My beautiful dissertation

Tobin South

January 7, 2020

Thesis submitted for the degree of
Masters of Philosophy
in
Applied Mathematics
at The University of Adelaide
Faculty of Engineering, Computer and Mathematical Sciences
School of Mathematical Sciences

ii

Contents

List of Tables

List of Figures

Signed Statement

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed: . . . .   Date: . . . .

Acknowledgements

Dedication

Abstract

In this thesis

Chapter 1

Background

1.0.1 Introduction

Natural Language Processing

Natural language processing (NLP) is the area of study in which 'natural' human language is examined via machine. Natural language refers to either spoken or written language, designed to be understandable to a human listener or reader. This language is not explicitly designed to be machine understandable, and machine comprehension of this language is a challenging problem [[TODO: cite: the challenges of NLP]].

NLP is a broad term covering many models and techniques to computationally extracting meaningful information from text, ranging from the simple extraction of individual words, to the extraction of deeper semantic meaning.

Early work in NLP focused around simple grammatical rules and small vocabularies, such as the work of Georgetown-IBM [[TODO: cite: John Hutchins. From first conception to first demonstration: the nascent years of machine translation, 1947–1954. a chronology. Machine Translation, 12(3):195–252, 1997.]] to translate 60 sentences from Russian to English in 1954. With the rapid increase in computational power and digital text corpuses, modern NLP has focused or deeper challenges of extracting meaning from text with tools such as Word2Vec [[TODO: cite: word to vec]] or deep learning methods such as Google's BERT [[TODO: cite: BERT]].

These methods face a daunting challenge, language is not only complex and often duplicitous, but contextual and ever-changing. [[TODO: end better]]

1.0.2 Entropy Basics

Entropy is a measure of the uncertainty of a random variable. In the context of information theory, this is defined by **??**, often refereed to as Shannon entropy, named after Claude Shannon for his work in 1948 studying the quantiles of information in transmitted messages [[TODO: cite Claude shannon 1948]].

Definition 1.0.1 (Shannon Entropy)**.** Let X be a discrete random variable with alphabet $\mathcal{X}$ and probability mass function $p(x) = P(X = x), x \in \mathcal{X}$. The entropy $H(X)$ of the discrete random variable X, measured in bits, is

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x) \tag{1.1}$$

The entropy of the random variable is measured in bits. A bit can have two states, typically 0 or 1. The entropy of a random variable is the number of bits on average that is required to describe the random variable in question. To measure the entropy in bits, we use a logarithm of base 2, and all logarithms throughout this work are assumed to be in based 2, unless otherwise specified.

To give a typical example of entropy, if a fair coin is tossed there are two equally probable outcomes, giving an entropy of 1 bit. Further, we use the convention of $0 \log 0 = 0$, which sensibly means that adding a state with 0 probably to the random variable does not change it's entropy.

Remark (Suprise)*.* The entropy of the random variable X can also be described in terms of the expected surprise, where the surprise of a state is $log \frac{1}{p(x)}$.

$$H(X) = \mathbb{E}\left[\frac{1}{p(x)}\right] \tag{1.2}$$

[[TODO: add some filler]]

Lemma 1.0.1. The entropy of a random variable is strictly non-negative, $H(X) \geq 0$.

Proof. $0 \leq p(x) \leq 1$ which implies that $log \frac{1}{p(x)} \geq 0$, hence the sum of products of strictly non-negative terms will always be non-negative. ∎

1.0.3 Predictability

Chapter 2

Introduction

Chapter 3

a second chapter

Chapter 4

What is this chapter about?

Appendix A

Tandem queue processes

Appendix B

second appendix

Bibliography