Text Summarization for Wikipedia Articles

Prasoon Goyal
Courant Institute of
Mathematical Sciences
New York University
pg1338@nyu.edu

Sahil Goel
Courant Institute of
Mathematical Sciences
New York University
sg4187@nyu.edu

Kshitiz Sethia
Courant Institute of
Mathematical Sciences
New York University
ks3839@nyu.edu

Abstract

With the increasing amount of textual data, automatic summarization has become an important task in natural language processing. In this work, we consider the summarization of Wikipedia articles, using graph-based and machine learning approaches. We obtain results comparable to an approximately-optimal summary.

1 Introduction

Automatic summarization is a key task in natural language processing. The problem of automatic summarization can be formulated as follows (taken from Wikipedia): Automatic summarization is the process of reducing a text document with a computer program in order to create a summary that retains the most important points of the original document.

Automatic summarization tasks can be divided into categories based on several criteria. If the summary is generated by selecting sentences from the given document, it is referred to as *extractive summarization*. On the other hand, if the summary is constructed by paraphrasing the most important content of the given document, it is referred to as *abstractive summarization*. If the summarization task involves summarizing the content of a single document, it is called *single document summarization*, while summarizing the content of multiple documents into a single summarized text is called *multi-document summarization*. In this work, we consider the problem of **single document extractive summarization**.

2 Problem definition

We consider the following problem of automatic summarization of Wikipedia articles. Given a Wikipedia article, generate an extractive summary of the article. To evaluate our system- generated summary, we compare it against the introduction section of the original Wikipedia article, which is roughly a summarization of the most important content of the entire article for a large number of Wikipedia articles. Therefore, our system takes a Wikipedia article *without the introduction section* and generates a summary for the same.

3 Overview of the pipeline

Our pipeline for the entire task can be summarized as follows:

1. **Procuring data**: We worked with `enwik9` which is the first 10^9 bytes of the dump of English Wikipedia on March 3, 2006, which contains 243,426 article titles, of which 85,560 are #REDIRECT to fix broken links, and the rest are regular articles.

2. **Cleaning up**: The data in the dump are present as a single file that has all the articles in XML format concatenated together. We convert this to plaintext format, removing all XML tags, and other redundant information.

3. **Filtering summarizable articles**: A large fraction of the articles are such that the introduction section is far from an approximate summary of the remainder of the article. We use a heuristic to filter out such articles. The details are described in section 4.

4. **Summarization**: We use two kinds of models to generate summaries. The first one involves sequentially selecting sentences that are maximally similar to the rest of the article. The second one involves generating a sentence score for each sentence that estimates the likelihood of the sentence being in the summary, using a regression model. The details of these models are given in sections 5 and 6.

5. **Evaluation**: We use ROUGE metric for evaluating our performance, by comparing our system generated summary with the introduction section of the article. For baseline, we use a summary that is generated by selecting random sentences from the given article.

4 Preprocessing

4.1 Data cleaning

The plaintext sources of the Wikipedia articles require significant preprocessing, including the following tasks:

- Removing XML tags
- Removing metadata such as sections "See also" and "References".
- Removing infoboxes, tables, lists
- Modifying links to retain only the text of the link
- Removing formatting markups

For preprocessing, we primarily used a Perl script [10] modified to suit our task. [1]

4.2 Filtering summarizable articles

We found that for several articles, the introduction section could not be considered a summary of the remainder of the article for one or more of the following reasons:

- The introduction section of the article is empty or contains very few sentences.
- The remainder of the article is empty or contains very few sentences.
- The size of the introduction section is not comparable to the expected size of the summary of the article.

We used the following approach to filter out the articles of the above types. First, any article with less than 200 words in the introduction section or the rest of the article is discarded. Next, if the ratio of the number of words in the introduction section to the number of words in the remainder of the article is not between 7% and 37%, we discard the article.

After this filtering operation, we end up with 3789 articles.

4.3 Sentence chunking

To obtain sentences from the processed text, we use the following regular expression:

$$([A-Za-z0-9]^{+}\backslash s^{+}[A-Za-z0-9]^{+})\backslash s^{*}([.!?])\backslash s^{+}[A-Z0-9]$$

We found the performance of this to be reasonably good.

5 Optimal extractive summary

Evaluating our generated summary with the first section of the articles using ROUGE measure did not give promising results. We hypothesize that many articles, even after our filtering process, have introduction paragraphs that cover significantly different topics than the rest of the article. Therefore, we attempted to find the best possible ROUGE score between an extractive summary from the given article and the original summary. For a summary of size k for an article with n sentences, this would require computing $\binom{n}{k}$ ROUGE values, which is clearly intractible. Thus, we came up with the following heuristic to approximate the best ROUGE score possible.

For each sentence in the original summary, we compute its ROUGE score with each sentence in the remainder of the article, and pick the sentence with the highest score. All these sentences are put together to create an approximately-optimal summary.

6 Graph-based approaches

We represent the article as a fully connected graph in which each sentence is represented as a node, and the weight of the edge between two sentences is the similarity between the two sentences. To compute the dissimilarity between the sentences, we used the following approach:

1. For each word in sentence 1, compute its dissimilarity to all words in sentence 2, to get mn dissimilarity values, where m and n are the number of words in sentences 1 and 2 respectively.
2. From these mn dissimilarity values compute the average of the minimum k values.

[1] We started by collecting the plaintext sources of 5000 most popular Wikipedia pages from [9], implemented data cleaning from scratch, but could not get a perfectly working system. As such, we found the Perl script along with the enwik9 dataset, and switched to the same.

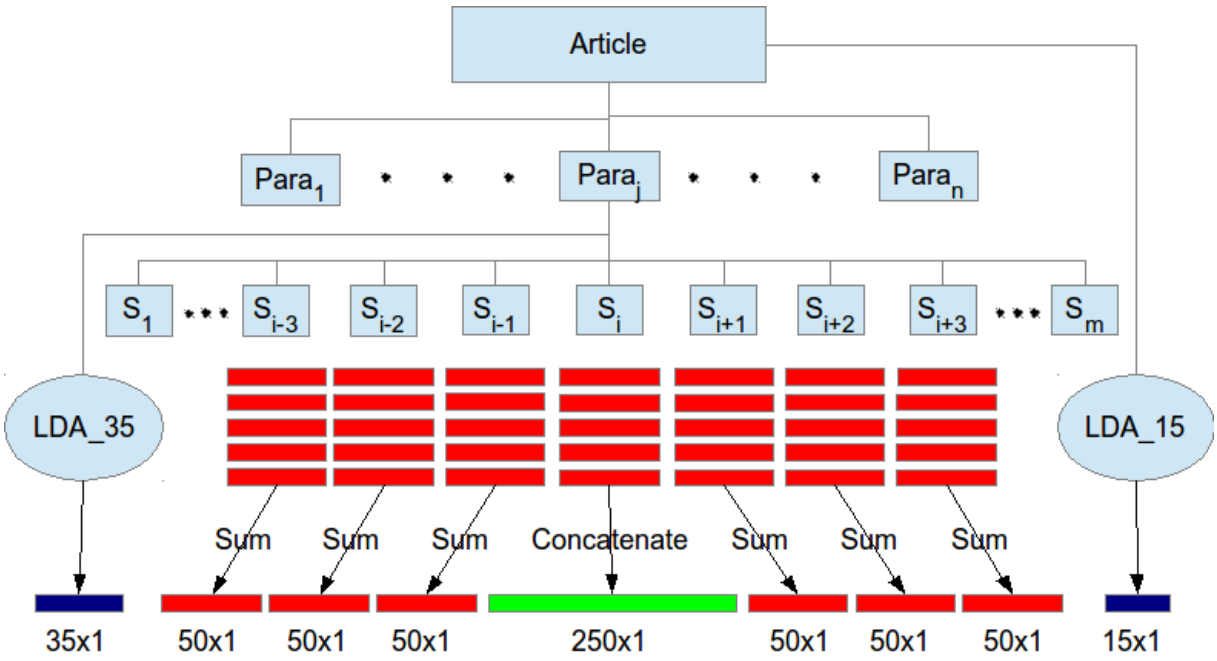

Figure 1: Schematic diagram of feature vector for machine learning based approach

For dissimilarity between words, we used the following two metrics

- Cosine distance between word embeddings of the words.
- Euclidean distance between word embeddings of the words

where we experimented with pretrained Google Word2Vec [11] and GloVe [5] to get word embeddings for words.

After constructing the graph, we incrementally construct the summary by picking the node with the highest sum of weights of outgoing edges, and modifying the graph at each step to remove the selected node and edges incident on it.

7 Machine learning-based approaches

We modelled the summarization task as a regression task in which, given a sentence from an article, the model seeks to predict the likelihood of the sentence being in the summary.

For these approaches, we performed a random split of our 3789 articles into 3000 training articles and 789 test articles.

We used the following features for a sentence in an article:

- **Sentence level features** : This included the word embedding representations of the top k words in the sentence ranked by their tf-idf scores. We took $k = 5$, and $50-$dimensional word embedding vectors to get a $250-$dimensional vector for the sentence level features.
- **Context level features** : This included the word embeddings of the neighboring sentences. We considered the previous three and the next three sentences. To get an overall sense of each of these neighboring sentences, we added the word embedding vectors of the top 5 words in each sentence to get a $50-$dimensional vector for each sentence. This resulted in a $300-$dimensional vector for context level features.
- **Paragraph level features** : To capture the overall topic of the paragraph to which the current sentence belongs, we used Latent Dirichlet Allocation (LDA), which is a well-known technique for topic modelling. We trained our LDA model on all paragraphs in the training corpus.
- **Article level features** : We used LDA to generate article level features as well, training a model on all articles in the training corpus.

A crucial parameter in LDA is the number of latent topics. We used the values (35, 15) for

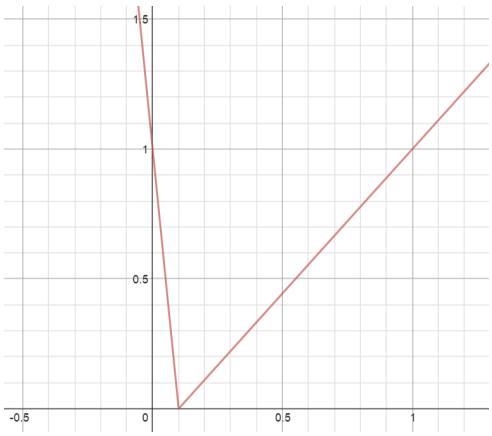

Figure 2: Plot of penalty as a function of the ratio of the length of the actual summary and the remainder of the article.

paragraph and sentence level features respectively, decided by some manual inspection. The values were chosen keeping in mind that the paragraph level topics should be much more fine-grained than the article level topics and that the number of LDA-based features in the feature vector of a sentence should be a significant fraction of the entire feature vector for the entire feature vector for the sentence.

To generate the target value of a sentence, which according to our model corresponds to the likelihood of the sentence being included in the summary, we computed the ROUGE measure of the sentence with all the sentences in the actual summary and took an average of the top k values for robustness. We chose $k = 3$ in our experiments.

We learn a support vector regression model with radial basis function kernel on the dataset thus generated.

8 Experiments

In this section, we briefly describe the approaches we tried and the problems encountered.

8.1 Preprocessing

The first step was to preprocess the articles removing any unwanted characters, and metadata.

We then implemented a method to filter summarizable articles from the rest. For this, we considered keeping articles that had a high similarity between the introduction section and the remainder of the article. To compute the similarity, we converted the introduction section and the remainder of the article to their respective bag of words models, scaling each word by its tf-idf score. We implemented the code to compute tf-idf scores. We tested this filtering on the 5000 most popular Wikipedia articles [9], but ended up with a set of articles that were still not all summarizable. In particular, many of these articles had very few sentences in the introduction section, or very few sentences in the remainder of the article or the size of the introduction section was too small or too large compared to the remainder of the article. Hence, in addition to the cosine similarity, we imposed a penalty term based on the lengths of these quantities as follows. We computed a summary by picking the first sentence of each paragraph from the remainder of the article, to get a summary of length l_s. Denote by l_d the number of words in the remainder of the article, and l_{os} the number of words in the original summary. Then, our penalty function for an article was given by

$$\sqrt{\frac{l_s - l_{os}}{l_d - l_{os}}}$$

which was subtracted from the cosine similarity computed earlier. Filtering out articles based on a reasonable threshold left us with only a few hundred articles to work with.

We looked for a larger collection of articles, and came across `enwik9`, which has about $130,000$ regular articles. Along with it, we also found a preprocessing script written in Perl that we found to be more robust than our implementation. We ran our filtering code on this larger dataset, but the tf-idf computation on such a large dataset was taking extremely long. We also tried using Gensim's implementation of tf-idf, but that was also found to be very slow for such a large dataset. Therefore, we modified our filtering scheme to consider only the penalty function. This was also found to work reasonably well, because the cosine similarity values that were used earlier were quite noisy in many cases.

We also experimented with the following penalty function

$$\max\left\{1 - 10x, \frac{10x - 1}{9}\right\}$$

where $x = \dfrac{l_{os}}{l_d}$. This function takes the value zero when $x = \frac{1}{10}$, and increases linearly as x deviates

from this value in either direction. This penalty function was found to be better than the previous one, and therefore was used as the final filtering function. Figure 2 shows the plot of this penalty function.

8.2 Graph-based approach

As described in section 6, to compute the similarity between sentences, we computed the similarity between the word embeddings of all pairs of words. For this, we used Google Word2Vec [11] that is pretrained on Google News dataset.

8.3 Machine learning-based approach

This step required using the following libraries:

- tf-idf: To compute the top k words in a sentence, we used `gensim` implementation of tf-idf trained on our corpus.
- GloVe: To compute the word embeddings of the top k words, we intended to use Google Word2Vec for consistency with our graph-based approach, but since the pretrained vectors were $300-$dimensional, it would have resulted in very large feature vectors. Therefore, we used GloVe word embeddings, which are $50-$dimensional vectors.
- LDA: We used `gensim` LDA implementation to train our LDA models on paragraph and document levels.
- ROUGE: We used Chin-Yew Lin's ROUGE implementation to generate labels.
- SVR: Finally, we used `sklearn` implementation for support vector regression using 5-fold cross-validation to train our model

8.4 Other comments

It is also worth mentioning that because our entire pipeline consists of several steps, and each step takes a quite a while to execute, a the need for a small change in the pipeline based on results from a later stage require running the entire pipeline over again, making the overall experimentation fairly time-consuming.

9 Results

We present the results obtained by our graph-based approach and our machine learning-based approach. As already stated, for our graph-based model, we experimented with two different distance functions - cosine distance and Euclidean distance, and found the Euclidean distance function to work better. We therefore only present the results of graph-based approach using Euclidean distance.

Because the feature generation and training on the machine-learning based model was taking fairly long, we ran our final set of experiments on 100 articles, chosen such that the ROUGE scores of the optimal summary and the random summary differ by at least 0.2.

Table 1 summarizes the mean and the median ROUGE scores obtained using various approaches. We represent the scores of all these models on all articles in Figure 4.

It can be seen that the graph-based model improves over the random summary by a fair amount, and the machine learning-based approach further improves over it significantly, reaching very close to the optimal.

It should also be noted here that from the graph, we see that while the optimal summary is the best out of the four models on most articles, it is sometimes outperformed by others, showing that it is only an approximation of the optimal summary as described in section 5.

	Mean	**Median**
Random summary	0.0886	0.08791
Optimal summary	0.3254	0.3245
Model summary Graph	0.1587	0.1507
Model summary ML	0.2972	0.2902

Table 1: Results comparison on 113 articles

We include some sample summaries in the appendix. It can be seen that while we do not explicitly give to either model the position of the sentence in the paragraph, both of the models pick the first or the last sentences of the paragraphs fairly often, which are usually the most important parts of a paragraph.

10 Conclusion and future work

We proposed two models - one based on graph-based approach and another on machine learning based approach for automatic text summarization. We showed that our models perform significantly better than a randomly generated summary, and the machine learning based model generates summaries that are fairly close to the optimal summary

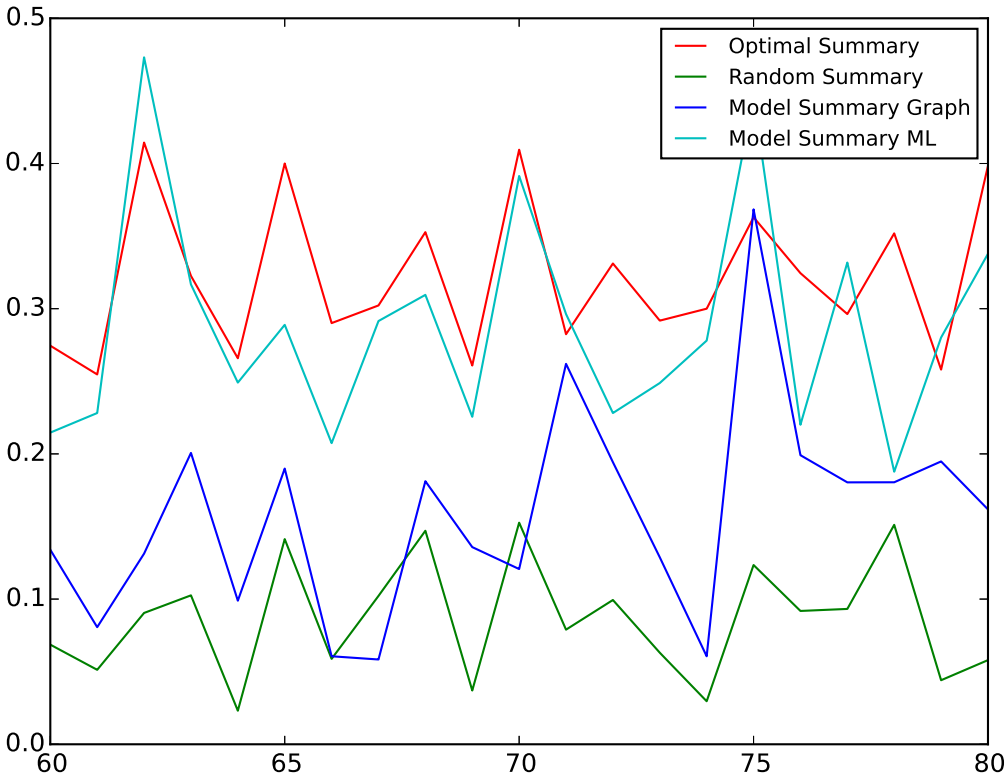

Figure 3: Results of the four models on 100 articles

in terms of ROUGE score.

The following are some directions to pursue as future work

- Coreference resolution and Name tagging: Using name tagging, the sentence chunking process would become more robust. Using coreference resolution, estimates of similarity scores are likely to be better.

- Better graph models: While greedily selecting sentences using graph based approach, we can penalize sentences that are similar to the summary constructed so far, to get more diversity in the summary.

- Better machine learning models: Instead of adding the word vectors of neighboring sentences, we can concatenate them to generate larger feature vectors, which can then be passed through some dimensionality reduction technique such as principle components analysis. We can also add features like number of names in the sentence, position of sentence in the paragraph and the position of sentence in the article.

- Post summary generation steps: We observed that the generated summaries had sentences selected from different parts of the article. So if one selected sentence talks about some event and the next selected sentence talks about another event, in the summary, it might appear that both the sentences talk about the same event. This needs to be disambiguated by some technique such as reference resolution.

References

[1] Nenkova, Ani, and Kathleen McKeown. "A survey of text summarization techniques." Mining Text Data. Springer US, 2012. 43-76.

[2] Das, Dipanjan, and Andre FT Martins. A survey on automatic text summarization. Literature Survey for the Language and Statistics II course at CMU 4 (2007): 192-195.

[3] Lin, Chin-Yew. "Rouge: A package for automatic evaluation of summaries."Text Summarization Branches Out: Proceedings of the ACL-04 Workshop. 2004.

[4] Rehuek, Radim, and Petr Sojka. "GensimStatistical Semantics in Python." (2011).

[5] Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. "Glove: Global vectors for word representation." Proceedings of the Empiricial Methods in Natural Language Processing (EMNLP 2014) 12 (2014).

[6] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." the Journal of machine Learning research 3 (2003): 993-1022.

[7] Huang, Eric H., et al. "Improving word representations via global context and multiple word prototypes." Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1. Association for Computational Linguistics, 2012.

[8] Radev, Dragomir R., et al. "Centroid-based summarization of multiple documents." Information Processing & Management 40.6 (2004): 919-938.

[9] http://en.wikipedia.org/wiki/User:West.andrew .g/Popular_pages

[10] https://cs.fit.edu/mmahoney/compression/text data.html

[11] https://code.google.com/p/word2vec/

Appendix

A.1 Sample summary: Graph-based and ML-based model

The following is the summary generated using graph based model and ML based model on `Charles Kennedy`.

==Early life==

Born in Inverness Scotland and raised as a Roman Catholic Charles Kennedy was educated at Lochaber High School Fort William and went on to study for a Master of Arts degree in Politics and Philosophy at the University of Glasgow. It was at university that he became politically active joining the Dialectic Society a debating club but he was not a member of the Labour Club. He was elected president of Glasgow University Union and won the British Observer Mace university debating award. GUU was the last all male student union in the UK during his time as president it is claimed that Kennedy opposed admitting women as members he has personally denied this. Shortly after he joined the Social Democrats SDP.

Upon graduation in 1982 he went to work for BBC Highland as a journalist and later received a Fulbright Fellowship allowing him to carry out research at Indiana University in the United States on the speeches and writings of Roy Jenkins.

==Political career==

While studying in America he also received the Social Democratic Party SDP nomination for the Scottish seat of Ross Cromarty and Skye which he went on to win in 1983 becoming the youngest sitting Member of Parliament. He has retained the seat and its successor Ross Skye and Inverness West at five subsequent general elections. He is the Liberal Democrat MP for the seat which replaced it in 2005 Ross Skye and Lochaber.

In the late 1980s the SDP and the Liberal Party which had been co operating in the SDP Liberal Alliance merged to form the Social and Liberal Democratic Party later renamed the Liberal Democrats. Kennedy was the only one of the five SDP MPs to support the merger from the outset.

Kennedy served as a frontbencher for the Lib Dems in a variety of posts including social security agriculture and rural affairs health Scotland and Europe. He was also party president for four years between 1990 and 1994.

==Initial period as party leader==

On 9 August 1999 Charles Kennedy was elected leader of the Liberal Democrats after the retirement of Paddy Ashdown he beat Jackie Ballard Malcolm Bruce Simon Hughes and David Rendel. He won 57 of the transferred vote under the Alternative Vote system Simon Hughes the runner up won 43 of the vote.

Kennedy's style of leadership differed from Ashdown's being regarded as more conversational and laid back. Although he has been dismissed as Chatshow Charlie by some observers as a result of his appearances on the satirical panel game Have I Got News For You opinion polls showed him to be regarded positively as a party leader and potential Prime Minister by a significant fraction of the British electorate.

Kennedy maintained the long standing aspiration for his party to break through to the status of official opposition. In his first major campaign the 2001 general election the Liberal Democrats improved their share of the vote to 18.3 1.5 more than in the 1997 election. Although this was a smaller share than the 25.4 the SDP Liberal Alliance achieved in 1983 the Lib Dems won 52 seats compared to the Alliance's 23. In his last General Election as leader in May 2005 the Liberal Democrats won 62 seats their greatest number of seats since the 1920s gaining 22.1 of the vote.

In 2001 Sir Ludovic Kennedy (no relation) condemned Kennedy a Roman Catholic for opposing euthanasia. He then resigned from the party to stand in the general election as an independent on a platform of legalising voluntary euthanasia but has since rejoined.

==2005 general election==

Kennedy along with his election guru Lord Rennard targeted the Lib Dems campaigning on a limited number of seats in such a way as to turn a lower level of national support into a greater number of Parliamentary seats. He extended this strategy at the 2005 General Election targeting the seats held by the most senior and or highly regarded Conservative MPs dubbed a decapitation strategy with the expectation that without these key figures the Conservatives would be discredited as the Official Opposition allowing Charles Kennedy and the Liberal Democrats to claim that they are the effective Opposition.

However this strategy is widely seen to have failed. At the 2005 General Election the Liberal Democrats failed to unseat leading Conservatives such as the Shadow Chancellor of the Exchequer Oliver Letwin Shadow Home Secretary David Davis Shadow Secretary of State for the Family Theresa May and the Leader of the Opposition Michael Howard. The biggest scalp the Liberal Democrats managed to claim was that of the Shadow Education Secretary Tim Collins in Westmorland and Lonsdale.

At the same time the Lib Dems also hoped to capture marginal Labour seats attracting Labour voters particularly Muslim voters who were dissatisfied because of the invasion of Iraq the party had succeeded with this tactic in by elections taking Brent East and Leicester South from Labour. The Party did succeed to some extent in this aim winning

particularly in student areas such as Bristol and Manchester but did not see the breakthrough some expected in areas with large Asian populations even losing Leicester South.

In the 2005 election the party succeeded in regaining the seat of Ceredigion its first gain from the Welsh Party Plaid Cymru. Overall Kennedy's party achieved a total of 62 seats their highest number since 1923 with 22 of the overall vote.

==Developments since the election==

In the wake of the General Election Kennedy's leadership came under increased criticism from those who felt that the Liberal Democrats could have surged forward at a time when arguably the Official Opposition the Conservative Party were in a relatively weak position. Many pointed the finger of blame at Kennedy for failing to widen the Party's appeal while others like the former Deputy Chairman of the Federal Liberal Democrat Party Donnachadh McCarthy resigned from the Party citing the party's shift to the right of the political spectrum under Kennedy in pursuit of Conservative votes. Under the party's rules a leader has to stand for re election within a year of a general election. Kennedy handed out the ballot papers to the parliamentary party within days of the 2005 election leaving no time for anyone to mount a challenge and allowing him to be re elected unopposed. There was much speculation at the time as to whether he would have survived a challenge.

In late 2005 the leadership speculation was renewed with the journalist Andrew Neil claiming to speak on good authority that Kennedy would announce his resignation at the 2006 spring conference of the Liberal Democrats. Kennedy's spokeswoman denied the report and complained against the BBC which had broadcast it. After the election of the more moderate David Cameron as Leader of the Conservative Party in December 2005 it was widely reported that senior members of the Liberal Democrats had told Kennedy that he must either raise his game or resign.

===Leadership contest===

On December 13 2005 the BBC s Political Editor Nick Robinson claimed that there were briefings against the leader with members of his party unhappy at what they saw as lack of leadership from Kennedy. A Kennedy Must Go petition was started by The Liberal magazine a publication which has no affiliation to the Liberal Democrats and allegedly had been signed by over 3,300 party members including 386 local councillors and two MPs by the end of 2005. A round robin letter signed by Liberal Democrat MPs rejecting his leadership received 23 signatures.

On January 5 2006 Kennedy was informed that ITN would be reporting that he had received treatment for alcoholism and called a sudden news conference to make a personal statement confirming the story. He stated clearly that over the past eighteen months he had been coming to terms with a drinking problem but has sought ongoing professional help. He told reporters that recent questions among his colleagues about his suitability as leader were partly as a result of the drinking problem but stated that he had been dry for the past two months and would be calling a leadership contest to resolve the issues surrounding his authority once and for all. It was later claimed that the source for ITN s story was his former press secretary turned ITV News correspondent Daisy McAndrew.

Responses to Kennedy s statement focused on his previous denials of any problems with alcohol. As recently as the Jonathan Dimbleby programme on ITV1 on December 18 2005 when asked Has it been a battle to stay off the booze have you had to have medical support in any way at all? Kennedy replied No no no that is not the case it is a matter on all fronts if there s something my doctor really wants me to do over this holiday period as a matter of fact is give up smoking and I think he s right. In 2002 the journalist Jeremy Paxman claimed Kennedy was often drunk and asked him if he drank privately by yourself a bottle of whisky late at night?. No I do not Kennedy replied. BBC apologised to Charles Kennedy Paxman refused to endorse the apology. In 2004 The Times published a clarification over a report it had made stating Kennedy had not taken part in that year s budget debate due to excessive drinking.

===Resignation===

Following Kennedy's admission a letter from twenty five Liberal Democrat MPs was delivered to him. It stated that the signatories could no longer serve as frontbench speakers under his leadership and gave a deadline of Monday January 9 for him to make a decision before they resigned. Despite a combative interview in The Independent at which Kennedy described a decision to resign as a dereliction of duty on January 6 a large number of senior Liberal Democrats stated that his position was untenable. Chris Davies leader of Liberal Democrat Members of the European Parliament described him as a dead man walking. A survey for BBC Newsnight found that more than half of Liberal Democrat MPs thought he should resign and only seventeen out of sixty two MPs positively wanted him to stay while eleven of his twenty three frontbenchers wanted him to leave. Among those who thought he should go were Norman Lamb and Andrew George who had served as his Parliamentary Private Secretary and Matthew Taylor the chairman of his 1999 leadership campaign.

At 3pm on January 7 Kennedy called a press conference where he announced that whilst he was buoyed by the supportive messages he had received from grassroot members he felt that he could not continue because of the lack of confidence of the parliamentary party. He said he would not be a candidate in the leadership election and that he would stand down as leader with immediate effect with Menzies Campbell acting as interim leader until a new leader has been elected. He also confirmed in his resignation speech that he does not have any expectations of remaining on the frontbench pledging his loyalty to a new leader as a backbench MP but that he wishes to remain active in the party and politics. His leadership lasted slightly less than six years and five months.

==Personal life==

In July 2002 Charles Kennedy married Sarah Gurling the sister of his best friend James Gurling.

Reports of Kennedy s ill health in 2003 at the time of crucial debates on Iraq and the budget his ill health meant he missed an entire budget speech were linked to the rumours of alcoholism which were strenuously denied by himself and the party but which he subsequently admitted.

Sarah gave birth to their first child Donald James Kennedy at 12 14 a.m. April 12 2005 at St Thomas Hospital London. Kennedy had been due to launch his party's manifesto for the

General Election which was subsequently delayed with Sir Menzies Campbell taking temporary charge as acting leader and covering Kennedy's campaign duties. During the manifesto launch on his first day back on the campaign trail after the birth Kennedy struggled to remember the details of a key policy replacing the Council Tax with a Local Income Tax at an early morning press conference which he later blamed on a lack of sleep due to his new child.

A.2 Sample summary: ML-based model

The following is the summary generated using ML based model on `William Rowan Hamilton`. The selected sentences are marked in blue.

=Biography=

William Rowan Hamilton?s mathematical career included the study of geometrical optics adaptation of dynamic methods in optical systems applying quaternion and vector methods to problems in mechanics and in geometry development of theories of conjugate algebraic couple functions in which complex numbers are constructed as ordered pairs of real numbers solvability of polynomial equations and general quintic polynomial solvable by radicals the analysis on Fluctuating Functions and the ideas from Fourier analysis linear operators on quaternions and proving a result for linear operators on the space of quaternions which is a special case of the general theorem which today is known as the Cayley Hamilton Theorem . Hamilton also invented Icosian Calculus which he used to investigate closed edge paths on a dodecahedron that visit each vertex exactly once.

===Early life===

Rowan Hamilton : A child prodigy Hamilton was born the son of Archibald Hamilton a solicitor in Dublin at 36 Dominick Street. He was subsequently educated by James Hamilton curate of Trim his uncle and an Anglican priest.

Hamilton's genius first displayed itself in the form of a power of acquiring languages. At the age of seven he had already made very considerable progress in Hebrew and before he was thirteen he had acquired under the care of his uncle a linguist almost as many languages as he had years of age. Among these besides the classical European languages and the modern European languages were included Persian Arabic Hindustani Sanskrit and even Malay. But though to the very end of his life he retained much of the singular learning of his childhood and youth often reading Persian and Arabic in the intervals of sterner pursuits he had long abandoned them as a study and employed them merely as a relaxation.

Hamilton was part of a small but well regarded school of mathematicians associated with Trinity College Dublin where he spent his life. He studied both classics and science and was appointed Professor of Astronomy in 1827 prior to his graduation.

===Mathematical studies===

Hamilton's mathematical studies seem to have been undertaken and carried to their full development without any assistance whatsoever and the result is that his writings belong to no particular school unless indeed we consider them to form as they are well entitled to do a school by themselves. As an arithmetical calculator Hamilton was not only an expert but he seems to have occasionally found a positive experience in working out to an enormous number of places of decimals the result of some irksome calculation. At the age of twelve Hamilton engaged Zerah Colburn the American calculating boy who was then being exhibited as a curiosity in Dublin and he had not always the worst of the encounter. But two years before he had accidentally fallen in with a Latin copy of Euclid which he eagerly devoured and at twelve Hamilton attacked Newton?s Arithmetica universalis . This was his introduction to modern analysis. Hamilton soon commenced to read the Principia and at sixteen Hamilton had mastered a great part of that work besides some more modern works on analytical geometry and the differential calculus.

About this period Hamilton was also engaged in preparation for entrance at Trinity College Dublin and had therefore to devote a portion of time to classics. In the summer of 1822 in his seventeenth year he began a systematic study of Laplace?s Mcanique Cleste . Nothing could be better fitted to call forth such mathematical powers as those of Hamilton for Laplace?s great work rich to profusion in analytical processes alike novel and powerful demands from the student careful and often laborious study.

It was in the successful effort to open this treasure house that Hamilton?s mind received its final temper Ds lors il commena marcher seul to use the words of the biographer of another great mathematician. From that time Hamilton appears to have devoted himself almost wholly to the mathematics investigation though he ever kept himself well acquainted with the progress of science both in Britain and abroad. Hamilton detected an important defect in one of Laplace?s demonstrations and he was induced by a friend to write out his remarks that they might be shown to Dr John Brinkley afterwards bishop of Cloyne but who was then the first royal astronomer for Ireland and an accomplished mathematician. Brinkley seems at once to have perceived the vast talents of young Hamilton and to have encouraged him in the kindest manner.

Hamilton?s career at College was perhaps unexampled. Amongst a number of competitors of more than ordinary merit he was first in every subject and at every examination. He achieved the rare distinction of obtaining an optime for both Greek and for physics. The amount of many more such honours Hamilton might have attained it is impossible to say but Hamilton was expected to win both the gold medals at the degree examination had his career as a student not been cut short by an unprecedented event. This was Hamilton?s appointment to the Andrews professorship of astronomy in the university of Dublin vacated by Dr Brinkley in 1827. The chair was not exactly offered to him as has been sometimes asserted but the electors having met and talked over the subject authorized one of their number who was Hamilton's personal friend to urge Hamilton to become a candidate a step which Hamilton?s modesty had prevented him from taking. Thus when barely twenty two Hamilton was established at the Dunsink Observatory near Dublin.

Hamilton was not specially fitted for the post for although he had a profound acquaintance with theoretical astronomy he had paid but little attention to the regular work of the practical astronomer. And it must be said that Hamilton?s time was better employed in original investigations than it would have been had he spent it in observations made even with the best of instruments. Hamilton was intended by the university authorities who elected him to the professorship of astronomy to spend his time as Hamilton best could for the advancement of science without being tied down to any particular branch. If Hamilton devoted himself to practical

astronomy the University of Dublin would assuredly have furnished him with instruments and an adequate staff of assistants.

In 1835 being secretary to the meeting of the British Association which was held that year in Dublin he was knighted by the lord lieutenant. But far higher honours rapidly succeeded among which his election in 1837 to the president?s chair in the Royal Irish Academy and the rare distinction of being made corresponding member of the academy of St Petersburg. These are the few salient points other of course than the epochs of Hamilton?s more important discoveries and inventions presently to be considered in the uneventful life of Hamilton.

===Optics and dynamics===

He made important contributions to optics and to dynamics. Hamilton's papers on optics and dynamics demonstrated theoretical dynamics being treated as a branch of pure mathematics. Hamilton's first discovery was contained in one of those early papers which in 1823 Hamilton communicated to Dr Brinkley by whom under the title of ? Caustics ? it was presented in 1824 to the Royal Irish Academy. It was referred as usual to a committee. Their report while acknowledging the novelty and value of its contents recommended that before being published it should be still further developed and simplified. During the time between 1825 to 1828 the paper grew to an immense bulk principally by the additional details which had been inserted at the desire of the committee. But it also assumed a much more intelligible form and the features of the new method were now easily to be seen. Hamilton himself seems not till this period to have fully understood either the nature or importance of optics as later Hamilton had intentions of applying his method to dynamics.

In 1827 Hamilton presented a theory that provided a single function that brings together mechanics optics and mathematics. It helped in establishing the wave theory of light. He proposed for it when he first predicted its existence in the third supplement to his Systems of Rays read in 1832. The Royal Irish Academy paper was finally entitled ? Theory of Systems of Rays ? April 23 1827 and the first part was printed in 1828 in the Transactions of the Royal Irish Academy . It is understood that the more important contents of the second and third parts appeared in the three voluminous supplements to the first part which were published in the same Transactions and in the two papers ? On a General Method in Dynamics ? which appeared in the Philosophical Transactions in 1834 and 1835.

The principle of ? Varying Action ? is the great feature of these papers and it is indeed that the one particular result of this theory which perhaps more than anything else that Hamilton has done something which should have been easily within the reach of Augustin Fresnel and others for many years before and in no way required Hamilton?s new conceptions or methods although it was by Hamilton?s new theoretical dynamics that he was led to its discovery. This singular result is still known by the name ? conical refraction ?.

The step from optics to dynamics in the application of the method of ? Varying Action ? was made in 1827 and communicated to the Royal Society in whose Philosophical Transactions for 1834 and 1835 there are two papers on the subject. These display like the ? Systems of Rays ? a mastery over symbols and a flow of mathematical language almost unequalled. But they contain what is far more valuable still the greatest addition which dynamical science had received since the strides made by Sir Isaac Newton and Joseph Louis Lagrange. C. G. J. Jacobi and other mathematicians have extended Hamilton?s processes and have thus made extensive additions to our knowledge of differential equations.

And though differential equations optics and theoretical dynamics of course are favored in which any such contribution to science can be looked at the other must not be despised. It is characteristic of most of Hamilton?s as of nearly all great discoveries that even their indirect consequences are of high value.

===Quaternions===

Quaternion Plague on Broome Bridge

The other great contribution made by Hamilton to mathematical science was the invention of quaternions which he discovered in 1843.

Hamilton was looking for ways of extending complex numbers which can be viewed as points on a plane to higher spatial dimensions. Hamilton could not do so for 3 dimensions but 4 dimensions produce quaternions. According to the story Hamilton told on October 16 Hamilton was out walking along the Royal Canal in Dublin with his wife when the solution in the form of the equation

$$i^2 = j^2 = k^2 = ijk = -1$$

suddenly occurred to him Hamilton then promptly carved this equation into the side of the nearby Broome Bridge which Hamilton called Brougham Bridge. Since 1989 the National University of Ireland Maynooth has organized a pilgrimage where mathematicians take a walk from Dunsink observatory to the bridge where unfortunately no trace of the carving remains though a stone plaque does commemorate the discovery.

The quaternion involved abandoning the commutative law a radical step for the time. Not only this but Hamilton had in a sense invented the cross and dot products of vector algebra. Hamilton also described a quaternion as an ordered four element multiple of real numbers and described the first element as the scalar part and the remaining three as the vector part.

In 1852 Hamilton introduced quaternions as a method of analysis. His first great work Lectures on Quaternions Dublin 1852 is almost painful to read in consequence of the frequent use of italics and capitals. Hamilton confidently declared that quaternions would be found to have a powerful influence as an instrument of research. He popularized quaternions with several books the last of which Elements of Quaternions had 800 pages and was published shortly after his death.

Peter Guthrie Tait among others advocated the use of Hamilton's quaternions. They were made a mandatory examination topic in Dublin and for a while they were the only advanced mathematics taught in some American universities. However controversy about the use of quaternions grew in the late 1800s. Some of Hamilton's supporters vociferously opposed the growing fields of vector algebra and vector calculus from developers like Oliver Heaviside and Willard Gibbs because quaternions provide superior notation. While

this is undeniable for four dimensions quaternions cannot be used with arbitrary dimensionality though extensions like Clifford algebras can . Vector notation largely replaced the space time quaternions in science and engineering by the mid 20th century.

Today the quaternions are in use by computer graphics control theory signal processing and orbital mechanics mainly for representing rotations orientations. For example it is common for spacecraft attitude control systems to be commanded in terms of quaternions which are also used to telemeter their current attitude. The rationale is that combining many quaternion transformations is more numerically stable than combining many matrix transformations. In pure mathematics quaternions show up significantly as one of the four finite dimensional normed division algebras over the real numbers with applications throughout algebra and geometry.

Hamilton also contributed an alternative formulation of the mathematical theory of classical mechanics. While adding no new physics this formulation which builds on that of Joseph Louis Lagrange provides a more powerful technique for working with the equations of motion. Both the Lagrangian and Hamiltonian approaches were developed to describe the motion of discrete systems were then extended to continuous systems and in this form can be used to define vector fields. In this way the techniques find use in electromagnetic quantum and relativity theory.

===Other originality===

Hamilton originally matured his ideas before putting pen to paper. The discoveries papers and treatises previously mentioned might well have formed the whole work of a long and laborious life. But not to speak of his enormous collection of books full to overflowing with new and original matter which have been handed over to Trinity College Dublin the previous mentioned works barely form the greater portion of what Hamilton has published. Hamilton developed the variational principle which was reformulated later by Carl Gustav Jacob Jacobi. He also introduced Hamilton's puzzle which can be solved using the concept of a Hamiltonian path.

Hamilton's extraordinary investigations connected with the solution of algebraic equations of the fifth degree and his examination of the results arrived at by N. H. Abel G. B. Jerrard and others in their researches on this subject form another contribution to science. There is next Hamilton's paper on Fluctuating Functions a subject which since the time of Joseph Fourier has been of immense and ever increasing value in physical applications of mathematics. There is also the extremely ingenious invention of the hodograph. Of his extensive investigations into the solutions especially by numerical approximation of certain classes of physical differential equations only a few items have been published at intervals in the Philosophical Magazine.

Besides all this Hamilton was a voluminous correspondent. Often a single letter of Hamilton's occupied from fifty to a hundred or more closely written pages all devoted to the minute consideration of every feature of some particular problem for it was one of the peculiar characteristics of Hamilton's mind never to be satisfied with a general understanding of a question Hamilton pursued the problem until he knew it in all its details. Hamilton was ever courteous and kind in answering applications for assistance in the study of his works even when his compliance must have cost him much time. He was excessively precise and hard to please with reference to the final polish of his own works for publication and it was probably for this reason that he published so little compared with the extent of Hamilton's investigations.

===Death and afterwards===

Hamilton retained his faculties unimpaired to the very last and steadily continued till within a day or two of his death which occurred on the 2nd of September 1865 the task of finishing the Elements of Quaternions which had occupied the last six years of his life.

Hamilton is recognized as one of Ireland s leading scientists and as Ireland becomes more aware of its scientific heritage he is increasingly celebrated. There is a research institute named for him at NUI Maynooth and the Royal Irish Academy holds an annual public Hamilton lecture at which Murray Gell Mann Andrew Wiles and Timothy Gowers have all spoken. 2005 is the 200th anniversary of Hamilton's birth and the Irish government has designated this the Hamilton Year celebrating Irish science. Trinity College Dublin intends to mark the year by launching the Hamilton Mathematics Institute TCD a mathematics institute modelled on for example the Isaac Newton Institute in Cambridge.

=== Commemorations of Hamilton===

Hamilton's equations are a formulation of classical mechanics. Hamiltonian is the name of both a function classical and an operator quantum in physics and a term from graph theory.

==Quotations==

Time is said to have only one dimension and space to have three dimensions. ... The mathematical quaternion partakes of both these elements in technical language it may be said to be time plus space or space plus time and in this sense it has or at least involves a reference to four dimensions. And how the One of Time of Space the Three Might in the Chain of Symbols girdled be. William Rowan Hamilton Quoted in Robert Percival Graves Life of Sir William Rowan Hamilton 3 vols. 1882 1885 1889

He used to carry on long trains of algebraic and arithmetical calculations in his mind during which he was unconscious of the earthly necessity of eating we used to bring in a ?snack? and leave it in his study but a brief nod of recognition of the intrusion of the chop or cutlet was often the only result and his thoughts went on soaring upwards. William Edwin Hamilton his elder son

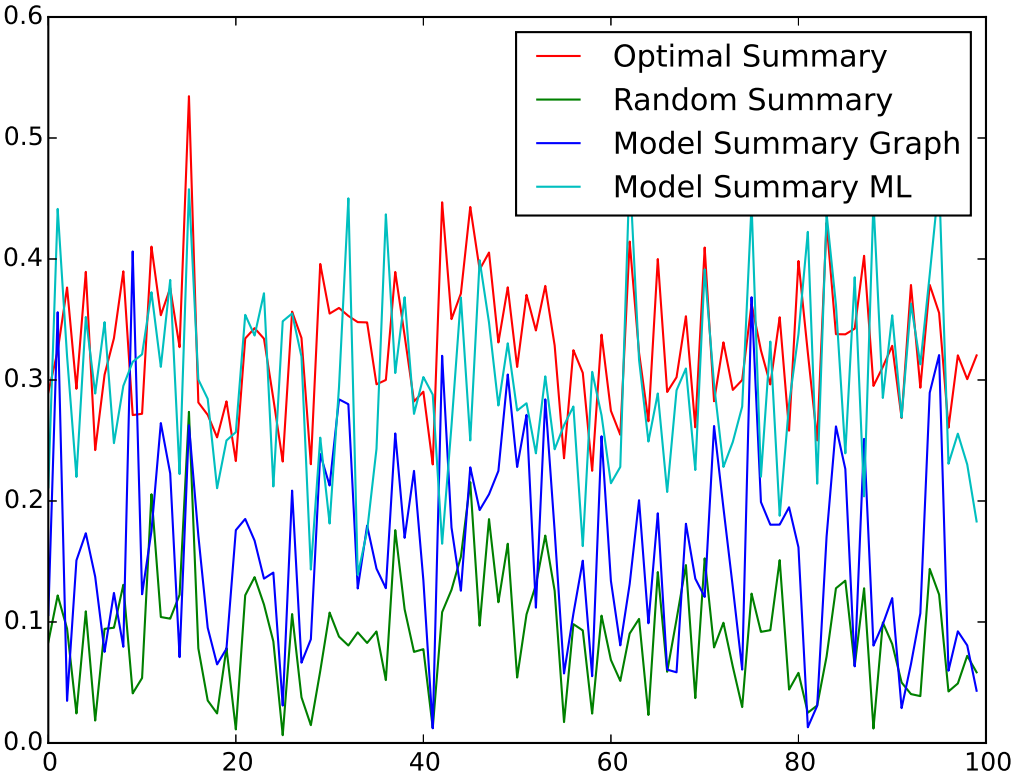

Figure 1: Results of the 4 models on 100 articles