Mod 4 - CO4 - 2024-09-03 - Apply Semantic Analysis

03 September 2024 09:22

→ Supervised Learning for WSD.

↳ Parametric Model (Naive Bayes)

$$\hat{s} = \underset{s \in S}{\operatorname{argmax}} \; P(s \mid f)$$

$$= \underset{s \in S}{\operatorname{argmax}} \; \frac{P(f \mid s) \cdot P(s)}{P(f)}$$

$$= \underset{s \in S}{\operatorname{argmax}} \; P(s) \cdot P(f \mid s)$$

$$= \underset{s \in S}{\operatorname{argmax}} \; P(s) \prod_{i=1}^{n} P(f_i \mid s)$$

$[f_1 \ldots\ldots f_n]$

↳ context

→ Collocation Vector (Set of words around it)

→ Setting parameters of Naive Bayes using MLE from training data:

$P(s_i) = count(s_i, w_j)$

$$P(s_i) = \frac{count(s_i, w_j)}{count(w_j)}$$

$$P(f_j \mid s_i) = \frac{count(f_j, s_i)}{count(s_i)}$$

→ Non-Parametric Method (Decision List)

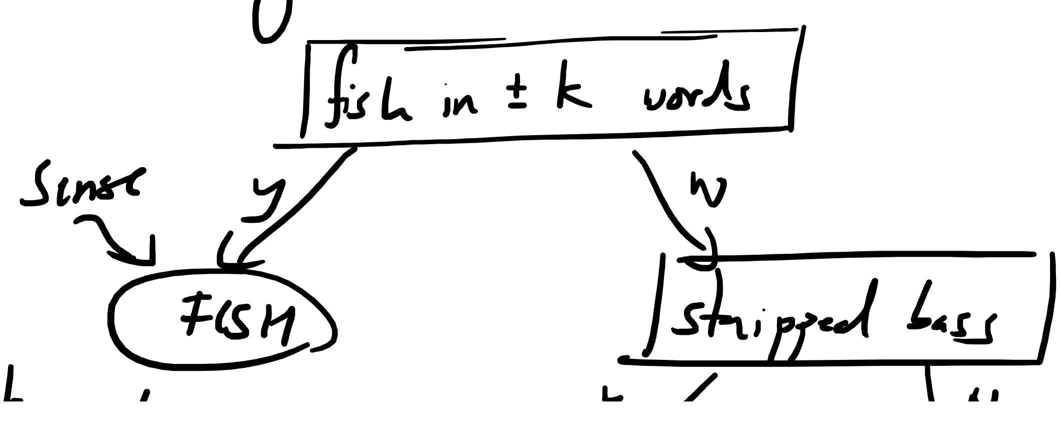
↳ One Sense per Collocation.

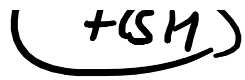
↳ Log-Likelihood Ratio:

$$\log \left\{ \frac{P(\text{Sense-}A \mid \text{Collocation}_i)}{P(\text{Sense-}B \mid \text{Collocation}_i)} \right\}$$

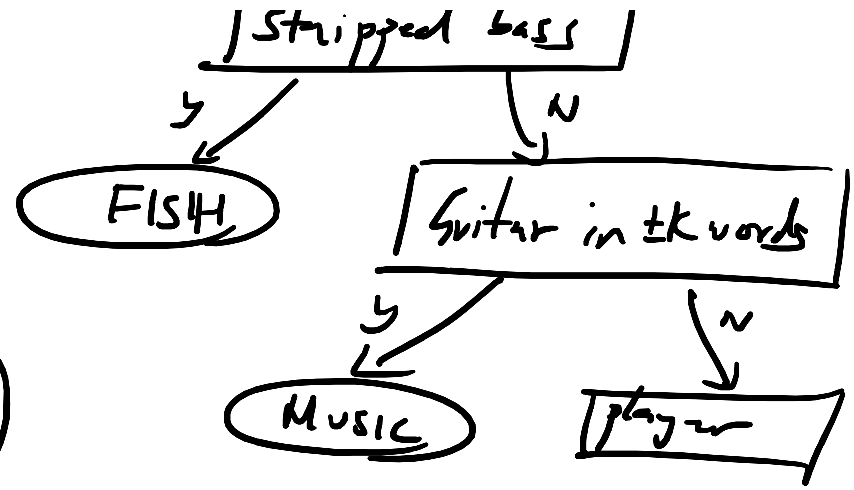
→ Higher Log-likelihood = More Predictive Evidence.

Eg: Discriminating b/w bass (fish / music)





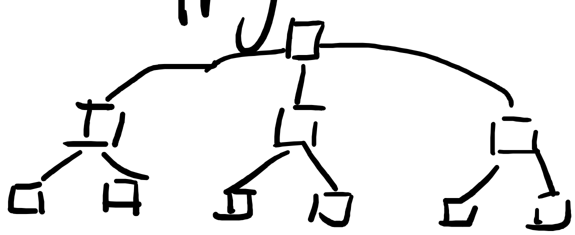
fish in ±k words
stripped bass
Guitar in ±k words
player



O(n log n)

→ Minimally Supervised WSD ~ Yarowsky.

↳ Bootstrapping or Co-training.



Random Forest

↳ Start with small seed, decision list.

↳ Use decision list to label corpus.

↳ Retain confident labels as annotated data to learn new decision list.

⇒ Heuristics (Derived from observations)

↳ One sense per discourse

↳ One sense per collocation.

Eg:

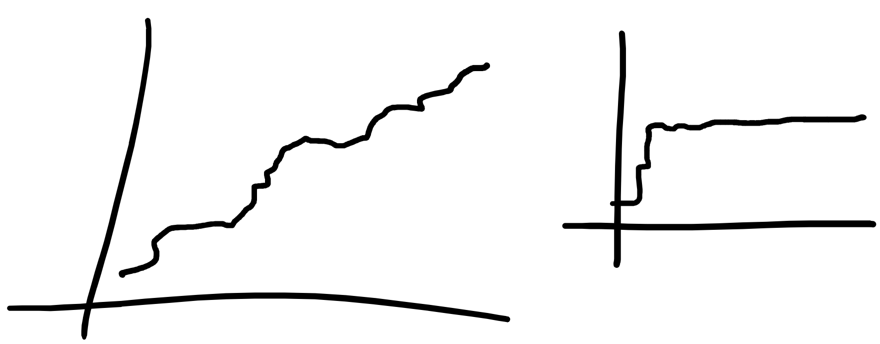
Eg:

-> Disambiguate plant (living thing) vs plant (industrial)

Intermediate State

Termination:

-> Stop when?

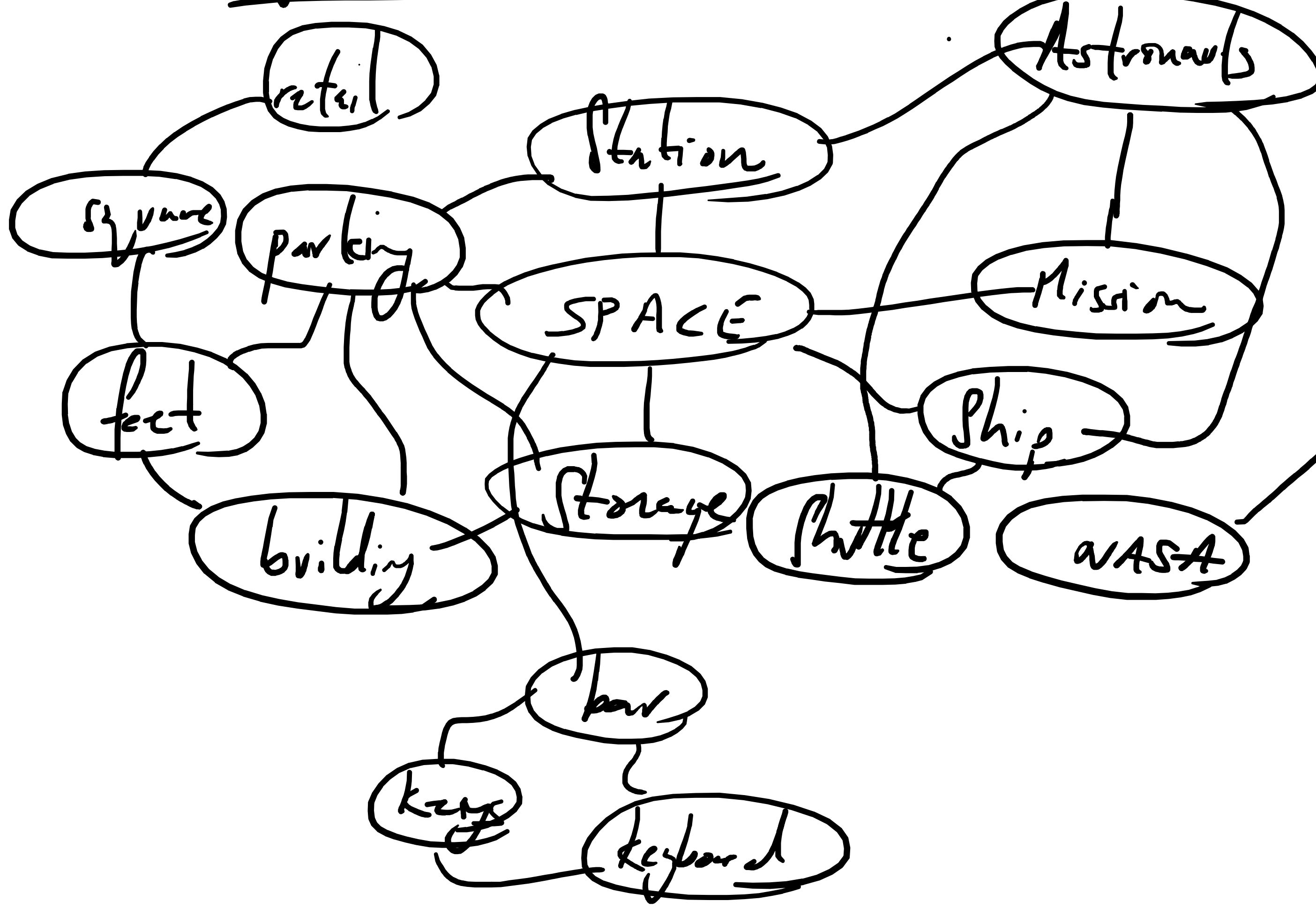


Advantages:

• Planning

-> Unsupervised:

-> Hyperlex: Word Sense Induction



-> Detecting Root Hubs:

-> Co-occurrence graph:

↳ Arrange nodes in decreasing order

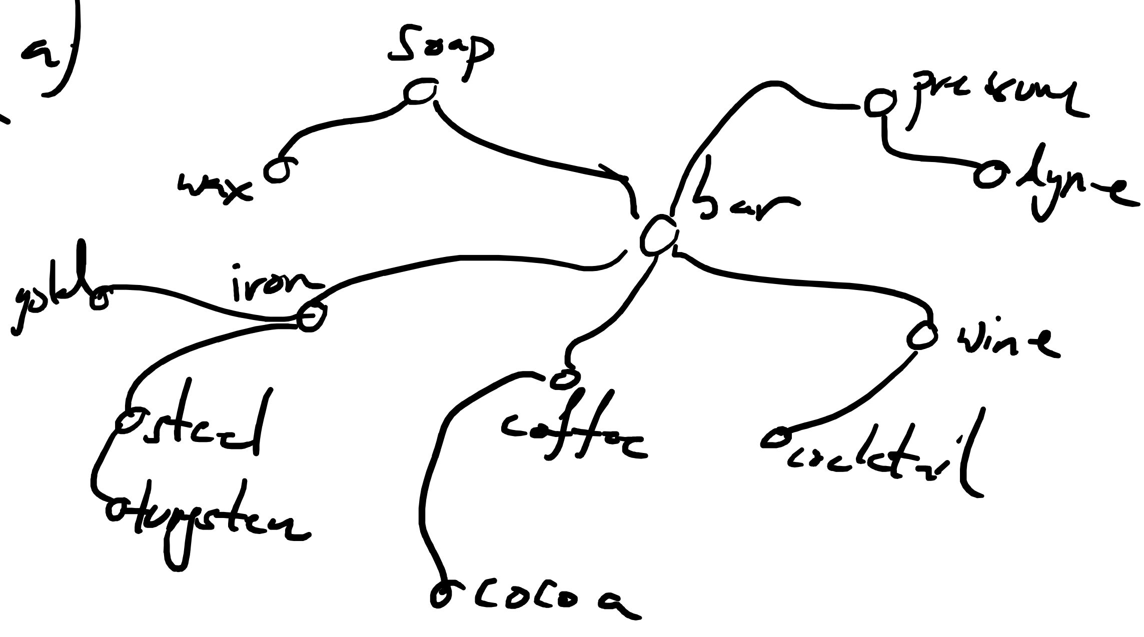
↳ Arrange nodes in decreasing order of degrees.

↳ Select the nodes from graph which has the highest degree. This node will be the hub of the first high density component.

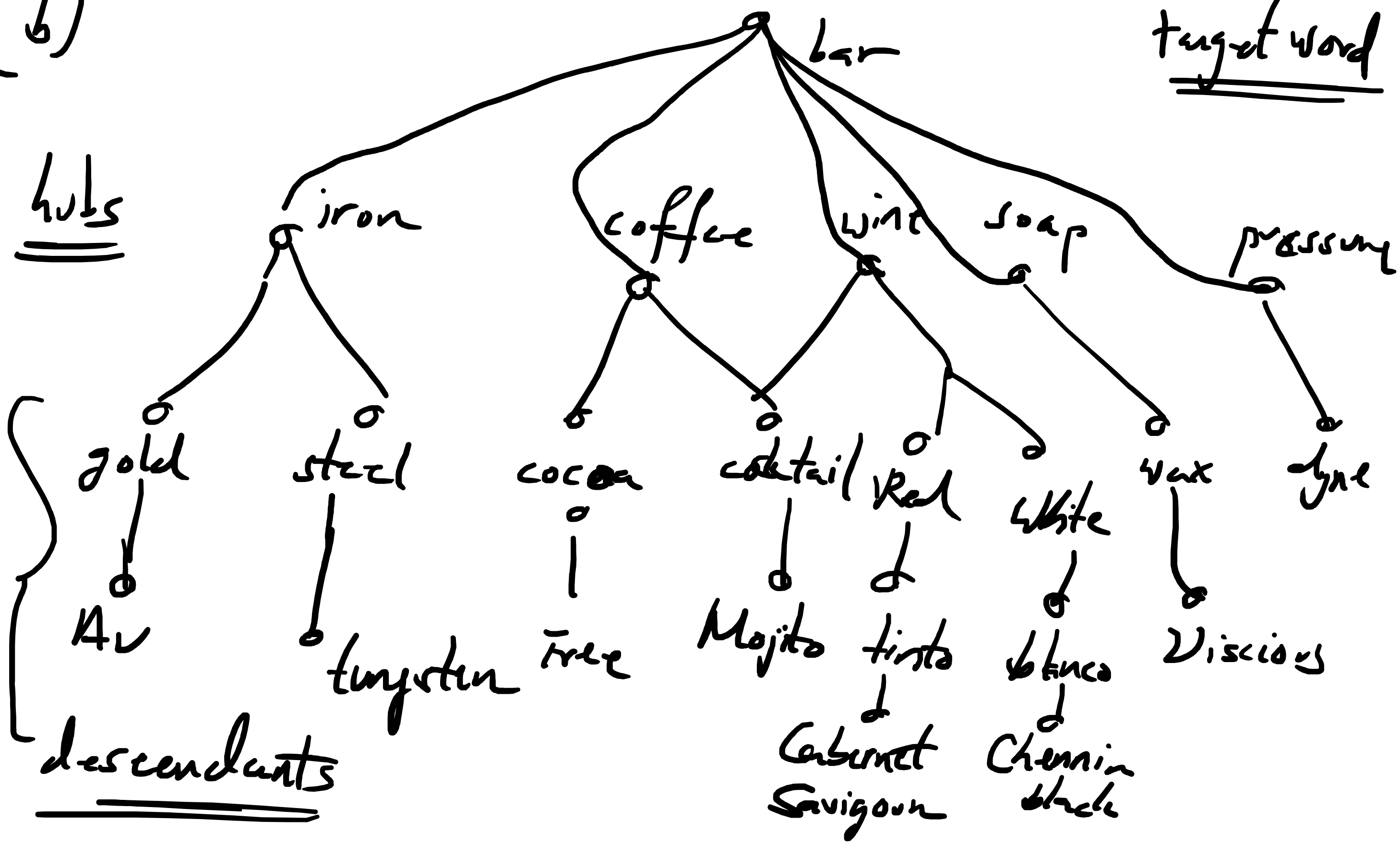
↳ Delete the hub & all its neighbors from graph.

↳ Repeat steps to detect hubs of other high degree components.

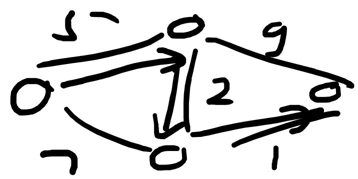
(a)



(b)



→ Dielineating Components :



Computing distance b/w 2 nodes

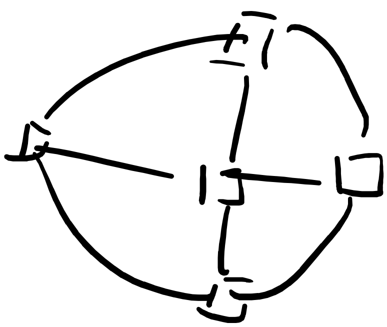
w_i & w_j

$$W_{ij} = 1 - \max \{ P(w_i|w_j), P(w_j|w_i) \}$$

where $P(w_i|w_j) =$

where $P(w_i|w_j) = freq_{ij} / freq_j$

→ <u>Disambiguation</u> : <u>Minimum Spanning Tree</u>



→ Let $W = (w_1 \ldots w_i \ldots w_n)$ be a context in which w_i is an instance of our target word.

→ Let w_i has k hubs in its MST.

→ A score vector $\underline{s}$ is associated with each $w_j \in W (j \neq i)$, such that s_k represents the contribution of the k^{th} hub as:

$$S_i = \frac{1}{\cdots}$$

$$s_k^j = \frac{1}{1 + d(h_k, w_j)}$$

if h_k is an ancestor of w_j

$= 0$ otherwise