Towards a Computational Model of Discourse Summarization

Samuel W. K. Chan
School of Computer Science & Engineering
The University of New South Wales
NSW 2052, Australia
swkchan@cse.unsw.edu.au

Abstract

Understanding a discourse is considered to involve a series of specific processing phases which final result is a complete semantic, mental representation (Johnson-Laird, 1983; van Dijk & Kintsch, 1983). This result is not only a representation of the text, but rather of what the text is about. When a reader is asked to summarize a discourse, vast amounts of information within the discourse are selectively ignored in order to produce a distilled version of the original text. This simplification process emphasizes central elements of the discourse while the peripheral details are neglected. It is further demonstrated that discourse can be represented as a skeleton in which the relationships among the clauses could be chunked in a way that replicated the semantic structure of the original discourse (Grosz & Sidner, 1986). Textual continuity, which differentiates a text from a random sequence of sentences, is a prime factor in discourse summarization (Ehrlich & Charolles, 1991).

In this paper, we present a computational model for transforming discourses into Quasi-Mental Clusters (QMCs) through a convergence process. The process is interpreted as a particular transformation of a given set of discourse segments and concepts by examining the textual continuity. Before the process, a cohesion parsing is first conducted in testing the local cohesion amongst the sentences. It is achieved in a constraint net which is formulated as a constraint satisfaction problem over a set of finite elements (Waltz & Pollack, 1985). The elements in the net may represent words, phrases, and more importantly, the buffers which are designated to carry each prior analyzed sentences over into the current processing cycle, in hope that they would serve as common bridging elements between the sentences. An efficient filtering algorithm is employed to reduce the incohesive ones while cohesive links amongst sentences are then defined. In the convergence process, sentences in a discourse are represented as nodes and connected by the links in a modified Brain-State-in-a-Box (BSB) network (Anderson & Murphy, 1986). They are highly interconnected and feedback upon themselves. The BSB network operates by accepting a pattern of activations and amplifying that pattern through the feedback loop. Competing coalitions of the nodes drive the network into a stable equilibration from which the QMCs are extracted. The strongest connection in the QMCs will arise from the pair of nodes that maintain high activations for a prolonged period of time. The model is tested using children's stories and the results attest its validity. The representation of texts in QMCs captures some important aspects of the memory representation of discourses. Thus, the resulting QMCs are the useful data structures in summarization, question answering and knowledge discovery in discourses.

References

Anderson, J.A., & Murphy, G.L. (1986). Psychological concepts in a parallel system. *Physica D*, 22, 318-336.

Ehrlich, M.-F., & Charolles, M. (1991). Aspects of textual continuity: Psycholinguistic approaches. In G. Denhiere & J.-P. Rossi (Eds.), *Text and Text Processing*, North-Holland, 269-285.

Grosz, B.J., & Sidner, C.L. (1986). Attention, intention, and the structure of discourse. *Computational Linguistics*, 12, 175-204.

Johnson-Laird, P.N. (1983). *Mental Models*. Cambridge, MA: Harvard University Press.

van Dijk, T.A., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. NY: Academic.

Waltz, D.L., & Pollack, J.B. (1985). Massively parallel parsing: a strongly interactive model of natural language interpretation. *Cognitive Science*, 9, 51-74.