Lecture 10: Sentiment Analysis

Pilsung Kang
School of Industrial Management Engineering
Korea University

AGENDA

Definition

Feldman (2013)

- What is Sentiment Analysis?
  - ✓ Computational study of opinions, sentiments, evaluations, attitudes, appraisal, affects, views, emotions, subjectivity, etc., expressed in text.
    - Text: Reviews, blogs, discussions, news, comments, feedback,…
  - ✓ Sentiment Analysis is sometimes called **Opinion Mining**

Backgrounds

Feldman (2013)

- Typical Usage of Sentiment Analysis
  - ✓ Extract from text how people feel about different products
  - ✓ Sentiment analysis can be tricky
    - **`Honda Accords and Toyota Camrys are nice sedans`**
    - **`Honda Accords and Toyota Camrys are nice sedans, but hardly the best car on the road`**

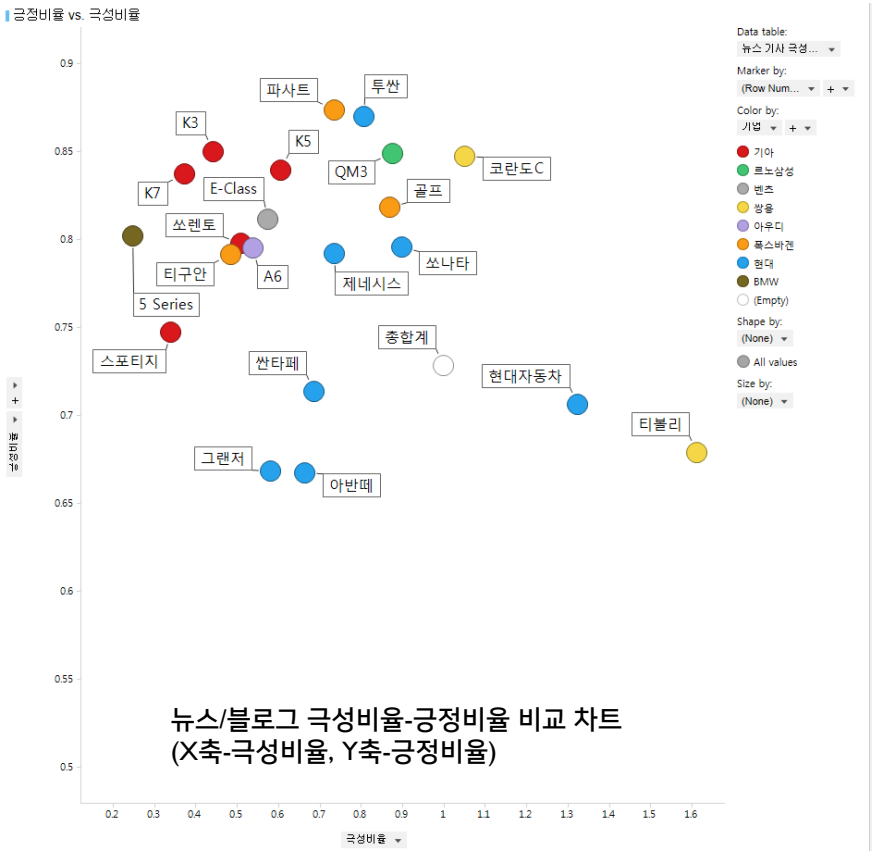
Backgrounds

Feldman (2013)

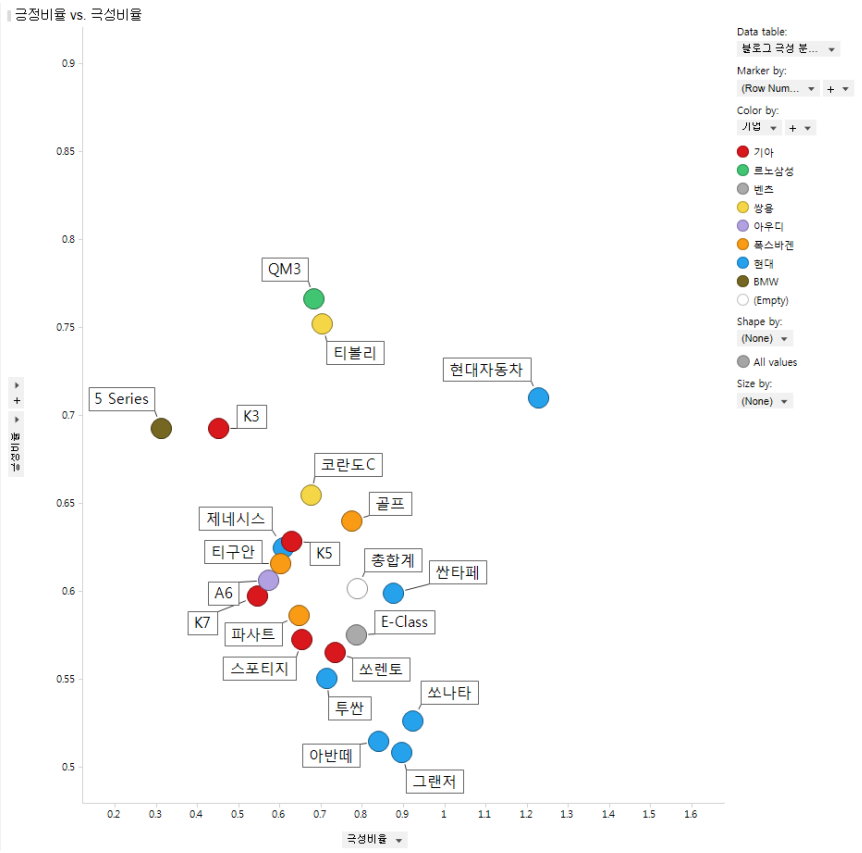
- Opinions are widely stated
  - ✓ Organization's internal data
    - Customer feedback from emails, call centers, etc.
  - ✓ News and reports
    - Opinions in news articles and commentaries
  - ✓ Word-of-mouth on the web
    - Personal experiences and opinions about anything in reviews, forums, blogs, Twitter, micro-blogs, etc.
    - Comments about articles, issues, topics, reviews, etc.
    - Posting at social networking sites (Facebook, Instagram, etc.)

Applications

- Business and organizations
  - ✓ Benchmark products and services; market intelligence
    - Business spend a huge amount of money to find customer opinions using consultants, surveys and focus groups, etc.



뉴스/블로그 극성비율-긍정비율 비교 차트
(X축-극성비율, Y축-긍정비율)



Applications

- Individuals
  - ✓Make decisions to purchase products or to use services
  - ✓Find public opinions about political candidates and issues

Applications

- Ad placement
  - ✓ Place an ad if one praises a product
  - ✓ Place an ad from a competitor if one criticizes a product

Guest

Why sentiment analysis is the future of ad optimization

PETER YARED MARCH 20, 2011 6:21 PM

TAGS: KLOUT, PEERINDEX

[Peter Yared is the vice presi
acquired Transpond, a socia

Sentiment analysis is a hot
the promise of helping bra
are thinking and saying ab
including early contender
BuzzLogic, and my own co
Measurement product are becoming pervasive
while consumer sentiment is important, what's
revenue.

First Workshop
on

Internet Advertising Using Sentiment Analysis (AdSent 2013)
with

IEEE International Conference on Data Mining series (ICDM), December 7, Dallas, Texas, USA

Applications

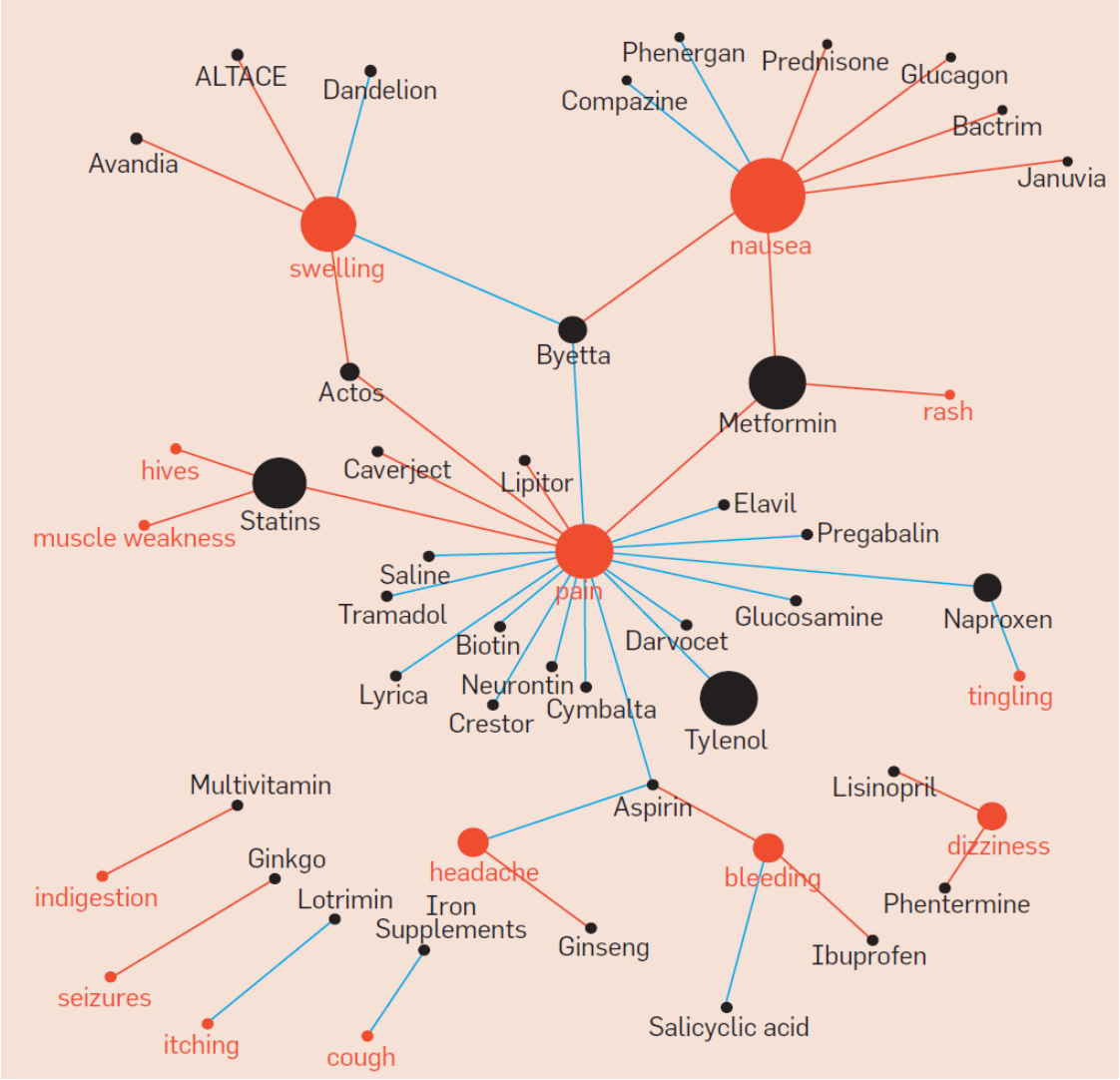
- Opinion Retrieval
  - ✓ Provide general search for opinions

Player	Line	Score	승	패배	피안타	피홈런	볼넷	실점	평균 자책점	이닝당 삼진	이닝당 볼넷	삼진 /볼넷	방어율
웨버	NC웨버는 7이닝 동안 1안타 3볼넷 10탈삼진으로 무실점 호투했다.	0.991	1	0	0	0	1	0	0	1	1	1	0
이재학	이재학 '10승', 3년만에 달라진 위상 증명하다	0.986	1	0	0	1	0	1	0	0	0	1	0
이태양	'이태양 완벽투' NC, LG 상대로 창단 첫 스윕	0.966	1	0	0	0	0	0	0	0	0	1	0
에릭	한편, 이날 7이닝 3실점으로 잘 던지며 7경기 만에 국내 첫 승을 거둔 에릭은 "굉장히 흥분된다.	0.803	1	0	1	0	0	0	0	0	0	0	0
찰리	어이없는 실책 2개가 나오자 찰리는 흔들리기 시작했다.	0.506	0	1	1	1	0	1	0	1	0	1	1
찰리	찰리가 대량 실점을 했지만 그 과정에서 3루수 모창민과 1루수 테임즈의 실책이 동반되면서 자책은 1점 밖에 되지 않았다.	0.506	0	1	1	1	0	1	0	1	0	1	1
원종현	NC 원종현이 패전투수가 됐다.	0.293	0	1	0	0	0	1	1	1	0	1	1
원종현	세 번째 투수로 등판한 원종현 역시 1사 후 8번 김성현에게 솔로포를 맞았다.	0.102	0	0	0	1	1	0	1	1	1	0	0
이재학	NC 선발 이재학은 8이닝 8피안타(2홈런) 5탈삼진 3볼넷 2실점을 기록했다.	0.079	0	0	1	1	0	1	1	0	0	0	1
이혜천	이혜천이 올라온 뒤 한꺼번에 5점을 내주면서 맥빠진 경기가 되고 말았다.	0.066	0	0	1	1	0	1	0	0	1	0	0

Applications

Feldman (2013)

- Relationships between Drugs and Symptoms



AGENDA

Problem Definition

Feldman (2013)

Problem Statement: Abstraction

- Consists of two parts
  - ✓ (1) Opinion definition
    - What is an opinion?
  - ✓ (2) Opinion summarization
    - Opinions are subjective
    - An opinion from a single person (unless a VIP) is often not sufficient for action
    - We need opinions from many people, and thus need to summarize those opinions

Problem Definition

Feldman (2013)

Entity and Aspect/Feature

> **Id: Abc123 on 5-1-2008** "*I bought an iPhone a few days ago. It is such a nice phone. The touch screen is really cool. The voice quality is clear too. It is much better than my old Blackberry, which was a terrible phone and so difficult to type with its tiny keys. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, …*"

- What do we see?
  - ✓ Opinion targets: entities and their features/aspects
  - ✓ Sentiments: positive and negative
  - ✓ Opinion holders: persons who hold the opinions
  - ✓ Time: when opinions are expressed

Problem Definition

Feldman (2013)

- Two main types of opinions
  - ✓ Regular opinions: Sentiment/opinion expressions on some target entities
    - Direct opinions: "The touch screen is really cool."
    - Indirect opinions: "After taking the drug, my pain has gone."
  - ✓ Comparative opinions: Comparisons of more than one entity
    - "iPhone X is better than Galaxy 9"
  - ✓ Focus more on regular opinions in this lecture

Problem Definition

Feldman (2013)

- A (Regular) Opinion
  - ✓ An opinion has the following basic components

$$(g_i, so_{ijk}, h_j, t_k)$$

where

- g_i is a target.
- so_{ijk} is the sentiment value (positive, negative, neutral, or a rating score) of the opinion from opinion holder h_j on target g_i at time t_k.
- h_j is an opinion holder.
- t_k is the time when the opinion is expressed.

Problem Definition

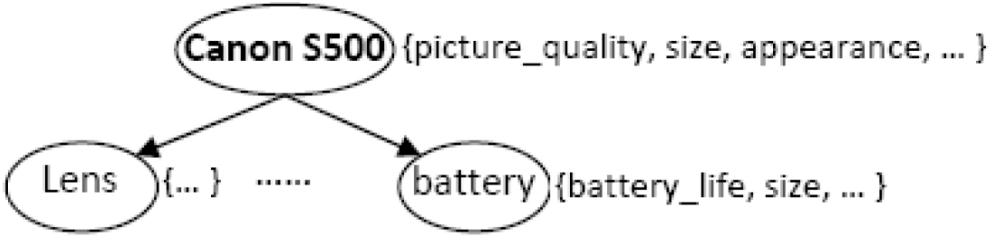
Feldman (2013)

- Opinion Target
  - ✓ In some cases, opinion target is a single entity or topic.
    - "I love G5" and "I support tax cut."
  - ✓ But in many other cases, it is more complex.
    - "I bought an iPhone a few days ago. It is such a nice phone. The touch screen is really cool."
    - Opinion target of the 3rd sentence is not just touch screen, but the "touch screen of iPhone"
    - "I support tax cut for the middle class, but not for the rich…"

Problem Definition

Feldman (2013)

- Entity and Aspect
  - ✓ Definition of Entity: An **entity e** is a product, person, event, organization, or topic. **e** is represented as
    - a hierarchy of **components**, **sub-components**, and so on.
    - Each node represents a component and is associated with a set of **attributes** of the component.



    - An opinion can be expressed on any node or attribute of the node.
    - For simplicity, we use the term **aspects (features)** to represent both components and attributes.

Problem Definition

Feldman (2013)

- Definition of an Opinion as a Quintuple

$$(e_i, a_{jl}, so_{ijkl}, h_j, t_k)$$

where

- e_i is a target entity/object
- a_{jl} is an aspect/feature/attribute/facet of the entity e_i
- so_{ijkl} is the sentiment value (positive, negative, neutral, or a rating score) of the opinion from opinion holder (source) h_j on aspect a_{jk} of entity e_i at time t_k
- h_j is an opinion holder
- t_k is the time when the opinion is expressed

Problem Definition

Feldman (2013)

- Quintuple Opinion Examples

> **Id: Abc123 on 5-1-2008** *"I bought an iPhone a few days ago. It is such a nice phone. The touch screen is really cool. The voice quality is clear too. It is much better than my old Blackberry, which was a terrible phone and so difficult to type with its tiny keys. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, …"*

✓ (iPhone, General, Positive, Abc123, 5-1-2008)

✓ (iPhone, touch_screen, Positive, Abc123, 5-1-2008)

✓ …

Problem Definition

Feldman (2013)

Purpose: Structure the unstructured

- Given an opinionated document,
  - ✓ Discover all quintuples $(e_i, a_{jl}, so_{ijkl}, h_j, t_k)$
  - ✓ Or, solve some simpler form of the problem
    - E.g. Classify the sentiment of the entire document
  - ✓ With the quintuples, unstructured texts transformed into structured data.

Rational vs. Emotional Evaluation

Feldman (2013)

- Rational Evaluation
  - ✓ Many evaluation/opinion sentences express no emotion
    - The voice of this phone is clear
- Emotional evaluation
    - I love this phone
- Some emotion sentences express no (positive or negative) opinion/sentiment
  - ✓ I am so surprised to see you

Abstraction: Opinion Summary

Feldman (2013)

- Feature-based opinion summary

""I bought an iPhone a few days ago. It is such a nice phone. The touch screen is really cool. The voice quality is clear too. It is much better than my old Blackberry, which was a terrible phone and so difficult to type with its tiny keys. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, …"

1.

….

Feature Based Summary of iPhone:

Feature1: **Touch screen**

Positive: 212

- *The touch screen was really cool.*
- *The touch screen was so easy to use and can do amazing things.*

…

Negative: 6

- The screen is easily scratched.
- I have a lot of difficulty in removing finger marks from the touch screen.

…

Feature2: **voice quality**

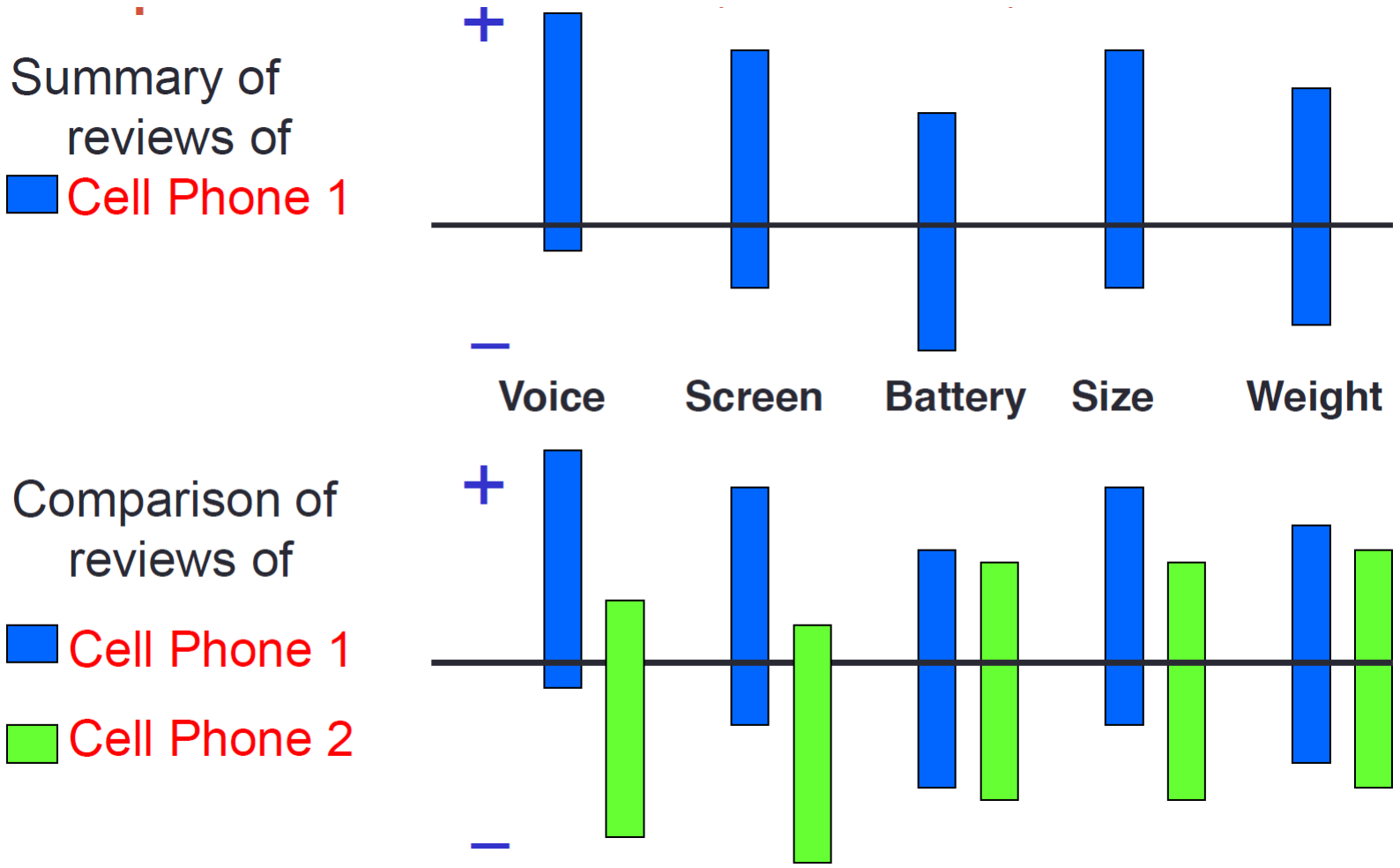
…

Note: We omit opinion holders

Abstraction: Opinion Summary

Feldman (2013)

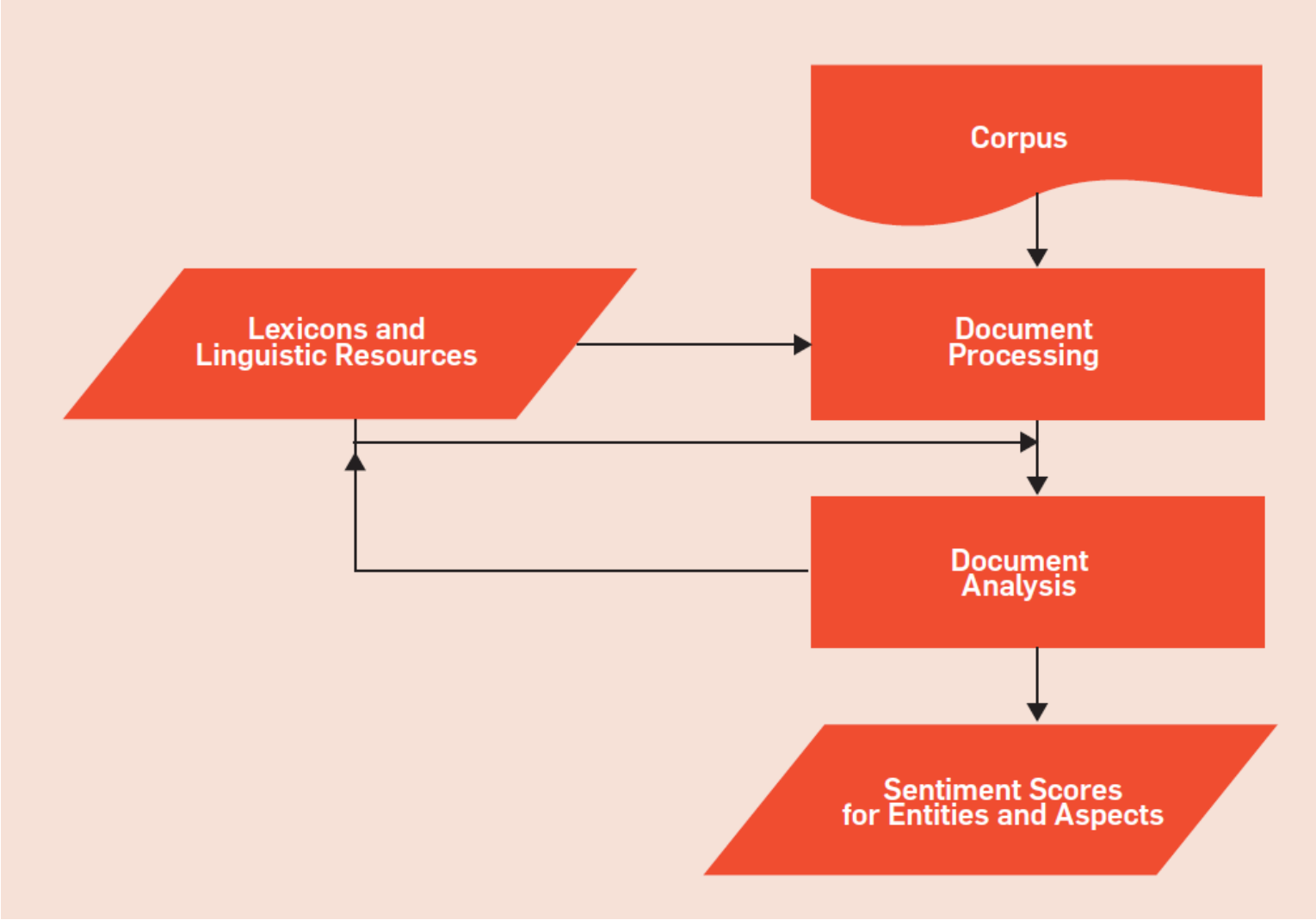
- Opinion Comparison



Sentiment Analysis Architecture

Feldman (2013)

- Architecture of generic sentiment analysis system



Document Sentiment Classification

Feldman (2013)

- Document Sentiment Classification
  - ✓ Classify a whole opinion document based on the overall sentiment of the opinion holder
    - Classes: **positive**, **negative** (possibly neutral)
    - Neutral or no opinion is hard. Most papers ignore it.
  - ✓ It is basically a text classification problem
  - ✓ Assumption: a document is written by **a single person** and expresses opinion/sentiment on **a single entity**.
  - ✓ Goal: discover **(_ , _ , so, _, _)** where e, a, h, and t are ignored.
    - Reviews usually satisfy the assumption (4/5 stars → positive, ½ stars → negative)
    - Many forum postings and blogs do not. They can mention and compare multiple entities or they express no sentiments.

Document Sentiment Classification

Feldman (2013)

- Supervised Learning Formulation
  - ✓ Training and test data: movie reviews with star ratings
    - 4-5 stars as positive and 1-2 stars as negative
    - Neutral is ignored
  - ✓ **Key: feature engineering**
    - Typically unigrams (bag of individual words)
    - Term frequency and different IR weighting schemes
    - Part of speech (POS) tags
    - Opinion words and phrases
    - Negations
    - Syntactic dependency
  - ✓ Too coarse for most applications

Sentence Sentiment Classification

Feldman (2013)

- Sentence Sentiment Classification
  - ✓ Assumes a single sentiment per sentence
  - ✓ Not always true, so one can classify clauses instead

- Two steps of sentence sentiment classification
  - ✓ Subjectivity classification
    - To identify subjective sentences
  - ✓ Sentence classification of subjective sentences
    - As positive or negative

Feature/Aspect-based Sentiment Classification

Feldman (2013)

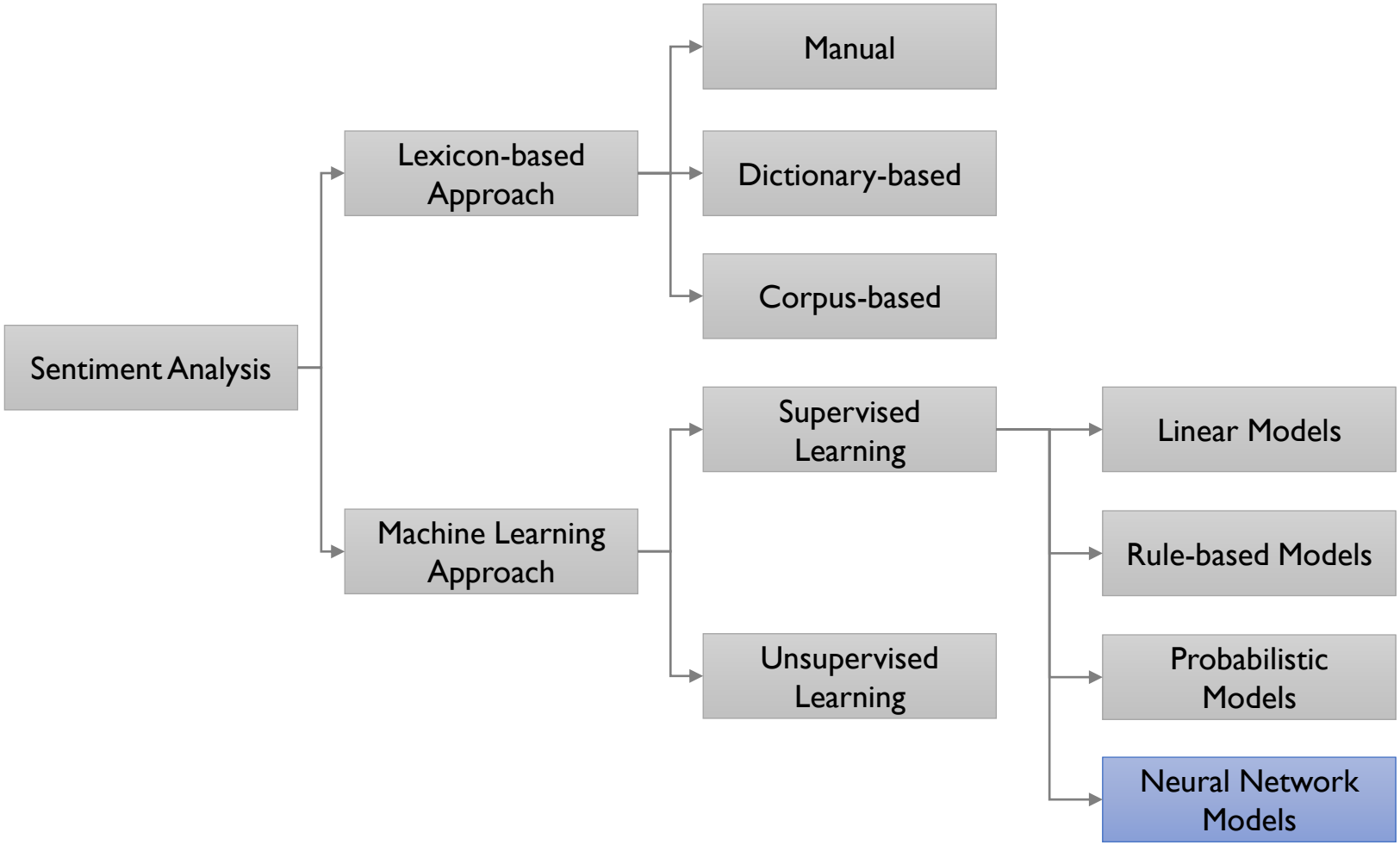
- Document/Sentence level sentiment classification
  - ✓ Useful but do not find what people liked and disliked.
  - ✓ They do not identify the targets of opinions: entities and their aspects
- Finding entities
  - ✓ Similar to but somewhat different from the traditional named entity recognition (NER)
    - Frequent nouns and noun phrases
    - Rule-based or supervised learning (HMM, CRFs)
    - Double propagation (DP): knowing one helps find the other (an opinion should have a target)
      - Input: a set of seed opinion words
      - The rooms are spacious
      - The phone has good screen

Sentiment Lexicon

- Sentiment Lexicon
  - ✓ Lists of words and expressions used to express people's subjective feelings and sentiment/opinions
    - Positive: beautiful, wonderful, good, amazing, …
    - Negative: bad, poor, terrible, …
    - Not just individual words, but also phrases and idioms
    - "cost an arm and a leg"
  - ✓ Many of them are context dependent, not just application domain dependent.
- Three main ways to compile such lists
  - ✓ Manual approach: not a bad idea for a one-time effort
  - ✓ Dictionary-based approach
  - ✓ Corpus-based approach

Sentiment Analysis Approaches

- Sentiment classification techniques




ANY
questions?