A COMPUTATIONAL MODEL FOR HUMAN AND MACHINE READING

Jonathan J. Hull
Department of Computer Science, SUNY Buffalo, Buffalo, New York 14260
hull@cs.buffalo.edu

INTRODUCTION

The reading of text by computer without human intervention remains an elusive goal of Artificial Intelligence research. Reading is the transformation of an arbitrary page of text, that could contain a mixture of machine-printed, hand-printed, or handwritten text, from its representation as a two-dimensional image into a form understandable by a computer, such as ASCII code. The current lack of a technique with these capabilities is interesting in light of the relative ease with which people read and the many years of investigation into computer reading algorithms, the methods people use to read text, and the long history of Artificial Intelligence research into computer vision.

The parallel between algorithms for reading text and explanations for human performance is most interesting. With some notable exceptions, most reading algorithms use a *character recognition* approach in which words are segmented into isolated characters that are individually recognized. For these algorithms reading is equivalent to a sequence of character recognitions.

The way people read is significantly different from character recognition. We bring to reading a wealth of information about the world and expectations about what we will read. This is mixed with knowledge about how text is arranged on a page, knowledge of the syntax and semantics of language, and visual knowledge about letters and words. The recognition processes that take place during reading use visual information from much more than just isolated characters. Whole words or groups of characters are recognized by processes that, in some cases, do not even require detailed visual processing. This is because human reading uses many knowledge sources to develop an understanding of a text while it is being recognized. This integration of understanding and recognition is responsible for human performance in reading.

The fact that few reading algorithms have utilized the many disparate knowledge sources or the recognition strategy of a human reader might explain the gap between the reading proficiency of algorithms and people. Although some character recognition techniques have been augmented with knowledge about words, no reading algorithm has been proposed that fully utilizes the knowledge that is routinely employed by a human reader [7]. Such an algorithm would have the potential of yielding substantial improvements in performance.

A COMPUTATIONAL THEORY AND ALGORITHM FOR READING

The computational theory of reading discussed here is derived from work on human reading that includes studies of human eye movements [6]. To a person who reads a line of text, it seems to them as if their eyes move smoothly from left to right. However, this is not completely true. In reality, our eyes move in ballistic jumps called *saccades* from one *fixation point* to the next. During a saccade the text is blurred and unreadable. (This is not apparent to the reader.) Therefore, most of the visual processing of reading takes place during the fixations. Usually there are about one to three fixations per word that occur near the beginning of the word. However, interestingly enough, some words are never fixated. The sequence of fixations is approximately from left to right across a line of text, however, regressions do occur frequently. An example is shown in Figure 1 of the sequence of fixations in a line of text [1].

There are two types of visual processing in reading. In the first type of processing, information from peripheral vision provides a gross visual description of words to the right of the current fixation point. This information is used to form expectations about the words. The second stage of processing occurs on a subsequent fixation when these expectations are integrated with other visual information.

The visual processing is influenced by many high-level factors that include the reason a person is reading the passage of text, as well as the familiarity of the reader with the subject and his or her skill level. A more skilled reader uses visual information more economically than a less skilled one [2]. That is, a skilled reader uses less visual processing than a poor reader. Recent work has also shown that syntactic processing also influences the visual processing of a text.

The computational theory of reading contains three stages that are similar to those of human reading. The first stage generates hypotheses about words from a gross visual description. This is similar to the visual processing of words to the right of a fixation point and is an essential component of one theory of human reading [3]. The second stage uses these hypotheses to determine a feature testing sequence that can be executed on the image to recognize the word. This sequence is adaptable to different high-level influences and can be executed at different physical locations in the word. This stage is similar to the detailed visual processing that takes place at a fixation. The third stage of the theory concerns high-level processing. This stage captures the influence of the

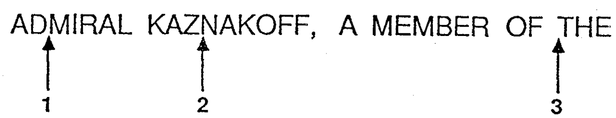

Figure 1. Sequence of fixations in a line of text [1].

various non-visual processes that influence reading such as syntax and semantics. These processes remove word-hypotheses from consideration that do not agree with the high-level constraints. This is a way to represent the influence of many high-level knowledge sources.

A set of algorithms that implement the three steps of the theory are discussed here. A routine for *hypothesis generation* implements the gross visual processing of the words in a text. This routine provides the expectations mentioned above by finding a set of words in a dictionary that have the same feature description as an input word. A technique for *global contextual analysis* uses global infomation about a text to reduce the number of words output by the hypothesis generation stage of the algorithm. A method for *hypothesis testing* implements the integration phase of visual processing. This routine determines which hypothesis is contained in the image.

HYPOTHESIS GENERATION

The hypothesis generation component of the algorithm uses a description of the gross visual characteristics of a word image to index into a dictionary and retrieve a subset of words called a *neighborhood* that have the same description. The visual description is the left-to-right sequence of occurrence of a small number of features. The features are simple and easy to extract to increase the reliability of the technique in the presence of noise. This approach is suitable for generating hypotheses about an input word since a small number of features can partition a large dictionary into a limited number of small neighborhoods [5]. This is less error-prone than using many features to carry out complete recognition.

GLOBAL CONTEXTUAL ANALYSIS

Global contextual analysis uses global information about a text to reduce the number of words in the neighborhoods determined by hypothesis testing. Various forms of such information have been tested, including tables of allowable transitions between words, and tables of allowable transitions between features of words. The second technique has the most generality since the features can be applicable to text in almost any font or script. For example, it was shown that features as simple as the number of characters in a word and the identity of its first character can substantially reduce the size of typical neighborhoods [4]. This is done by a-priori construction of a table of allowable pairs of these features for two words in specified locations, e.g., the city and state name on an envelope. Given that the length and identity of the first character in the city name have been determined, members are removed from the neighborhood of the state name that do not have a length and first character specified in the table.

HYPOTHESIS TESTING

The hypothesis testing strategy uses the words in a neighborhood to determine a feature testing sequence. The testing sequence is structured as a tree that specifies the order features are tested in a word image. The results of those tests determine successive tests and reduce the number of words that could match the input image.

At each step in hypothesis testing, different discrimination tests can be applied to the input word. Each discrimination test refers to a location between two adjacent features in the output of the neighborhood calculation. Given that the word in the image is a word in the neighborhood, the features at any one location are constrained to be in a fixed set of alternatives. Therefore, the discrimination tests tell the difference between a, usually small, number of alternative features.

If the number of alternative features that can occur between adjacent features in the output of the neighborhood calculation is much less than 26, a simpler discrimination is possible than if a character recognition methodology were employed.

DISCUSSION AND CONCLUSIONS

A computational theory and algorithm for reading was presented. The work presented in this paper sought to bridge the gap between theory and methods and to bring to reading algorithms the benefits of many years of psychological investigation of human reading. The three-stage algorithm presented here implemented three of the basic processes in human reading. This is an example of applying the theoretical constructs of Artificial Intelligence to an engineering problem.

REFERENCES

1. W. K. Estes, "On the interaction of perception and memory in reading," in *Basic Processes in Reading: Perception and Comprehension*, D. LaBerge and S. J. Samuels (editor), Lawrence Erlbaum Associates, Hillside, New Jersey, 1977.
2. R. N. Haber and L. R. Haber, "Visual components of the reading process," *Visible Language XV*, 2 (1981), 147-181.
3. J. Hochberg, "Components of literacy: Speculations and exploratory research," in *Basic Studies on Reading*, H. Levin and J. P. Williams (editor), Basic Books, Inc., New York, 1970, 74-89.
4. J. J. Hull and S. N. Srihari, "Use of external information in ZIP code recognition," *United States Postal Service Advanced Technology Conference*, October 21-23, 1986, 361-370.
5. J. J. Hull, "Hypothesis generation in a computational model for visual word recognition," *IEEE Expert*, August, 1986, 63-70.
6. K. Rayner, *Eye movements in reading: perceptual and language processes*, Academic Press, New York, 1983.
7. J. Schurmann, "Reading machines," *Proceedings of the 6th International Conference on Pattern Recognition*, Munich, West Germany, October 19-22, 1982, 1031-1044.

Advanced Printing of Symposium Summaries

SPSE's 40th ANNUAL CONFERENCE
And
SYMPOSIUM ON
HYBRID IMAGING SYSTEMS

May 20-21, 1987
Genesee Plaza/Holiday Inn
Rochester, New York

Published by:

SPSE - THE SOCIETY FOR IMAGING SCIENCE AND TECHNOLOGY
7003 Kilworth Lane
Springfield, Virginia USA
(703) 642-9090