Random Walks on Text Structures

Rada Mihalcea

University of North Texas
Computer Science Department
rada@cs.unt.edu

Abstract. Since the early ages of artificial intelligence, associative or semantic networks have been proposed as representations that enable the storage of language units and the relationships that interconnect them, allowing for a variety of inference and reasoning processes, and simulating some of the functionalities of the human mind. The symbolic structures that emerge from these representations correspond naturally to graphs – relational structures capable of encoding the meaning and structure of a cohesive text, following closely the associative or semantic memory representations. The activation or ranking of nodes in such graph structures mimics to some extent the functioning of human memory, and can be turned into a rich source of knowledge useful for several language processing applications. In this paper, we suggest a framework for the application of graph-based ranking algorithms to natural language processing, and illustrate the application of this framework to two traditionally difficult text processing tasks: word sense disambiguation and text summarization.

1 Introduction

Many language processing applications can be modeled by means of a graph. These data structures have the ability to encode in a natural way the meaning and structure of a cohesive text, and follow closely the associative or semantic memory representations. For instance, Figure 1 shows examples of graph representations of textual units[1] and the relationships that interconnect them: 1(a) (adapted from [6]) shows a network of concepts related by semantic relations – simulating a fragment of human memory, on which reasoning and inferences about various concepts represented in the network can be performed; 1(b) shows a network with similar structure, this time automatically derived via definitional links in a dictionary; finally, 1(c) is a graph representation of the cohesive structure of a text, by encoding similarity relationships between textual units.

Provided a graph representation of the text, algorithms for the activation or ranking of nodes in such structures can be used to simulate the functioning of human memory, consequently resulting in solutions for a variety of natural language processing tasks that can be modeled by means of a graph. In this paper, we suggest a framework for the application of graph-based ranking algorithms to text-based graph structures, and show how two text processing applications: word sense disambiguation and text summarization, can find successful solutions within this framework.

[1] We use the term *textual unit* to refer to the textual representation of a *cognitive unit* as defined by Anderson [1]. It can be a word, a concept, a sentence, or any other unit that can find a representation in language.