CNN-Random	<u>0.378</u>	0.527	0.583	0.653
CNN-Pretrain	**0.444**	**0.620**	**0.684**	**0.761**

Conclusion and Future Work

In this paper, we reported the results of conventional and deep learning approaches for IT tickets categorization. As far as we understand, this is the first work applying a deep neural network model on such task. In future works, we plan to explore in more neural models and join the results from multiple classifiers for accurate modeling of the task.

Acknowledgements

The first author is in SAP Industrial Ph.D Program, partially funded by the Economic Development Board and the National Research Foundation of Singapore. The authors thank Kazuhiro Inaba for his assistance during the research.

References

Agarwal, S.; Sindhgatta, R.; and Sengupta, B. 2012. Smartdispatch: enabling efficient ticket dispatch in an it service environment. In *KDD*, 1393–1401.

Aggarwal, C. C., and Zhai, C. 2012. A survey of text classification algorithms. *Mining text data.* 163–222.

Bojanowski, P.; Grave, E.; Joulin, A.; and Mikolov, T. 2016. Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606*.

Diao, Y.; Jamjoom, H.; and Loewenstern, D. 2009. Rule-based problem classification in it service management. In *International Conference of Cloud Computing*, 221–228.

Kim, Y. 2014. Convolutional neural networks for sentence classification. In *EMNLP*.

Sun, A. 2012. Short text classification using very few words. In *SIGIR*, 1145–1146.

Zhang, X.; Zhao, J.; and LeCun, Y. 2015. Character-level convolutional networks for text classification. In *NIPS*, 649–657.