

# RESEARCH PAPERS(NLP)

## 1. A NEURAL PROBABILISTIC LANGUAGE MODEL

### PROBLEM STATEMENT

**ABSTRACT:** A goal of statistical language modeling is to learn the joint probability function of sequences of words in a language. This is difficult because of the curse of dimensionality. a word sequence on which the model will be tested is likely to be different from all the word sequences seen during training. We propose to fight the curse of dimensionality by learning a distributed representation for words which allows each training sentence to inform the model about an exponential number of semantically neighboring sentences. This experiments using neural networks for the probability function, showing on two text corpora that the proposed approach significantly improves on state and that the proposed approach allows to take advantage of longer contexts.

**CONCLUSION:** The experiments on two corpora, one with more than a million examples, and a larger one with above 15 million words. It shown that the proposed approach yields much better perplexity than a state-of-the-art method, the smoothed trigram, with differences between 10 and 20% in perplexity. This approach allows to take advantage of the learned distributed representation to fight the curse of dimensionality with its own weapons. More generally, the work presented here opens the door to improvements in statistical language models brought by replacing “tables of conditional probabilities” by more compact and smoother representations based on distributed representations.

## 2: BIDIRECTIONAL ENCODER REPRESENTATION FROM TRANSFORMERS

### PROBLEM STATEMENT

**ABSTRACT:** The purpose of the study is to investigate the relative effectiveness of four different sentiment analysis techniques. They are unsupervised lexicon-based model using Sent WordNet, traditional supervised machine learning model using logistic regression, supervised deep learning model using Long Short-Term Memory (LSTM) and advanced supervised deep learning models using Bidirectional Encoder Representations from Transformers (BERT). We use publicly available labeled corpora of 50,000 movie reviews originally posted on internet movie database (IMDB) for analysis using Sent WordNet lexicon, logistic regression, LSTM, and BERT. This is to compare the advanced pre-trained supervised deep learning model of BERT via other sentiment analysis models of LSTM, logistic regression, and Sent WordNet.

**CONCLUSION:** The study examined the different sentiment analysis techniques on a publicly available labeled dataset of 50,000 IMDB movie reviews. Sentiment classification was carried out using unsupervised lexicon based model using Sent WordNet, supervised machine learning model using logistic regression, supervised deep learning model using LSTM, and advanced supervised deep learning model using pre-trained BERT. A comparative analysis of the four models reveal undisputed superiority of the pre-trained BERT model in sentiment classification. The study shows comparative classification performance evaluation of various sentiment analysis techniques on a labeled corpora. The insight generated study can be used by academicians and industry experts executing sentiment analysis for improved sentiment classification using a proven superior technique.

### **3.DEEP CONTEXTUALIZED WORD REPRESENTATIONS**

#### **PROBLEM STATEMENT**

**ABSTRACT:** We use a new type of *deep contextualized* word representation that models both complex characteristics of word use e.g., syntax and semantics and how these uses vary across linguistic contexts. The word vectors are learned functions of the internal states of a deep bidirectional language model which is pre trained on a large text corpus. These representations can be easily added to existing models and significantly improve the state of the art across six challenging NLP problems, including question answering, textual and sentiment analysis. It also present an analysis showing that exposing the deep internals of the pre-trained network is crucial, allowing downstream models to mix different types of semi-supervision signals.

**CONCLUSION:** This is a general approach for learning high-quality deep context-dependent representations from BILMS and shown large improvements when applying ELMO to a broad range of NLP tasks. Through ablations and other controlled experiments, we have also confirmed that the BILM layers efficiently encode different types of syntactic and semantic information about words in-context, and that using all layers improves overall task performance.

### **4.DEEP RESIDUAL LEARNING FOR IMAGE CLASSIFICATION**

#### **PROBLEM STATEMENT**

**ABSTARCT:** Deeper neural networks are more difficult to train. We present a residual learning framework to ease the training of networks that are substantially deeper than those used previously. We explicitly reformulate the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions. We provide comprehensive empirical evidence showing that these residual networks are easier to optimize, and can gain

accuracy from considerably increased depth. The depth of representations is of central importance for many visual recognition tasks.

CONCLUSION: This deep convolutional neural networks have led to a series of breakthroughs for image classification. Deep networks naturally integrate low/mid/high level features and classifiers in an end-to-end multilayer fashion, and the levels of features can be enriched by the number of stacked layers . The training accuracy indicates that not all systems are similarly easy to optimize. We have tested various plain/residual nets, and have observed consistent phenomena. So by this the image will be classified.

## **5.FINE TUNED LANGUAGE MODELS FOR HINGLISH SENTIMENT DETECTION**

### **PROBLEM STATEMENT**

ABSTRACT: Sentiment analysis for code-mixed social media text continues to be an under-explored area. This work adds two common approaches. They are fine-tuning large transformer models and sample efficient methods like ULMFIT. Prior work demonstrates the efficacy of classical ML methods for polarity detection. Fine-tuned general-purpose language representation models, such as those of the BERT family are benchmarked along with classical machine learning and ensemble methods.

CONCLUSION: It demonstrate that ensembles of classical Machine Learning models, even NB-SVM exhibit competitive performance and can in fact be better than some Transformer baselines. It is still worthwhile to implement simple classical baselines. Additionally, we hope that the released dataset and models 2 will encourage readers to investigate this further.

## **6. Enhance Text-to-Text Transfer Transformer with Generated Questions for Thai Question Answering**

### **PROBLEM STATEMENT**

ABSTRACT: Question Answering is a natural language processing task that enables the machine to understand a given context and answer a given question. There are several QA research trials containing high resources of the English language. However, Thai is one of the languages that have low availability of labeled corpora in QA studies. According to previous studies, while the English QA models could achieve more than 90% of F1 scores, Thai QA models could obtain only 70% in our baseline. In this study, we aim to improve the performance of Thai QA models by generating more question-answer pairs with Multilingual Text-to-Text Transfer Transformer along with data preprocessing methods for Thai.

**CONCLUSION:** In this paper, we propose to employ transformer-based models for Thai QA, which aims to improve the performance of the Thai question answering model. The limitation of the low resource Thai QA corpora can be overcome by using a data generation composed of three steps: answer generation, question generation, and question filtration. However, a limitation of our work is that our data generation technique is appropriate with a span-extraction question answering task; the answer of the given question is part of the given context only.

## **7. STRUCTURED NEURAL SUMMARIZATION**

### **PROBLEM STATEMENT**

**ABSTRACT:** Summarization of long sequences into a concise statement is a core problem in natural language processing, requiring non-trivial understanding of the input. Based on the promising results of graph neural networks on highly structured data, we develop a framework to extend existing sequence encoders with a graph component that can reason about long-distance relationships in weakly structured data such as text. In an extensive evaluation, we show that the resulting hybrid sequence-graph models outperform both pure sequence models as well as pure graph models on a range of summarization tasks.

**CONCLUSION:** A framework for extending sequence encoders with a graph component that can give additional structure. In an evaluation on three different summarization tasks, we have shown that this augmentation improves the performance of a range of different sequence models across all tasks. We are excited about this initial progress and look forward to deeper integration of mixed sequence-graph modeling in a wide range of tasks across both formal and natural languages. The key insight, which we believe to be widely applicable, is that inductive biases induced by explicit relationship modeling are a simple way to boost the practical performance of existing deep learning systems.

## **8. Semi-supervised Sequence Learning**

### **PROBLEM STATEMENT**

**ABSTRACT:** The two approaches that use unlabeled data to improve sequence learning with recurrent networks. The first approach is to predict what comes next in a sequence, which is a conventional language model in natural language processing. The second approach is to use a sequence autoencoder, which reads the input sequence into a vector and predicts the input sequence again. These two algorithms can be used as a pre training step for a later supervised sequence learning algorithm.

CONCLUSION: With pretraining, we are able to train long short term memory recurrent networks up to a few hundred timesteps, thereby achieving strong performance in many text classification tasks, such as IMDB and 20 Newsgroups.

## **9. GLOVE: Global Vectors for Word Representation**

### **PROBLEM STATEMENT**

ABSTRACT: Recent methods for learning vector space representations of words have succeeded in capturing fine-grained semantic and syntactic regularities using vector arithmetic, but the origin of these regularities has remained opaque. We analyze and make explicit the model properties needed for such regularities to emerge in word vectors. The result is a new global bilinear regression model that combines the advantages of the two major model families in the literature. Our model efficiently leverages statistical information by training only on the nonzero elements in a word-word cooccurrence matrix, rather than on the entire sparse matrix or on individual context windows in a large corpus. It also outperforms related models on similar tasks and named entity recognition.

CONCLUSION: Recently, considerable attention has been focused on the question of whether distributional word representations are best learned from count-based. In contrast, noise-contrastive estimation is an approximation which improves with more negative samples. The result, GLOVE, is a new global log-bilinear regression model for the unsupervised learning of word representations that outperforms other models on word analogy, word similarity, and named entity recognition tasks.

## **10. Neural Machine Translation by Jointly Learning to Align and Translate**

### **PROBLEM STATEMENT**

ABSTRACT: Neural machine translation is a recently proposed approach to machine translation. Unlike the traditional statistical machine translation, the neural machine translation aims at building a single neural network that can be jointly tuned to maximize the translation performance. The models proposed recently for neural machine translation often belong to a family of encoder-decoders and encode a source sentence into a fixed-length vector from which a decoder generates a translation. Furthermore, qualitative analysis reveals that the soft-alignments found by the model agree well with our intuition.

CONCLUSION: The conventional approach to neural machine translation, called an encoder-decoder approach, encodes a whole input sentence into a fixed-length vector from which a

translation will be decoded. Extended the basic encoder–decoder by letting a model search for a set of input words, or their annotations computed by an encoder, when generating each target word. One of challenges left for the future is to better handle unknown, or rare words. This will be required for the model to be more widely used and to match the performance of current state-of-the-art machine translation systems in all contexts.

## **11. Sequence to Sequence Learning with Neural Networks**

### **PROBLEM STATEMENT**

**ABSTRACT:** Deep Neural Networks (DNNs) are powerful models that have achieved excellent performance on difficult learning tasks. Although DNNs work well whenever large labeled training sets are available, they cannot be used to map sequences to sequences. In this paper, we present a general end-to-end approach to sequence learning that makes minimal assumptions on the sequence structure. Our method uses a multilayered Long Short-Term Memory (LSTM) to map the input sequence to a vector of a fixed dimensionality.

**CONCLUSION:** In this work, we showed that a large deep LSTM, that has a limited vocabulary and that makes almost no assumption about problem structure can outperform a standard SMT-based system whose vocabulary is unlimited on a large-scale MT task. The success of our simple LSTM-based approach on MT suggests that it should do well on many other sequence learning problems, provided they have enough training data. These results suggest that our approach will likely do well on other challenging sequence to sequence problems.

## **12.WORD REPRESENTATION IN VECTOR SPACE**

### **PROBLEM STATEMENT**

**ABSTRACT:** The model architectures for computing continuous vector representations of words from very large data sets. The quality of these representations is measured in a word similarity task, and the results are compared to the previously best performing techniques based on different types of neural networks. We observe large improvements in accuracy at much lower computational cost. Furthermore, we show that these vectors provide state-of-the-art performance on our test set for measuring syntactic and semantic word similarities.

**CONCLUSION:** In this paper we studied the quality of vector representations of words derived by various models on a collection of syntactic and semantic language tasks. We observed that it is possible to train high quality word vectors using very simple model architectures, compared to the popular neural network .. We believe that our comprehensive test set will help the research community to improve the existing techniques for estimating the word vectors.

## **13. A survey of named entity recognition and classification**

### **PROBLEM STATEMENT**

**ABSTRACT:** In defining the task, people noticed that it is essential to recognize information units like names, including person, organization and location names, and numeric expressions including time, date, money and percent expressions. Identifying references to these entities in text was recognized as one of the important sub-tasks of IE and was called “Named Entity Recognition and Classification. The first section of this survey presents some observations on published work from the point of view of activity per year, supported languages, preferred textual genre and domain, and supported entity types.

**CONCLUSION:** The Named Entity Recognition field has been thriving for more than fifteen years. It aims at extracting and classifying mentions of rigid designators, from text, such as proper names, biological species, and temporal expressions. In this survey, we have shown the diversity of languages, domains, textual genres and entity types covered in the literature. More than twenty languages and a wide range of named entity types are studied. However, most of the work has concentrated on limited domains and textual genres such as news articles and web pages.

## **14. The Impact of AI Language Models on the Future of White-Collar Jobs: A Comparative Study of Job Projections in Developed and Developing Countries**

### **PROBLEM STATEMENT**

**ABSTRACT:** The rapid advancements in artificial intelligence (AI) and natural language processing, particularly the development of AI language models, have raised concerns about their potential impact on the future of white-collar jobs. This research paper presents a comparative study of job projections in developed and developing countries, focusing on the implications of AI language models for white-collar occupations. Based on the findings, policy recommendations are provided for both developed and developing countries to ensure a smooth transition to a future workforce increasingly influenced by AI language models

**CONCLUSION:** The advent of AI language models, such as GPT-4, has the potential to significantly impact white-collar jobs across various sectors in both developed and developing countries. While these technologies offer numerous benefits, such as increased efficiency, cost savings, and improved decision-making, they may also give rise to challenges like job displacement, skill gaps, and exacerbating existing inequalities. To address these challenges and harness the potential of AI language models, policymakers and stakeholders in both

developed and developing countries must adopt a proactive and collaborative approach to policy development.

## **15. Attention Is All You Need**

### **PROBLEM STATEMENT**

**ABSTRACT:** The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train.

**CONCLUSION:** In this work, we presented the Transformer, the first sequence transduction model based entirely on attention, replacing the recurrent layers most commonly used in encoder-decoder architectures with multi-headed self-attention the Transformer can be trained significantly faster than architectures based on recurrent or convolutional layers. We plan to extend the Transformer to problems involving input and output modalities other than text and to investigate local, restricted attention mechanisms to efficiently handle large inputs and outputs such as images, audio and video.