ASPECT-LEVEL SENTIMENT ANALYSIS:

Implicit Features Detection in Consumers Reviews

Kim Schouten

(schouten@ese.eur.nl)

Flavius Frasincar

(frasincar@ese.eur.nl)

Franciska de Jong

(f.m.g.dejong@eshcc.eur.nl)

ERASMUS UNIVERSITEIT ROTTERDAM

SIKS

erasmus studio

COMMIT/ infiniti

Introduction

- People like to voice their opinion
- This is especially true on the Web
- We specifically focus on user-generated reviews
  - (e.g., Amazon, Yelp, etc.)
- These data can be harnessed for business purposes
  - For consumers, it is an important source of information when looking to purchase something
  - For producers, it is a valuable source of consumer feedback
    - Usually honest
    - Free!

Introduction

- The aim of Aspect-Level Sentiment Analysis is to find a quadruple

$$(s, g, h, t)$$

  - s = sentiment score
  - g = the target on which the sentiment is expressed
  - h = the holder: the one expressing the sentiment
  - t = the time when the sentiment was expressed

Introduction

- Traditionally, sentiment analysis is performed at document or sentence level
- Assumption that only one topic is discussed there
- Why not look for the actual topics being discussed and attach sentiment scores to those?
- This is called aspect-based sentiment analysis

Introduction

- Usually one document or sentence describes one entity
- However, multiple facets or aspects are described for that entity within a document, often with conflicting sentiment scores:
  - “The pizza was perfect, but the waiters were rude”
- Aspects can be fine-grained (pro: very detailed)
  - “pizza” and “waiters”
- Or coarse-grained (pro: easier to compare across reviews)
  - “food” and “service”
- Aspect-based sentiment analysis has two main tasks: finding aspects, and finding their sentiment scores

Introduction

- Both fine-grained and coarse-grained aspects can be implicit
- Since coarse-grained aspects are more general and abstract, these tend to be implied more often than fine-grained aspects
- Coarse-grained aspects are also referred to as aspect categories

ASPECT-LEVEL SENTIMENT ANALYSIS:

Coarse-grained Aspect Detection in Consumers Reviews

Kim Schouten

(schouten@ese.eur.nl)

Flavius Frasincar

(frasincar@ese.eur.nl)

Franciska de Jong

(f.m.g.dejong@eshcc.eur.nl)

ERASMUS UNIVERSITEIT ROTTERDAM

SIKS

erasmus studio

COMMIT/ infinity

Coarse-grained Aspect Detection Method

- Our intuition is that certain words are expected to co-occur a lot with certain coarse aspects (e.g., ‘delicious’ -> food, ‘expensive’ -> price)
- Hence, if we encounter those words, we can predict that these related coarse aspects are present in that sentence
- By looking at the co-occurrence of each word in a sentence with each of the possible coarse aspects, we can compute a score for each aspect

Coarse-grained Aspect Detection Method

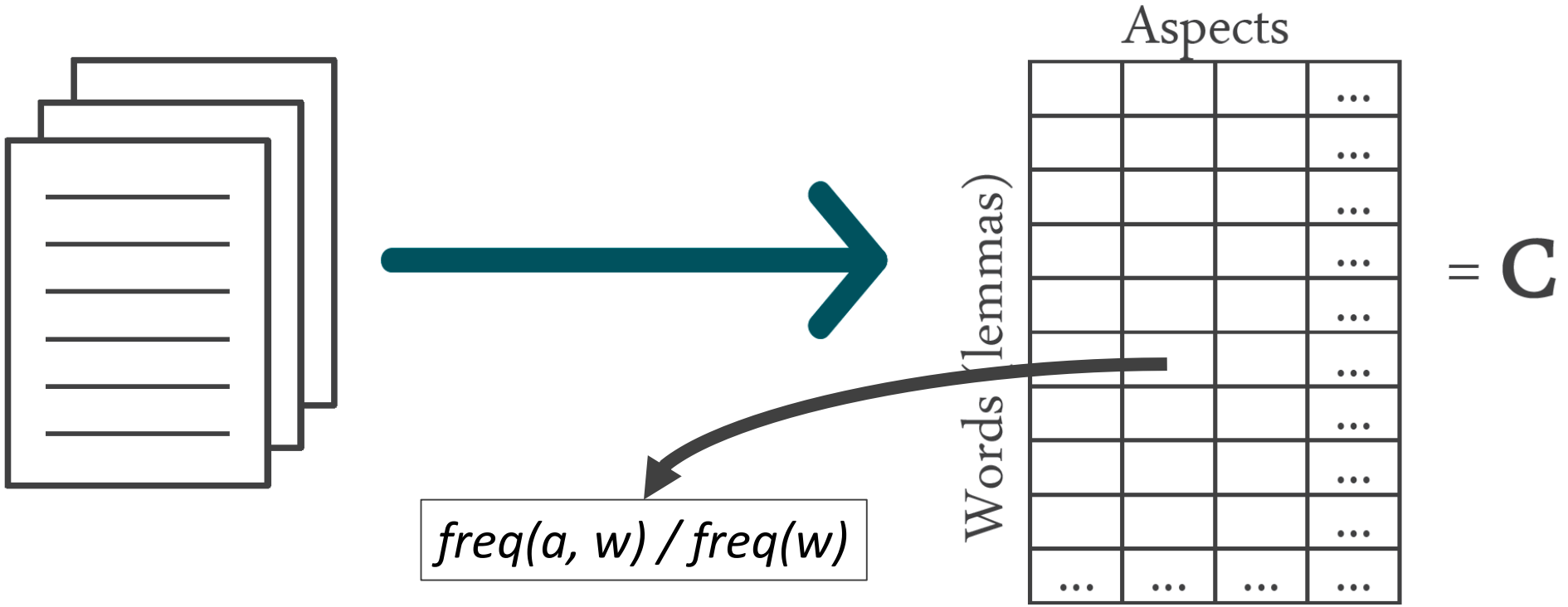
Annotated sentences

```
<sentence id="1458">
    <text>Our agreed favorite is the orrechiete with sausage and chicken (usually
    the waiters are kind enough to split the dish in half so you get to sample both
    meats).</text>
    <aspectTerms>
        <aspectTerm term="orrechiete with sausage and chicken"
    polarity="positive" from="27" to="62"/>
        <aspectTerm term="waiters" polarity="positive" from="76" to="83"/>
        <aspectTerm term="meats" polarity="neutral" from="152" to="157"/>
        <aspectTerm term="dish" polarity="neutral" from="113" to="117"/>
    </aspectTerms>
    <aspectCategories>
        <aspectCategory category="food" polarity="positive"/>
        <aspectCategory category="service" polarity="positive"/>
    </aspectCategories>
</sentence>
```

Training Phase – Creating **C**

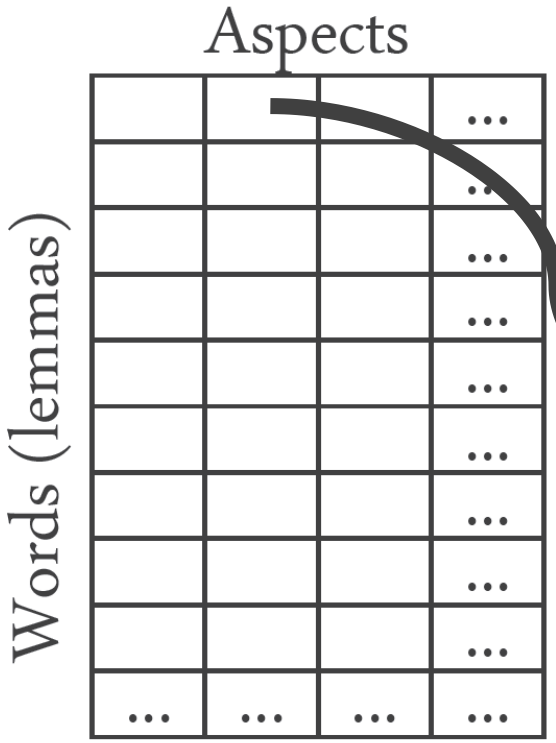
Annotated sentences

Co-occurrence matrix



Training Phase – Pruning **C**

Co-occurrence matrix



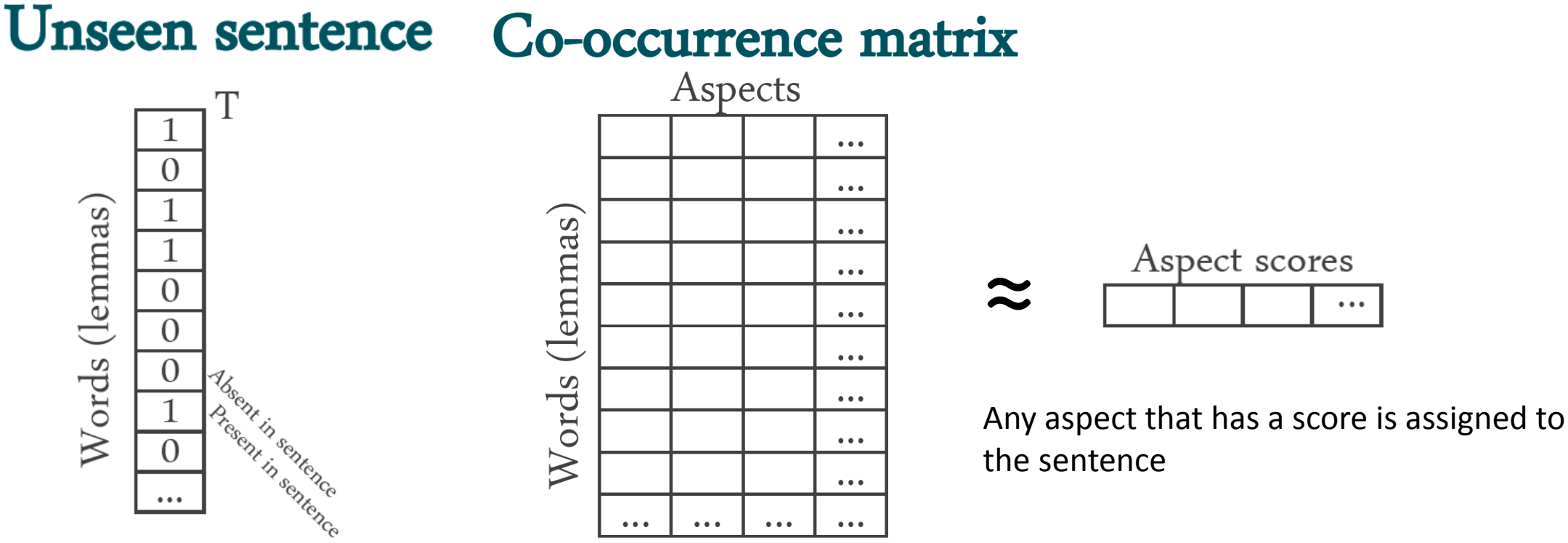
Using simple linear search, two threshold variables (*min_cooc* and *min_freq*) are optimized

freq(a, w) / freq(w) > *min_cooc*
freq(w) > *min_freq*

Only entries that satisfy both constraints are retained in **C**

Processing unseen data

Simply put: **A** = **SC**



Any aspect that has a score is assigned to the sentence

Results

- Two data sets with restaurant reviews
  - First one has ~3800 sentences with 4 different aspect categories
  - Second one has ~1300 sentences with 13 different aspect categories
- One data set with laptop reviews
  - (~1700 sentences with 82 different aspect categories)

	Precision	Recall	F_1	SemEval best F_1
Restaurants 1	66.7%	73.9%	70.1%	85.3%
Restaurants 2	64.7%	62.2%	63.4%	61.9%
Laptops	40.1%	42.3%	41.4%	49.6%

- Very good performance for a relatively simple method that uses no advanced machine learning techniques

Alternative SVM

- Trained a basic bag-of-words model SVM classifier
  - One (binary) SVM model for each aspect
  - Per sentence determine whether that coarse aspect is present or not
- Results on the same data:

	Precision	Recall	F_1	Co-occurrence F_1	SemEval best F_1
Restaurants 1	81.3%	75.2%	78.1%	70.1%	85.3%
Restaurants 2	77.5%	57.7%	66.1%	63.4%	61.9%
Laptops	60.9%	34.7%	44.2%	41.4%	49.6%

- Very good performance for a relatively simple method that uses no advanced machine learning techniques

On-going and future work

- Use of ontologies or other knowledge bases to make use of domain knowledge
  - Move towards a more concept-driven, or semantics-driven form of aspect-level sentiment analysis
- Use of more advanced machine learning techniques
  - Latent Dirichlet Allocation might be used to find coarse aspect
  - Recurrent Neural Networks (with LSTM)

QUESTIONS?