

Textual Distraction as a Basis for Evaluating Automatic Summarisers

Antoinette Renouf and Andrew Kehoe

University of Liverpool
19 Abercromby Square, L69 7ZG, UK
{ajrenouf, ajkehoe}@liv.ac.uk

Abstract

Our summarisation tool, SEAGULL (Summary Extraction Algorithm Generated Using Lexical Links), is a sentence extractor which exploits the patterns of lexical repetition across a text and creates abridgements which express non-trivially the conceptual content and development of topic. In this paper, we report on a test devised to assess its performance against other summarisers. This involves the introduction of progressive batches of unrelated sentences into a source text. Targeted distraction reveals the relative degrees of robustness in summariser performance. The tests show that our system functions best.

Introduction

Our interest in the evaluation of automatic summarisers stems from our development of such tools, specifically here the SEAGULL system. This is a sentence extractor which exploits the patterns of lexical repetition across a text (Hoey, 1991) and which, by virtue of this informed linguistic approach, creates an abridgement which expresses non-trivially the conceptual content and development of a text.

Summaries are used in textual database management to aid both broad-coverage browsing and focussed search. In the latter, the user needs to retrieve only documents which contain a specific thread (support or opposition for a particular line of argument, etc). Our system is particularly valuable for this and more exacting search, such as reviewing scientific papers, or other text containing highly granular information (for example, argumentation, or multiple-topic).

Criteria for a good-quality summary

For the purposes of evaluating competing summarising tools in the above light, our definition of a good-quality sentence-based summary is as follows:

- It represents the key points of the topic
- It follows the flow of the argument/story
- It represents a cohesive and readable mini-text
- It is sufficiently robust to perceive the topic and ignore totally-unrelated material
- It is sufficiently robust to perceive the topic and ignore extraneous but similar topic material
- It is sufficiently robust to perceive the topic and ignore extraneous but similar topic material at start of source text
- It is sufficiently robust to cope with words with ambiguous reference
- It is sufficiently robust to cope with homographs

Data

Although it is generic, for ready tailoring, our tool (<http://www.rdues.liv.ac.uk/abrdemo.html>) is currently calibrated for use with texts of newspaper article length and construction. We thus took as our test text an article (appended) from the *Guardian* Website, 02.07.2003 about the banning of fox-hunting, of 730 running words in length. We then selected 3 leading commercial summarisers that were freely available, for comparison:

*Microsoft Word 2000 Summarizer

*Copernic Summarizer 2.0

(www.copernic.com/en/products/summarizer/)

*Sinope Summarizer 1.2 (www.sinope.nl/en/sinope/)

Method

We first devised a 'distraction' measure, based on the cumulative introduction into the fox-hunting text of sentences from unrelated texts, similar texts, and sentences containing homographs and ambiguous referents which could confuse a less linguistically-informed summariser.

We then devised 8 tests, to compare the performance of our system with others. The first test was a control, simply running the four summarisers on the original 730-word text to compare standard performance. The next 7 tests introduced a series of distractors at various stages, as reported below, so that a cumulative level of difficulty was built into the summarisation task. Each test was run on the four summarisers in turn.

Results and Assessments

Test 1: Does the summariser meet the following criteria for success in a four-sentence summary of Text 1 on fox-hunting (see Appendix 1), by:

- representing the key points of the topic
- following the flow of the argument/story
- producing a cohesive, readable minitext

Microsoft Word summary

- Labour whips last night played down the defeat, which came a day after ministers warned that a vote for a total ban would "wreck" the government's hunting bill.
- Five ministers and whips backed the prime minister's preferred compromise.
- Mr Banks warmly welcomed the vote.
- Anti-hunting groups also welcomed the vote.

Copernic summary

- Tony Blair last night suffered a humiliating rebuff when more than 300 Labour MPs, including seven members of the cabinet, ignored Downing Street to vote overwhelmingly in favour of an outright ban on foxhunting.
- Labour whips last night played down the defeat, which came a day after ministers warned that a vote for a total ban would "wreck" the government's hunting bill.
- A government source said: "This shows that the government is listening to the Labour party.
- Such remarks show that ministers are hoping that last night's vote will improve the government's relations

with backbenchers, thereby reducing the chances of a rebellion over foundation hospitals in the Commons next week.

Sinope summary

- That will help us in the weeks ahead.
- The bill will first have to return to a Commons standing committee because the measure has been so radically altered.
- The government has promised to use the Parliament Act to enforce the will of the Commons if peers vote down the bill.
- The government have made a firm commitment to allow this legislation to be enacted, and we fully expect it to meet its promise to end hunting with dogs.

SEAGULL summary

(1) Tony Blair last night suffered a humiliating rebuff when more than 300 Labour MPs, including seven members of the cabinet, ignored Downing Street to vote overwhelmingly in favour of an outright ban on foxhunting.

(4) Labour whips last night played down the defeat, which came a day after ministers warned that a vote for a total ban would "wreck" the government's hunting bill.

(8) Downing Street was warned by government business managers that any pressure on Labour MPs, who were entitled to a free vote last night, would have caused the government a headache.

(17) Mr Blair, who did not vote last night, will dismiss talk of a rebellion because he will insist that he was living up to his pledge to give Labour MPs a free vote.

According to our manual assessment of the content supplied by each summariser above, based on key points, flow and cohesiveness, we rank performance success for Test 1 as follows: 1. Seagull, 2. Copernic, 3. Word, 4. Sinope.

The next 7 tests are run on Text 1, and consist of the progressive addition of 7 sentence batches from unrelated texts. The results are below (more easily understandable in conjunction with the compendium text in Appendix 1). The percentages refer to summary length as a percentage of the length of the particular text version at which errors start appearing; thus the lower this percentage, the less robust we hold the summariser to be. By contrast, a high percentage is a good score, showing that a summariser is applying an algorithm capable of identifying and following the main thread of the story, and sufficiently robust to ignore irrelevance.

Test 2: Can the summariser ignore sentences from totally unrelated texts?

Addition of 3 sentences on human fertilisation (HF) to Text 1 (i.e. **9% distraction**, or 76 extraneous wds)

System	Word	Copernic*	Sinope	SEAGULL**
Max % (HF)	41	24	24	(2:2) 82

Test 3: Can the summariser ignore further sentences from different unrelated texts?

Addition of 4 sentences on Iraq conflict (IC) to Text 2 (i.e. **16% distraction**, or 144 unrelated wds)

System	Word	Copernic*	Sinope	SEAGULL**
Max % (HF)	32	49	21	(2:2) 77
Max % (IC)	71	24	50	(1:3) 92
Avge	52	37	36	85

Test 4: Can the summariser ignore sentences from unrelated texts placed in article initial position?

Addition of 4 sentences on Iraq dossiers (ID) to start of Text 3 (i.e. **27% distraction**, or 265 unrelated wds)

System	Word	Copernic*	Sinope	SEAGULL**
Max % (HF)	30	49	37	(2:1) 86
Max % (IC)	79	24	18	(1:2) 97
Max % (ID)	49	24	47	(2:3) 64
Avge	53	32	34	82

Test 5: Can the summariser ignore sentences from unrelated texts but on similar topics?

Addition of 3 sentences with ambiguous word 'hunt' (AH) to Text 4 (i.e. **33% distraction**, or 357 unrelated wds)

System	Word	Copernic*	Sinope	SEAGULL**
Max % (HF)	37	49	22	(2:1) 84
Max % (IC)	68	24	46	(1:2) 98
Max % (ID)	49	24	14	(2:3) 64
Max % (AH)	18	49	51	(2:3) 64
Avge	43	37	33	78

Test 6: Can the summariser ignore an unrelated sentence containing a related homograph?

Addition of 1 sentence with ambiguous wd 'ban' (AB) to Text 5 (i.e. **35% distraction**, or 390 unrelated wds)

System	Word	Copernic*	Sinope	SEAGULL**
Max % (HF)	35	49	22	(2:1) 85
Max % (IC)	70	24	47	(1:2) 99
Max % (ID)	49	21 (250wds)	14	(2:3) 62
Max % (AH)	16	49	45	(2:3) 62
Max % (AB)	59	49	74	(2:1) 85
Avge	46	38	40	79

Test 7: Can the summariser ignore sentences from texts with similar topics containing referentially ambiguous words?

Addition of 6 sentences with ambiguous words and similar topics (AS) to Text 6 (i.e. **44% distraction**, or 585 extraneous wds)

System	Word	Copernic *	Sinope	SEAGULL **
Max % (HF)	37	49	19	(2:1) 86
Max % (IC)	69	24	45	(1:2) 99
Max % (ID)	40	19 (250wds)	33	(2:3) 68
Max % (AH)	24	49	46	(2:3) 68
Max % (AB)	52	49	72	(3:2) 30
Max % (AS)	19	5		(2:3) 68
Avge	40	33	36	70

Test 8: Can the summariser ignore sentences from unrelated texts containing related homographs?

Addition of 6 sentences with homographs (HM) to Text 7 (i.e. **51% distraction**, or 757 extraneous wds)

System	Word	Copernic *	Sinope	SEAGULL **
Max % (HF)	34	49	19	(2:2) 67
Max % (IC)	70	24	44	(1:2) 99
Max % (ID)	39	16 (250wds)	25	(2:3) 61
Max % (AH)	16	49	41	(2:3) 61
Max % (AB)	47	49	68	(2:1) 83
Max % (AS)	14	4	5	(3:2) 27
Max % (HM)	37	7 (100wds)	1	(3:2) 27
Avge	37	28	29	61

*Copernic allows summaries of 5, 10, 25 or 50%, or of 100, 250 or 1000 words. For our texts, its 100-word summary \approx 10%, and its 250-word summary \approx 25%.

**SEAGULL operates according to thresholds of repetition between words ('links') and sentences ('bonds') rather than numbers or percentages of words. The bracketed numbers next to SEAGULL's percentages represent thresholds used.

Conclusion and Prospects

The distraction-based evaluation criteria tested above provide an objective means of comparing summariser performance. The progressive introduction of cumulative targeted distraction proves to be capable of revealing relative degrees of robustness. The tests also show that our summariser functions best.

Future refinements will include user-specification of results in terms of number of words/sentences or percentages; and system tailoring for other text types.

Reference

- Hoey, M. (1991) *Patterns of Lexis in Text*. Oxford: OUP.

Appendix: Compendium Text 8

Text 1 **in bold**, plus 7 batches of unrelated sentences, marked **(2)** for the sentences added for the 2nd test, etc.

- (4) A leading member of the foreign affairs select committee today dismissed its forthcoming report into the government's claims about Iraqi weapons, saying it would not be able to come to any "definitive conclusions".
- (4) John Maples, a former shadow foreign secretary, claimed that without seeing "all the papers" - such as the earlier drafts of the September dossier - their investigation had "a real problem".
- (4) His comments come on the day that the Liberal Democrat leader, Charles Kennedy, slammed the prime minister for failing to give evidence to the committee.
- (8) The BBC said the programme would try to ease Mr Blair's worries about turning 50 by seeking advice from TV chef Gary Rhodes and wine aficionado Oz Clarke on creating the perfect birthday party.
- (8) Stephen Downing could still receive more than £1m in compensation for being wrongly jailed, despite remaining the prime suspect for the murder of Wendy Sewell.
- (8) Banks can turn over a dirty, defaced, worn, or ripped dollar bill to a Federal Reserve Bank to be retired or destroyed, or if more than half of the dollar is clearly visible, you can go to any U.S. bank to exchange it for a fresh one.
- (8) She carried on to row for Glasgow University where she won the Scottish University Championships in the coxless four.
- (8) I had a meeting with Lionel Blair, who wanted me to act in a special for him, and he opened up his wardrobe and it was absolutely full of our costumes.
- (8) Once the down pillow is unpacked, simply re-plump and watch it return to its former fullness
- 10. Tony Blair last night suffered a humiliating rebuff when more than 300 Labour MPs, including seven members of the cabinet, ignored Downing Street to vote overwhelmingly in favour of an outright ban on foxhunting.**
- (6) But Woodward's sister, Patricia Bennetts, who has petitioned Mr Blair to ban the ship, said: "It was a floating jail and a place where torture was inflicted on victims of the Pinochet dictatorship.
- (4) This refusal to explain his actions may be behind a poll in today's Financial Times, which shows that almost two thirds of the public have lost personal trust in the Tony Blair.
- (7) Delegates are set to vote tomorrow on a motion to cut Labour funding by 25% in protests at the government's failure to adopt EU labour laws.
- (7) Thanks to last night's 11th hour compromise with Lord Puttnam's band of rebel peers, the Lords is expected to allow the bill a clear passage through to next week's vote, when the Commons is expected to give the bill final clearance.
- 15. In an embarrassment to the prime minister, who had been pressing for a compromise to preserve a handful of hunts, MPs across the chamber voted by 362 to 154 to ban the ancient sport.**
- 16. The rebels' majority was 208.**

17. **Labour whips last night played down the defeat, which came a day after ministers warned that a vote for a total ban would "wreck" the government's hunting bill.**
18. (2) Babies created from the tissue of an "unborn mother" came a step closer yesterday after scientists said they had made progress in producing eggs from an aborted foetus.
19. **A government source said: "This shows that the government is listening to the Labour party.**
20. **That will help us in the weeks ahead."**
21. (7) The government will have its bill in place by the autumn, while Lord Puttnam and his supporters will claim a major victory in safeguarding the diversity of media outlets in the UK.
22. (5) Blair will tackle the Glasgow conference, hoping for ammunition from Blix's pivotal progress report on the hunt for weapons of mass destruction - due on Friday.
23. **Such remarks show that ministers are hoping that last night's vote will improve the government's relations with backbenchers, thereby reducing the chances of a rebellion over foundation hospitals in the Commons next week.**
24. **Downing Street was warned by government business managers that any pressure on Labour MPs, who were entitled to a free vote last night, would have caused the government a headache.**
25. (3) The US military headquarters in Baghdad said that it had no information on the incident.
26. (7) And it is fitting that the issue that proved a stumbling block right up to the climax of the report stage in the Lords, was the same one that had dominated the headlines all the way through the process.
27. (7) Mr Woodley queried why Mr Blair's was the first Labour government to extend privatisation, do nothing to prevent job losses and preside over an increase in the gap between rich and poor.
28. (3) It was unclear who the Iraqi victims were.
29. **The government signalled it was in trouble when Alun Michael, the rural affairs minister, withdrew a compromise motion minutes before a lengthy debate was due to end.**
30. **This would have tightened up the bill to ensure that all but a handful of hunts would have been banned.**
31. **Under the original bill, some hunts would have been allowed to continue if they could show they were not cruel and were the only means of pest control.**
32. **The government performed its climbdown when it became clear that Labour MPs were abandoning the government in droves.**
33. **Charles Clarke, the education secretary, and Ian McCartney, the Labour chairman, led the unofficial cabinet rebellion on the free vote.**
34. **In all, 62 ministers and whips voted in favour of the outright ban, which was proposed by the former minister Tony Banks.**
35. (5) Staggering around accidentally breaking things while on the hunt for a bag of crisps is alien to the political classes, more interested as they are in policy, public service and acts of auto-eroticism.
36. **Five ministers and whips backed the prime minister's preferred compromise.**
37. **In an embarrassment to the environment secretary, Margaret Beckett, whose department is piloting the bill, the "rebels" included her deputy, Elliot Morley, who sat on the frontbench during the debate.**
38. **Mr Blair, who did not vote last night, will dismiss talk of a rebellion because he will insist that he was living up to his pledge to give Labour MPs a free vote.**
39. (2) The human fertilisation and embryology authority, which regulates fertilisation procedures in Britain, said such a procedure for producing eggs for IVF would not be allowed in Britain.
40. **Last night's vote guarantees a bruising row with the Lords, where peers are determined to preserve fox-hunting.**
41. **But the bill will first have to return to a Commons standing committee because the measure has been so radically altered.**
42. **Ministers had hoped that the prospect of this unusual move would persuade Labour opponents of hunting to back their compromise because the delay might have jeopardised the entire bill.**
43. **But this tactic collapsed when Mr Michael said the amended bill would be finished in the Commons by the July recess.**
44. (3) Fallujah, which lies 35 miles west of Baghdad, has been a hotbed of anti-US activity, and the scene of several confrontations between US troops and insurgents.
45. **The government has promised to use the Parliament Act to enforce the will of the Commons if peers vote down the bill.**
46. **Initial doubts that the act could be used were allayed when the Speaker indicated the delay is unlikely to cause a problem.**
47. **The bill has to be presented to the Lords a month before the end of the parliamentary session in November for the act to be enforced.**
48. **Mr Banks warmly welcomed the vote.**
49. **He said: "It is excellent news.**
50. **We have got some tidying up to do, which we will do in committee, and hopefully we have achieved a total ban on the hunting of wild mammals with dogs."**
51. (2) Last month, Suzi Leather, the HFEA chairman, acknowledged that rules made over a decade ago may be past their sell-by date.
52. (3) Despite the unrest, tribal elders in the city insist that people want to cooperate with the occupying US authorities.
53. **But he warned: "We have promises that the Parliament Act will be used, but until the piece of legislation is signed off by the secretary of state, we can never be certain of anything."**
54. (5) And what Blix had to say last week was that the intelligence he had been given by the US and the UK was fundamentally useless in the hunt for weapons of mass destruction.
55. (7) The call to arms is the latest onslaught in the long running battle over the future political direction of the government and the Labour party.
56. **Anti-hunting groups also welcomed the vote.**
57. **Phyllis Campbell-McRae, the UK director of the International Fund for Animal Welfare, said: "The vote means that the House of Commons has signalled a total end to this barbaric activity.**
58. **"The government have made a firm commitment to allow this legislation to be enacted, and we fully expect it to meet its promise to end hunting with dogs."**
59. **But Simon Hart, head of the Countryside Alliance's campaign for hunting, said: "Rather than the death of hunting, this is the death of the hunting bill.**
60. **The bill is in complete chaos and is now terminally damaged."**