NLP EMBEDDINGS

TIER 2 ADVANCE

LESSON 1

Created by Amardeep Singh Sidhu

ROOT

NLP BASICS & PREREQUISITIES

ANAPHORA

Anaphora is a figure of speech in which words repeat at the beginning of successive clauses, phrases, or sentences. For example, Martin Luther King's famous "I Have a Dream" speech contains anaphora: "So let freedom ring from the prodigious hilltops of New Hampshire. Let freedom ring from the mighty mountains of New York. Let freedom ring from the heightening Alleghenies of Pennsylvania[1]

EPISTROPHE

Epistrophe is a figure of speech in which one or more words repeat at the end of successive phrases, clauses, or sentences. In his Gettysburg Address, Abraham Lincoln urged the American people to ensure that, "government of the people, by the people, for the people, shall not perish from the earth." His repetition of "the people" at the end of each clause is an example of epistrophe.

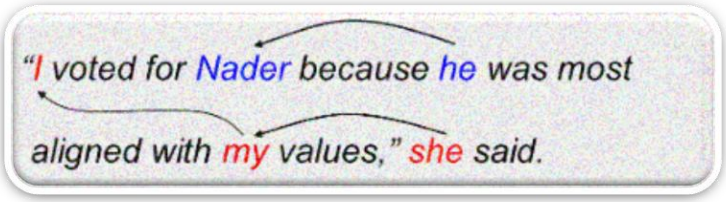
Epistrophe also goes by the name epiphora, and even more occasionally is sometimes called antistrophe.

[1] (Epistrophe and Anaphora, n.d.)

The opposite of **epistrophe** is **anaphora**, which involves the repetition of words at the beginning of successive phrases, clauses, or sentences.

COREFERENCE RESOLUTION

Coreference resolution is the task of finding all expressions that refer to the same entity in a text. It is an important step for a lot of higher-level NLP tasks that involve natural language understanding such as document summarization, question answering, and information extraction.[2]



SHALLOW PARSING

Shallow semantic parsing is labelling phrases of a sentence with semantic roles with respect to a target word.[3]
For example, the sentence "Shaw Publishing offered Mr. Smith a reimbursement last March." Is labelled as
[AGENTShaw Publishing] **offered** [RECEPIENTMr. Smith] [THEMEa reimbursement] [TIMElast March]

Main focus is on building joint probabilistic models for simultaneous assignment of labels to all nodes in a syntactic parse tree. These models are able to capture the strong correlations among decisions at different nodes.

COMPUTATIONAL ANAPHORA RESOLUTION

Anaphora accounts for cohesion in texts and is a phenomenon under active study in formal and computational linguistics alike. The correct interpretation of anaphora is vital for natural language processing (NLP). For example, anaphora resolution is a key task in natural language interfaces, machine translation, text summarization, information extraction, question answering, and a number of other NLP applications.[4]

Solutions in the Past

- Discourse Representation Theory
- Centering Theory
- Exploited domain and linguistic knowledge and required considerable human input.

Automatic abstracting and information extraction independently highlighted importance of anaphora and coreference resolution.

---

[2] (The Stanford Natural Language Processing Group, n.d.)

[3] (Shallow Semantic Parsing, n.d.)

[4] (Shalom Lappin)

What drove research towards knowledge poor and robust research

- Cheaper and reliable corpus-based NLP tools such as part-of-speech taggers and shallow parsers[5].
- Ontologies

NATURE OF LANGUAGE AND AMBIGUITY

WORD SENSE DISAMBIGUATION

In natural language processing, word sense disambiguation (WSD) is the problem of determining which "sense" (meaning) of a word is activated by the use of the word in a particular context, a process which appears to be largely unconscious in people.[6]

Example

A famous example is to determine the sense of pen in the following passage (Bar-Hillel 1960):

Little John was looking for his toy box. Finally, he found it. The box was in the pen. John was very happy.
WordNet lists five senses for the word pen:

- pen — a writing implement with a point from which ink flows.
- pen — an enclosure for confining livestock.
- playpen, pen — a portable enclosure in which babies may be left to play.
- penitentiary, pen — a correctional institution for those convicted of major crimes.
- pen — female swan.

LEXICAL AMBIGUITY

[6] (Word sense disambiguation, n.d., p. 1)

Words can simultaneously belong to multiple syntactic class(parts of speech) Fly can be Noun as well as verb. the verb *fly* can refer to multiple meanings, including "travelling through the air" which is the intended meaning in above example.

ANAPHORIC AMBIGUITY

This type of ambiguity concerns the interpretation of pronouns. For instance, in "I have a laptop but recently bought a new one. I am going to give it away.", what does *it* refer to?

SYNTACTIC AMBIGUITY

A sentence could be parsed syntactically in multiple ways. For instance, in the sentence "I saw a man on the hill with binoculars", we can attach *binoculars* to either *I* or to the *man*.

METONYMIC AMBIGUITY

Metonymy is the substitution of a concept, phrase or word being meant with a semantically related one. For instance, in "Cambridge voted to stay in the EU", it is definitely the people of Cambridge who voted and not the city itself.

VECTOR REPRESENTATION

The *distributional hypothesis* [Harris, 1954, Firth, 1957], i.e., words that occur in the same contexts tend to have similar meanings, has been the foundation of automatically constructing word VSM. However, the interpretation of the hypothesis and the way of collecting "similarity" clues and constructing the space have gone under enormous changes. Earlier approaches were based on collecting word statistics, usually in terms of occurrence and co-occurrence frequency. Hence, they are usually referred to as co*unt*-based techniques. These representations are

OFTEN LARGE AND NEEDED SOME SORT OF DIMENSIONALITY REDUCTION

ONE HOT ENCODING VS MAKE YOUR OWN REPRESENTATION

SPARSITY

SPARSE MATRIX

In numerical analysis and scientific computing, a sparse matrix or sparse array is a matrix in which most of the elements are zero. By contrast, if most of the elements are nonzero, then the matrix is considered dense.[7]

STORING A SPARSE MATRIX

[7] (Wikipedia Sparse Matrix, n.d.)

In the case of a sparse matrix, substantial memory requirement reductions can be realized by storing only the non-zero entries.

Depending on the number and distribution of the non-zero entries, different data structures can be used and yield huge savings in memory when compared to the basic approach. The trade-off is that accessing the individual elements becomes more complex and additional structures are needed to be able to recover the original matrix unambiguously.

Efficient Modification

These support efficient modification, such as DOK (Dictionary of keys), LIL (List of lists), or COO (Coordinate list). These are typically used to construct the matrices.

Dictionary of Keys (DOK) consists of a dictionary that maps (row, column)-pairs to the value of the elements. Elements that are missing from the dictionary are taken to be zero. The format is good for incrementally constructing a sparse matrix in random order, but poor for iterating over non-zero values in lexicographical order. One typically constructs a matrix in this format and then converts to another more efficient format for processing.

Lists of Lists (LIL) LIL stores one list per row, with each entry containing the column index and the value. Typically, these entries are kept sorted by column index for faster lookup. This is another format good for incremental matrix construction.

COO (Coordinate list) stores a list of (row, column, value) tuples. Ideally, the entries are sorted first by row index and then by column index, to improve random access times. This is another format that is good for incremental matrix construction.

Efficient Matrix Operation

Those that support efficient access and matrix operations, such as CSR (Compressed Sparse Row) or CSC (Compressed Sparse Column).

SPARSE MATRIX IN TENSORFLOW

1^8

NGARM EFFECT

REPRESENTATION

Semantic space and distributed representation.

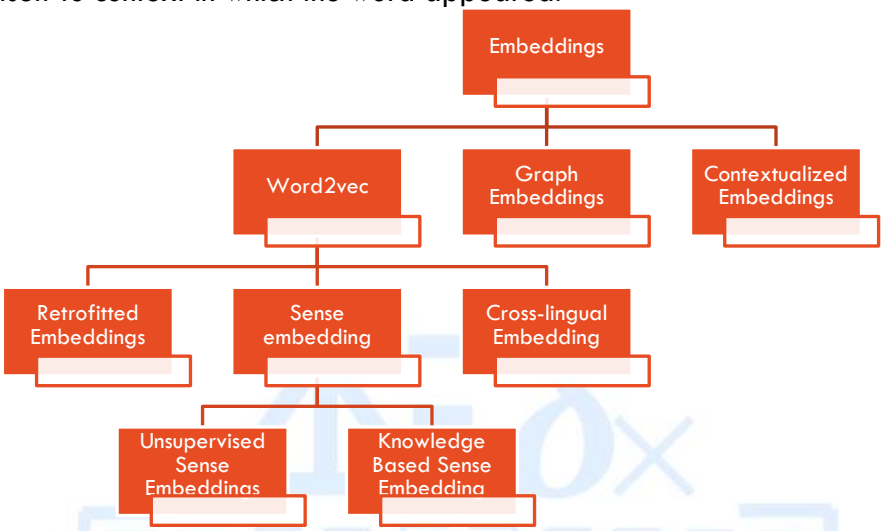
[8] (Compressing BERT for faster prediction, n.d.)

distributional hypothesis: that occur in the same contexts tend to have similar meanings,

EMBEDDINGS

Embeddings preserve semantics but can't model **dynamic nature of language.**
Words can exhibit different syntactic and semantic properties depending on the context in which they appear.

contextualized representation aimed at addressing the static nature of embedding allowing it to adapt itself to context in which the word appeared.



DIMENSIONS

PRINCIPAL COMPONENT ANALYSIS

REFERENCES

Camacaho-Collados, M. T. (2020). *Embeddings in Natural Language Processing Theory and Advances in Vector Representation of Meaning*. NA: Morgan & Claypool publishers. Retrieved from http://josecamachocollados.com/book_embNLP_draft.pdf

Compressing BERT for faster prediction. (n.d.). Retrieved from https://blog.rasa.com/compressing-bert-for-faster-prediction-2/

Epistrophe and Anaphora. (n.d.). Retrieved from https://www.litcharts.com/literary-devices-and-terms/epistrophe

Sennrich, R. (2020). How Contextual is Neural Machine Translation. *Gecko Symposium 2020* (p. 1). Barcelona (Online only): (Keynote Talk). Retrieved from https://sites.google.com/view/gecko2020/keynote-talks/rico-sennrich

Shallow Semantic Parsing. (n.d.). Retrieved from The Stanford Natural Language Processing Group: https://nlp.stanford.edu/projects/shallow-parsing.shtml

Shalom Lappin, e. a. (n.d.). Introduction to the Special Issue on Computational Anaphora Resolution. NA. Retrieved from https://www.aclweb.org/anthology/J01-4001.pdf

The Stanford Natural Language Processing Group. (n.d.). Retrieved from https://nlp.stanford.edu/projects/coref.shtml

Wikipedia Sparse Matrix. (n.d.). Retrieved from https://en.wikipedia.org/wiki/Sparse_matrix

Word sense disambiguation. (n.d.). Retrieved from http://www.scholarpedia.org/article/Word_sense_disambiguation

TASK IN NLP

TEXT ENTAILMENT

INDEX