



DEPARTMENT OF ENGINEERING MATHEMATICS

Bayesian Deep Learning For Extractive Test Summarisation

James Stephenson

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree
of Master of Science in the Faculty of Engineering.

Monday 22nd May, 2023

Supervisor: Dr. Edwin Simpson

Declaration

This dissertation is submitted to the University of Bristol in accordance with the requirements of the degree of MSc in the Faculty of Engineering. It has not been submitted for any other degree or diploma of any examining body. Except where specifically acknowledged, it is all the work of the Author.

James Stephenson, Monday 22nd May, 2023

Contents

1	Introduction	1
1.1	Current Approaches	2
1.2	Proposed Approach	2
2	Background	3
2.1	Active Learning	3
2.2	Interactive Learning	5
2.3	Deep Learning	5
2.4	Bayesian Deep Learning	6
2.5	Deep Active Learning	8
2.6	Text Summarisation Models	8
3	Execution	9
3.1	Example Section	9
4	Critical Evaluation	13
5	Conclusion	14
A	An Example Appendix	15

List of Figures

2.1	Typical deep learning architecture [?]	6
2.2	Neural Network undergoing Monte Carlo Dropout [?]	7
3.1	This is an example figure.	10

List of Tables

3.1 This is an example table.	10
---------------------------------------	----

List of Algorithms

3.1 This is an example algorithm.	11
---	----

List of Listings

3.1 This is an example listing.	11
---	----

Abstract

Text summarisation is a valuable technique that facilitates the computational processing of documents, saving users hours in manual processing. Users have different summary requirements; however, current extractive summarisation systems construct generic summaries that are not tailored to the user's needs. Asking users for feedback is one solution to combat this problem. Thus we aim to find an approach that allows the summary to be tailored to the users whilst minimising requests for user feedback.

Legacy approaches use Bayesian optimisation [?] strategies to achieve minimal user feedback; however, this strategy is blocked since modern summarisation techniques involve deep neural networks which cannot effectively express uncertainty and are typically overconfident when encountered by new topics [?]. This poses an issue in utilising the feedback strength of Bayesian optimisation. This project will investigate the feasibility of applying newly-developed techniques from Bayesian deep learning [?] to rank summary instances or *passages*. Bayesian deep learning allows us to generate significant estimates of the model confidence, so we can use Bayesian optimisation to determine which instances to ask the user for feedback on.

Specifically, we look to utilise pre-trained deep learning models such as BERT [?] to ascertain instances in a vector format to be used in an active learning component. Monte-Carlo Dropout [?] techniques appear to be proficient approximations for parameter posterior distributions. Thus, we will look to utilise this approach to calibrate our model.

It is common in passage ranking active learning solutions to use a pool-based strategy to query unlabelled instances [?]. However, this requires excess computational processing. Thus, we will examine using a stream-based approach to identify instances to query since it provides a computationally-cheaper framework for an interactive setting. Query-by-committee acquisition functions are popular for stream-based active learning; however, since Simpson et al. [?] found Bayesian optimisation strategies effectively minimised user feedback, we will look to utilise strategies such as expected improvement since Bayesian deep learning will provide a higher level of model confidence.

Supporting Technologies

A compulsory chapter, roughly 10% of the total page-count

This chapter should describe the project context, and motivate each of the proposed aims and objectives. Ideally, it is written at a fairly high-level, and easily understood by a reader who is technically competent but not an expert in the topic itself.

In short, the goal is to answer three questions for the reader. First, what is the project topic, or problem being investigated? Second, why is the topic important, or rather why should the reader care about it? For example, why there is a need for this project, who will benefit from the project and in what way (e.g., clients/end-users who needed some analysis done, or other data scientists who might need the tools you have developed), what work does the project build on and why is the selected approach either important and/or interesting (e.g., fills a gap in literature, applies results from another field to a new problem). Finally, what are the central challenges involved and why are they significant?

The chapter should conclude with a concise bullet point list that summarises the aims, objectives, **and achievements** of your work.

Notation and Acronyms

A compulsory chapter, roughly 10% of the total page-count

This chapter should describe the project context, and motivate each of the proposed aims and objectives. Ideally, it is written at a fairly high-level, and easily understood by a reader who is technically competent but not an expert in the topic itself.

In short, the goal is to answer three questions for the reader. First, what is the project topic, or problem being investigated? Second, why is the topic important, or rather why should the reader care about it? For example, why there is a need for this project, who will benefit from the project and in what way (e.g., clients/end-users who needed some analysis done, or other data scientists who might need the tools you have developed), what work does the project build on and why is the selected approach either important and/or interesting (e.g., fills a gap in literature, applies results from another field to a new problem). Finally, what are the central challenges involved and why are they significant?

The chapter should conclude with a concise bullet point list that summarises the aims, objectives, **and achievements** of your work.

Acknowledgements

A compulsory chapter, roughly 10% of the total page-count

This chapter should describe the project context, and motivate each of the proposed aims and objectives. Ideally, it is written at a fairly high-level, and easily understood by a reader who is technically competent but not an expert in the topic itself.

In short, the goal is to answer three questions for the reader. First, what is the project topic, or problem being investigated? Second, why is the topic important, or rather why should the reader care about it? For example, why there is a need for this project, who will benefit from the project and in what way (e.g., clients/end-users who needed some analysis done, or other data scientists who might need the tools you have developed), what work does the project build on and why is the selected approach either important and/or interesting (e.g., fills a gap in literature, applies results from another field to a new problem). Finally, what are the central challenges involved and why are they significant?

The chapter should conclude with a concise bullet point list that summarises the aims, objectives, **and achievements** of your work.

Chapter 1

Introduction

Text summarisation is the process of condensing a passage of text into a shorter version whilst retaining the necessary information in the text. This is a valuable research area since summarisation massively reduces the comprehension time of large pieces of text. Moreover, it has applications in many different domains, both public and professional: academics are required to read extensive research papers, individuals read long articles to keep up to date with the world news, and individuals read books to learn about various topics from history to science.

There are two approaches to text summarisation: extractive and abstractive summarisation. Extractive summarisation is a summarisation technique which focuses on selecting particular words and sentences to convey the meaning of the original text; one would consider a system whereby several sentences are randomly selected to generate summaries an extractive model, albeit not a smart one. Whereas abstractive summarisation techniques look to understand the semantics of the text before generating text to summarise what the model has learnt; an example of such a model is Google's Pegasus text summarisation model [?].

Extractive summarisation models are more common as most practical text summarisation models are extractive [?]. The basic structure of these models is made up of three stages [?]: first, capturing the key aspects of the text; secondly, using these aspects to score the importance of each sentence; thirdly, to create a summary using the highest scoring sentences. Abstractive summarisation is, naturally, harder, and more computationally costly to perform [?]. The difficulty is centred on learning the semantics of the text; different texts can have many different structures and models find it difficult to learn such variety [?].

Passage ranking is a popular technique that is used in various natural language processing (NLP) domains, such as search engine queries [?], community question answering models [?] and text summarisation models [?]. In fact, on the 10th of February 2021, Google introduced an update to Google Search which moved their algorithm from a passage indexing approach to a passage ranking approach [?]. We will be utilising this powerful approach to rank summaries by how appropriate they are for the user's requirements. We intend to build up the passage ranking mindset outlined by Simpson et al. [?] who found this approach useful, but established that a complete ranking was not required. Instead, they concluded that an iterative comparison between the current best summary and a suitable proposed summary was all that was required.

This project focuses on assessing a new approach that uses Bayesian deep learning techniques to rank text summaries generated using extractive text summarisation models. It is necessary for our approach to achieve the following requirements:

- The ability to tailor summaries to the user's preference since there is a range of requirements that different users have.
- Of all proposed summaries, the highest-ranked summary should be the most effective at conveying the information in the original text.
- The framework should be used in an interactive setting which minimises user interactions and a timely processing speed.

1.1 Current Approaches

From current literature, models have been proposed to capture the preference of one summary to another such as the Bradley-Terry model [?] and the Thurstone-Mosteller model [?, ?]. Both the Bradley-Terry and Thurstone-Mosteller models are linear models that takes the "value" of two instances, a and b and represents the probability that a is more comparative than b using a monotonic, increasing function: $\mathcal{P}(a > b) = H(V_a - V_b)$ for "value" variable V_i [?]. These models differ in the functions that are used. The Bradley-Terry model uses a simple additive fraction, $\mathcal{P}(a > b) = \frac{p_a}{p_a + p_b}$ [?]; whereas the Thurstone-Mosteller model uses the normal cumulative distribution function for comparison [?].

These models provide good solutions; however, they fail to differentiate between aleatoric and epistemic uncertainty. This limits the models' ability to determine where there is weakness in the model and leads to reduced performance. Alternative approaches use deep learning techniques to rank passages which beat state-of-the-art performance [?]. However, such models are limited by their requirement of large training data. Which, in the context of text summarisation, comes at a high cost as human annotators are required to manually produce and evaluate summaries. Moreover, these models are unable to account for user preferences which limit the models' ability to tailor summaries to the user.

1.2 Proposed Approach

1.2.1 Research Aim

We aim to develop and evaluate a Bayesian deep learning framework for passage ranking text summaries which incorporates an active learning component to allow for user influence on the summary rankings.

Typical deep learning approaches demand vast amounts of training data; however, the active learning component will minimise the number of user interactions required, whilst maintaining high performance. During iterations within the active learning component, we use a stream-based strategy to minimise the amount of overhead processing as, for this strategy, summary instances are evaluated sequentially by the active learner as opposed to requiring a pool of unlabelled instances. As initiated by Simpson et al. [?], we will use a Bayesian optimisation acquisition function to determine if an unlabelled instance should be queried by an oracle - an all-knowing information source - or not. We also aim to use Monte Carlo Dropout [?] to approximate the posterior distribution across the model weights and calibrate our model.

Within our framework development and evaluation, we wish to establish if the proposed Bayesian deep learning approach provides a sufficient passage ranking solution in comparison to a classical deep learning model. Moreover, we aim to determine if a stream-based active learning strategy is appropriate for such a problem.

This research project will include an experimentation stage; whereby we test the proposed framework, discuss the results and draw conclusions. A range of data sets that have been used for experimentation in summary passage ranking literature: Simpson et al. use the DUC 2001, 2002 and 2004 datasets [?]; whereas, common benchmarking datasets are the CNN/Daily Mail and GovReport datasets [?, ?]. We will use one of these datasets to benchmark our results based on which extractive text summarisation model we choose - this will be discussed in Section 2.6. Since we aim to utilise an active learning component, we will use a noisy random selector to mimic user summary preferences and provide answers to queries; a similar approach was taken by Simpson et al. [?].

1.2.2 Research Concerns

The central challenge within the project is effectively combining a Bayesian deep learning model with an active learning component. Firstly, it is a concern as to whether stream-based learning is an appropriate active learning strategy for a passage ranking problem since there is minimal current documentation; pool-based active learning is the most common strategy used [?]. Secondly, consideration needs to be made with regard to the number of user interactions. It is necessary for the model to be interactive; thus, it is important to examine the number of interactions that are required and if this is a reasonable level for an interactive setting. Finally, it is a concern as to whether the noisy random simulator effectively represents a human annotator in experimentations. As it will not have a preference for a particular type of summary, it provides little information on whether the framework learns the attributes of a preferred summary and starts to regularly produce summaries of such a nature.

Chapter 2

Background

Since the crux of this project is to assess the suitability of applying Bayesian deep learning (BDL) techniques to passage ranking (PR) problems, this chapter starts by defining the key concepts that underpin BDL techniques before exploring the relevant literature that discusses previous approaches to passage ranking solutions. Once this assessment has been done, we will then also examine literature that assesses BDL as opposed to classical deep learning techniques.

2.1 Active Learning

Alongside unsupervised and supervised learning, active learning (AL) is a machine learning framework whereby queries are asked of an oracle – such as a human annotator – in the form of labelling unlabelled observations [?]. The active interactions with oracles allow better performance with few labelled data points. AL is beneficial in the cases where labelled data is scarce due to high costs; for speech recognition problems [?] details a scale factor of ten times between the length of a speech extract and the time taken to annotate such as extract.

2.1.1 Active Learning Strategies

Settles [?] describes three scenarios that are considered in the literature to categorise AL problems: membership query synthesis, stream-based selective sampling and pool-based active learning.

Membership query synthesis. Labels are requested by the learner for any unlabelled instance in the input space. This includes queries that are generated as if for the first time rather than from some causal distribution [?]. A considerable limitation of this scenario occurs when the oracle is a human annotator. Baum and Lang [?] employed membership query learning to classify handwritten characters using a human oracle. They found that many query images that were generated were unrecognisable symbols. This limitation could feasibly produce nonsense summaries when tasked with a PR situation; something we should be cautious of.

Stream-based selective sampling. In this setting, unlabelled observations are selected sequentially and the learner determines if to query or discard each instance; this is done to reduce annotation effort [?]. This is under the major assumption that acquiring unlabelled instances is low-cost since the learner needs to be able to decide it can discard the unlabelled observation with minimal opportunity cost. The most common way of defining if a sample should be queried or discarded is by creating a *version space* [?] using two models with different parameter choices; for those instances that the models agree on, we can discard as there is little uncertainty. However, with regards to the cases of disagreement, these unlabelled instances fall in the region of uncertainty [?]. This region of uncertainty is computationally expensive to calculate; thus, it is common to use approximations in practice [?, ?, ?].

Pool-based active learning. A common approach for many real-world examples such as text classification [?], information extraction [?] and speech recognition [?] since it is common to find large groups of unlabelled data collected at once. The *pool-based active learning* workflow starts with a learner trained on a small set of labelled data, \mathcal{D}_{lab} , which is then used to *greedily* rank instances in a large collection of unlabelled instances, \mathcal{D}_{unlab} [?]. The highest-ranked instance is then labelled by an oracle and then

used within the learner retrain. In comparison to a stream-based active learner, a greater computational cost is associated with a pool-based learner since it ranks the entire set \mathcal{D}_{unlab} before making a query as opposed to making sequential decisions.

2.1.2 Acquisition Functions

Whilst introducing *AL*, we have spoken about measuring the usefulness of each instance and whether to query it or not. We measure how informative an instance is using *acquisition functions*. Naturally, there is a trade-off between two types of approaches: exploration and exploitation. Exploitative strategies search the area of the current best instances; whereas exploration strategies look at instances that have greater levels of uncertainty. As expected, there are many acquisition functions currently researched; we will cover a few important ones from the areas of uncertainty sampling and Bayesian optimisation.

Uncertainty Sampling. Posed by Lewis and Gale [?], it is an explorative query framework which focuses on querying instances that have the most uncertainty. A common strategy used to calculate uncertainty for probabilistic learning models is by using Shannon’s entropy [?] given by the formula below.

$$x_{ENT}^* = \operatorname{argmax}_x - \sum_i \mathbb{P}(y_i | x; \omega) \log [\mathbb{P}(y_i | x; \omega)]$$

for y_i across the range of possible labels; in the context of whether to query or not, $y_i \in \{0, 1\}$ since we have a binary decision to make. Entropy-based acquisition functions have been generalised for more complex models so they are suitable for tree-based or multi-label classification models [?, ?]. However, uncertainty sampling suffers from a lack of sensitivity to noise and outliers as it can get very easily distracted. Uncertainty sampling also does not consider *why* the model holds uncertainty for a particular instance [?].

Expected Improvement (EI). This is an alternative, Bayesian optimisation, approach that has a strong focus on the exploitation of good instances [?]. The basic idea is that it provides an estimation of the *expected improvement* of a proposed candidate over our current best candidate. Simpson et al. [?] find this an effective acquisition function for their interactive PR model with a minimal number of user interactions. To outline how we calculate EI, we must first define *improvement* as $\max\{0, f_a - f_b\}$ with a our candidate instance, b our current best instance, and f the utility of a given instance [?]. The first assumption we make is that $\mathcal{N}(\hat{f}, \mathcal{C})$ is a good estimate for the posterior distribution of candidate utilities; second that the difference in utilities $f_a - f_b$ is Gaussian-distributed. With these assumptions, we can derive the following equation for expected improvement with $z = \frac{\hat{f}_a - \hat{f}_b}{\sqrt{v}}$ and posterior standard deviation \sqrt{v} :

$$\operatorname{Imp}(a; \mathcal{D}) = \sqrt{v} [z\Phi(z) + \mathcal{N}(z; 0, 1)]$$

Some limitations of EI is that it has been found to over-exploit in some cases [?]; since it takes a very exploitative sample, if there are inaccuracies in the estimation of the mean or variance, it does not have the explorative capabilities to find the optimal instance area.

Query-By-Committee (QBC). This acquisition function utilises a committee of models with different parameters, $\omega^1, \dots, \omega^c$, that are all trained on the same labelled dataset [?]. Optimal research size has been researched; however, even a committee size of two or three models has shown positive results in practice [?, ?, ?] providing no agreement on an appropriate committee size. QBC looks for instances that the models disagree on, making this acquisition function have a strong emphasis on exploration. It is a common acquisition function for stream-based learning [?] as it does not require a batch of unlabelled instances to make a decision. QBC does require a measure of disagreement among committee models. The two main approaches are *vote entropy* and *average Kullback-Leibler (KL) divergence*.

Vote Entropy. This is a QBC generalisation of entropy-based uncertainty sampling, defined by the following structure [?]:

$$x_{VE}^* = \operatorname{argmax}_x - \sum_i \frac{V(y_i)}{c} \log \left[\frac{V(y_i)}{c} \right]$$

where y_i ranges across all possible labels and $V(y_i)$ is the number of votes that the instance receives to be assigned label i .

Average KL divergence. This is built on KL divergence [?] to measure the average difference between two probability distributions as detailed below [?]:

$$x_K^* L = \operatorname{argmax}_x \frac{1}{C} \sum_{c=1}^C \sum_i \mathbb{P}(y_i | x; \omega^c) \log \left[\frac{\mathbb{P}(y_i | x; \omega^c)}{\mathbb{P}(y_i | x; C)} \right]$$

where $\omega^{(c)}$ represents the parameters of a particular model in the committee, C represents the committee as a whole.

Unlike vote entropy, Average KL divergence is known to miss instances when committee members disagree; whereas, a limitation of vote entropy is that it can miss informative instances since it is rooted in uncertainty sampling [?]. A shared weakness is that both metrics often fail to select enough valuable instances to achieve the same classification accuracy as passive learning.

2.2 Interactive Learning

Interactive learning is a machine learning workflow involving directed experimentation with inputs and output [?]. A rapid change in response to user input facilitates interactive inspection of the impact of the user's input. This workflow format is commonly used to solve NLP problems; related works include literature in PR in the context of translations, question answering and text summarisation [?, ?, ?]. These works had a focus on interactionally-expensive uncertainty sampling to learn the rankings of *all* candidate passages [?]. Gao et al. [?] researched how to reduce the number of user interactions for uncertainty sampling techniques with some success using an active learner. A positive step towards reasonable interactive learning.

Simpson et al. [?] take an alternative approach to uncertainty sampling by proposing a Bayesian optimisation (BO) strategy instead. With Gaussian process (GPs) displaying some success in error reduction for NLP tasks with noisy labels [?, ?], Simpson and Gurevych [?] proposed using Gaussian process preference learning (GPPL) with uncertainty sampling. This approach has been further built upon by Simpson et al. [?] to a BO framework. This approach showed a marked improvement in the accuracy of chosen answers in a community question answering (cQA) task with a small number of interactions required. The methodology used Expected Improvement (IMP) as the acquisition function for AL which twisted the focus of optimisation to find the best candidate, as opposed to the ranks of all candidates. The switch to the exploitation of promising candidates showed to be massively influential on performance. Simpson et al. [?] furthered the performance enhancement gained from using the BO framework by using prior predictions from a state-of-the-art scoring method, SUPERT [?], as an informative prior for GPPL to address the cold-start problem for recommender systems [?].

2.3 Deep Learning

Deep learning methods form a subset of machine learning, based on neural networks with at least three hidden layers. These techniques have dramatically increased the capabilities of model recognition in many domains including visual object recognition, question answering and text summarisation [?, ?, ?]. In classical training, one typically uses maximum a-posteriori (MAP) optimisation to choose the set of parameters, \hat{w} , for our model that maximises the posterior probability from our parameter distribution [?]. MAP does not require computationally-costly calculations of the marginal distribution [?]; however, since MAP is a point estimate, it cannot be fully considered a Bayesian approach [?].

2.3.1 Pre-trained Models

Pre-trained, deep learning, language models are useful in unsupervised learning problems due to the lack of major architectural modifications required and the high-performance levels that are delivered [?]. One popular pre-trained language model is the Bidirectional Encoder Representations from Transformers (BERT) which takes an entire sequence of words, bidirectionally, to produce significantly improved results. The input is augmented by three embeddings – position, segment and token embeddings – and padded by a [CLS] token at the beginning of the first sentence to ensure BERT has lots of useful information [?].

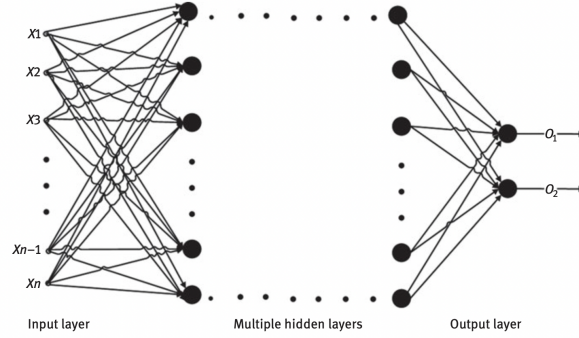


Figure 2.1: Typical deep learning architecture [?]

BERT is trained on two tasks in parallel: Masked Language Modelling, prediction of hidden words in sentences, and Next Sentence Prediction [?]. However, BERT can be applied to many NLP tasks [?] such as question answering and text classification tasks with some minor fine-tuning; we add a small layer on the top of the transformer output for the [CLS] token [?] to adapt the core model to different tasks. Recent publications have found BERT-based models [?] to be extremely effective when tasked with PR situations across the question answering and text summarisation disciplines [?, ?]. Xu et al. [?] explored a query-passage set-up when applying BERT to cQA such that the BERT final hidden state fed into an MLP module to produce relevance scores in a supervised way. Since finding this technique outperformed the baseline, it may be a useful structure to consider adapting to the text summarisation domain.

The limitation of utilising an interactive learning framework such as one outlined by Simpson et al. [?] is that it does not utilise the vast performance capabilities of newer, pre-trained techniques such as BERT. Although the framework presented does limit the number of interactions required from a user – allowing the user to tailor the summary – Ein-Dor et al. [?] look to take this idea further with the incorporation of a BERT component in an AL framework.

2.4 Bayesian Deep Learning

Bayesian Deep Learning is a deep learning approach which uses a probabilistic framework – whether that be in the model acquisition function or model parameters – to improve model performance. Bayesian acquisition functions are something we have mentioned previously; however, concerning a probabilistic approach to the selection of model parameters, ω , marginalisation is used to replace optimisation. This is so we can utilise the effect of several models using different ω with probability distribution $p(\omega)$. To allow us to marginalise over ω , we require Bayes Theorem to link the *prior distribution*, $p(\omega)$, for parameters ω ; the likelihood, $p(\mathcal{D} | \omega)$ of such parameters being suitable for data, \mathcal{D} ; and the *posterior distribution*, $p(\omega | \mathcal{D})$, of the parameters.

$$p(\omega | \mathcal{D}) = \frac{p(\mathcal{D}_y | \mathcal{D}_x, \omega)p(\omega)}{\int_{\omega'} p(\mathcal{D}_y | \mathcal{D}_x, \omega')p(\omega')d\omega'}$$

The marginalisation stage forms the integral over all possible ω on the numerator. This is important since the posterior distribution is incredibly useful to calculate the predictive distribution (or marginal probability distribution) of the output. The *predictive distribution*, $p(y | \mathcal{D}, x)$, defines the probability of label y given additional input x and dataset \mathcal{D} [?].

$$p(y | x, \mathcal{D}) = \int_{\omega} p(y | x, \omega)p(\omega | \mathcal{D})d\omega$$

This integral is called the *Bayesian Model Average (BMA)* and can be thought of as the weighted average (using probability distributions) of all parameters and defines the probability for label y given input x and data \mathcal{D} [?]. Wilson and Izmailov [?] argue that using a BMA increases accuracy as well as obtaining a realistic expression of uncertainty with classical neural networks exhibiting overconfident predictions [?]. Unfortunately, calculating the posterior distribution is a computationally expensive task, due to the marginalisation step in the denominator, so approximate posterior distributions are used.

2.4.1 Bayesian Deep Learning Strategies

Firstly, Wilson and Izamailov [?] comment that taking a selection of possible ω and combining the resulting models to approximate BMA – named Monte Carlo approximation – evocative of frequentist deep ensembles. However, there are modern approaches one can take.

A common practical method is using Monte Carlo Markov Chains (MCMC) to approximate the posterior [?]. MCMCs are used to approximate variable distributions for an idealised system [?] and two common algorithms have been tailored to approximate posterior distributions: Gibbs Sampling and the Metropolis-Hastings Algorithm. However, Gibbs Sampling is not appropriate for neural networks with conditional posterior distributions due to the interdependency of parameters [?]. Simple forms of the Metropolis-Hastings algorithm (MH) can be more appropriate; however, again due to the high interdependence of states, MH can be costly and prone to random walks [?]. Duane et al. [?] propose an alternative *hybrid Monte Carlo* which is a combination of MH with sampling techniques from dynamical simulation.

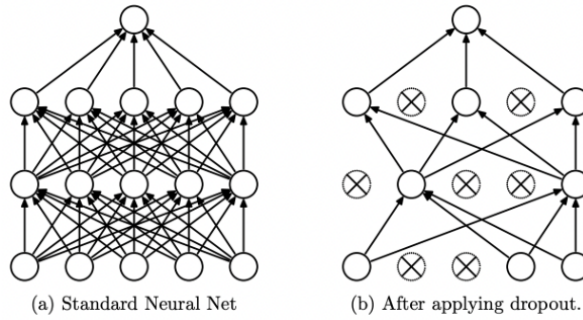


Figure 2.2: Neural Network undergoing Monte Carlo Dropout [?]

Second, Graves [?] proposed fitting a Gaussian variational posterior approximation over the parameters of neural networks and optimising over the parameters to ensure the variational distribution is as good an estimate of the posterior distribution as possible. This method works well for networks of moderate size, but supplies training difficulties when working with larger architectures [?].

Thirdly, Gal and Ghahramani [?] present Monte Carlo Dropout (MC Dropout); a dropout framework which integrates stochasticity into a neural network, by randomly removing parameters *during training*. We can interpret dropout as approximate Bayesian inference, leading to a range of different parameters. It is intuitive to see the link between this and sampling parameters from a posterior to approximate a predictive distribution.

Denoting the neural network parameter matrices for layer i as W_i alongside input and output sets \mathcal{D}_{in} and \mathcal{D}_{out} , we again suffer from an intractable posterior distribution: $p(y | x, \mathcal{D}_{out}, \mathcal{D}_{in})$. Thus $q(\omega)$ is an approximation defined below

$$\begin{aligned} z_{i,j} &\sim \text{Bernoulli}(p_i) \\ W_i &= M_i \cdot \text{diag}(z_{i,j}) \end{aligned}$$

A simple Bernoulli distribution is used to determine which states are set to zero given some probability p_i and variational parameters M_i . Note here that $z_{i,j}$ denotes unit j in layer $i - 1$. To obtain the model uncertainty obtained through dropout in neural networks, we take our approximate predictive distribution. Through T sample sets of realisations from our posterior distribution $z_{i,j}$, we get T parameter matrices $\{W_1^t, W_2^t, \dots, W_L^t\}_{t=1}^T$ and the following estimate by which we call our Monte Carlo Dropout.

$$\mathbb{E}_{q(y^*|x^*)}(y^*) \approx \frac{1}{T} \sum_{t=1}^T \hat{y}^*(x^*, W_1^t, \dots, W_L^t)$$

Another popular technique is *Stochastic Weight Averaging – Gaussian (SWAG)* [?]. This builds on the idea of *Stochastic Weight Averaging (SWA)* which combines parameters of the same neural network at different stages in training [?]. SWAG uses Stochastic Gradient Descent (SGD) information to estimate

the shape of the posterior distribution by fitting a Gaussian distribution to the first two moments of the SGD iterate [?]. We use these fitted Gaussian distributions for BMA. The benefits of SWAG are grounded in its practicality, stability and accuracy which are essential attributes when working with large neural networks [?].

2.5 Deep Active Learning

Ideally, our solution would retain the AL component that exists in the PR framework proposed by Simpson et al. [?] since it allows us to tailor generated summaries to the user preferences; an essential aspect of summary ranking.

Zhang and Zhang also explored an ensemble of AL strategies to build a deep active learning framework [?]. This was a composition of a BERT-based classifier and an ensemble sampling method to choose valuable data for training. They found that this alternative approach only required half the training data to attain state-of-the-art performance. However, the framework proposed by Ein-Dor et al. [?] may be of more use since experimentation was constructed on data with high class imbalance, scarce labelling and a small annotation budget: attributes of an interactive PR context.

Ein-Dor et al. [?] developed a framework that used an AL approach with BERT-based classification. This structure consisted of pool-based AL in batch mode in conjunction with BERT as the classification scheme. Different AL strategies were examined – MC Dropout, a Bayesian approach, and Discriminative Active Learning (DAL) [?] – with AL proving an excellent boost to helping BERT emerge from its poor initial model [?]. Although DAL would not be appropriate for the PR context due to its focus on querying batches, using MC Dropout as a strategy seems to be effective for PR.

Gal and Ghahramani [?] also present an AL framework which incorporates recent BDL techniques. AL is limited by its ability to scale to high-dimensional datasets [?], key for deep learning scenarios. Thus, Gal and Ghahramani proposed an approach that used specialised Bayesian convolutional neural networks (BCNNs) whereby Gaussian, prior probability distributions are used to describe a set of parameters as a basepoint to start inference. Like Ein-Dor et al., they also introduce MC Dropout to sample the approximate posteriors; however, Gal and Ghahramani do take a different approach by using the BALD acquisition function [?]. BALD chooses pool sizes with the greatest expectation of the information gained from the model parameters [?]; chosen since it demonstrates a small test error in experimentation.

2.6 Text Summarisation Models

Since an extractive text summarisation method to generate candidate summaries is not the central focus of our research, it suffices to use an off-the-shelf solution. Simpson et al. [?] simply take random sentences from the base text to create summaries which they found to be a sufficient approach to test their proposed PR framework since it is lightweight and can produce test summaries quickly. However, a modern, alternative model such as MemSum, proposed by Gu et al. [?] is likely to produce more realistic summaries for popular use and would be a positive replacement. The final model to mention is HAHSum: an extractive text summarisation model proposed by Jia et al. [?] which provides us with a different option to the two alternative models since it is trained on shorter documents and so is likely to produce practical summaries for a different context.

Each approach has its strengths and limitations. Namely, the approach that takes sentences randomly is unlikely to produce a practically useful summary as there is no analysis, and subsequent weighting, into the importance of each sentence. Nonetheless, it will produce a range of different summaries to test our PR framework with low computational cost [?]. Contrastingly, using a model such as MemSum gives us the opposite situation to consider; this model produces more accurate summaries, but with greater computational cost [?]. The key limitation of MemSum is that it is trained to summarise *long* documents such as ones taken from PubMed, arXiv and GovReport so its viability will depend on the data we use in experimentation. If MemSum does not prove viable, there are comparable extractive models trained on shorter documents such as HAHSum [?]. Unfortunately, since MemSum and HAHSum were trained on different datasets, we do not have a direct performance comparison; however, HAHSum outperforms previous extractive summarisers on the CNN/Daily Mail dataset [?, ?] so is likely to provide suitable summaries to rank.

Chapter 3

Execution

A topic-specific chapter, roughly 30% of the total page-count

This chapter is intended to describe what you did: the goal is to explain the main activity or activities, of any type, which constituted your work during the project. The content is highly topic-specific. For some projects it will make sense to split the content into two main sections, or maybe even into two separate chapters: one will discuss the design of something, including any rationale or decisions made, and the other will discuss how this design was realised via some form of implementation. You could instead give this chapter the title “Design and Implementation”; or you might split this content into two chapters, one titled “Design” and the other “Implementation” .

Note that it is common to include evidence of “best practice” project management (e.g., use of version control, choice of programming language and so on). Rather than simply a rote list, make sure any such content is useful and/or informative in some way: for example, if there was a decision to be made then explain the trade-offs and implications involved.

3.1 Example Section

This is an example section; the following content is auto-generated dummy text. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus

foo

Figure 3.1: This is an example figure.

foo	bar	baz
0	0	0
1	1	1
⋮	⋮	⋮
9	9	9

Table 3.1: This is an example table.

convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

3.1.1 Example Sub-section

This is an example sub-section; the following content is auto-generated dummy text. Notice the examples in Figure 3.1, Table 3.1, Algorithm 3.1 and Listing 3.1. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat

```
for i = 0 upto n do
  | ti ← 0
end
```

Algorithm 3.1: This is an example algorithm.

```
for( i = 0; i < n; i++ ) {
  t[ i ] = 0;
}
```

Listing 3.1: This is an example listing.

magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Example Sub-sub-section

This is an example sub-sub-section; the following content is auto-generated dummy text. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis

dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Example paragraph. This is an example paragraph; note the trailing full-stop in the title, which is intended to ensure it does not run into the text.

Chapter 4

Critical Evaluation

A topic-specific chapter, roughly 30% of the total page-count

This chapter is intended to evaluate what you did. The content is highly topic-specific, but for many projects will have flavours of the following:

1. functional testing, including analysis and explanation of failure cases,
2. behavioural testing, often including analysis of any results that draw some form of conclusion wrt. the aims and objectives, and
3. evaluation of options and decisions within the project, and/or a comparison with alternatives.

This chapter often acts to differentiate project quality: even if the work completed is of a high technical quality, critical yet objective evaluation and comparison of the outcomes is crucial. In essence, the reader wants to learn something, so the worst examples amount to simple statements of fact (e.g., “graph X shows the result is Y”); the best examples are analytical and exploratory (e.g., “graph X shows the result is Y, which means Z; this contradicts [1], which may be because I use a different assumption”). As such, both positive *and* negative outcomes are valid *if* presented in a suitable manner.

Chapter 5

Conclusion

A compulsory chapter, roughly 10% of the total page-count

The concluding chapter(s) of a dissertation are often underutilized because they're too often left too close to the deadline: it is important to allocate enough time and attention to closing off the story, the narrative, of your thesis.

Again, there is no single correct way of closing a thesis.

One good way of doing this is to have a single chapter consisting of three parts:

1. (Re)summarise the main contributions and achievements, in essence summing up the content.
2. Clearly state the current project status (e.g., “X is working, Y is not”) and evaluate what has been achieved with respect to the initial aims and objectives (e.g., “I completed aim X outlined previously, the evidence for this is within Chapter Y”). There is no problem including aims which were not completed, but it is important to evaluate and/or justify why this is the case.
3. Outline any open problems or future plans. Rather than treat this only as an exercise in what you *could* have done given more time, try to focus on any unexplored options or interesting outcomes (e.g., “my experiment for X gave counter-intuitive results, this could be because Y and would form an interesting area for further study” or “users found feature Z of my software difficult to use, which is obvious in hindsight but not during at design stage; to resolve this, I could clearly apply the technique of Bloggs *et al.*”).

Alternatively, you might want to divide this content into two chapters: a penultimate chapter with a title such as “Further Work” and then a final chapter “Conclusions”. Again, there is no hard and fast rule, we trust you to make the right decision.

And this, the final paragraph of this thesis template, is just a bunch of citations, added to show how to generate a BibTeX bibliography. Sources that have been randomly chosen to be cited here include: [?, ?, ?, ?, ?, ?, ?, ?, ?, ?].

Appendix A

An Example Appendix

Content which is not central to, but may enhance the dissertation can be included in one or more appendices; examples include, but are not limited to

- lengthy mathematical proofs, numerical or graphical results which are summarised in the main body,
- sample or example calculations, and
- results of user studies or questionnaires.

Note that in line with most research conferences, the examiners are not obliged to read such appendices.