Unsupervised WSD with the Naïve Bayes model

The classical clustering technique is represented by the Naïve Bayes model trained with the EM algorithm. When the Naïve Bayes model is applied to supervised disambiguation, the actual words occurring in the context window are usually used as features. This type of framework generates a great number of features and, implicitly, a great number of parameters. This can dramatically decrease the model performance since the available data is usually insufficient for the estimation of the great number of resulting parameters. This situation becomes even more drastic in the case of unsupervised disambiguation, where parameters must be estimated in the presence of missing data (the sense labels). In order to overcome this problem, the various existing unsupervised approaches to WSD implicitly or explicitly perform a feature selection.

Semantic WN-based feature selection

- **This approach is based on a set of features formed by the actual words occurring near the target word (within the context window) and reduces the size of this feature set by performing knowledge-based feature selection that relies entirely on WN.**

- **The WN semantic network provides the words considered as relevant for the set of senses taken into consideration corresponding to the target.**

- **The value of a feature is given by the number of occurrences of the corresponding word in the given context window.**

Semantic WN - based Feature Selection

- **We start by choosing words occurring in the same WN synsets as the target word (WN synonyms), corresponding to all senses of the target. Additionally, the words occurring in the synsets related (through explicit relations provided in WN) to those containing the target word are also considered as part of the so-called “disambiguation vocabulary". Synsets and relations are restricted to those associated with the part of speech of the target. The content words of the glosses and the example strings of all types of synsets participating in the disambiguation process are equally considered. Usage of the example string as well is in accordance with previous studies performed for knowledge-based WSD.**
- **Existing problem:**

- **One of the problems that semantic WN-based feature selection is prone to, is the possibility of achieving <u>reduced corpus coverage</u>, especially in the case of those synsets having few semantic relations in WN. This is the reason why other types of features, of completely different natures, are studied as well.**

- **Please read the uploaded paper "Performing word sense disambiguation at the border between unsupervised and knowledge-based techniques" for this type (WN-based) of feature selection.**

Syntactic Dependency-based Feature Selection

- **This type of features augments the role of linguistic knowledge.**

- **Syntactic features are provided by dependency relations as defined by the classical Dependency Grammar formalism.**

- **The semantic space that is proposed to the Naïve Bayes model for unsupervised WSD is entirely based on syntactic knowledge, more precisely on dependency relations, extracted from natural language texts via a syntactic parser. The parser extracts dependency relations that will indicate the disambiguation vocabulary required by Naïve Bayes. This is how the disamb. voc. is formed:**

- **It is various combinations of dependencies, pointed out by the obtained <u>dependency structure</u>, that form the <u>context</u> over which the semantic space for WSD is constructed. <u>The number of features</u> (coming from all resulted dependencies) <u>is decreased</u> by taking into account only specific dependency relations. The <u>disambiguation vocabulary</u> is formed by considering all words that participate in the considered dependencies.**

Syntactic Dependency-based Feature Selection

In order to inform the construction of the semantic space for WSD with syntactic knowledge of this type, several elements must be taken into consideration. At the first stage of the experiments, we make no qualitative distinction between the different relations, by not taking into account the type of the involved dependencies. At the second stage of the experiment, we can inform the construction of the semantic space in a more linguistically rich manner, by considering the dependency type. Therefore we make use only of specific dependency relations, which are considered more informative than others relatively to the studied part of speech.

Syntactic Dependency-based Feature Selection

While the classical dependency-based linguistic theory does not allow the arches denoting the dependency relations to intersect (thus leading to an oriented graph which has no cycles), the dependency analysis performed by the Stanford parser, for instance, can be either projective (disallowing crossing dependencies) or non-projective (permitting crossing dependencies). The number of resulting features is then decreased by taking into account only dependency relations of specific types. Both types of analyses (non-projective and projective) have been performed and discussed in the literature, relatively to the presented case studies, with the projective syntactical analysis leading to the best results so far.

Syntactic Dependency-based Feature Selection

- **In defining the syntactic context of the target word, we can first take into consideration direct relationships (first order dependencies) between the target and other words, where the target can be either the head or a dependent, and which correspond to paths of length 1 anchored at the target in the associated dependency graph. At the next stage we can consider indirect relationships between the target and other words, by taking into account paths of length 2 (second order dependencies) in the same associated graph. The order of these dependencies (lengths of the paths in the associated graph) represents a parameter that can vary and which can be considered analogous to the classical "window size" parameter. Just as in the case of the window size, this parameter should have relatively small values,**

since it is a known fact that linguistically interesting paths are of limited length. This justifies the usual choice of limiting the investigation to second order dependencies.

- **Separate experiments, which view the <u>target word</u> as <u>head</u> and as <u>dependent</u>, respectively, can be carried out both corresponding to first order and to second order dependencies.**

Syntactic Dependency-based Feature Selection

- **When using the Naïve Bayes model as clustering technique, it is observed that a small number of dependency types should be considered, if possible just one, in order to decrease the number of parameters that must be estimated by the EM algorithm.**

  This conclusion would not hold in the case of a clustering technique requiring a high number of features (such as spectral clustering, for instance). In such a case, several types of dependencies would probably have to be considered, which would create a more linguistically informed semantic space for WSD – that could be even more beneficial.

- ✓ **Please read the paper "On a Dependency-based Semantic Space for Unsupervised Noun Sense**

Disambiguation with an Underlying Naïve Bayes Model" for an example of usage of this type of features.