Lecture -19 WSD

Word Sense Disambiguation (WSD) is the task of identifying the correct meaning (sense) of a word in a given context. Words often have multiple meanings; WSD aims to resolve this ambiguity automatically.

WSD with Naive Bayes

Word Sense Disambiguation (WSD) aims to determine the correct sense of a word in a given context. Naive Bayes is a probabilistic algorithm that can be applied to WSD.

Naive Bayes Approach

The Naive Bayes approach to WSD involves these key steps:

1. Training Data: A labeled dataset is required, where each instance of the ambiguous word is tagged with its correct sense in context.
2. Feature Extraction: Contextual information around the ambiguous word is extracted as features. Common features include:
    - Bag of Words: The surrounding words.
    - Collocations: Frequent word combinations.
3. Naive Bayes Classifier: A Naive Bayes classifier is trained on the labeled data. It learns to predict the sense of a word based on the observed features.

Formula

The Naive Bayes classifier uses Bayes' Theorem. For WSD, we want to find the probability of a specific sense (S) given a context (C):

$$P(S \mid C) = [P(C \mid S) * P(S)] / P(C)$$

Where:

- $P(S \mid C)$: Posterior probability of the sense given the context.
- $P(C \mid S)$: Likelihood probability of the context given the sense.
- $P(S)$: Prior probability of the sense.
- $P(C)$: Probability of the context.

The "naive" assumption is that the features in the context are conditionally independent of each other, given the sense. This simplifies the calculation of $P(C \mid S)$. If the context C is represented by words w1, w2, ..., wn, then:

$$P(C \mid S) = P(w1 \mid S) * P(w2 \mid S) * ... * P(wn \mid S)$$

Scenario:

We want to disambiguate the word "sound," which can have two senses:

- **Sense 1: Sound-Noun-Acoustic:** Vibrations that travel through a medium (like air) and are capable of being heard.
- **Sense 2: Sound-Noun-BodyOfWater:** A large body of water connecting two larger bodies of water.

1. Creating a Sample Dataset:

Here's a small dataset with sentences containing "sound," designed to highlight the importance of repeated context words:

1. "The **sound** of the music was loud and clear." (Sense: Sound-Noun-Acoustic)
2. "The **sound** of the waves crashing was very soothing." (Sense: Sound-Noun-Acoustic)
3. "The ship sailed across the **sound**." (Sense: Sound-Noun-BodyOfWater)
4. "The **sound** is a large body of water." (Sense: Sound-Noun-BodyOfWater)
5. "I could hear the **sound** of the wind. The **sound** of the music was also soothing." (Sense: Sound-Noun-Acoustic)
6. "The boat trip through the **sound** was beautiful. The ship sailed the **sound**." (Sense: Sound-Noun-BodyOfWater)

2. Stop Word Removal:

First, we'll remove common stop words (like "the," "of," "was," "is," "a," "and", "I", "could", "also") because they often don't contribute much to distinguishing between word senses.

Here's the dataset after stop word removal:

1. "**sound** music loud clear" (Sense: Sound-Noun-Acoustic)
2. "**sound** waves crashing very soothing" (Sense: Sound-Noun-Acoustic)
3. "ship sailed across **sound**" (Sense: Sound-Noun-BodyOfWater)
4. "**sound** large body water" (Sense: Sound-Noun-BodyOfWater)
5. "hear **sound** wind **sound** music soothing" (Sense: Sound-Noun-Acoustic)
6. "boat trip through **sound** beautiful ship sailed **sound**" (Sense: Sound-Noun-BodyOfWater)

3. Feature Extraction (Bag of Words):

We'll use a simple bag-of-words approach, extracting the words in the context (excluding the target word "sound" after stop word removal) as features.

Sentence	Target Word	Sense	Features
"sound music loud clear"	sound	Sound-Noun-Acoustic	{music, loud, clear}
"sound waves crashing very soothing"	sound	Sound-Noun-Acoustic	{waves, crashing, very, soothing}
"ship sailed across sound"	sound	Sound-Noun-BodyOfWater	{ship, sailed, across}
"sound large body water"	sound	Sound-Noun-BodyOfWater	{large, body, water}
"hear sound wind sound music soothing"	sound	Sound-Noun-Acoustic	{hear, wind, music, soothing}
"boat trip through sound beautiful ship sailed sound"	sound	Sound-Noun-BodyOfWater	{boat, trip, through, beautiful, ship, sailed}

4. Naive Bayes Calculation (Simplified):

Let's walk through a simplified Naive Bayes calculation.

A. Calculate Prior Probabilities:

- P(Sense = Sound-Noun-Acoustic)=3/6=0.5
- P(Sense = Sound-Noun-BodyOfWater)=3/6=0.5

B. Calculate Likelihood Probabilities :

Let's calculate the likelihood for a few key words. Assume our vocabulary is: {music, loud, clear, waves, crashing, very, soothing, ship, sailed, across, large, body, water, hear, wind, boat, trip, through, beautiful}. Vocabulary size = 20

- **For Sense = Sound-Noun-Acoustic (Count = 3):**
    - P(music | Sound-Noun-Acoustic)=(2+1)/(3+20)=3/23
    - P(loud | Sound-Noun-Acoustic)=(1+1)/(3+20)=2/23
    - P(clear | Sound-Noun-Acoustic)=(1+1)/(3+20)=1/23
    - P(waves | Sound-Noun-Acoustic)=(1+1)/(3+20)=1/23
    - P(hear | Sound-Noun-Acoustic)=(1+1)/(3+20)=1/23
    - P(soothing | Sound-Noun-Acoustic)=(1+1)/(3+20)=1/23
    - P(large | Sound-Noun-Acoustic)=(0+1)/(3+20)=1/23
- **For Sense = Sound-Noun-BodyOfWater (Count = 3):**
    - P(ship | Sound-Noun-BodyOfWater)=(2+1)/(3+20)=3/23
    - P(sailed | Sound-Noun-BodyOfWater)=(2+1)/(3+20)=3/23
    - P(across | Sound-Noun-BodyOfWater)=(1+1)/(3+20)=1/23
    - P(large | Sound-Noun-BodyOfWater)=(1+1)/(3+20)=1/23
    - P(music | Sound-Noun-BodyOfWater)=(0+1)/(3+20)=1/23
    - P(boat | Sound-Noun-BodyOfWater)=(1+1)/(3+20)=1/23
    - P(trip | Sound-Noun-BodyOfWater)=(1+1)/(3+20)=1/23
    - P(beautiful | Sound-Noun-BodyOfWater)=(1+1)/(3+20)=1/23
    - P(through | Sound-Noun-BodyOfWater)=(1+1)/(3+20)=1/23

C. Disambiguating a New Sentence:

Let's disambiguate: "The sound of the loud music." After removing "the" and "of" our features are {loud, music}

P(S | C) = [P(C | S) * P(S)] / P(C)

Where:

- S = Sense of the word we're trying to disambiguate
- C = Context of the word

In our example, we have:

- Word to disambiguate: "sound"
- Context (C): {loud, music} (after removing stop words)
- Possible senses (S):
    - S1 = Sound-Noun-Acoustic
    - S2 = Sound-Noun-BodyOfWater

We want to calculate:

- P(S1 | C) = P(Sound-Noun-Acoustic | {loud, music})
- P(S2 | C) = P(Sound-Noun-BodyOfWater | {loud, music})

Applying the Formula to the Example:

Let's break down the calculation for P(S1 | C):

P(S1 | C) = P(Sound-Noun-Acoustic | {loud, music})

Using Bayes' Theorem:

P(Sound-Noun-Acoustic | {loud, music}) = [P({loud, music} | Sound-Noun-Acoustic) * P(Sound-Noun-Acoustic)] / P({loud, music})

Naive Bayes Assumption: We assume "loud" and "music" are independent given the sense.

Therefore: P({loud, music} | Sound-Noun-Acoustic) = P(loud | Sound-Noun-Acoustic) * P(music | Sound-Noun-Acoustic)

Putting it together:

P(Sound-Noun-Acoustic | {loud, music}) = [P(loud | Sound-Noun-Acoustic) * P(music | Sound-Noun-Acoustic) * P(Sound-Noun-Acoustic)] / P({loud, music})

P(Sound-Noun-Acoustic) = 0.5 (Prior probability of the sense)

P(loud | Sound-Noun-Acoustic) = 2/23 (Likelihood of "loud" given the sense)

P(music | Sound-Noun-Acoustic) = 3/23 (Likelihood of "music" given the sense)

$=0.5*(2/23)*(3/23)=0.5*6/529=3/529$

Since we're comparing the probabilities of the two senses, the denominator P({loud, music}) is the same for both. We can ignore it for the comparison:

P(Sound-Noun-BodyOfWater | {loud, music}) = [P(loud | Sound-Noun-BodyOfWater) * P(music | Sound-Noun-BodyOfWater) * P(Sound-Noun-BodyOfWater)] / P({loud, music})

P(Sound-Noun-BodyOfWater) = 0.5 (Prior probability of the sense)

P(loud | Sound-Noun-BodyOfWater) = 1/23 (Likelihood of "loud" given the sense)

P(music | Sound-Noun-BodyOfWater) = 1/23 (Likelihood of "music" given the sense)

$=1/1058$

- **Prediction:** Since 3/529 > 1/1058, the Naive Bayes classifier predicts **Sound-Noun-Acoustic**.

Benefits of Naive Bayes for WSD

- Simplicity: Naive Bayes is easy to implement.
- Efficiency: It can be trained efficiently, even with a moderate amount of training data.
- Effectiveness: It often performs surprisingly well in text classification tasks, including WSD, despite its naive assumption.

Drawbacks of Naive Bayes for WSD

- Naive Independence Assumption: The assumption that contextual words are independent of each other given the sense is rarely true in natural language. This oversimplification can limit accuracy.
- Feature Dependence: Performance heavily relies on the quality of the selected features.
- Zero-Frequency Problem: If a feature doesn't occur with a particular sense in the training data, its likelihood probability becomes zero, which can skew the results. Smoothing techniques (like Laplace smoothing) are used to mitigate this.

Note:

Understanding the Formula

- **P(music | Sound-Noun-Acoustic):** This represents the probability of encountering the word "music" given that the word "sound" has the sense "Sound-Noun-Acoustic" in a sentence. It answers the question: "Out of all the times 'sound' has the meaning of acoustic sound, how often is the word 'music' present in its surrounding context?

(2 + 1) / (3 + 20):

2: This is the number of times the word "music" appears in the context of "sound" when it has the sense "Sound-Noun-Acoustic" in our training data.

+ 1: This addition of '1' is Laplace smoothing (or add-one smoothing). It's a technique used to avoid zero probabilities. If a word doesn't appear at all in the context of a particular sense in the training data, its likelihood probability would be zero. Laplace smoothing adds 1 to all word counts to prevent this, ensuring that every word has a non-zero probability.

3: This is the total number of times the sense "Sound-Noun-Acoustic" appears in our training data. The word "sound" has the sense "Sound-Noun-Acoustic" in the first, second, and fifth sentences.

+ 20: This is the total number of unique words (the vocabulary size) in our training data's context words (after stop word removal).