Word Sense Disambiguation

Supervised vs Unsupervised Methods

Sydney Richards

Division of Science and Mathematics
University of Minnesota, Morris
Morris, Minnesota, USA

18 Nov 2017

Natural Language is Ambiguous

One word can have multiple senses

"Plant"

- Noun: Facilities for production
- Noun: Living organism of the kingdom Plantae
- Verb: sow; place seed in ground to grow

https://www.merriam-webster.com/dictionary/plant

Word Sense Disambiguation

Word Sense Disambiguation (WSD) is the task of identifying which sense of an ambiguous word is being used in a given context.

"I am going to plant an apple tree"

- Noun: Facilities for production
- Noun: Living organism of the kingdom Plantae
- Verb: sow; place seed in ground to grow

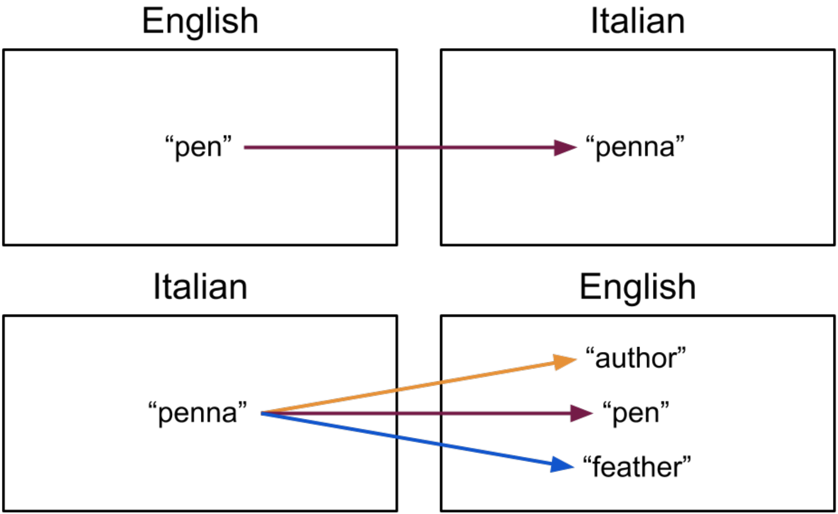
https://www.merriam-webster.com/dictionary/plant

Applications of WSD

Intermediate task for Natural Language Processing applications

- Information Retrieval
- Content Analysis
- Word Processing
- Machine Translation

Machine Translation

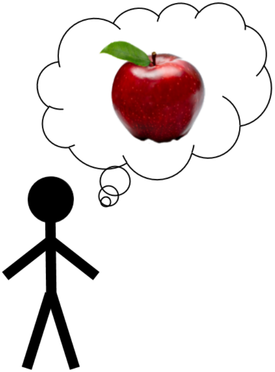


Outline

Outline

Why is WSD Difficult?

"apple"



Humans can read a string of letters and understand what it represents

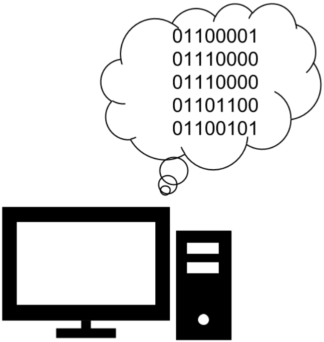
http://www.pngmart.com/image/tag/apple-fruit

This is not trivial for computers

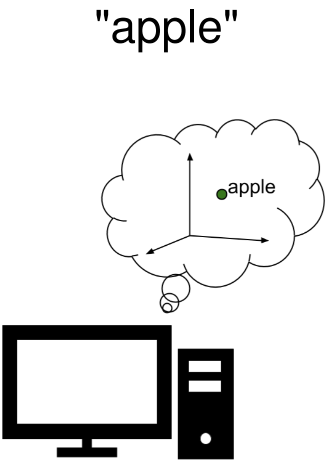
"apple"

If computers process words as strings of letters

- Information will be lost
- Not useful for WSD



Solution



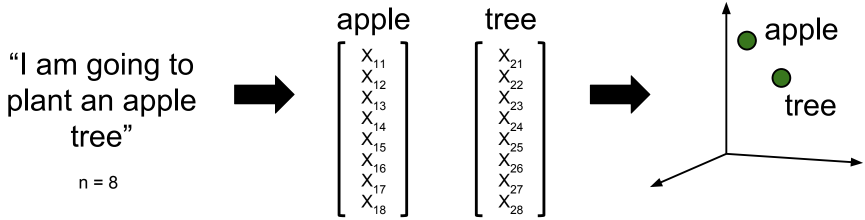
New word representation

- Useful to computers
- Preserves information about words

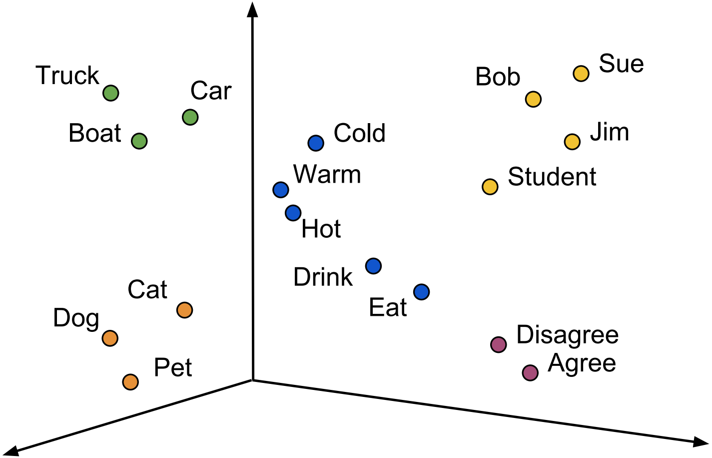
Word Embeddings

Unique mappings of words to vectors

- Each word represented by one vector in a continuous vector space
- n-dimensional
- Usage of a word defines its meaning

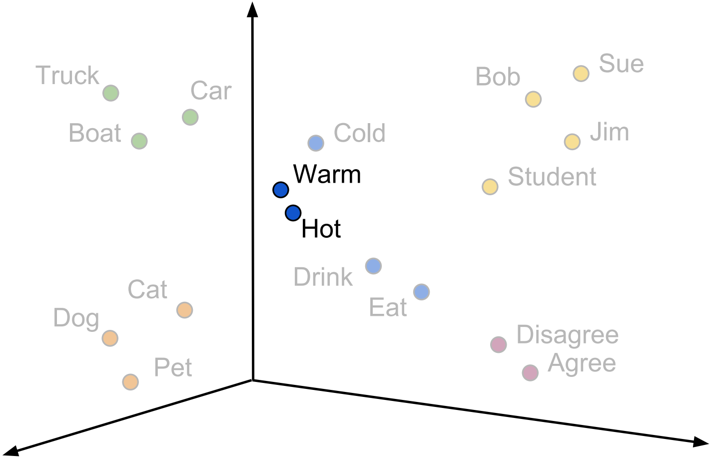


Word Embeddings



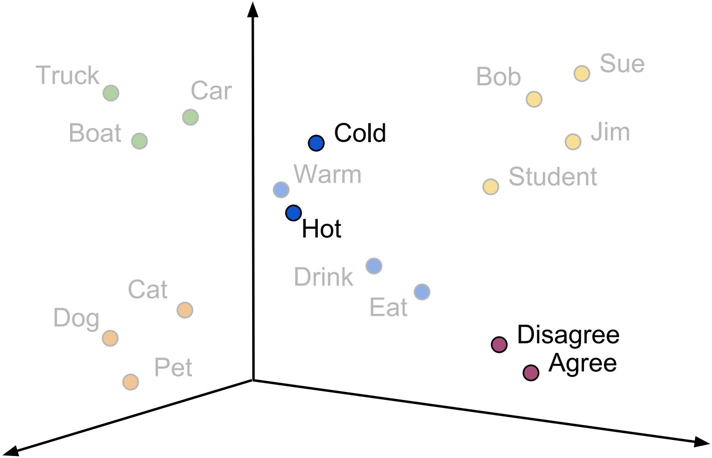
https://www.youtube.com/watch?v=Eku_pbZ3-Mw

Synonym Relationship



https://www.youtube.com/watch?v=Eku_pbZ3-Mw

Antonym Relationship



https://www.youtube.com/watch?v=Eku_pbZ3-Mw

Machine Learning

Development of computer programs that can:

- Extract information from data
- Learn patterns within the data
- Without explicit programming

Categorized into:

- Supervised
- Semi-supervised
- Unsupervised

Outline

Supervised Machine Learning

Receive labeled training data

- Inputs
- Expected Outputs

Training process

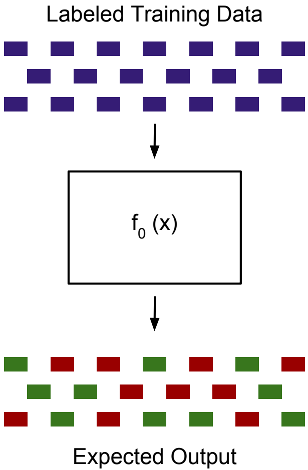
- Infer function to map input to output

Advantage

- Highly accurate

Disadvantage

- Reliant on training data



Supervised Machine Learning

Receive labeled training data

- Inputs
- Expected Outputs

Training process

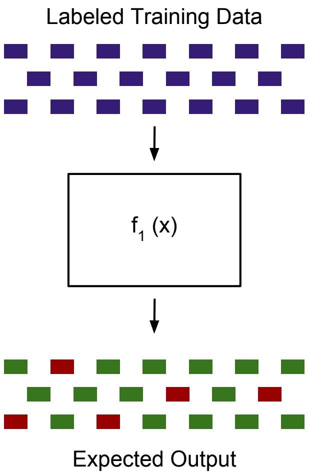
- Infer function to map input to output

Advantage

- Highly accurate

Disadvantage

- Reliant on training data



Supervised Machine Learning

Receive labeled training data

- Inputs
- Expected Outputs

Training process

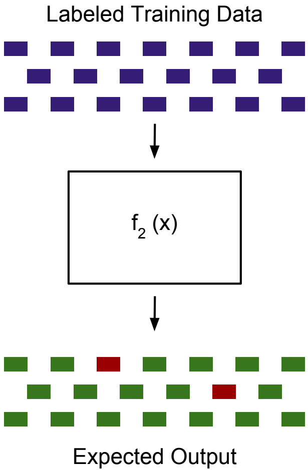
- Infer function to map input to output

Advantage

- Highly accurate

Disadvantage

- Reliant on training data



Supervised Method

"It Makes Sense" (IMS) supervised WSD software, Zhong et al. [2010]

- Takes in labeled training data
- Extracts three features from this data
- Trains Support Vector Machines classifiers

IMS

Takes in labeled training data

"I am going to plant/sow an apple tree"

IMS

Extracts three features from text

- Context words
- Part of Speech tags of context words
- Local collocations

IMS

Extracts three features from text

- Context words
- Part of Speech tags of context words
- Local collocations

"I am going to" plant "an apple tree"

IMS

Extracts three features from text

- Context words
- Part of Speech tags of context words
- Local collocations

"I/pronoun am going/verb to/preposition"

"an/determiner apple/adjective tree/noun"

It Makes Sense

Extracts three features from text

- Context words
- Part of Speech tags of context words
- Local collocations

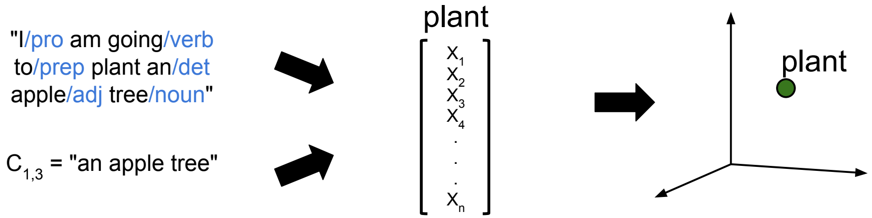
"I am going to plant an apple tree"

$C_{1,3}$ = "an apple tree"

Feature Vectors

Features from previous step converted to feature vectors

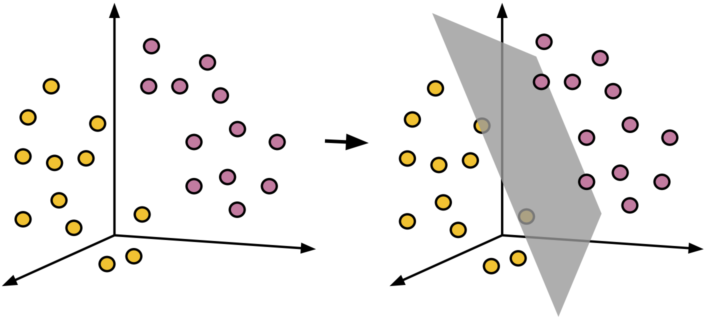
Feature vectors used to train Support Vector Machines



Support Vector Machines (SVM)

Supervised machine learning algorithm

- Takes labeled groups as training data
- Outputs separating hyperplane classifier

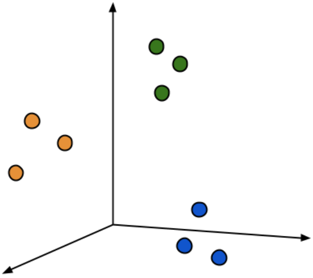


Training SVMs for WSD

One classifier must be trained to classify each possible sense of each ambiguous word

"Plant"

- Noun: Facilities for production
- Noun: Living organism of the kingdom Plantae
- Verb: sow; place seed in ground to grow



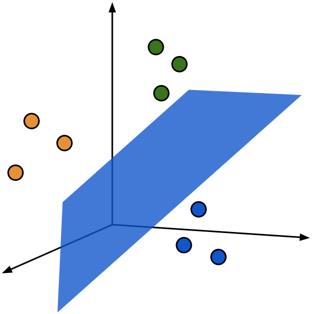
Feature Vectors

Training SVMs for WSD

One classifier must be trained to classify each possible sense of each ambiguous word

"Plant"

- Noun: Facilities for production
- Noun: Living organism of the kingdom Plantae
- Verb: sow; place seed in ground to grow



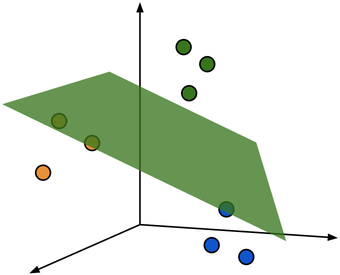
Feature Vectors

Training SVMs for WSD

One classifier must be trained to classify each possible sense of each ambiguous word

"Plant"

- Noun: Facilities for production
- Noun: Living organism of the kingdom Plantae
- Verb: sow; place seed in ground to grow



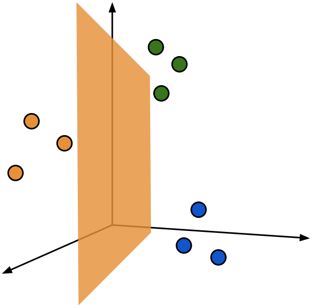
Feature Vectors

Training SVMs for WSD

One classifier must be trained to classify each possible sense of each ambiguous word

"Plant"

- Noun: Facilities for production
- Noun: Living organism of the kingdom Plantae
- Verb: sow; place seed in ground to grow



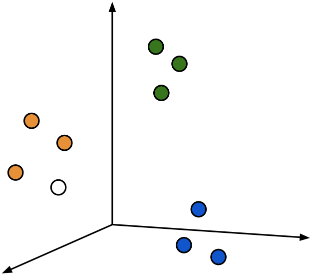
Feature Vectors

Testing

"Did you plant squash this year?"

"Plant"

- Noun: Facilities for production
- Noun: Living organism of the kingdom Plantae
- Verb: sow; place seed in ground to grow



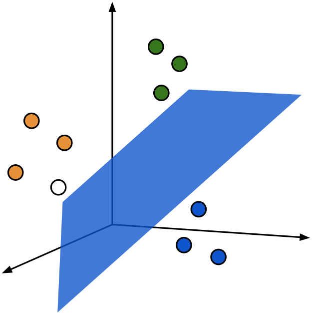
Feature Vectors

Testing

"Did you plant squash this year?"

"Plant"

- Noun: Facilities for production
- Noun: Living organism of the kingdom Plantae
- Verb: sow; place seed in ground to grow



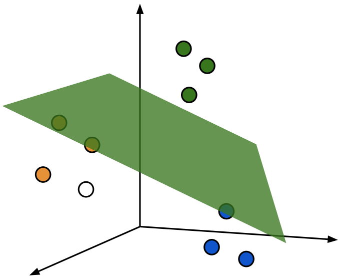
Feature Vectors

Testing

"Did you plant squash this year?"

"Plant"

- Noun: Facilities for production
- Noun: Living organism of the kingdom Plantae
- Verb: sow; place seed in ground to grow



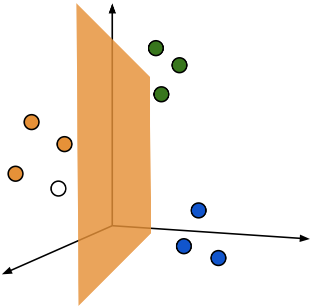
Feature Vectors

Testing

"Did you plant squash this year?"

"Plant"

- Noun: Facilities for production
- Noun: Living organism of the kingdom Plantae
- Verb: sow; place seed in ground to grow



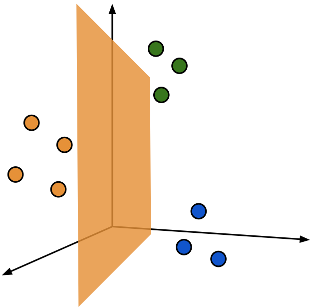
Feature Vectors

Testing

"Did you plant squash this year?"

"Plant"

- Noun: Facilities for production
- Noun: Living organism of the kingdom Plantae
- Verb: sow; place seed in ground to grow



Feature Vectors

Testing and Results

Taghipour et al. [2015] test IMS on all-words data set SE3

If a word has not been seen in training, IMS will output the first sense from WordNet

Compared to WNs1 baseline: first sense from WordNet

Evaluation measure: accuracy - the number of correct answers over the total number of answers to be given

Method	SE3
WNs1 baseline	62.40%
IMS	67.60%

Outline

Semi-Supervised Machine Learning

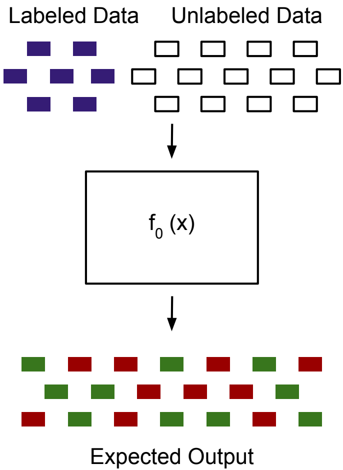
Similar to supervised machine learning

- Labeled training data
- Unlabeled data

Training process

- Infer function to map input to output

"Compromise" between supervised and unsupervised machine learning



Semi-Supervised Machine Learning

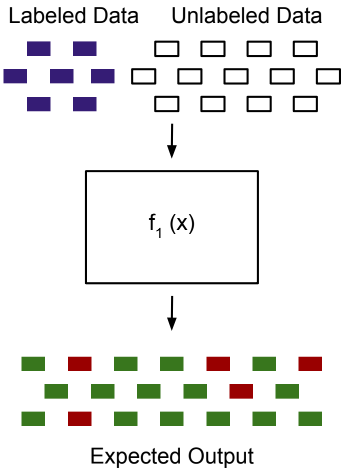
Similar to supervised machine learning

- Labeled training data
- Unlabeled data

Training process

- Infer function to map input to output

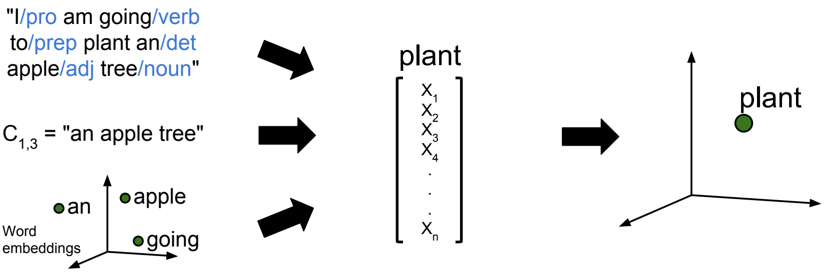
"Compromise" between supervised and unsupervised machine learning



Semi-Supervised Method

Taghipour et al. [2015] adapt IMS

- Additional feature: word embeddings of context words
- All-words data
- Considered unsupervised



Testing and Results

Taghipour et al. [2015] test IMS + word embeddings on all-words data set SE3

If a word has not been seen in training, IMS will output the first sense from WordNet

Compared to WNs1 baseline: first sense from WordNet

Evaluation measure: accuracy - the number of correct answers over the total number of answers to be given

Method	SE3
WNs1 baseline	62.40%
IMS	67.60%
IMS + word embeddings	68.00%

Outline

Unsupervised Machine Learning

Does not receive labeled training data

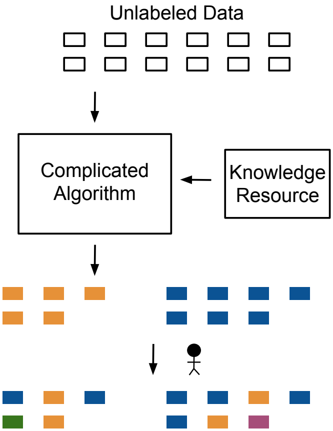
- Tries to identify patterns

Advantage

- Does not require labeled training data

Disadvantage

- Less accurate



Unsupervised Machine Learning

Does not receive labeled training data

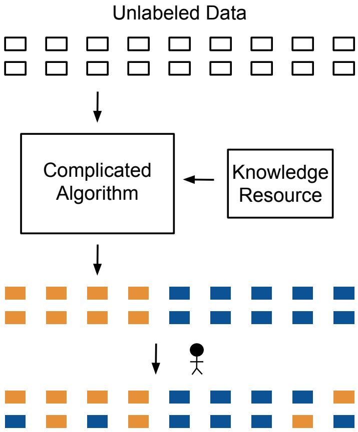
- Tries to identify patterns

Advantage

- Does not require labeled training data

Disadvantage

- Less accurate



Unsupervised Method

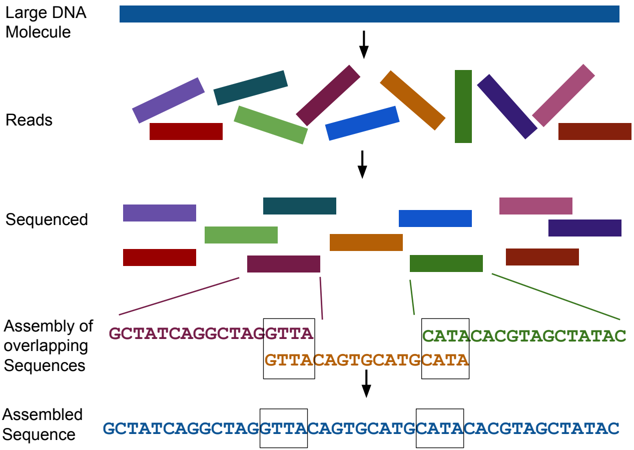
ShotgunWSD unsupervised WSD algorithm by Butnaru et al. [2017]

Knowledge-based

- Sense embeddings
- WordNet

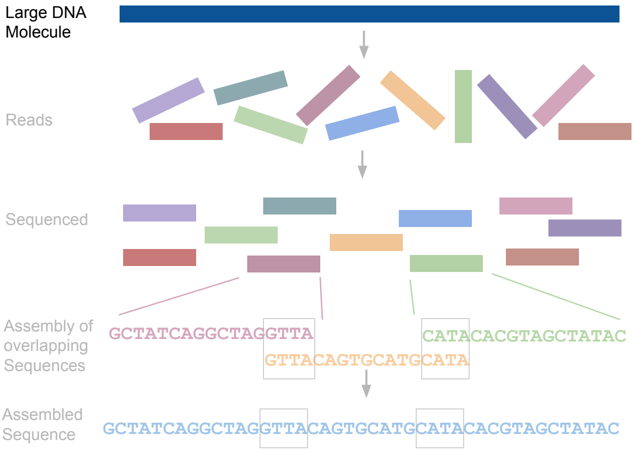
Based on Shotgun DNA sequencing technique

Shotgun DNA Sequencing



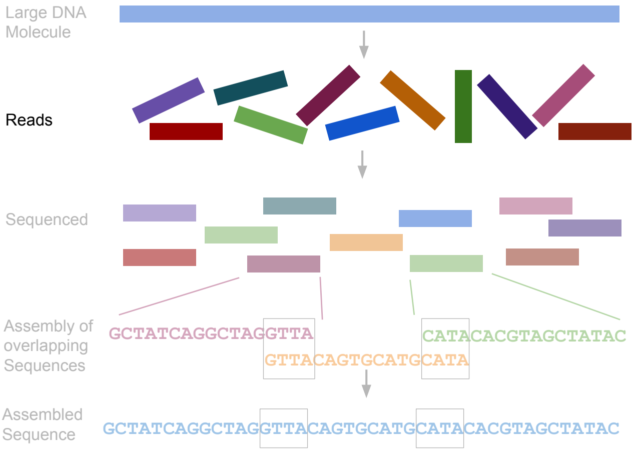
http://knowgenetics.org/whole-genome-sequencing/

Shotgun DNA Sequencing



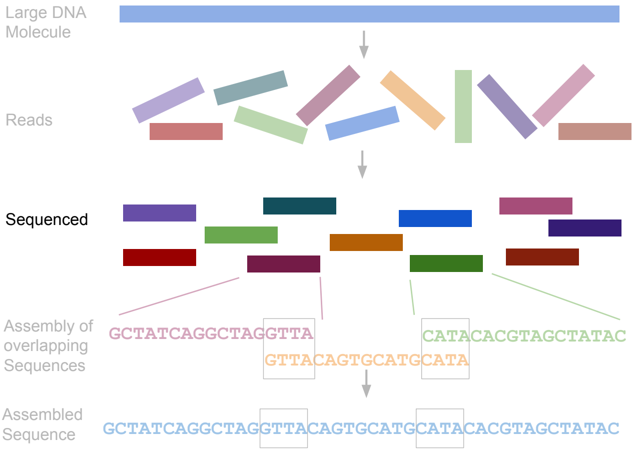
http://knowgenetics.org/whole-genome-sequencing/

Shotgun DNA Sequencing



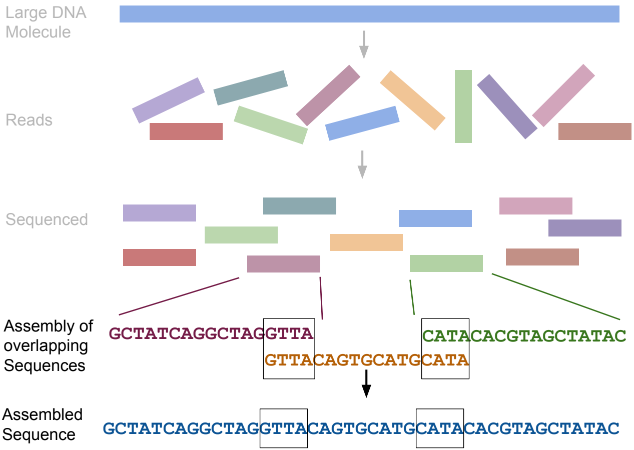
http://knowgenetics.org/whole-genome-sequencing/

Shotgun DNA Sequencing



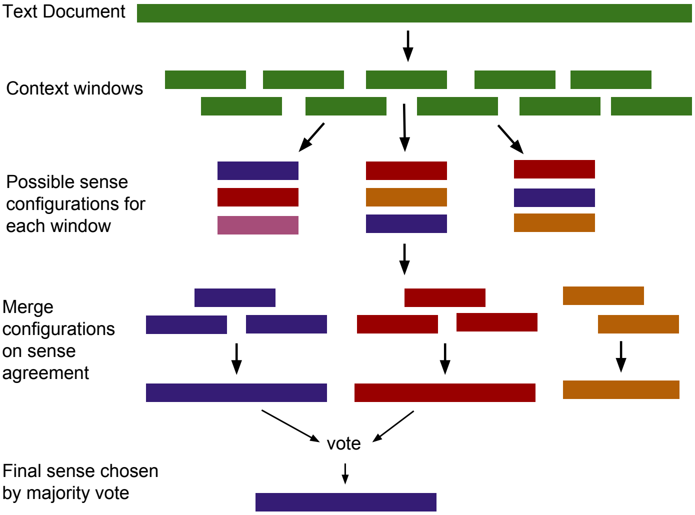
http://knowgenetics.org/whole-genome-sequencing/

Shotgun DNA Sequencing

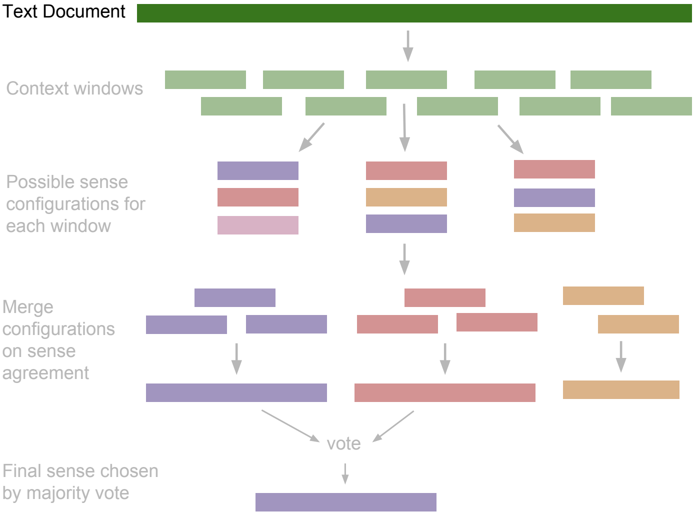


http://knowgenetics.org/whole-genome-sequencing/

ShotgunWSD


Text Document
Context windows
Possible sense configurations for each window
Merge configurations on sense agreement
vote
Final sense chosen by majority vote

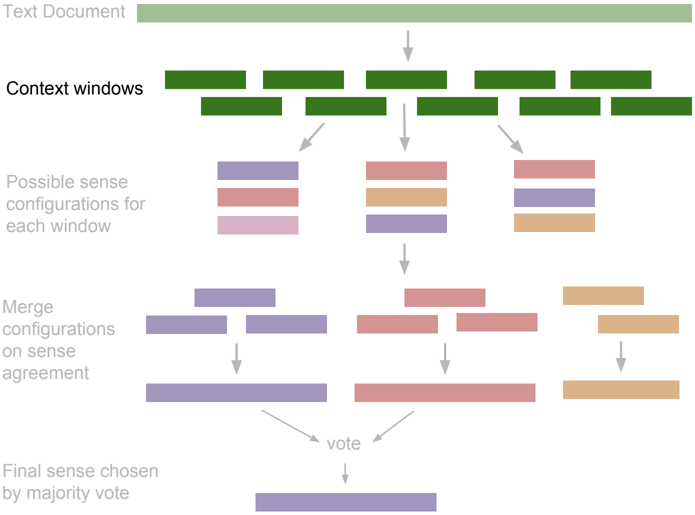
Text Document



Text Document

"I am going to plant an apple tree"

Context Windows



Context Windows

"I am going to plant an apple tree"

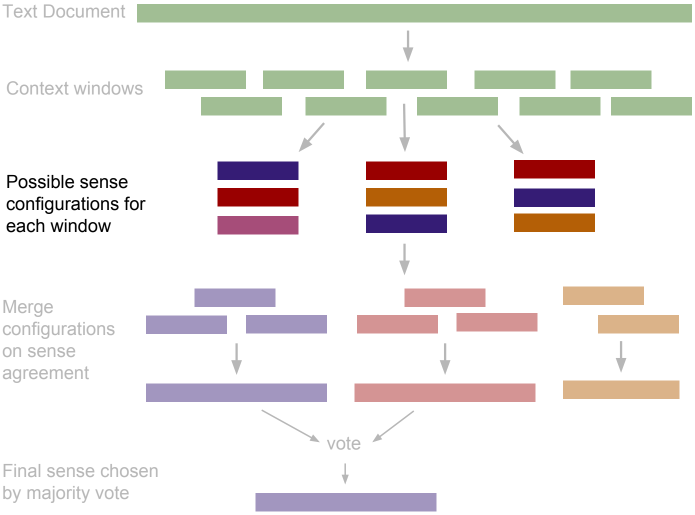
Context Windows

"I am going to plant an apple tree"

"going to plant an apple"

"plant an apple tree"

Possible Sense Configurations



Possible Sense Configurations

"going to plant an apple"

[1] going to [living organism] an [fruit]
[0] going to [factory] an [fruit]
[1] going to [sow] an [fruit]
[0] going to [living organism] an [Apple Inc]
[1] going to [factory] an [Apple Inc]
[0] going to [sow] an [Apple Inc]

"plant an apple tree"

[1] [living organism] an [fruit] tree
[0] [factory] an [fruit] tree
[1] [sow] an [fruit] tree
[0] [living organism] an [Apple Inc] tree
[1] [factory] an [Apple Inc] tree
[0] [sow] an [Apple Inc] tree

Possible Sense Configurations

"going to plant an apple"

[1] going to [living organism] an [fruit]
[0] going to [factory] an [fruit]
[1] going to [sow] an [fruit]
[0] going to [living organism] an [Apple Inc]
[1] going to [factory] an [Apple Inc]
[0] going to [sow] an [Apple Inc]

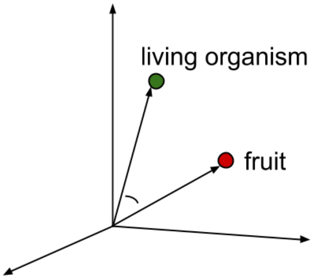
"plant an apple tree"

[1] [living organism] an [fruit] tree
[0] [factory] an [fruit] tree
[1] [sow] an [fruit] tree
[0] [living organism] an [Apple Inc] tree
[1] [factory] an [Apple Inc] tree
[0] [sow] an [Apple Inc] tree

Sense Embeddings

Similar to word embeddings

Relatedness between senses found by computing similarity of their sense embeddings



Sense Embeddings

Possible Sense Configurations

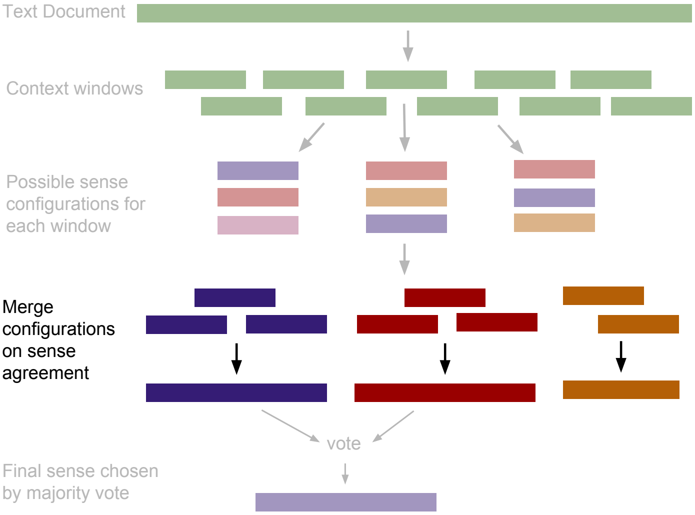
"going to plant an apple"

[1] going to [living organism] an [fruit]
[1] going to [sow] an [fruit]
[1] going to [factory] an [Apple Inc]

"plant an apple tree"

[1] [living organism] an [fruit] tree
[1] [sow] an [fruit] tree
[1] [factory] an [Apple Inc] tree

Merging Configurations



Merging Configurations

"going to plant an apple"

[1] going to [living organism] an [fruit]
[1] going to [sow] an [fruit]
[1] going to [factory] an [Apple Inc]

"plant an apple tree"

[1] [living organism] an [fruit] tree
[1] [sow] an [fruit] tree
[1] [factory] an [Apple Inc] tree

[1] going to [living organism] an [fruit]
[1] [living organism] an [fruit] tree

Merging Configurations

"going to plant an apple"

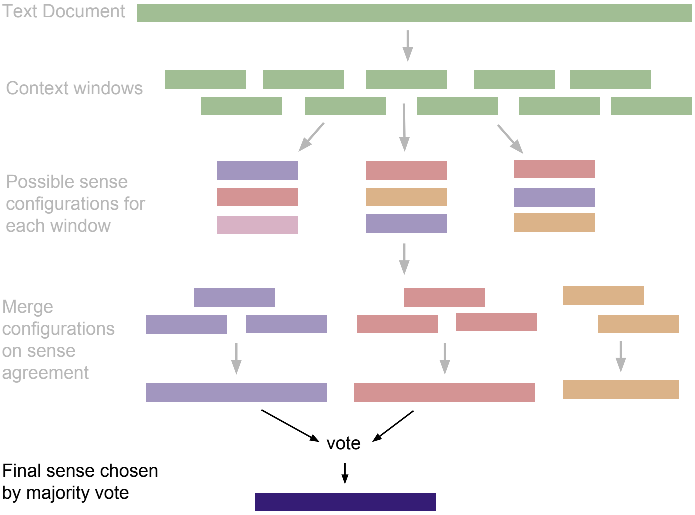
[1] going to [living organism] an [fruit]
[1] going to [sow] an [fruit]
[1] going to [factory] an [Apple Inc]

"plant an apple tree"

[1] [living organism] an [fruit] tree
[1] [sow] an [fruit] tree
[1] [factory] an [Apple Inc] tree

[2] going to [living organism] an [fruit] tree
[2] going to [sow] an [fruit] tree
[2] going to [factory] an [Apple Inc] tree

Majority Vote



Majority Vote

"I am going to plant an apple tree"

[2] going to [living organism] an [fruit] tree
[2] going to [sow] an [fruit] tree
[2] going to [factory] an [Apple Inc] tree

"I am going to plant/living organism an apple/fruit tree"

Testing

Butnaru et al. [2017] tested ShotgunWSD on three all-words datasets

- SemEval2007
- Senseval-2
- Senseval-3

ShotgunWSD does not go through training before testing

Compared to "Most Common Sense" (MCS) baseline

Results

$$F1 = \frac{2PR}{P + R}$$

Precision (P) - the number of true positives over the number of true positives and true negatives

Recall (R) - the number of true positives over the number of true positives and false negatives

Data set	ShotgunWSD	MCS baseline
SemEval2007	79.68%	78.89%
Senseval-2	57.55%	60.10%
Senseval-3	59.82%	62.30%

Outline

Summary

WSD is an active area of research

Supervised Methods

- Accurate
- Difficult to train

Unsupervised Methods

- Less accurate
- Easier to train

Semi-Supervised Methods

- Compromise

References

[1] Butnaru, A. M., Ionescu, R. T., and Hristea, F.
ShotgunWSD: An unsupervised algorithm for global word sense disambiguation inspired by DNA sequencing.
CoRR abs/1707.08084 (2017).

[2] Taghipour, K., and Ng, H. T.
Semi-Supervised Word Sense Disambiguation Using Word Embeddings in General and Specific Domains.
In *HLT-NAACL* (2015), pp. 314–323.

[3] Zhong, Z., and Ng, H. T.
It Makes Sense: A Wide-coverage Word Sense Disambiguation System for Free Text.
In *Proceedings of the ACL 2010 System Demonstrations* (Stroudsburg, PA, USA, 2010), ACLDemos '10, Association for Computational Linguistics, pp. 78–83.

Questions?

Special thanks to Elena Machkasova, Skatje Myers,
Kristin Lamberty, and Peter Dolan