Wiktionary
(& Dictionaries)
Lexical Resources

November 6th, 2019

Contents of the lecture

Basic concepts (contains Latin!)

Basic concepts

Definitions

- a dictionary is a collection of definition
- a definition couples a definiendum with definientia
- the definiendum (pl. definienda) is the word for which a definition is given
- a definiens (pl. definientia) is a word that occurs in order to define a definiendum
- technically, one definition corresponds to one sense.

Here's a concrete example:

- if the definition is `random : lacking a definite plan, purpose, or pattern`
- the definiendum is `random`
- the definientia are `lacking`, `a`, `definite`, `plan`, `,`, `purpose`, `,`, `or`, `pattern`
- there might be other definitions for the definiendum `random`, eg. `random: relating to, having, or being elements or events with definite probability of occurrence`

Basic concepts

Definitions

- definitions (more precisely, definientia) are generally organized according to a 'Genus + Differentium' pattern
- the Genus (pl. genera) corresponds to a definiens (or a series of definientia) that gives the broad semantic category of the definiendum
- the Differentium (pl. differentia) corresponds to a definiens (or a series of definientia) that give the finer semantics within the broad category that is the Genus
- the genus is *generally* a hypernym of the definiendum.

Not all definitions follow this pattern. Other patterns include: synonymy, meta-linguistic reference ("used to denote ...")...

Who cares about dictionaries?

Example applications

WSD & WSI

- WSD stands for ‘word sense disambiguation’. Given an ambiguous word in context and an inventory of meanings, find which meaning applies to the word.
- WSI for ‘word sense induction’. Given an ambigous word in context, find its meaning (without a precomputed inventory).

Dictionaries can be used either as an inventory of meanings for WSD, or as labels to learn a WSI architecture in a supervised setup.

- in those WSI setups, the basic idea is to transform definitions into labels or targets.
- (Chang and Chen, 2019): a ‘simple’ way is to convert the definientia into a single sentence embedding, and then learn the mapping from the (contextual) embedding onto the definientia embedding.

Example applications

'Old school' WSD

(Gaume, Hathout, and Muller, 2004)

The basic idea is to turn the dictionary itself into a graph, and explore the graph structure

- We noted that genera are hypernyms of their respective definienda
- Therefore, by transitivity, the genus of the genus of word w should also be a hypernym of w
- More generally, if a definiens contributes to the meaning of its definiendum w ,then the definiens of a definiens contributes indirectly to the meaning of w. We can think of it as "taking one more step" in the graph linking definienda to their definientia.

Example applications

'Old school' WSD

- Let a set D of definitions $\langle w^i, \langle w^i_1 \ \ldots \ w^i_n \rangle\rangle$, where w^i is a definiendum and $\langle w^i_1 \ \ldots \ w^i_n \rangle$ associated definientia. Let the vocabulary $V = \bigcup_{\langle w^i, \langle w^i_1 \ \ldots \ w^i_n \rangle\rangle \in D} \{w^i, w^i_1 \ \ldots \ w^i_n\}$
- Let the graph $\mathscr{G}$ defined over $V \cup D \times V \cup D$, such as $\langle w^i, d^j \rangle$ and $\langle d^j, w^k \rangle$ are edges of $\mathscr{G}$ iff. $\exists d^j$ such as $d^j = \langle w^i, \langle w^i_1 \ \ldots w^k \ldots \ w^i_n \rangle\rangle \in D$. Let the connectivity $|V \cup D| \times |V \cup D|$ matrix $\hat{\mathscr{G}}$ the matrix notation of $\mathscr{G}$, where $\hat{\mathscr{G}}_{ij} = 1$ iff. $\langle v^i, v^j \rangle$ is an edge of $\mathscr{G}$.
- Let the matrix $\tilde{\mathscr{G}}$ the normalized version of $\hat{\mathscr{G}}$. Hence all rows of $\tilde{\mathscr{G}}$ sum up to 1 and can be seen as probability distributions. If the row corresponds to a definition d, it defines how likely words from V are its definientia. If the row corresponds to a word w, it simply links to its possible definitions in D.
- We can transcribe hypernymy transitivity using the chain rule. The probability that word w^j occurs within k steps as a definiens for the definition d^i will be given by $\left(\tilde{\mathscr{G}}^k\right)_{ij}$.

Disambiguate a word α by retrieving the node for one of its definition d^j whose probability distribution, according to $\left(\tilde{\mathscr{G}}^k\right)$, matches best with the context of α.

Example applications

Crossword solver

(Hill et al., 2016): You can use definitions as meaning inventories, so you can learn to retrieve words based on what how they are defined, which may be useful for Information retrieval

- as a proof of concept, a crossword solver
- train a model to retrieve or recompute the embedding of the definiendum, based on the sequence of definientia
- test on crosswords: for each cross word definition, select the most probable embedding that fits in the grid

Example applications

Definition Modeling

(Noraset et al., 2017): can you evaluate word embeddings content using a dictionary? You can see it as the inverse function of the previous crossword solver

- Given a set of embeddings $E = \vec{w}^1 \ \ldots \ \vec{w}^n$, we want to assert that the vector representations adequately capture the meaning of the words they are purported to represent
- If $\vec{w}^i$ is a good representation of w^i, then the dictionary definition of w^i can be reconstructed from $\vec{w}^i$
- How can you test that? Initialize a recurrent neural network with $\vec{w}^i$, and have it produce the corresponding definientia $\langle w^i_1 \ \ldots \ w^i_n \rangle$.
- You can see that as a basic sequence-to-sequence task, mapping $\langle w^i \rangle$ to $\langle w^i_1 \ \ldots \ w^i_n \rangle$
- From this basic formulation, you can derive representation for words in context, simply by extending the set containing the definiendum with its context words.

Working with Wiki

Working with Wiki

A closer look at Wiktionary

- Wiktionary is a collaboratively edited multilingual web-based project.
- The aim is to produce dictionaries for all the world's languages, currently it covers $\approx$ 171 languages
- Wiktionary data is frequently used in NLP, both in multilingual and in monolingual contexts
- As a consequence of the collaborative nature of the project, Wiktionary is generally deemed to have broad coverage, but unsystematic definitions.

Working with Wiki

"It's complicated"

- Wiki-markup is notably messy. As a potpourri to give you a taste: includes recursive markups, implicit section and subsection handling, template and boilerplate elements...
- Wiki-markup is doubly nightmarish for Wiktionaries across languages, which generally do not have the same conventions
- The raw data can be found here: https://dumps.wikimedia.org/
- But parsing it correctly takes time...

Working with Wiki

Solutions exist.

- Some languages have NLP-friendly wiktionary, or wiktionary ports
- A very good example is GLAWI, a French wiktionary dump parsed in a XML format, developed by CLLE-ERSS (Toulouse)
- Another such example is Dbnary, which has coverage for multiple languages (mostly European, but also Indonesian, Japanese, Turkish); however the format is not especially user-friendly (intended as a SPARQL endpoint for a webservice)
- There are some libraries dedicated to wiki-markup handling, so some (or a lot) of dedication using these may suffice to retrieve the information you care about.

References I

Chang, Ting-Yun and Yun-Nung Chen (Nov. 2019). "What Does This Word Mean? Explaining Contextualized Embeddings with Natural Language Definition". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, pp. 6066–6072. doi: 10.18653/v1/D19-1627. url: https://www.aclweb.org/anthology/D19-1627.

Gaume, Bruno, Nabil Hathout, and Philippe Muller (2004). "Word Sense Disambiguation Using a Dictionary for Sense Similarity Measure". In: *Proceedings of the 20th International Conference on Computational Linguistics*. COLING '04. Geneva, Switzerland: Association for Computational Linguistics. doi: 10.3115/1220355.1220528. url: https://doi.org/10.3115/1220355.1220528.

Hill, Felix et al. (2016). "Learning to Understand Phrases by Embedding the Dictionary". In: *Transactions of the Association for Computational Linguistics* 4, pp. 17–30. issn: 2307-387X. url: https://transacl.org/ojs/index.php/tacl/article/view/711.

Noraset, Thanapon et al. (2017). "Definition Modeling: Learning to define word embeddings in natural language". In: *AAAI*.