

Automatic Summarization (Mani) Book Review

Hal Daumé III

March 16, 2004

Researchers in automatic document summarization have already adopted many techniques from existing machine translation literature. Likewise, there is much that the machine translation community can learn from current research in summarization. *Automatic Summarization*, by Inderjeet Mani, provides a firm grounding in the primary techniques that have been applied to the summarization task, so that researchers or students unfamiliar with the topic will be sufficiently versed in its primary techniques to begin their own work. The book provides a logically structured, comprehensive survey of what the field of summarization was, up through the time of the book's publication.

Overall, *Automatic Summarization* provides a strong and effective comparison between techniques that professional human summarizers employ and those used by automatic summarization systems. This is one of the aspects that differentiates the book from *Advances in Automatic Text Summarization*, edited by Mani and Maybury. For a beginner in the summarization field, the introductory coverage of the various dimensions along which summarization systems can lie, and the coverage of basic definitions, is essential. The format is effective for readers of varying levels of sophistication: for students, the chapter conclusions help solidify the main points; for researchers, they provide a good reference into summarization literature.

The book covers much ground: the content flows from simple extraction-based systems, through revision and discourse, to abstraction, multi-document summarization and multi-media summarization. This content is bookended by an introductory chapter on professional summarization and a concluding chapter on summary evaluation.

The introductory chapter on professional summarization sets the framework for the remainder of the book, in which each task is considered, as it relates to the methods humans use for producing summaries. The automatic text summarization discussed ranges from simple single-document sentence extraction systems to systems that produce abstracts of document collections.

The research presented in the chapters on extraction, revision and discourse has reached some reasonable level of stability, and accordingly, the chapter organization is easy to follow. The quality of the organization decreases in the chapters on abstraction, multi-document summarization and multi-media summarization; this reflects both that the later topics cover research that is, itself, inconclusive and that the topics are, themselves, less interrelated than those discussed in the first few chapters.

The concluding chapter covers the problem of summary evaluation. Evaluation is largely an unsolved problem in summarization, due greatly to the fact that different humans typically create very different summaries of the same document. This pre-

vents the establishment of a clear metric against which to judge machine-generated summaries. Mani discusses several alternative approaches to performing both human-assisted and automatic evaluations.

Researchers in machine translation will immediately recognize similarities between semantic-based summarization and interlingua-based MT, which Mani points out. In both cases, the textual input is parsed into a semantic representation, manipulated (i.e., summarized or translated), and then output is generated. However, there are also many approaches to summarization that differ dramatically from research in MT (and vice-versa). Both summarization and pure natural language generation (see, for instance, the work of Robin and McKeown [RM96]) have benefited from revision operators, and it is likely that such techniques could improve the quality of MT output. Mani dedicates an entire chapter to revision-based summarization.

Another familiar topic to summarization researchers is discourse theory. A substantial portion of the book is devoted to discourse analysis (partitioned into the coherence/cohesion distinction) and its applications to summarization. There has been relatively little work in applying discourse knowledge to the task of machine translation (with the exception of anaphora resolution – see, for instance, the work of Peral [PF03] – which isn't discussed in much depth in this book). There are a few exceptions (for instance, Marcu et. al. [MCW00]) that attempt to translate discourse structures from English to Japanese), but for the most part, this is an untapped resource in machine translation. One advantage that this book presents over reading a book or paper on the topic of discourse, is that it presents discourse theories in a highly practical light.

The range of systems referenced in this book is wide enough that the primary techniques of the field are well represented, but not so wide as to make the discussion watered-down. The book correctly emphasizes the importance of evaluation, despite the fact that, as a community, we have not yet reached the point where we can do this reliably. The discussion of each system flows logically into the next, and the historical discussion provides a useful paradigm for considering current problems in the field.

The strong comparisons drawn between human summarization and automatic summarization, which is certainly one of the strengths of this book, is also one of its largest downfalls. The book emphasizes that human summarizers never read the entire document that they are summarizing, usually because this would take far too long. On the other hand, for a machine to “read” an entire document is trivial. It is certainly worth modeling summarization systems after humans, but there are also many capabilities available to a machine that are not available to a human; we would be remiss to ignore them. This dichotomy is even more pronounced in multi-document summarization and seems to limit the information that could be gained from observing humans performing the task.

The intended audience is perhaps too broad: in order to benefit maximally from this book, some background in computational linguistics is desirable. It would be very difficult for someone not minimally versed in these topics to understand which topics discussed are difficult and which are not, as the book tends not to ascribe level-of-difficulty measures to its varied foci.

In the postscript, Mani discusses the various directions in which he thinks the field will go. This section is only two pages long and, while the book does primarily focus on past work, rather than forecasting, the book would have benefited from more detail

here. Nevertheless, he calls for many of the obvious things: corpus-based abstraction, multi-document summarization, annotated corpora and more human studies.

Another form of summarization – cross-lingual summarization – is also an active field of research, though it is not discussed in any depth in this book. However, for researchers in the field of machine translation, this book provides a sufficient introduction to summarization to enable them to begin work in cross-lingual summarization, should they desire.

This book should make apparent to a researchers in machine translation the various similarities between their field and automatic document summarization: both have used the analysis/process/synthesize paradigm extensively; extrinsic evaluations can be carried out in largely the same manner; machine-aided human translation parallels machine-aided human summarization (and, of course, the same is true of human-aided machine translation and summarization); and in a small domain, the techniques used for abstraction and translation are largely the same. Statistical parameter estimation techniques originating in machine translation literature have been applied to the task of headline generation by Banko et. al. [BMW00] and Schwartz et. al. [RSD02]. Lin and Hovy [LH03] use n-gram statistics to attack the problem of summary evaluation in the style of Papineni et. al. [PRWZ02]. Mani’s book highlights the opportunity for cross-fertilization between the fields of machine translation and summarization; he shows that, just as researchers in summarization have come to adopt techniques from the MT community, so can the MT community learn from the research efforts in the field of summarization.

References

- [BDDM93] Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311, 1993.
- [BMW00] Michele Banko, Vibhu Mittal, and Michael Witbrock. Headline generation based on statistical translation. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL-2000)*, pages 318–325, Hong Kong, October 1–8 2000.
- [LH03] Chin-Yew Lin and Eduard Hovy. Automatic evaluation of summaries using n-gram co-occurrence statistics. In *Proceedings of the 2003 Language Technology Conference (HLT-NAACL-2003)*, Edmonton, Canada, May 27 – June 1 2003.
- [MCW00] D. Marcu, L. Carlson, and M. Watanabe. The automatic translation of discourse structures. In *Proceedings of NAACL, 2000*.
- [MM99] Inderjeet Mani and Mark Maybury, editors. *Advances in Automatic Text Summarization*. The MIT Press, 1999.

- [PF03] J. Peral and A. Ferrandez. Translation of pronominal anaphora between English and Spanish: Discrepancies and evaluation. *Journal of Artificial Intelligence Research*, 18, 2003.
- [PRWZ02] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 311–318, 2002.
- [RM96] Jacques Robin and Kathleen McKeown. Empirically designing and evaluating a new revision-based model for summary generation. *Artificial Intelligence*, 85(1-2):135–179, 1996.
- [RSD02] D. Zajic R. Schwartz and B. Door. Automatic headline generation for newspaper stories. In *Proceedings of the Document Understanding Conference (DUC)*, 2002.