UNIVERSITÀ DEGLI STUDI DI MILANO
LA STATALE

Master Degree in Data Science and Economics

Text Mining and Sentiment Analysis

Aspect Based Sentiment Analysis

Sentence- and Phrase-Level Analysis

Prof. Alfio Ferrara

Department of Computer Science, Università degli Studi di Milano
Room 7012 via Celoria 18, 20133 Milano, Italia alfio.ferrara@unimi.it

sed noli modo

Problem definition

In AbSA, sentiment is a **subjective consciousness of human beings towards an aspect (objective existence)**. In this light: “A sentiment is basically an opinion that a person expresses towards an aspect, entity, person, event, feature, object, or a certain target.”

- Schouten, K., & Frasincar, F. (2015). Survey on aspect-level sentiment analysis. IEEE Transactions on Knowledge and Data Engineering, 28(3), 813-830.

- Rana, T. A., & Cheah, Y. N. (2016). Aspect extraction in sentiment analysis: comparative analysis and survey. Artificial Intelligence Review, 46(4), 459-483.

- Nazir, A., Rao, Y., Wu, L., & Sun, L. (2020). Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. IEEE Transactions on Affective Computing.

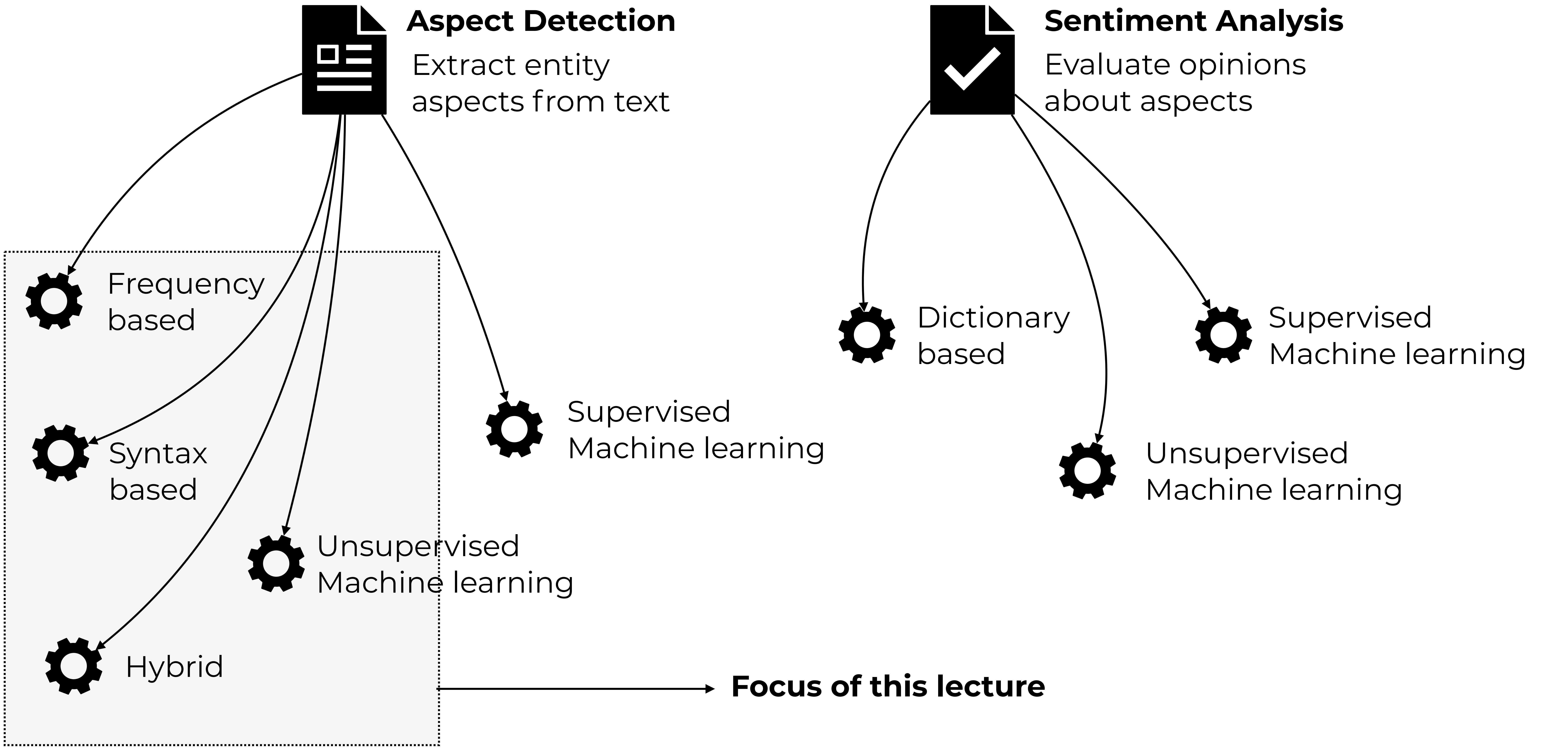
Tasks in AbSA

Given an entity, e.g., a product, the tasks of AbSA are:

i. Identify **entity features**.
ii. Identify **opinions regarding entity features**.
iii. Determine the **polarity of opinions**.
iv. Rank opinions based on their **strength**


```
<sentence id="1634">
  <text>
     The food is uniformly exceptional, with a very capable
     kitchen which will proudly whip up whatever you feel like
     eating, whether it's on the menu or not.
  </text>
  <aspectTerms>
    <aspectTerm term="food" polarity="positive"
        from="4" to="8"/>
    <aspectTerm term="kitchen" polarity="positive"
        from="55" to="62"/>
    <aspectTerm term="menu" polarity="neutral"
        from="141" to="145"/>
    </aspectTerms>
  <aspectCategories>
    <aspectCategory category="food" polarity="positive"/>
  </aspectCategories>
</sentence>
<sentence id="2534">
  <text>
     Where Gabriela personaly greets you and recommends you what
     to eat.
  </text>
  <aspectCategories>
     <aspectCategory category="service" polarity="positive"/>
  </aspectCategories>
</sentence>
```

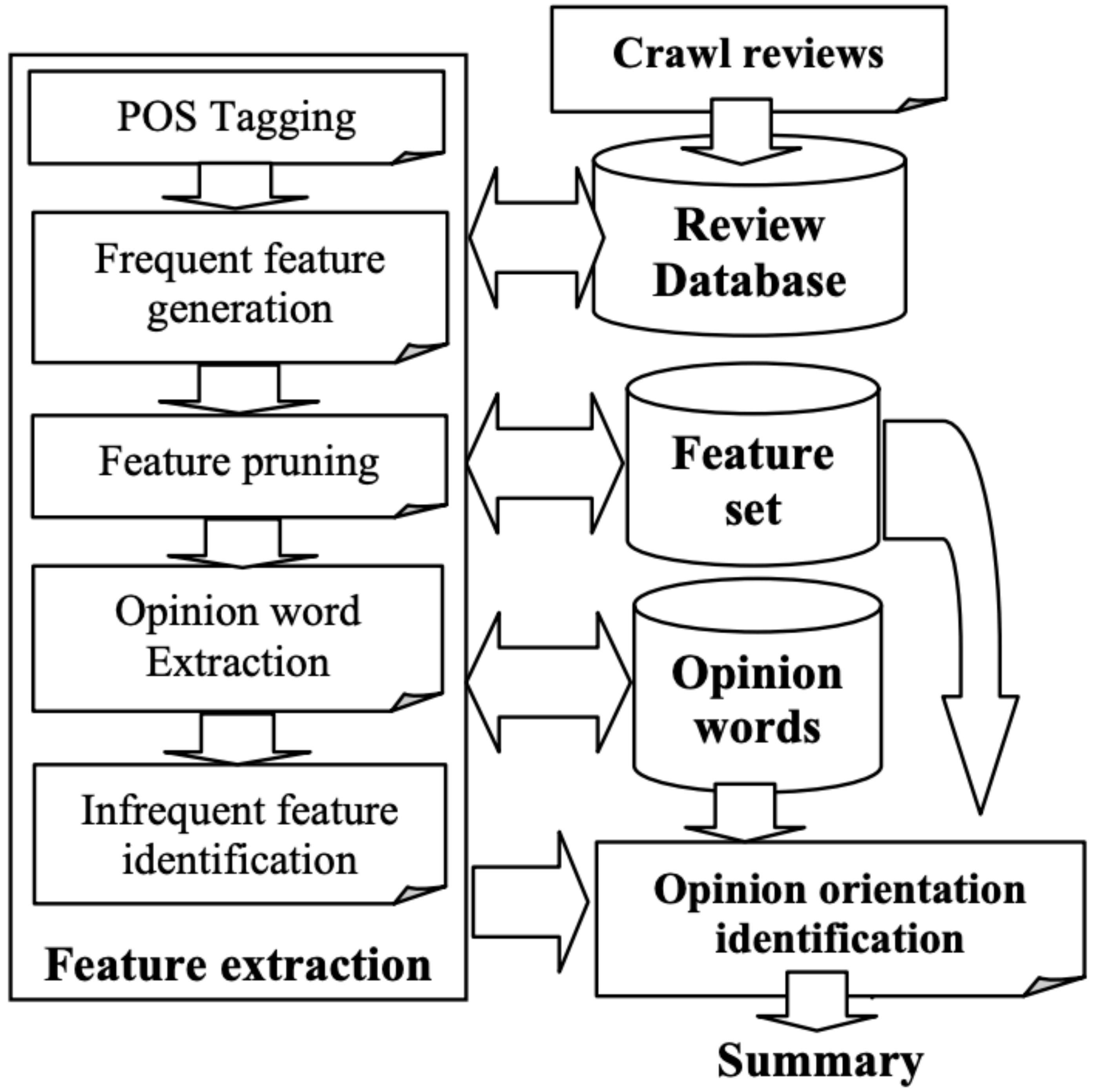
Overview of the main approaches



Frequency-based approaches (I)

It has been observed that in reviews, a **limited set of words is used much more often than the rest of the vocabulary**. These frequent words (usually only single nouns and compound nouns are considered) are likely to be aspects.

- Hu M, Liu B (2004b) Mining opinion features in customer reviews. AAAI 4:755–760



Association rule mining

Let $I = \{i_1, \ldots, i_n\}$ be a set of items (words), and D be a set of transactions (sentences).

Each transaction consists of a subset of items in I.

An association rule is an implication of the form $X \rightarrow Y$, where $X \subset I$, $Y \subset I$, and $X \cap Y = \emptyset$.

The rule $X \rightarrow Y$ holds in D with **confidence** c if c% of transactions in D that support X also support Y.

The rule has **support** s in D if s% of transactions in D contain $X \cup Y$.

Frequency-based approaches (II)

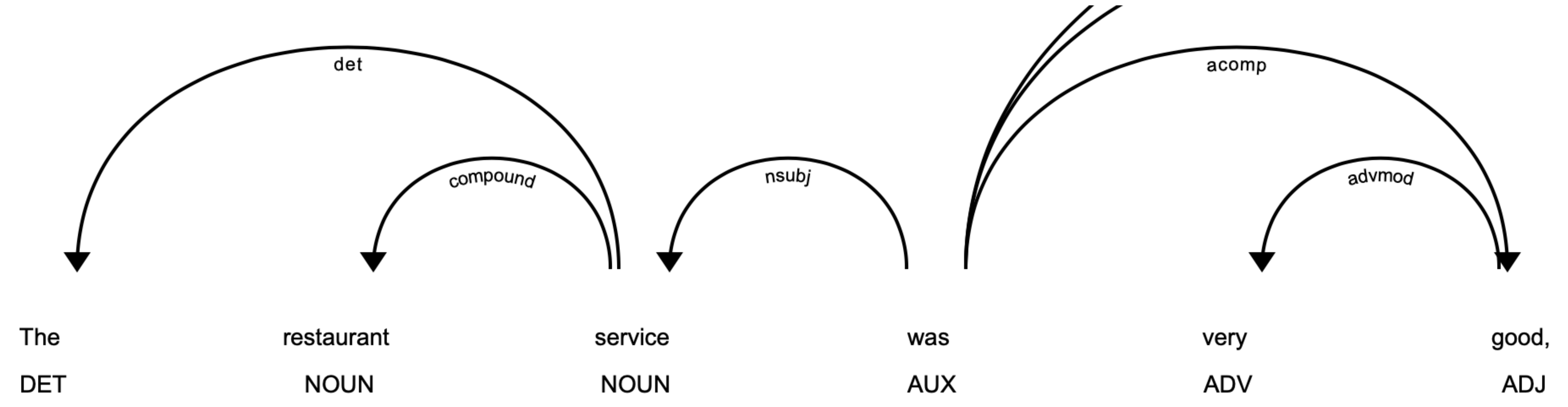
Raju, S., Pingali, P., & Varma, V. (2009, April). An unsupervised approach to product attribute extraction. In European Conference on Information Retrieval (pp. 796-800). Springer, Berlin, Heidelberg.

Pre-processing. Identify **noun phrases** which are then given as input to clustering.

Product descriptions contain phrases which begin with a determiner word like "your favorite music", [...] and other single word noun phrases like "comfort", [...] which often explain an attribute of the product rather than define it.
*We employ two **pruning methods** to eliminate the above noun phrases. [i] discard all the noun phrases which begin with a determiner word. [ii] assume that **single word noun phrases mentioned above occur more frequently in general English than in product descriptions**. Let p and q be the unigram probability distributions of input document set and a general English corpus respectively. Now we compute pointwise KL divergence score* δ_w *for each unigram w in the input documents which gives the **relative importance of the unigram in the input document set compared to the generic corpus**.*

```
import spacy

nlp = spacy.load("en_core_web_sm")
doc = nlp("The restaurant service was very good, but I'm not
sure about the food quality")
for chunk in doc.noun_chunks:
    print(chunk.text, chunk.root.text, chunk.root.dep_,
            chunk.root.head.text)
```



$$\delta_w(p \mid\mid q) = p(w)\log\frac{p(w)}{q(w)}$$

Frequency-based approaches (II)

Raju, S., Pingali, P., & Varma, V. (2009, April). An unsupervised approach to product attribute extraction. In European Conference on Information Retrieval (pp. 796-800). Springer, Berlin, Heidelberg.

The noun phrases obtained from the previous step are **clustered** so that **noun phrases describing the same attribute are grouped together in the same cluster**.

We calculate ***N gram overlap to measure the similarity between two noun phrases****. We consider unigram and bigram overlap for this. Bigrams are ordered pairs of words co-occurring within five words of each other. Let S_i and S_j be the sets of uni-grams, bi-grams belonging to two noun phrases P_i and P_j respectively. Now we define the similarity between the two noun phrases P_i and P_j using* ***Dice's Coefficient similarity***

$$\sigma(P_i, P_j) = \frac{2 \mid S_i \cap S_j \mid}{\mid S_i \mid + \mid S_j \mid}$$

Frequency-based approaches (II)

UNIVERSITÀ DEGLI STUDI DI MILANO
LA STATALE

Raju, S., Pingali, P., & Varma, V. (2009, April). An unsupervised approach to product attribute extraction. In European Conference on Information Retrieval (pp. 796-800). Springer, Berlin, Heidelberg.

Assuming that each cluster has noun phrases that contain instances of same attribute, an **attribute is extracted from each cluster** (I.e., select the best n-gram for each cluster).

We define an Attribute Scoring Function AS to score each of these n-grams. We declare that the n-gram with highest score is the attribute.

$$AS(x) = \frac{PKL(x)}{AHD(x)}$$

PKL: Let P be the probability distribution of a cluster and Q be the probability distribution of the rest of the clusters together.

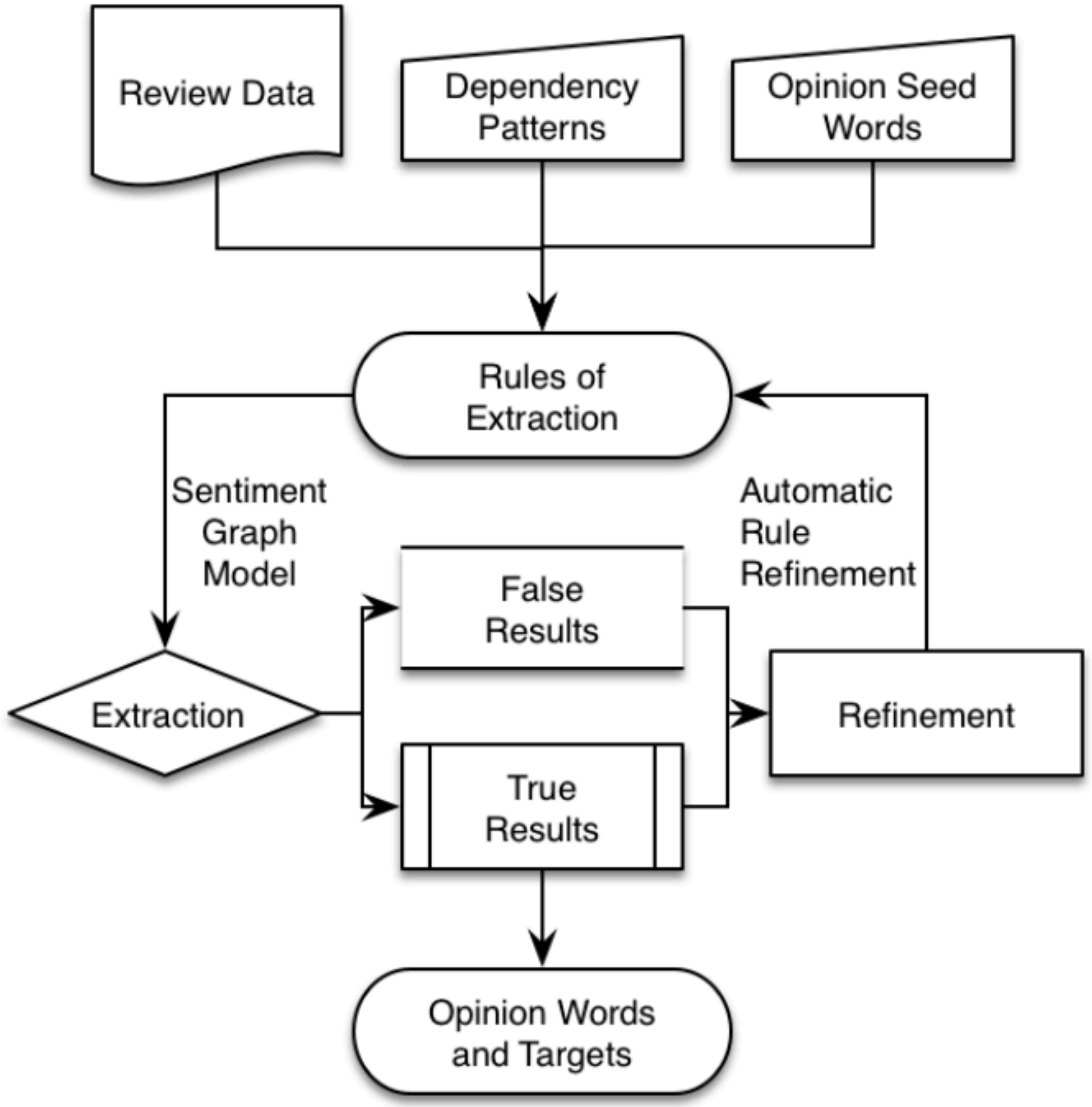
$$PKL(x) = P(x)\log\frac{P(x)}{Q(x)}$$

AHD: The average head noun distance of the n-gram x in its instances. Head Noun Distance is the distance of the n-gram x from the right most word (head noun) in the noun phrase.

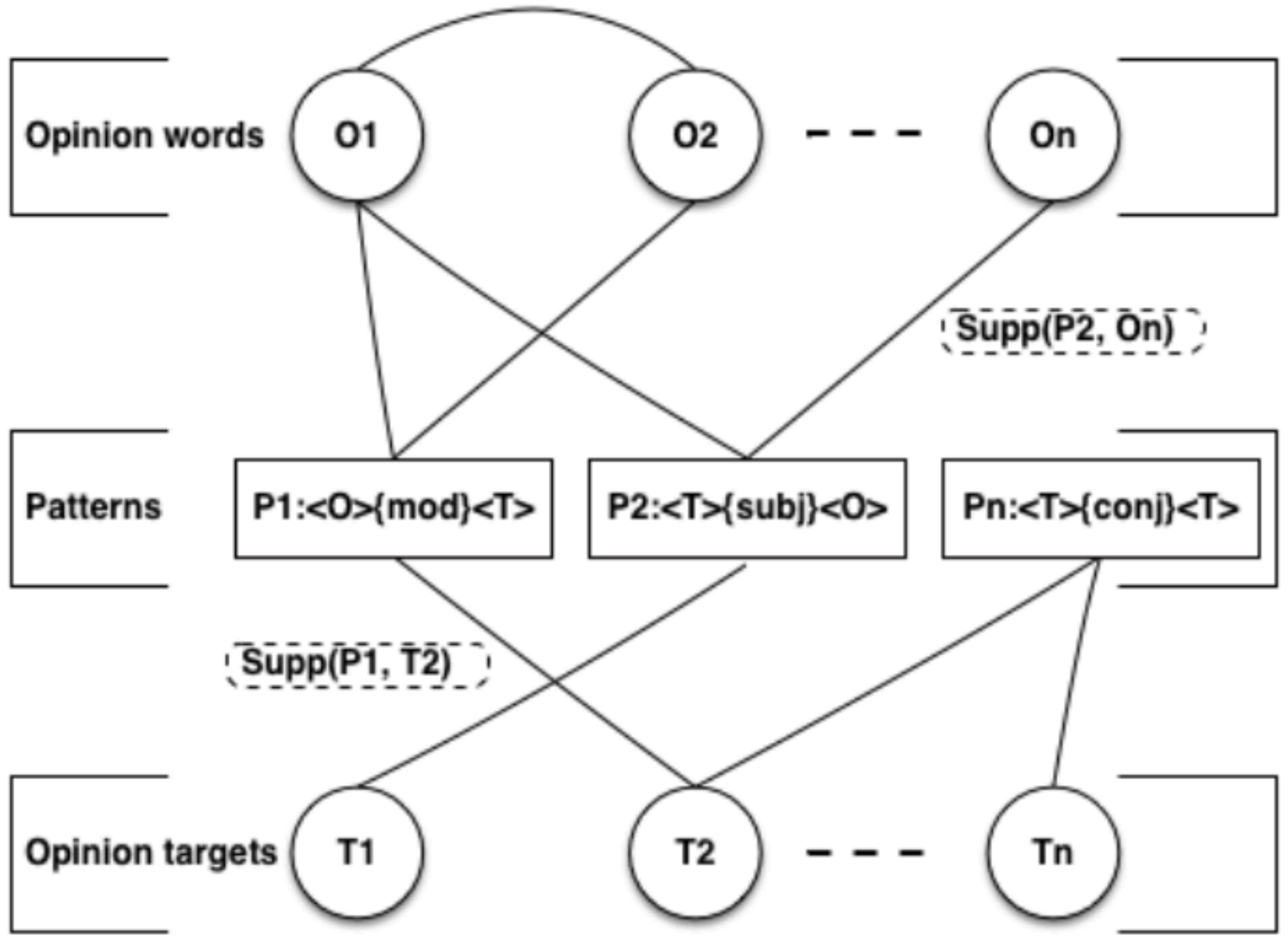
$$AHD(x) = \frac{1}{N(x)}\sum_i D(x,i)$$

Semi-supervised approaches

Zhao, Q., Wang, H., Lv, P., & Zhang, C. (2014, November). A bootstrapping based refinement framework for mining opinion words and targets. In Proceedings of the 23rd ACM international conference on conference on information and knowledge management (pp. 1995-1998).



Sentiment graph



$$\begin{aligned} supp(i,j) = 2[&\log L(P(i \mid j), N_{i,j}, N_j) \\ &+\log L(P(i \mid \bar{j}), N_{i,\bar{j}}, N_{\bar{j}}) \\ &-\log L(P(i), N_i, N)] \end{aligned}$$

$$P(i,j) = \frac{N_{i,j}}{N_j} \;;\; L(p,n,k) = p^k(1-p)^{n-k}$$

$Supp(o, t)$ is the weight of the edge e: $v_o \rightarrow v_t$ on the Sentiment Graph.

Ontology based (wordnet)

Another option to search for entity aspects in text is to exploit relations in a knowledge base, such as WordNet

```
photo_camera = wn.synset('camera.n.01')
photo_camera.part_meronyms()
```

```
[Synset('aperture.n.01'),
 Synset('camera_lens.n.01'),
 Synset('delayed_action.n.01'),
 Synset('diaphragm.n.01'),
 Synset('finder.n.03'),
 Synset('hood.n.04'),
 Synset('magazine.n.04'),
 Synset('shutter.n.01'),
 Synset('sprocket.n.01')]
```

```
candidates = ['photograph.n.01', 'quality.n.01',
'price.n.02', 'food.n.01']
for candidate in candidates:
    print(candidate,
          photo_camera.path_similarity(
             wn.synset(candidate)))
```

```
photograph.n.01 0.125
quality.n.01 0.083
price.n.02 0.062
food.n.01 0.09
```

Ontology based (knowledge base)

Example using Wikidata (https://www.wikidata.org/)

```
select ?aspect ?name
where {
  wd:Q15328 wdt:P527 ?aspect.
  ?aspect rdfs:label ?name.
  FILTER(lang(?name)='en')
}
```


```
{'head': {'vars': ['aspect', 'name']},
 'results': {'bindings': [{'aspect':
{'type': 'uri',
     'value': 'http://www.wikidata.org/
entity/Q192234'},
    'name': {'xml:lang': 'en', 'type':
'literal', 'value': 'camera lens'}},
   {'aspect': {'type': 'uri',
     'value': 'http://www.wikidata.org/
entity/Q209871'},
    'name': {'xml:lang': 'en', 'type':
'literal', 'value': 'system camera'}},
   {'aspect': {'type': 'uri',
     'value': 'http://www.wikidata.org/
entity/Q4410572'},
    'name': {'xml:lang': 'en',
     'type': 'literal',
     'value': 'photosensitive
materials'}}]}}
```

Datasets

UNIVERSITÀ DEGLI STUDI DI MILANO
LA STATALE

Mudalige, C. R., Karunarathna, D., Rajapaksha, I., de Silva, N., Ratnayaka, G., Perera, A. S., & Pathirana, R. (2020). **SigmaLaw-ABSA**: Dataset for Aspect-Based Sentiment Analysis in Legal Opinion Texts. arXiv preprint arXiv:2011.06326. https://osf.io/37gkh/

ASPECT BASED SENTIMENT ANALYSIS DATASET

Brun, C., & Nikoulina, V. (2018, October). Aspect based sentiment analysis into the wild. In Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (pp. 116-122). https://europe.naverlabs.com/research/natural-language-processing/aspect-based-sentiment-analysis-dataset/

ABSITA

Aspect-based Sentiment Analysis at EVALITA. http://sag.art.uniroma2.it/absita/data/