W

Web Advertising

►Text Mining for Advertising

Weight

Risto Miikkulainen
The University of Texas at Austin
Austin, TX, USA

Synonyms

Connection strength; Synaptic efficacy

Definition

In a ►neural network, connections between neurons typically have weights that indicate how strong the connection is. The neuron computes by forming a weighted sum of its input, i.e., the activation of each input neuron is multiplied by the corresponding connection weight. Adapting such weights is the most important way of learning in neural networks. Connection weights are loosely modeled after the synaptic efficacies in biological neurons, where they determine how large a positive or negative change in the membrane potential each input spike generates (see ►Biological Learning). In most models, all connection parameters are abstracted into a weight: attenuation or interaction of the potentials and connection delays are usually not taken into account. The weights are usually real-valued numbers $(-\infty .. \infty)$, although in some algorithms, intended for VLSI implementation, the range and precision of these values can be restricted (or weights eliminated altogether). Weights in some methods can be restricted to positive values if the inputs are known to be positive and the method is based on comparing the similarity to the weights (as in e.g., ►Self-Organizing Maps, ►Adaptive Resonance Theory, and ►Radial Basis Function networks). Most learning methods are based on adjusting the weight values. The weights are often initialized to small random values, although if enough is known about the input space and the task, more systematic initialization can improve performance significantly. The weights are then adjusted based on local information that is available on either side of the connection. Usually, only small modifications are made in each learning step to avoid disrupting what the network already knows, and learning converges over time to a setting of values that solves the task.

Within-Sample Evaluation

►In-Sample Evaluation

Word Sense Disambiguation

Rada Mihalcea
University of North Texas
Denton, TX, USA

Synonyms

Learning word senses; Solving semantic ambiguity

Definition

Ambiguity is inherent to human language. In particular, word sense ambiguity is prevalent in all natural languages, with a large number of the words in any given language carrying more than one meaning. For instance, the English noun *plant* can mean *green plant* or *factory*; similarly the French word *feuille* can mean *leaf* or *paper*. The correct sense of an ambiguous word can be selected based on the context where it occurs, and correspondingly the problem of word

Claude Sammut & Geoffrey I. Webb (eds.), *Encyclopedia of Machine Learning*, DOI 10.1007/978-0-387-30164-8,

sense disambiguation is defined as the task of automatically assigning the most appropriate meaning to a polysemous word within a given context.

Motivation and Background

Word sense disambiguation is considered one of the most difficult problems in natural language processing, due to the high semantic ambiguity that is typically associated with language. It was first noted as a problem in the context of machine translation, when Warren Weaver, in his famous 1949 memorandum, pointed out word ambiguity as one of the problems that needed to be solved in order to enable automatic translations between the languages of the world (Weaver, 1995). More than 50 years later, word sense ambiguity is still regarded as an important and difficult research problem, and it has been demonstrated to have a potentially significant impact on several natural language processing applications.

Applications

In addition to machine translation, the role of word sense disambiguation has also been explored in connection to other applications, such as monolingual information retrieval, cross-language information retrieval, question answering, knowledge acquisition, information extraction, text classification, and others. In particular, a significant amount of work has been carried out in areas related to information retrieval, where the resolution of word ambiguity has been shown to have an impact on both the precision of the system (by allowing for matches only between identical word meanings in the query and in the documents), as well as the recall of the system (by performing query expansion using synonyms of selected word meanings).

Brief History

Over the years, the field of word sense disambiguation has undergone steady improvements in both quality and scope, moving from the rule-based systems using hand crafted knowledge that were popular in the 1970s and 1980s, to the more advanced corpus-based methods used in the 1990s, and to the current hybrid systems that rely on a mix of knowledge-based and corpus-based resources, minimizing the need of sense annotated data and taking advantage of the Web. The shift from small-scale rule-based systems to large-scale data-driven methods has also implied an increase in coverage, with early systems typically addressing a handful of ambiguous words for which hand-coded rules were available, while many of the current systems have the ability to address all or almost all content words in unrestricted text.

Methods

Current word sense disambiguation systems are divided into three main categories:

Knowledge-based: These systems rely mainly on information drawn from lexical resources, such as dictionaries or thesauruses. The Lesk algorithm (Lesk, 1986) is one of the most well-known knowledge-based word sense disambiguation methods. It decides the meaning of a word based on a measure of similarity among the definitions provided by a dictionary. For instance, for the phrase *pine cone*, the algorithm will select the meaning of *kind of evergreen tree* for *pine*, and *fruit of evergreen tree* for *cone*, as these are the definitions with the highest lexical overlap among all the possible definitions provided by a dictionary.

Unsupervised corpus-based: These approaches typically consist of algorithms for clustering word sense occurrences in a corpus, without making explicit reference to a sense inventory. The clustering can be performed in a monolingual environment, in which case different word occurrences are represented by features derived from their immediate context (Schutze, 1998). Alternatively, a clustering of word senses can also be performed using cross-lingual evidence drawn from the translations observed in a parallel corpus (Ng, Wang, & Chan, 2003). This line of work is often referred to as word sense discrimination, as the word meanings are not disambiguated against a sense inventory, but are discriminated against each other.

Supervised corpus-based: These methods are the focus of the current chapter, and they consist primarily of machine learning algorithms applied on large sense-annotated corpora. Supervised algorithms have been typically applied to one word at a time, although experiments have also been carried out for their application to all words in unrestricted text. While sense-annotated corpora have usually been constructed by hand, recent work has also explored various approaches for the automatic generation of such data, which has

been used successfully in conjunction with machine learning algorithms.

Structure of the Learning System

Among the various knowledge-based and data-driven word sense disambiguation methods that have been proposed to date, supervised systems have been constantly observed as leading to the highest performance. In these systems, the sense disambiguation problem is formulated as a supervised learning task, where each sense-tagged occurrence of a particular word is transformed into a feature vector, which is then used in an automatic learning process.

Given a target word and a set of examples where this word occurs, each occurrence being annotated with the correct sense, a supervised system will attempt to learn how to automatically annotate occurrences of the given word in new, previously unseen, contexts. This process is accomplished in two steps. First, representative features are extracted from the context of the ambiguous word; this step is applied to the annotated examples (training) as well as the unlabeled examples (test). Second, a machine learning algorithm is applied on the feature vectors, and consequently the most likely sense is assigned to the test occurrences of the target word.

Features

Research in supervised word sense disambiguation has considered two main types of featutlllrestomodel occurrences of ambiguous words:

Contextual features, which are extracted tfrom the immediate vicinity of the ambiguous word. These features usually consist of the words before and after the target word (a window size of 3–10 words is typical), their parts of speech, words in a syntactic dependency with the target word (e.g., the subject of the verb, the noun modified by an adjective), position in the sentence, and the like. For instance, the adjective *green* could be one of the contextual features extracted from the context *the green plant* for the ambiguous word *plant*.

Topical features, which are represented by the words most frequently co-occurring with a given meaning of the target word. These words are usually determined by counting the number of times each word occurs in the context of a word meaning, divided by the total number of occurrences in the context of the word regardless of its meaning. For instance, the *factory* meaning of *plant* could have topical features such as *industrial* and *work*, whereas the *green plant* meaning of *plant* might have features such as *animal* and *water*.

As an example of feature vector construction, consider the following two contexts provided for the ambiguous word *plant*:

The/det growth/noun of/prep a/det seedling/noun into/prep a/det flowering/adj **plant**/noun helps/verb children/noun investigate/verb the/det conditions/noun that/prep plants/noun need/verb for/prep growth/noun.

The/det operations/noun staff/noun in/prep an/det industrial/adj **plant**/noun is/verb typically/adv measured/verb in/prep asset/noun utilization/noun.

The following two feature vectors are constructed:

Machine Learning

Provided a set of feature vectors representing different occurrences of an ambiguous target word, the goal of the machine learning system is to learn how to predict the most likely sense for a new occurrence. The word sense disambiguation literature describes experiments with a large number of machine learning algorithms, including decision lists (Yarowsky, 2000), instance-based learning (Ng & Lee, 1996), Naïve Bayes and decision trees (Pedersen, 1998), support vector machines (Lee & Ng, 2002), and others. A comparison of several machine learning algorithms for word sense disambiguation is provided in Lesk (1986) and Mooney (1996).

W-1	W + 1	P-1	P+1	Growth	Flowering	Industrial	Staff	Sense
Flowering	Helps	Adj	Verb	Y	Y	N	N	Green plant
Industrial	Is	Adj	Verb	N	N	Y	Y	Factory

Generation of Sense-Tagged Corpora

One of the main drawbacks associated with the supervised methods for word sense disambiguation is the cost incurred in the process of building sense-tagged corpora. Despite their high performance, the applicability of these supervised systems is limited to those few words for which sense-tagged data is available, and their accuracy is strongly connected to the amount of labeled data available at hand.

Sense annotations have been typically carried out by humans, which resulted in several publicly available data sets, such as those made available during the Senseval evaluations (http://www.senseval.org). However, despite the effort that went into the construction of these data sets, their applicability is limited to a handful of approximately 100 ambiguous words.

To address the sense-tagged data bottleneck problem, different methods for automatic sense-tagged data annotation have been proposed in the past, with various degrees of success. One such method relies on monosemous relatives extracted from dictionaries, which can be used to identify ambiguity-free occurrences in large corpora (Leacock, Chodorow, & Miller, 1998; Mihalcea, 1999). Another method relies on automatically bootstrapped disambiguation patterns, which can be used to generate a large number of sense-tagged examples (Mihalcea, 2002; Yarowsky, 1995). The use of volunteer contributors to create sense-annotated corpora has also been explored in the Open Mind Word Expert system (Chklovski and Mihalcea, 2002). Finally, in recent work, Wikipedia was identified as a rich source of word sense annotations, which can be used to build supervised word sense disambiguation systems (Mihalcea, 2007).

Cross References

►Semi-Supervised Text Processing

Recommended Reading

Agirre, E., & Edmonds, P. (2006). *Word sense disambiguation: Algorithms and applications*. Berlin: Springer. http://www.wsdbook.org

Chklovski, T., & Mihalcea, R. (2002). Building a sense tagged corpus with open mind word expert. In *Proceedings of ACL 2002 workshop on WSD*. Philadelphia, PA.

Leacock, C., Chodorow, M., & Miller, G. A. (1998). Using corpus statistics and wordnet relations for sense identification. *Computational Linguistics*, *24*(1), 147–165.

Lee, Y. K., & Ng, H. T. (2002). An empirical evaluation of knowledge sources and learning algorithms for word sense disambiguation. In *Proceedings of EMNLP 2002*. Philadelphia, PA.

Lesk, M. (1986). Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *SIGDOC 1986*. Toronto, ON, Canada.

Mihalcea, R. (1999). An automatic method for generating sense tagged corpora. In *Proceedings of AAAI 1999*. Orlando, FL.

Mihalcea, R. (2002). Bootstrapping large sense tagged corpora. In *Proceedings of LREC 2002*. Las Palmas, Spain.

Mihalcea, R. (2007). Using wikipedia for automatic word sense disambiguation. In *Proceedings of NAACL 2007*. Rochester, NY.

Mihalcea, R., & Pedersen, T. (2005). Advances in word sense disamhbiguation. Tutorial presented at IBERAMIA 2004, ACL 2005, AAAI 2005. http://www.d.umn.edu/~tpederse/WSDTutorial.html

Mooney, R. (1996). Comparative experiments on disambiguating word senses: An illustration of the role of bias in machine learning. In *Proceedings of EMNLP*. Philadelphia, PA.

Ng, H. T. & Lee, H. B. (1996). Integrating multiple knowledge sources to disambiguate word sense: An examplar-based approach. In *Proceedings of ACL*. Santa Cruz, CA.

Ng, H. T., Wang, B., & Chan, Y. S. (2003). Exploiting parallel texts for word sense disambiguation: an empirical study. In *Proceedings of ACL*. Sapporo, Japan.

Pedersen, T. (1998). *Learning probabilistic models of word sense disambiguation*. Ph.D. Dissertation. Southern Methodist University.

Schutze, H. (1998). Automatic word sense discrimination. *Computational Linguistics*, *24*(1), 97–123.

Weaver, W. (1995). Translation. In W. N. Locke, & A. D. Booth (Eds.), *Machine translation of languages: Fourteen essays*. Cambridge, MA: MIT Press.

Yarowsky, D. (1995). Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of ACL*. Cambridge, MA.

Yarowsky, D. (2000). Hierarchical decision lists for word sense disambiguation. *Computers and the Humanities*, *34*(1–2), 179–186.

Word Sense Discrimination

Word sense discrimination is sometimes used as a synonym for ►word sense disambiguation. Note, however, that these two terms refer to somewhat different problems, as word sense discrimination implies a distinction between different word meanings in a corpus (without reference to a sense inventory), whereas word sense disambiguation refers to a sense assignment using a given sense inventory.