

# test

Adrian Eriksen



Thesis submitted for the degree of  
Master in Language Technology  
60 credits

Department of Informatics  
Faculty of mathematics and natural sciences

UNIVERSITY OF OSLO

Autumn 2020



test

© 2020 Adrian Eriksen

test

<http://www.duo.uio.no/>

Printed: Reprosentralen, University of Oslo

## Introduction

In this essay I will explain some of the main challenges of adversarial data augmentation for intent classification for chatbots. Intent classification is the process of trying to understand the intent behind a statement or document using machine learning. Chatbots use these intents to try and select the response that is most likely to correspond with the users query. There are many different ways to formulate an intent, and even among these we find other challenges like grammatical errors, slang and abbreviations. One of the most promising ways to combat these challenges is adversarial data augmentation. In short, we want to find a way to make the chatbots more robust against human error, by training the chatbots on data containing mistakes. There has been a lot of recent research on adversarial data augmentation as it's relevancy has grown considerably, but we have yet to find a good solution. In my thesis, I will look at different strategies for adversary generation and measure their effectiveness across neural architectures and chatbot domains, focusing primarily on the Norwegian language.

- Explain structure of essay -

Kindly: What's relevant?, Background, Platform, How will our work benefit their goals?

Chatbots: Dialogue Systems, Reading comprehension (SQUAD - Stanford Question Answering Dataset), Response selection

Pretraining: BERT/ELMO, task specific

Adversarial Learning: Data augmentation

# Kindly

## Background

My thesis will be written in close cooperation with Kindly. Kindly is a language technology group based in Oslo. Kindly was founded in 2016 and has since grown to a team of over 40 employees, several of whom have graduated from the Language Technology Group (LTG) at the University of Oslo with master's and doctoral degrees. Natural language processing and machine learning are at the core of the Kindly chatbot platform, which powers the conversational agents for some of the leading enterprises in the Nordics, such as Norwegian Air Shuttle, Elkjøp, Kahoot!, Thon Hotels, and Finn. Kindly aims to make an easy and accessible platform for businesses to create their own chatbots tailored to their needs. In my thesis I will rely heavily on research already made by Kindly, and I will make use of Kindly's existing platform for chatbots.

## Kindly's chatbot platform

Chatbots for businesses in general want to provide a user with answers to their questions, much like an agent working at customer service. For a given business there is normally not too many different questions that a customer can ask, but for each one of the questions there is a lot of different ways to formulate them. Also, humans are faulty, which may lead to grammatical errors and badly formulated questions that confuses the chatbot. The way Kindly's chatbot works is that for each intent (question), they have a collection of samples that triggers a reply. The collection of samples consist of different ways to formulate a certain question. Kindly also include a simple system for data augmentation that let's their users expand their samples by adding synonyms. If a customer wants to rent a room, they might use the word "rent", "purchase", "buy" and so on. By letting the users append all of these words to the collection of samples, the chatbot will treat the words as if they were identical to their given synonym.

## Intent Classification

Intent classification is one of the most widely used ways in which to train a chatbot. Intent classification is the practice of assigning the most likely class to a given input of text. In simpler tasks like deciding whether a movie review is positive or negative, we only have two classes. In more complex tasks like replacing human-operated customer service with a chatbot, we usually need a large collection of predefined task-specific classes. After assigning a class, we then choose the appropriate response and return this to the user.

Machine learning is a natural way to train a chatbot for intent classification, as this is a classification problem. Kindly uses response selection for intent classification, which given an input, retrieves the most relevant response from a collection of samples. One of the benefits of using response selection is that we can train on a large unlabelled corpora of conversational data, which is easily accessible on the Internet. [?].

First, we train a language model (LM). The goal of the LM is to try and comprehend what the customer is saying, by understanding the context in which it is said.

## Pretraining

Pretraining language models is one of the key components of neural language understanding. Historically we have pretrained a model for a single task. If we wanted to make a model that can predict whether a movie review is positive or negative, we would feed the model a large amount of labelled reviews so that it may learn the differences. The challenges this raised lies in how exactly this data should be stored in the model, and figuring out what to focus on when learning. In NLP there are a lot of ways to approach text processing. The most intuitive way might be to split the words by sentences (punctuations) and whitespaces (words) and have the model learn which words and sentences are the most recurring in the different reviews, disregarding universally common words like 'the'. In addition, n-grams have been used to try and get a deeper understanding of context. Instead of only looking at one word at a time, we look at the current word as well as the n previous words. This is just one of many techniques that can be applied to improve the model. Others include POS-tagging, NP chunking and Named Entity Recognition to mention some.

There are however a couple of challenges when considering these different types of text processing for pretraining language model. A lot of text processing requires a large amount of manual labour beforehand. While splitting sentences and words in most cases are a trivial task for a program to solve, POS-tagging and NER are not. There is a fairly large amount of data that already have this, namely the Brown Corpus, but limiting language models to only learn from one

source would be crippling in regards to making newer and better models. In addition, if we would feed all the additional information about every word and sentence to the model, the computational power required would be insurmountable.

Solution:

Word vectorization

Pretraining word representations has become (one of the key components) of neural language understanding models in recent years. The main goal of pretraining word representations is to model complex characteristics of word use and how these uses vary across linguistic contexts. [?] When we pretrain word representations, we usually do so on a large text corpus.

## Transfer Learning

Transfer learning is a means to extract knowledge from a source setting and apply it to a different target setting