

UNIVERSITY of DELAWARE

UNIVERSITY of DE

Problem of Word Association Mining

Hui Fang
Department of Electrical and Computer Engineering
University of Delaware

1

UNIVERSITY of DELAWARE

UNIVERSITY of DE

What is word association mining?

- Two types of word relations
 - Paradigmatic:
 - A and B have paradigmatic relation if they can be substituted for each other (i.e., A and B are in the same semantic class)
 - E.g., "cat" and "dog"; "Monday" and "Tuesday"
 - Syntagmatic:
 - A and B have syntagmatic relation if they can be combined with each other (i.e., A and B are related semantically)
 - E.g., "cat" and "sit"; "car" and "drive"

2

UNIVERSITY of DELAWARE

UNIVERSITY of DE

Why is it important?

- They are useful for improving accuracy of many NLP tasks
 - POS tagging, parsing, entity recognition
 - Grammar learning
- They are directly useful for many applications in text retrieval and mining
 - Search: e.g., Use word associations for query expansion
 - Summarization: e.g., what words are most strongly associated with "battery" in positive and negative reviews about iPhone 6?

3

UNIVERSITY of DELAWARE

UNIVERSITY of DE

Intuition for paradigmatic relation discovery

Paradigmatic: similar context

My cat eats fish on Saturday
His cat eats turkey on Tuesday
...
My dog eats meat on Sunday
His dog eats turkey on Tuesday
...

cat:

My : eats fish on Saturday
His : eats turkey on Tuesday
...

dog:

My : eats meat on Sunday
His : eats turkey on Tuesday
...

Similar left context

Similar right context

Similar general context

How similar are context ("cat") and context ("dog")?
How similar are context ("cat") and context ("computer")?

4

UNIVERSITY of DELAWARE

UNIVERSITY of DE

Intuition for syntagmatic relation discovery

Syntagmatic: correlated occurrences

My cat eats fish on Saturday
His cat eats turkey on Tuesday
My dog eats meat on Sunday
His dog eats turkey on Tuesday
...

My : eats : on Saturday
His : eats : on Tuesday
My : eats : on Sunday
His : eats : on Tuesday
...

What words tend to occur to the left of "eats"?

What words tend to occur to the right of "eats"?

Whenever "eats" occurs, what other words also tend to occur?
How helpful is the occurrence of "eats" for predicting occurrence of "meat"?
How helpful is the occurrence of "eats" for predicting occurrence of "text"?

5

UNIVERSITY of DELAWARE

UNIVERSITY of DE

Mining Word Associations: General Ideas

- Paradigmatic
 - Represent each word by its context
 - Compute context similarity
 - Words with high context similarity likely have paradigmatic relation
- Syntagmatic
 - Count how many times two words occur together in a context (e.g., sentence or paragraph)
 - Compare their co-occurrences with their individual occurrences
 - Words with high co-occurrences but relatively low individual occurrences likely have syntagmatic relation

6

1