Chapter 11
Measuring Textual Context Based on Cognitive Principles

Ning Fang
Shanghai University, China

Xiangfeng Luo
Shanghai University, China

Weimin Xu
Shanghai University, China

ABSTRACT

Based on the principle of cognitive economy, the complexity and the information of textual context are proposed to measure subjective cognitive degree of textual context. Based on minimization of Boolean complexity in human concept learning, the complexity and the difficulty of textual context are defined in order to mimic human's reading experience. Based on maximal relevance principle, the information and cognitive degree of textual context are defined in order to mimic human's cognitive sense. Experiments verify that more contexts are added, more easily the text is understood by a machine, which is consistent with the linguistic viewpoint that context can help to understand a text; furthermore, experiments verify that the author-given sentence sequence includes the less complexity and the more information than other sentence combinations, that is to say, author-given sentence sequence is more easily understood by a machine. So the principles of simplicity and maximal relevance actually exist in text writing process, which is consistent with the cognitive science viewpoint. Therefore, this chapter's measuring methods are validated from the linguistic and cognitive perspectives, and it could provide a theoretical foundation for machine-based text understanding.

DOI: 10.4018/978-1-4666-0261-8.ch011

INTRODUCTION

According to Cognitive Informatics (Wang, 2002a, 2007b), a *concept* is defined as a cognitive unit to identify and/or model a real-world concrete entity and a perceived-world abstract object. According to Wang (2006b), the formal treatment of concepts and a new mathematical structure known are defined as Concept Algebra in Cognitive Informatics. The semantic environment or context (Ganter & Wille, 1999; Hampton, 1997; Hurley, 1997; Medin & Shoben, 1988) in a given language is denoted as a triple, i.e.: a finite or infinite nonempty set of objects, a finite or infinite nonempty set of attributes, and their relations. On the basis of the Object-Attribute-Relation model and the definition of context, an abstract concept is a composition of the above three elements. Although Concept Algebra includes the operations of objects, attributes, and relations in an abstract concept, the measure of the context of textual concept is not given in Cognitive Informatics, and it is a key issue to understand the textual concept.

At present, the measure in Cognitive Informatics mainly include as follows: the theorem (Wang, 2007d) indicates that the complexities of the syntactic rules (or grammar) and of the semantic rules are inversely proportional; the cognitive complexity of software systems presented is a measure of cognitive and psychological complexity of software as a human intelligent artifact, which takes into account of both internal structures of software and the I/O data objects under processing (Wang, 2006a).

According to the Object-Attribute-Relation model (Wang, 2003, 2007b, 2007c), the semantics of a sentence may be considered having been understood when: a) The logical relations of parts of the sentence are clarified; and b) All parts of sentence are reduced to the terminal entities, which are either a real-world image or a primitive abstract concept (Wang, 2007d). Furthermore, national language is context sensitive. That is to say, context can help us to understand a textual concept. Therefore, to understanding a textual concept, it is necessary of 1) obtaining a primitive abstract concept from textual sentences; 2) computing the weights of logical relations among sentences according to cognitive principles; 3) measuring the textual context in sentential granularity based on the weights.

Rosch (1975) indicates that *every organism hopes that the more information is acquired, the less energy is consumed in the surrounding environment*, i.e. the principle of cognitive economy. This principle enlightens us a new approach to measure textual context from subjective cognitive perspectives so as to help machine understand a text, if we regard the text understanding process as a special cognitive process. So we propose two criteria to measure textual context based on cognitive economical principle, i.e. the complexity and the information of textual context (here the complexity is considered as the energy approximately).

A *concept* in linguistics is a noun or noun-phrase that serves as the subject or object of a *to-be* statement (Hurley, 1997; Wang, 2002b, 2007a). Therefore, in this article, we only discuss concept generated from nouns in a sentence. Commonly, context refers to a wide range that includes textual semantic relations, common sense, knowledge, communication background, and culture background, etc. However, in this article we only consider simple textual context, i.e. co-text that includes textual semantic relations generated from related keywords and sentences in a text, some common sense and knowledge.

FORMULIZATION OF TEXTUAL CONTEXT

Simply speaking, textual context is constructed by sequential sentences, a sentence constructed by some significant word's combinations. If a keyword is regarded as an attribute and a sentence as an object in a text, the understanding of the