A Deep Learning Approach to Understanding Cloud Service Level Agreements

Srishty Saha, Karuna P. Joshi, Aditi Gupta, Renee Frank
Computer Science and Electrical Engineering University of Maryland, Baltimore County
{srishty1, kjoshi1, adigupta, rmfrank}@umbc.edu

INTRODUCTION

Educational organizations, like Universities and School Systems, are rapidly adopting Cloud based services to provide Information Technology (IT) infrastructure to their students. These include course offerings, class materials**,** data storage, emailing and collaboration software, virtual computing environment, etc. Moreover, cloud providers, like Amazon, are also providing free computing credits targeted to students. The legal documents associated with cloud based services, such as Service Level Agreements (SLAs), provide information regarding quality and use of cloud services. These documents are often long text-based documents containing domain specific terminology. In addition, this terminology varies from one document or service provider to another. We propose a framework to extract semantically similar terms and entities across cloud service documents using word embeddings and neural networks. Our work is intended to aid cloud service consumers across a variety of fields by providing the ability to understand the services and requirements offered by large-scale commercial cloud services. In some of our previous papers, we have used semantic web and natural language processing to analyze SLA and privacy policy documents for cloud services [1, 2, 3]. In this work, we extend our approach to propose a deep learning-based technique to analyze these documents and populate cloud service ontologies. The preliminary analysis of cloud SLAs documents performed by us showed that deep learning techniques are useful in context disambiguation and identifying semantically similar terminology across services.

METHODOLOGY

This section describes the modules for our proposed framework for automatic analysis of cloud SLA documents. Figure 1 shows the architecture of the overall system. In our previous work [1, 2, 3], we developed ontologies for cloud services, and now are using deep learning techniques, like vectorization, to populate and reason over the ontologies and create vectorized knowledge graphs.

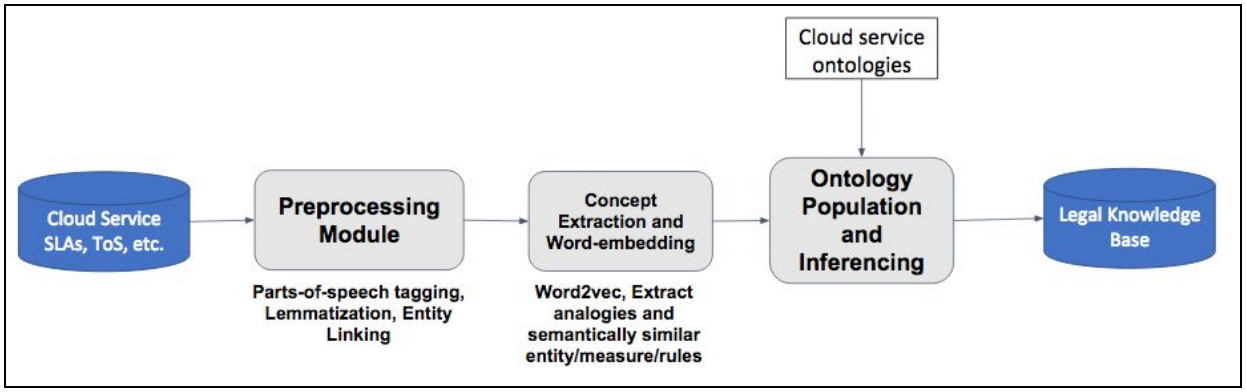

Figure1: Overall system architecture

We have created a repository of various Cloud legal documents like SLAs, Terms of Service (ToS) and Privacy policy documents of prominent cloud providers such as IBM, Google, Microsoft and Amazon. We extract text and preprocess (parts-of-speech tagging, lemmatization, entity linking) these documents before building word embedding models using the deep learning architecture of TensorFlow [6]. Obtaining analogies and semantically similar words using deep learning models will help users in understanding terminology and definitions across service providers. Figure 2 demonstrates the expected output from our proposed architecture. For example, one service provider may define the performance of cloud services in terms of metric such as *availability*, while another uses the *uptime* or *downtime* metrics, in these scenarios the word-embeddings will be able to extract such analogies across service providers. We performed preliminary analysis with a few SLA documents. We used the word2vec model for word embeddings which is essentially a neural network architecture utilizing a continuous bag-of-words model or skip-gram model to predict contextually similar words [4].

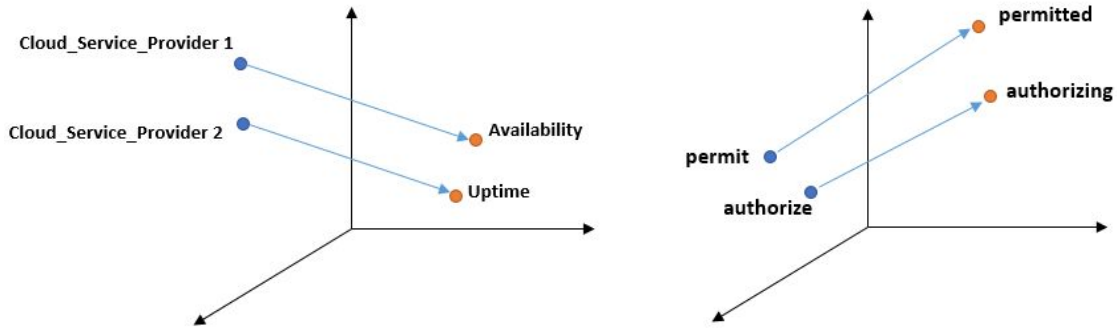

Figure 2: Expected output from the word-embedding models for cloud services.

CONCLUSION AND FUTURE WORK

In conclusion, with rapid adoption of cloud based services to educate students, it is essential for educational institutions to continuously monitor the performance of these cloud services to ensure data security and privacy. This work presented a deep learning and semantic web based approach for automating monitoring of Cloud legal documents. We have obtained preliminary results by building word-embedding models and these are being evaluated by a legal expert on our team. We will also enhance our dataset to add more open source SLAs, Terms of Services and Privacy Policies as well as other cloud services documents. The long term goal is to develop a question and answer system to respond to natural language-based questions of cloud service consumers.

REFERENCES

[1] Karuna P Joshi, Aditi Gupta, Sudip Mittal, Claudia Pearce, Anupam Joshi, and Tim Finin.Semantic Approach to Automating Management of Big Data Privacy Policies. InProceedings, IEEE BigData, 2016.

[2] Karuna P. Joshi, Aditi Gupta, Sudip Mittal, Claudia Pearce, Anupam Joshi and Tim Finin. ALDA : Cognitive Assistant for Legal Document Analytics. AAAI Fall Symposium, 2016.

[3] Aditi Gupta, Sudip Mittal, Karuna P. Joshi, Claudia Pearce and Anupam Joshi Streamlining Management of Multiple Cloud Services. In Proceedings, IEEE International Conference on Cloud Computing, June 2016.

[4] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781

[5] Quoc Le,Tomas Mikolov. Distributed Representations of Sentences and Documents. International Conference on Machine Learning, 2014.

[6] Martin Abadi et. al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems