Extract Clinical Lab Tests From Electronic Hospital Records Through Featured Transformer Model

Lucy M. Lu, Arkansas Bioscience Institute, USA

Richard S. Segall, Arkansas State University, USA*

iD https://orcid.org/0000-0001-7627-2609

ABSTRACT

Natural language, as a rich source of information, has been used as the foundation of the product review, the demographic trend, and the domain specific knowledge bases. To extract entities from texts, the challenge is, free text is so sparse that missing features always exist which makes the training processing incomplete. Based on attention mechanism in deep learning architecture, the authors propose a featured transformer model (FTM) which adds category information into inputs to overcome missing feature issue. When attention mechanism performs Markov-like updates in deep learning architecture, the importance of the category represents the frequency connecting to other entities and categories and is compatible with the importance of the entity in decision-making. They evaluate the performance of FTM and compare the performance with several other machine learning models. FTM overcomes the missing feature issue and performs better than other models.

KEYWORDS

Attention Mechanism, Conditional Models, Missing Features, Natural Language Processing, Transformer

1. INTRODUCTION

Natural language is the most common means of providing information to different people, to different organizations, and to different subjects; therefore, text has become a rich source of information. However, with the development of information technology, the amount of documentation has become so great that it is impossible for humans to handle. We need to take advantage of the capacity of computers to automatically understand documentation and to group documentation into different categories, topics, and subjects so that the total amount of information becomes manageable for humans.

Text mining can be applied to question answering, spam detection, semantic analysis, news categorization, and content classification, to name a few uses. Depending on the application, text data can come from different sources, such as web pages, emails, chats, social media, tickets, product

descriptions, invoices, insurance claims, user reviews, and so on. Due to the unstructured nature of the medium, it is challenging and time-consuming to extract information from text.

However, there are many open issues and challenges in text mining, such as synonyms, long-range dependencies, and multiple interacting features. Besides these challenges in the perspective of semantic analysis, we can also see issues from the perspective of data quality, such as missing features, lack of samples, and imbalanced classes. When these issues can be solved only through data modeling, not through data collection, the problem-specific data processing needs to be added to algorithms to overcome the limits on data quality. Another challenge is that the integration of domain knowledge could play an important role in text mining. Domain knowledge can help speed up text processing and increase the precision of the results. Domain-specific knowledge extraction requires semantic analysis to extract the association between the objects or concepts in the documentation. It is still challenging to make the semantic analysis efficient and scalable.

Feature engineering is an important step in machine learning. Typical text classification uses machine learning technology to perform Natural Language Processing (NLP) and to assign labels or tags to textual units, such as terms, sentences, paragraphs, documents, and queries. Normally, machine learning-based methodology performs classification in two steps: first selecting interesting features and then feeding features into classifiers to make predictions. In the training set, because the feature set plays the role of a shortcut of the context, it needs to be complete so that trained models can be used to retrieve results from new data. In other words, when the feature set is incomplete, only partial results can be retrieved by the trained models. Unlike traditional machine learning, deep learning trains word embeddings as the starting point of the classification. Feature engineering is done during model training by adjusting feature weights. However, the feature set still needs to be complete.

In terms of data modeling, text classification can be divided into two categories, one based on maximum likelihood and one based on minimum energy. For maximum likelihood-based methodology, we have Naive Bayes, Support Vector Machines (SVM), etc. For energy-based methodology, we have the Hidden Markov Model (HMM), the Conditional Random Fields model (CRF), etc. The difference between the two categories is not only in the technical details but also in how many language patterns can be modeled. Normally, maximum likelihood-based methodologies consider words to be independent tokens and use Bag of Words (BoW) to build sample sets. Minimum energy-based methodologies can fit models not only with individual words but also with the associations between words, making it possible to conduct semantic analysis. Deep learning is a separate architecture that trains word embeddings, and the classification layer is the last layer in the architecture.

Deep learning architecture is built upon neural networks. Neural approaches have the advantage of overcoming the limitations of feature engineering. Word embeddings convert input texts into an importance vector in which some words have higher significance and some words have lower significance. In this way, the words with higher significance can contribute more to the classification process and the words with lower significance can contribute less to classification process. It is optional to reduce the number of dimensions, but when the word embeddings are built, feature engineering is done.

Embedding models have a long history. The earliest embedding model is Latent Semantic Analysis (LSA; Dumais et al., 1988). LSA is based on dimension reduction. The dimensions with larger Eigenvalues are considered more significant and can be kept. The dimensions with smaller Eigenvalues are considered less significant and can be removed. The neural network model was proposed by Bengio et al. in 2001 and was based on a feed-forward neural network. However, the early embedding models underperformed classical models, so they were not generally used. In 2013 Google proposed word2vec models that were trained on billions of words and could be applied to many NLP tasks. In 2017, a contextual embedding model based on a three-layer bidirectional Long Short-Term Memory (LSTM) was trained on one billion words; this model performed better than word2vec because it could capture context and perform semantic analysis. The same year Google

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/extract-clinical-lab-tests-from-electronic-hospital-records-through-featured-transformer-model/336529

Related Content

The Need to Transform the Core Values of Medical Care and Health Organizations
Shaista Tayabaliand Carmel M. Martin (2013). *Clinical Solutions and Medical Progress through User-Driven Healthcare (pp. 85-92).*
www.irma-international.org/chapter/need-transform-core-values-medical/67739

Incremental Neural Network Training for Medical Diagnosis
Sheng-Uei Guan, Ji Hua Ang, Kay Chen Tanand Abdullah Al Mamun (2008). *Encyclopedia of Healthcare Information Systems (pp. 720-731).*
www.irma-international.org/chapter/incremental-neural-network-training-medical/13005

Post Thoracic Surgery Life Expectancy Prediction Using Machine Learning
Akshaya Ravichandran, Krutika Mahulikar, Shreya Agarwaland Suresh Sankaranarayanan (2021). *International Journal of Healthcare Information Systems and Informatics (pp. 1-20).*
www.irma-international.org/article/post-thoracic-surgery-life-expectancy-prediction-using-machine-learning/279344

Four Different Types of Classification Models
Hans Rudolf Straub (2002). *Knowledge Media in Healthcare: Opportunities and Challenges (pp. 57-82).*
www.irma-international.org/chapter/four-different-types-classification-models/25407

Incisional Hernia in Pregnancy
Sujoy Dasgupta (2012). *International Journal of User-Driven Healthcare (pp. 60-62).*
www.irma-international.org/article/incisional-hernia-pregnancy/75183