Artificial Intelligence 228 (2015) 95–128

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

From senses to texts: An all-in-one graph-based approach for measuring semantic similarity

Mohammad Taher Pilehvar *, Roberto Navigli

Department of Computer Science, Sapienza University of Rome, Italy

ARTICLE INFO

Article history:
Received 3 September 2014
Received in revised form 30 June 2015
Accepted 9 July 2015
Available online 15 July 2015

Keywords:
Semantic similarity
Lexical semantics
Semantic Textual Similarity
Personalized PageRank
WordNet graph
Semantic networks
Word similarity
Coarsening WordNet sense inventory

ABSTRACT

Quantifying semantic similarity between linguistic items lies at the core of many applications in Natural Language Processing and Artificial Intelligence. It has therefore received a considerable amount of research interest, which in its turn has led to a wide range of approaches for measuring semantic similarity. However, these measures are usually limited to handling specific types of linguistic item, e.g., single word senses or entire sentences. Hence, for a downstream application to handle various types of input, multiple measures of semantic similarity are needed, measures that often use different internal representations or have different output scales. In this article we present a unified graph-based approach for measuring semantic similarity which enables effective comparison of linguistic items at multiple levels, from word senses to full texts. Our method first leverages the structural properties of a semantic network in order to model arbitrary linguistic items through a unified probabilistic representation, and then compares the linguistic items in terms of their representations. We report state-of-the-art performance on multiple datasets pertaining to three different levels: senses, words, and texts.

1. Introduction

The measurement of semantic similarity is an essential component of many applications in Natural Language Processing (NLP) and Artificial Intelligence (AI). Measuring the semantic similarity of text pairs enables the evaluation of the output quality of machine translation systems [1] or the recognition of paraphrases [2], while laying the foundations for other fields, such as textual entailment [3,4], information retrieval [5,6], question answering [7,8], and text summarization [9]. At the word level, semantic similarity can have direct benefits for areas such as lexical substitution [10] or simplification [11], and query expansion [12], whereas, at the sense level, the measurement of semantic similarity of concept pairs can be utilized as a core component in many other applications, such as reducing the granularity of lexicons [13,14], Word Sense Disambiguation [15], knowledge enrichment [16], or alignment and integration of different lexical resources [17–20].

As a direct consequence of their design, most of the current approaches to semantic similarity are limited to operating at specific linguistic levels. For instance, similarity approaches for large pieces of texts, such as documents, usually utilize the statistics obtained from the input items [21–24] and, therefore, are inapplicable for pairs of linguistic items with small contextual information, such as words or phrases. A unified approach that can enable the efficient comparison of linguistic items at different linguistic levels would be able to free downstream NLP applications from needing to consider the type of

* Corresponding author.
E-mail addresses: pilehvar@di.uniroma1.it (M.T. Pilehvar), navigli@di.uniroma1.it (R. Navigli).

http://dx.doi.org/10.1016/j.artint.2015.07.005
0004-3702/

items being compared. However, despite the potential advantages, very few approaches have attempted to cover different linguistic levels: most previous work has focused on tuning or extending existing approaches to other linguistic levels, rather than proposing a unified similarity measurement method. For instance, sense-level measures have been extended to the word level by assuming the similarity of word pairs as being that of the closest senses of the two words [25], whereas word-level approaches have been utilized for measuring the similarity of text pairs [26]. However, these approaches do not usually work on the extended levels as effectively as they do on the original ones. For instance, measures for concept semantic similarity often fall far behind the state of the art when extended for use in measuring the similarity of word pairs [27–29].

In this article, we propose a unified approach to semantic similarity that can handle items from multiple linguistic levels, from sense to text pairs. The approach brings together two main advantages: (1) it provides a unified representation for all linguistic items, enabling the meaningful comparison of arbitrary items, irrespective of their scales or the linguistic levels they belong to (e.g., the phrase *take a walk* to the verb *stroll*); (2) it disambiguates linguistic items to a set of intended concepts prior to modeling and, hence, it is able to identify the semantic similarities that exist at the deepest sense level, independently of the text's surface forms or any semantic ambiguity therein. For example, consider the following two pairs of sentences:

a1. *Officers fired.*
a2. *Several policemen terminated in corruption probe.*

b1. *Officers fired.*
b2. *Many injured during the police shooting incident.*

Surface-based approaches that are merely based on string similarity cannot capture the similarity between any of the above pairs of sentences as there exists no lexical overlap. In addition, a surface-based semantic similarity approach considers both *a1* and *b1* as being identical sentences, whereas we know that different meanings of the verb *fire* are triggered in the two contexts.

In our recent work [30] we presented *Align, Disambiguate, and Walk* (ADW), a graph-based approach for measuring semantic similarity that can overcome both these deficiencies: firstly, it transforms words to senses prior to modeling, hence providing a deeper measure of similarity comparison and, secondly, it performs disambiguation by taking into account the context of the paired linguistic item, enabling the same linguistic item to have different meanings when paired with different linguistic items. Our technique models arbitrary linguistic items through a unified representation, called *semantic signature*, which is a probability distribution over concepts, word senses, or words in a lexicon. Thanks to this unified representation, our approach can compute the similarity of linguistic items at and across arbitrary levels, from word senses to texts. We also proposed a novel approach for comparing semantic signatures which provided improvements over the conventional cosine measure. Our approach for measuring semantic similarity obtained state-of-the-art performance on several datasets pertaining to different linguistic levels. In this article, we extend that work as follows:

1. we propose two novel approaches for injecting out-of-vocabulary (OOV) words into the semantic signatures, obtaining a considerable improvement on datasets involving many OOV entries while calculating text-level semantic similarity;
2. we provide an approach for creating a semantic network from Wiktionary and show that it can be used effectively for generating semantic signatures and for comparing pairs of items;
3. we re-design experiments in the sense and text levels in order to have a more meaningful comparison of different similarity measurement techniques and also perform evaluation on more datasets at the word level.

The rest of this article is organized as follows. We first introduce in Section 2 the three main linguistic levels upon which we focus in this article. We then provide an overview of the related work in Section 3. Section 4 explains how we constructed different semantic networks to be used as underlying resources of our approach. A detailed description of our similarity measurement approach, i.e., ADW, is provided in Section 5, followed by our experiments for evaluating the proposed technique at different linguistic levels in Section 6. Finally, we provide the concluding remarks in Section 7.

2. Semantic similarity at different levels

Measuring the semantic similarity of pairs of linguistic items can be performed at different linguistic levels. In this work, we focus on three main levels: senses, words, and sentences. Table 1 lists example semantic similarity judgments for pairs belonging to each of these three linguistic levels.[1] In our example in Table 1(a), which is based on the WordNet 3.0 sense inventory [31], the precious stone sense of the noun *jewel* (jewel^1_n) is paired with three senses of the noun *gem*: *gem*^5_n, which is synonymous to *jewel*^1_n being the stone used in jewelry, *gem*^3_n, which refers to a brilliant and precious person, and *gem*^4_n, which is synonymous to *muffin* that is a sweet baked bread. In the sense-level similarity, the task is to compute the

[1] Following [15], we denote the ith sense of the word w with the part of speech p as w^i_p in the reference inventory.