

Lecture 3: Natural Language Processing (III) Sentiment Analysis and Recommendation System

Data Science, Fall 2020

Hong-Han Shuai

Thanks to Prof. Bing Liu from UIC, Prof. Hung-yi Lee from NTU, and Prof. Dan Jurafsky from Stanford for the slides.

The NLP Research Community

- Papers

- [ACL Anthology](#) has nearly everything, free!
 - Over 36,000 papers!
 - Free-text searchable
 - Great way to learn about current research on a topic
 - New search interfaces currently available in beta
 - Find recent or highly cited work; follow citations
 - Used as a dataset by various projects
 - Analyzing the text of the papers (e.g., parsing it)
 - Extracting a graph of papers, authors, and institutions (Who wrote what? Who works where? What cites what?)

The NLP Research Community

<https://acl2020.org/blog/general-conference-statistics/#:~:text=General%20Statistics,than%20half%20of%20this%20year!>

- **Conferences**

- Most work in NLP is published as 8-page conference papers with 3 double-blind reviewers.
- Main annual conferences: ACL, EMNLP, NAACL
 - Also EACL, IJCNLP, COLING
 - + various specialized conferences and workshops
- Big events, and growing fast! ACL 2020:

	Total Submissions	Accepted	% Accepted
Total	3429	779	22.7%
Long	2244	571	25.4%
Short	1185	208	17.6%

The NLP Research Community

- **Institutions**

- **Universities:** Many have 2+ NLP faculty

- Several “big players” with many faculty
 - Some of them also have good linguistics, cognitive science, machine learning, AI

- **Companies:**

- Old days: AT&T Bell Labs, IBM
 - Now: Google, Microsoft, IBM, many startups ...
 - Speech: Nuance, ...
 - Machine translation: Language Weaver, Systran, ...
 - Many niche markets – online reviews, medical transcription, news summarization, legal search and discovery ...

The NLP Research Community

- **Software**

- Lots of people distribute code for these tasks
 - Or you can email a paper's authors to ask for their code
- Some [lists](#) of software, but no central site ☹
- Some [end-to-end pipelines](#) for text analysis
 - “One-stop shopping”
 - Cleanup/tokenization + morphology + tagging + parsing + ...
 - [NLTK](#) is easy for beginners and has a [free book](#) (intersession?)
 - [GATE](#) has been around for a long time and has a bunch of modules

The NLP Research Community

- **Software**

- To find good or popular tools:
 - Search current papers, ask around, use the web
- Still, often hard to identify the **best** tool for your job:
 - Produces appropriate, sufficiently detailed output?
 - Accurate? (on the measure you care about)
 - Robust? (accurate on your data, not just theirs)
 - Fast?
 - Easy and flexible to use? Nice file formats, command line options, visualization?
 - Trainable for new data and languages? How slow is training?
 - Open-source and easy to extend?

The NLP Research Community

- **Datasets**

- Raw text or speech corpora
 - Or just their [n-gram counts](#), for super-big corpora
 - Various languages and genres
 - Usually there's some metadata (each document's date, author, etc.)
 - Sometimes \exists licensing restrictions (proprietary or copyright data)
- Text or speech with manual or automatic annotations
 - What kind of annotations? That's the rest of this lecture ...
 - May include translations into other languages
- Words and their relationships
 - [Morphological](#), [semantic](#), translational, evolutionary
- [Grammars](#)
- [World Atlas of Linguistic Structures](#)
- Parameters of statistical models (e.g., grammar weights)

The NLP Research Community

- **Datasets**

- Read papers to find out what datasets others are using
 - [Linguistic Data Consortium](#) (searchable) hosts many large datasets
 - Many projects and competitions post data on their websites
 - But sometimes you have to email the author for a copy
- [CORPORA mailing list](#) is also good place to ask around
- [LREC Conference](#) publishes papers about new datasets & metrics
- [Amazon Mechanical Turk](#) – pay humans (very cheaply) to annotate your data or to correct automatic annotations
 - **Old task, new domain:** Annotate parses etc. on *your* kind of data
 - **New task:** Annotate something new that you want your system to find
 - **Auxiliary task:** Annotate something new that your system may benefit from finding (e.g., annotate subjunctive mood to improve translation)
- Can you make annotation so much [fun](#) or so [worthwhile](#) that they'll do it for free?

The NLP Research Community

- **Standard data formats**

- Often just simple *ad hoc* text-file formats
 - Documented in a README; easily read with scripts
- Some standards:
 - [Unicode](#) – strings in any language (see [ICU](#) toolkit)
 - PCM (.wav, .aiff) – uncompressed audio
 - BWF and AUP extend w/metadata; also many compressed formats
 - [XML](#) – documents with embedded annotations
 - [Text Encoding Initiative](#) – faithful digital representations of printed text
 - [Protocol Buffers](#), [JSON](#) – structured data
 - [UIMA](#) – “unstructured information management”; Watson uses it
- Standoff markup: raw text in one file, annotations in other files (“ \exists noun phrase from byte 378—392”)
 - Annotations can be independently contributed & distributed

The NLP Research Community

- Survey articles

- May help you get oriented in a new area
- Synthesis Lectures on Human Language Technologies
- Handbook of Natural Language Processing
- Oxford Handbook of Computational Linguistics
- Foundations & Trends in Machine Learning
- Survey articles in journals – JAIR, CL, JMLR
- ACM Computing Surveys?
- Online tutorial papers
- Slides from tutorials at conferences
- Textbooks

Sentiment Analysis and Opinion Mining

Positive or negative movie review?



- unbelievably disappointing



- Full of zany characters and richly applied satire, and some great plot twists



- this is the greatest screwball comedy ever filmed



- It was a philosophic story. I think of who I am and where I am after the movie.

Google Product Search



HP Officejet 6500A Plus e-All-in-One Color Ink-jet - Fax / copier / printer / scanner
\$89 online, \$100 nearby ★★★★★ 377 reviews
September 2010 - Printer - HP - Inkjet - Office - Copier - Color - Scanner - Fax - 250 shi

Reviews

Summary - Based on 377 reviews



What people are saying

ease of use	<div><div></div><div></div></div>	"This was very easy to setup to four computers."
value	<div><div></div><div></div></div>	"Appreciate good quality at a fair price."
setup	<div><div></div><div></div></div>	"Overall pretty easy setup."
customer service	<div><div></div><div></div></div>	"I DO like honest tech support people."
size	<div><div></div><div></div></div>	"Pretty Paper weight."
mode	<div><div></div><div></div></div>	"Photos were fair on the high quality mode."
colors	<div><div></div><div></div></div>	"Full color prints came out with great quality."

Bing Shopping

HP Officejet 6500A E710N Multifunction Printer

[Product summary](#) [Find best price](#) **Customer reviews** [Specifications](#) [Related items](#)



\$121.53 - \$242.39 (14 stores)

☐ Compare

Average rating ★★★★★ (144)



Most mentioned

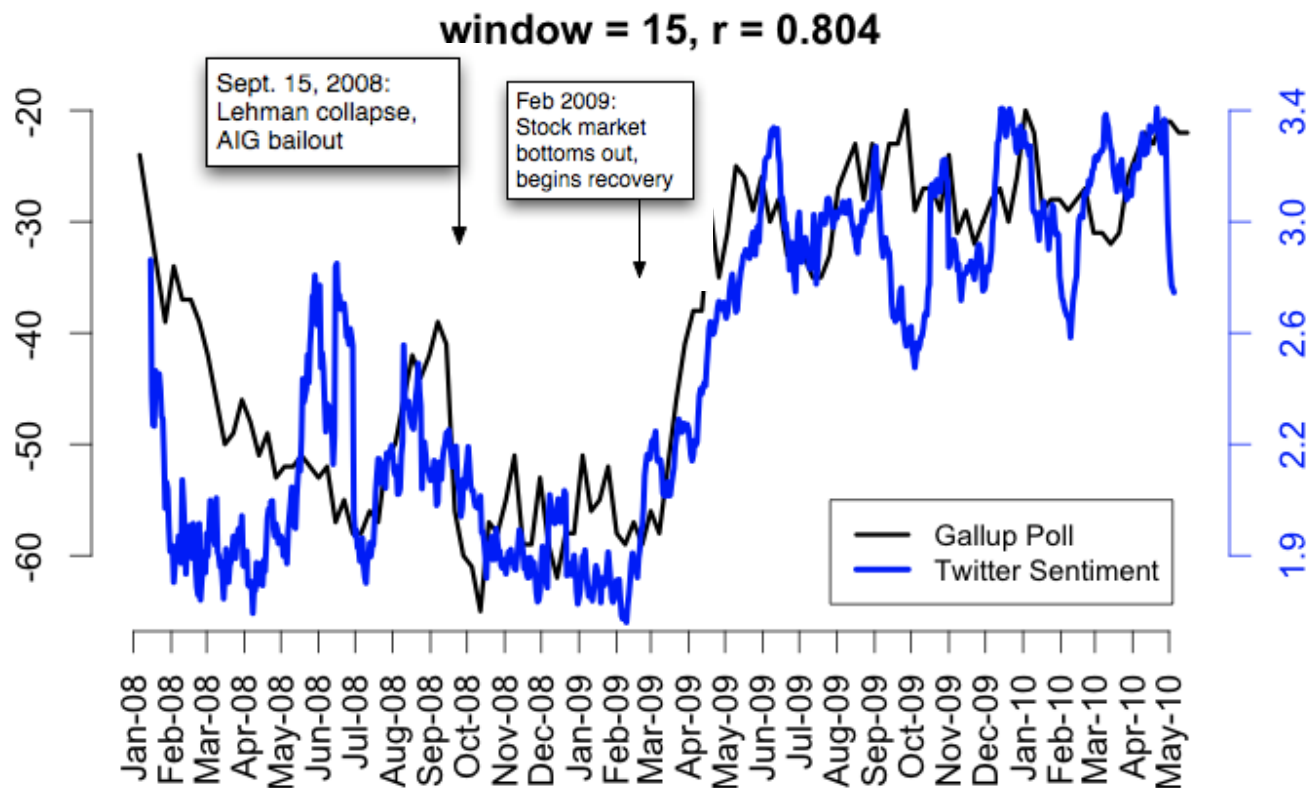


Show reviews by source

Best Buy (140)
CNET (5)
Amazon.com (3)

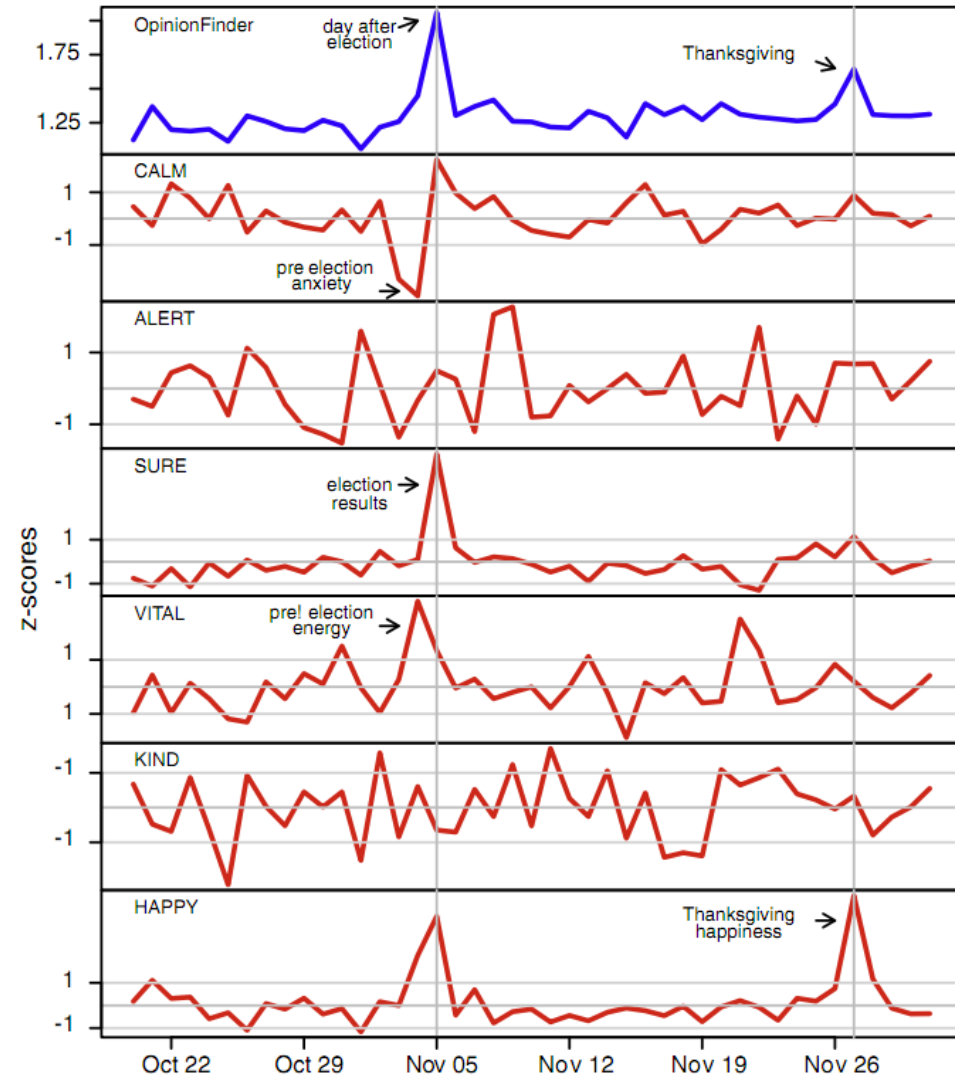
Twitter sentiment versus Gallup Poll of Consumer Confidence

Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. 2010. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In ICWSM-2010



Twitter sentiment:

Johan Bollen, Huina Mao, Xiaojun Zeng. 2011.
[Twitter mood predicts the stock market](#),
Journal of Computational Science 2:1, 1-8.
10.1016/j.jocs.2010.12.007.



Target Sentiment on Twitter

- [Twitter Sentiment App](#)

- Alec Go, Richa Bhayani, Lei Huang. 2009. Twitter Sentiment Classification using Distant Supervision

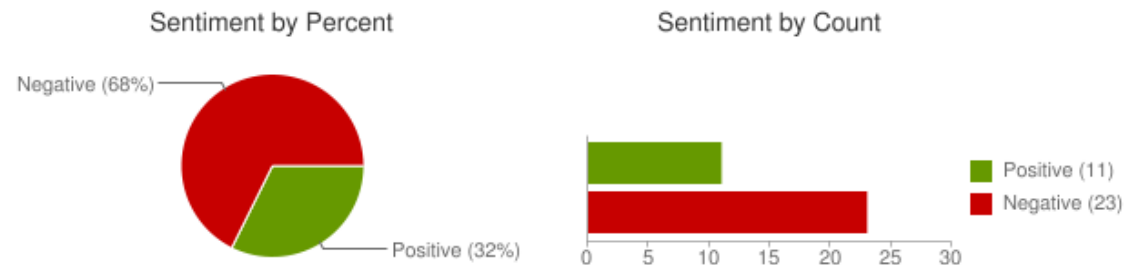
Type in a word and we'll highlight the good and the bad

"united airlines"

Search

[Save this search](#)

Sentiment analysis for "united airlines"



[iljacobson](#): OMG... Could **@United airlines** have worse customer service? W8g now 15 minutes
Posted 2 hours ago

[12345clumsy6789](#): I hate **United Airlines** Ceiling!!! Fukn impossible to get my conduit in this d...
Posted 2 hours ago

[EMLandPRGbelgiu](#): EML/PRG fly with Q8 **united airlines** and 24seven to an exotic destination
Posted 2 hours ago

[CountAdam](#): FANTASTIC customer service from **United Airlines** at XNA today. Is tweet more...
Posted 4 hours ago

Introduction

- Sentiment analysis (SA) or opinion mining
 - computational study of opinion, sentiment, appraisal, evaluation, and emotion.
- Why is it important?
 - Opinions are key influencers of our behaviors.
 - Our beliefs and perceptions of reality are conditioned on how others see the world. Whenever we need to make a decision we often seek out the opinions from others.
 - Rise of social media → opinion data
 - Rise of AI and chatbots:
 - Emotion and sentiment are key to human communication

Terms defined - Merriam-Webster

- **Sentiment**: an attitude, thought, or judgment prompted by feeling.
 - A sentiment is more of a feeling.
 - *"I am concerned about the current state of the economy."*
- **Opinion**: a view, judgment, or appraisal formed in the mind about a particular matter.
 - a concrete view of a person about something.
 - *"I think the economy is not doing well."*

SA: A fascinating problem!

- Intellectually challenging & many applications.
 - A popular research area in NLP, and data mining (Shanahan, Qu, and Wiebe, 2006 (edited book); Surveys - Pang and Lee 2008; Liu, 2006, 2012, and 2015)
 - spread from CS to management and social sciences (Hu, Pavlou, Zhang, 2006; Archak, Ghose, Ipeirotis, 2007; Liu Y, et al 2007; Park, Lee, Han, 2007; Dellarocas, Zhang, Awad, 2007; Chen & Xie 2007).
 - A large number of companies in the space globally
 - > 300 in the US alone.
- It touches every aspect of NLP & also is confined.
 - A “simple” semantic analysis problem.
- A major technology from NLP.
 - But it is hard.

Roadmap

- **Sentiment analysis problem**
- Document sentiment classification
- Sentence subjectivity & sentiment classification
- Aspect-based sentiment analysis
- Mining comparative opinions
- Opinion lexicon generation
- Some interesting sentences
- Summary

Two main types of opinions

(Jindal and Liu 2006; Liu, 2010)

- **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:** Comparison of more than one entity.
 - E.g., “iPhone is better than Blackberry.”
- We focus on regular opinions first, and just call them opinions.

(I): Definition of opinion

- **Id:** **Abc123** on **5-1-2008** -- “I bought an *iPhone* yesterday. It is such a nice *phone*. The *touch screen* is really *cool*. The *voice quality* is *great* too. It is much *better* than my *Blackberry*. However, *my mom* was *mad* with *me* as I didn’t tell her before I bought the *phone*. She thought the *phone* was too *expensive*”
- **Definition:** An *opinion* is a quadruple (Liu, 2012),
(*target*, *sentiment*, *holder*, *time*)
- This definition is concise, but not easy to use.
 - Target can be complex, e.g., “I bought an iPhone. The *voice quality* is amazing.”
 - *Target* = *voice quality*? (not quite)

A more practical definition

(Hu and Liu 2004; Liu, 2010, 2012)

- An *opinion* is a quintuple

(*entity*, *aspect*, *sentiment*, *holder*, *time*)

where

- *entity*: target entity (or object).
 - *Aspect*: aspect (or feature) of the entity.
 - *Sentiment*: +, -, or neu, a rating, or an emotion.
 - *holder*: opinion holder.
 - *time*: time when the opinion was expressed.
- *Aspect-based sentiment analysis*

Our example blog in quintuples

- **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 great 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**, ...”*
- **In quintuples**
 - (iPhone, GENERAL, +, Abc123, 5-1-2008)
 - (iPhone, touch_screen, +, Abc123, 5-1-2008)
 -
- We will discuss comparative opinions later.

(II): Opinion summary (Hu and Liu 2004)

- With a lot of opinions, a summary is necessary.
 - Not traditional text summary: from long to short.
 - Text summarization: defined operationally based on algorithms that perform the task
- Opinion summary (OS) can be defined precisely,
 - not dependent on how summary is generated.
- Opinion summary needs to be quantitative
 - 60% positive is very different from 90% positive.
- Main form of OS: *Aspect-based opinion summary*

Opinion summary

(Hu and Liu, 2004)

Aspect/feature Based Summary of opinions about iPhone:

Aspect: **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**.

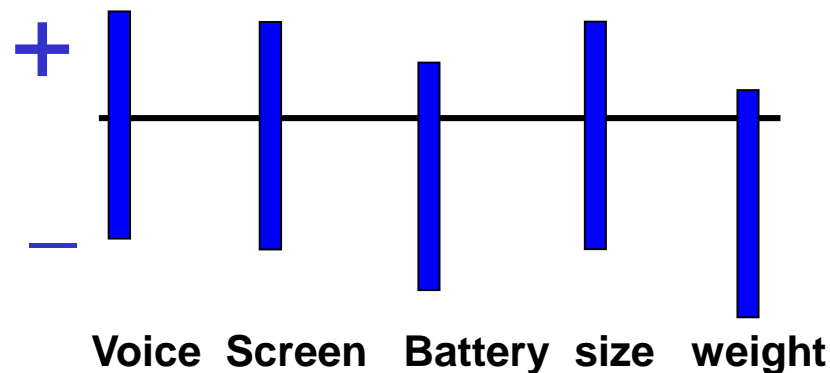
...

Aspect: **voice quality**

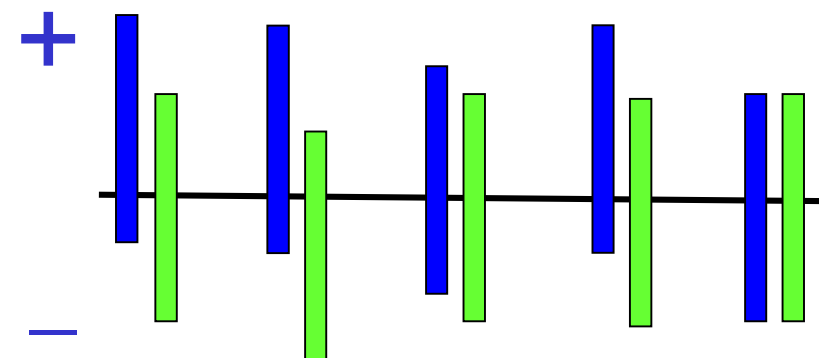
...

(Liu et al. 2005)

■ Opinion Summary of 1 phone



■ Opinion comparison of 2 phones



Aspect-based opinion summary

bing

HP printer

ALL RESULTS

Shopping

POPULAR FEATURES

- all
- Affordability
- Speed
- Print Quality
- Reliability
- Ease Of Use
- Brand
- Installation
- Size
- Compatibility

SHOPPING

HP LaserJet 1020 - prin



user reviews

user reviews

speed 96%

The quality is as good as any laserjet printer I've used and the speed is fast.
Love Reading www.amazon.com 3/17/2006 more...

Quick and fast transaction.
Arthur L. Taylor www.amazon.com 2/5/2008 more...

It's small and fast and very reliable.
Muffinhead's mom www.amazon.com 1/9/2007 more...

Google products sony camera Search Products

Sony Cyber-shot DSC-W370 14.1 MP Digital Camera (Silver)

[Overview](#) - [Online stores](#) - [Nearby stores](#) - [Reviews](#) - [Technical specifications](#) - [Similar items](#) - [Accessories](#)

 **\$140 online, \$170 nearby**

★★★★☆ 159 reviews +1 0

Reviews

Summary - Based on 159 reviews

1 2 3 stars 4 stars 5 stars

What people are saying

pictures		"We use the product to take quickly photos."
features		"Impressive panoramic feature."
zoom/lens		"It also record better and focus better on sunny days."
design		"It has the slightest grip but it's sufficient."
video		"Video zoom is choppy."
battery life		"Even better, the battery lasts long."
screen		"I Love the Sony's 3" screen which I really wanted."

view: positive comments (44)

Summarization

(AddStructure.com)



(1,043 customer reviews)

👍 Pros

- Great Price (518)
- Good Sound Quality (895)
- Easy Setup (138)

👎 Cons

- Remote (9)
- Inputs (8)
- Little product flaws (8)

"The only down side is there is no **input** to connect to a computer."



(435 customer reviews)

👍 Pros

- Great Picture Quality (256)
- Good Sound Quality (77)
- Easy Setup (60)

👎 Cons

- Speakers (5)
- Changing Channels (4)
- Volume (3)

"The only "bad" thing we have noticed is that there is quite a delay when you **change channels**."

Not just ONE problem

- (*entity*, *aspect*, *sentiment*, *holder*, *time*)
 - target *entity*: Named entity extraction, more
 - *aspect* of *entity*: Aspect extraction
 - *sentiment*: Sentiment classification
 - opinion *holder*: Information/data extraction
 - *time*: Information/data extraction
- There are more problems ...
- Other NLP problems
 - Synonym grouping (voice = sound quality)
 - Coreference resolution
 -

Reason for Opinion/Sentiment

- **Definition:** A reason for an opinion is the justification or explanation of the opinion.
- There are two main cases, e.g.,
- (1). “*I hate this car as it eats too much gas.*”
 - Negative about an entity due to a bad aspect.
 - This can be identified by negative aspects
- (2). “*This car is too small.*” (Shuai et al. 2016)
 - Negative about an aspect because of a reason
 - This can be identified by aspect specific sentiment

Qualifier of Opinion

- **Definition:** A qualifier of an opinion limits or modifies the meaning of the opinion.
- It tells what an opinion is good for, e.g.,
 - *“This car is too small **for a tall person.**”*
 - *“The picture quality of **night shots** is bad”*
- Not every opinion comes with an explicit reason and/or an explicit qualifier.
 - No reason and no qualifier, e.g.,
 - “This car is bad.”

Roadmap

- Sentiment analysis problem
- **Document sentiment classification**
- Sentence subjectivity & sentiment classification
- Aspect-based sentiment analysis
- Mining comparative opinions
- Opinion lexicon generation
- Some interesting sentences
- Summary

Sentiment classification

- **Classify a whole opinion document** (e.g., a review) based on the overall sentiment of the opinion holder (Pang et al 2002; Turney 2002)
 - **Classes**: Positive, negative (possibly neutral)
- **An example review**:
 - *“I bought an iPhone a few days ago. It is such a nice phone, although a little large. The touch screen is cool. The voice quality is great too. I simply love it!”*
 - **Classification**: positive or negative?
- **It is basically a text classification problem**

Assumption and goal

- **Assumption**: The doc is written by a single person and express opinion/sentiment on a single entity.
- **Reviews usually satisfy the assumption.**
 - Almost all research papers use reviews
 - Positive: 4 or 5 stars, negative: 1 or 2 stars
- Forum postings and blogs do not
 - They may mention and compare multiple entities
 - Many such postings express no sentiments

Supervised learning (Pang et al, 2002)

- Directly apply supervised learning techniques to classify reviews into positive and negative.
- Three classification techniques were tried:
 - Naïve Bayes, Maximum Entropy, Support Vector Machines (SVM)
- Features: negation tag, unigram (single words), bigram, POS tag, position.
- SVM did the best based on movie reviews.

Features for supervised learning

- The problem has been studied by numerous researchers.
- **Key:** feature engineering. A large set of features have been tried by researchers. E.g.,
 - Terms frequency and different IR weighting schemes
 - Part of speech (POS) tags
 - Opinion words and phrases
 - Negations
 - Syntactic dependency

Lexicon-based approach (Taboada *et al.* (2011))

- Using a set of sentiment terms, called the **sentiment lexicon**
 - Positive words: great, beautiful, amazing, ...
 - Negative words: bad, terrible awful, unreliable, ...
- The SO value for each sentiment term is assigned a value from $[-5, +5]$.
 - Consider *negation*, *intensifier* (e.g., very), and *diminisher* (e.g., barely)
- Decide the sentiment of a review by aggregating scores from all sentiment terms

Deep learning

- Recently, deep neural networks have been used for sentiment classification. E.g.,
 - Socher et al (2013) used deep learning to work on the sentence parse tree based on words/phrases compositionality in the framework of distributional semantics
 - Many papers ...
 - Also related
 - Irsoy and Cardie (2014) extract opinion expressions
 - Xu, Liu and Zhao (2014) identify opinion & target relations

Aspect Sentiment Classification with Document-level Sentiment Preference Modeling (ACL'20)

Inter-Aspect Sentiment Tendency

Document 2:

S1: If you've ever been along with the river in Weehawken you have an idea of the top of view the chart house has to offer.

- Category = *LOCATION#GENERAL*, polarity = *positive*

S2: Add to that great service and great food at a reasonable price and you have yourself the beginning of a great evening.

- Category = *SERVICE#GENERAL*, polarity = *positive*

- Category = *FOOD#QUALITY*, polarity = *positive*

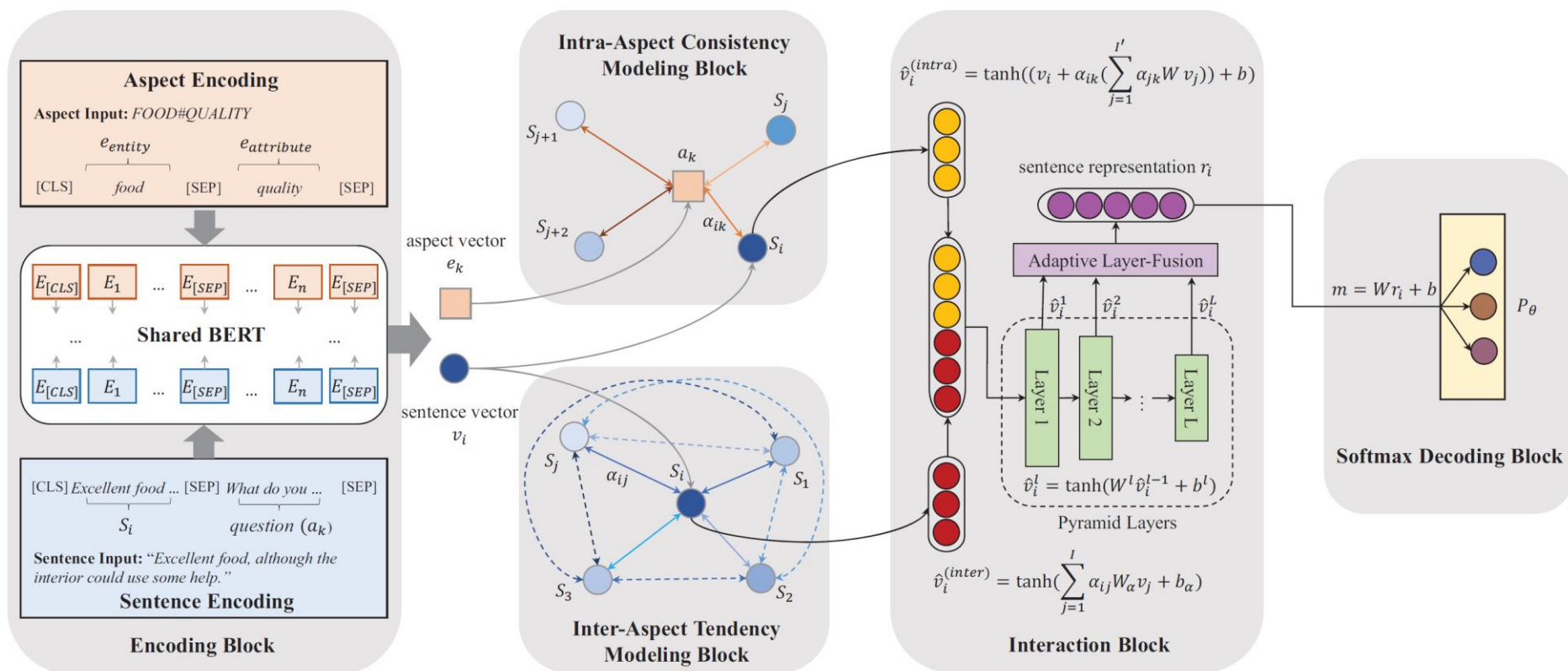
- Category = *FOOD#PRICES*, polarity = *positive*

S3: The lava cake dessert was incredible and I recommend it.

- Category = *FOOD#QUALITY*, polarity = *positive*



Architecture



Review rating prediction

- Apart from classification of positive or negative sentiments,
 - research has also been done to **predict the rating scores** (e.g., 1–5 stars) of reviews (Pang and Lee, 2005; Liu and Seneff 2009; Qu, Ifrim and Weikum 2010; Long, Zhang and Zhu, 2010).
 - Training and testing are reviews with star ratings.
- **Formulation:** The problem is formulated as regression since the rating scores are ordinal.
- Again, feature engineering and model building.

Roadmap

- Sentiment analysis problem
- Document sentiment classification
- **Sentence subjectivity & sentiment classification**
- Aspect-based sentiment analysis
- Mining comparative opinions
- Opinion lexicon generation
- Some interesting sentences
- Summary

Sentence sentiment analysis

- Usually consist of two steps
 - Subjectivity classification (Wiebe et al 1999)
 - To identify subjective sentences
 - Sentiment classification of subjective sentences
 - Into two classes, positive and negative
- But bear in mind
 - Many objective sentences can imply sentiments
 - Many subjective sentences do not express positive or negative sentiments/opinions
 - E.g., "I believe he went home yesterday."

Assumption

- **Assumption**: Each sentence is written by a single person and expresses a single positive or negative opinion/sentiment.
- True for simple sentences, e.g.,
 - “I like this car”
- But not true for many compound and “complex” sentences, e.g.,
 - “I like the picture quality but battery life sucks.”
 - “Apple is doing very well in this poor economy.”

Subjectivity and sentiment classification

(Yu and Hazivassiloglou, 2003)

- **Subjective sentence identification**: a few methods were tried, e.g.,
 - Sentence similarity.
 - Naïve Bayesian classification.
- **Sentiment classification (positive, negative or neutral)** (also called **polarity**): it uses a similar method to (Turney, 2002), but
 - with more seed words (rather than two) and based on log-likelihood ratio (LLR).
 - For classification of each word, it takes the average of LLR scores of words in the sentence and use cutoffs to decide positive, negative or neutral.

Roadmap

- Sentiment analysis problem
- Document sentiment classification
- Sentence subjectivity & sentiment classification
- **Aspect-based sentiment analysis**
- Mining comparative opinions
- Opinion lexicon generation
- Some interesting sentences
- Summary

We need to go further

- Sentiment classification at both the document and sentence (or clause) levels are useful, but
 - They do not find what people liked and disliked.
- They do not identify the targets of opinions, i.e.,
 - Entities and their aspects
 - Without knowing targets, opinions are of limited use.
- We need to go to the entity and aspect level.
 - *Aspect-based opinion mining and summarization* (Hu and Liu 2004).
 - We thus need the full opinion definition.

Recall the opinion definition

(Hu and Liu 2004; Liu, 2010, 2012)

- An *opinion* is a quintuple

(*entity*, *aspect*, *sentiment*, *holder*, *time*)

where

- *entity*: target entity (or object).
 - *Aspect*: aspect (or feature) of the entity.
 - *Sentiment*: +, -, or neu, a rating, or an emotion.
 - *holder*: opinion holder.
 - *time*: time when the opinion was expressed.
- *Aspect-based sentiment analysis*

Aspect extraction

- **Goal:** Given an opinion corpus, extract all aspects
- Four main approaches:
 - (1) Finding frequent nouns and noun phrases
 - (2) Exploiting opinion and target relations
 - (3) Supervised learning
 - (4) Topic modeling

(1) Frequent nouns and noun phrases

(Hu and Liu 2004)

- Nouns (NN) that are frequently mentioned are likely to be true **aspects** (frequent aspects).
- **Why?**
 - Most aspects are nouns or noun phrases
 - When product aspects/features are discussed, the words they use often converge.
 - Those frequent ones are usually the main aspects that people are interested in.

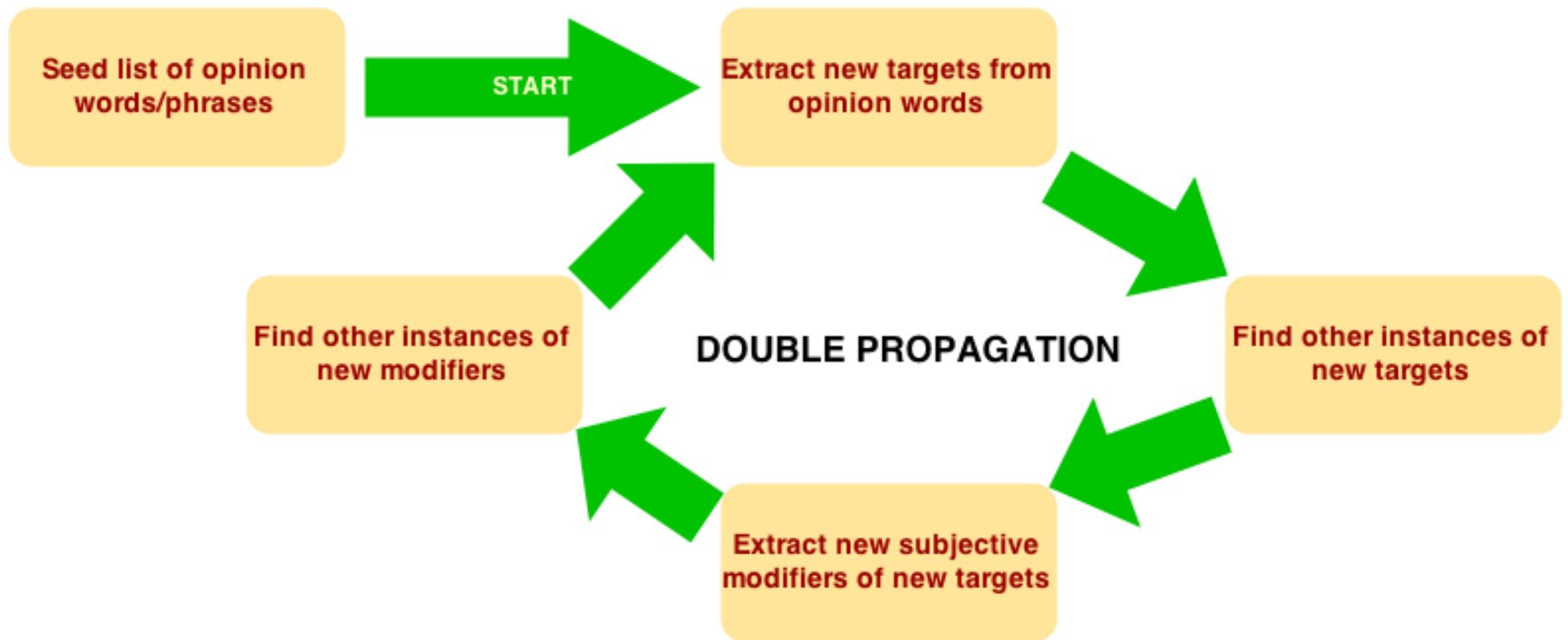
(2) Exploiting opinion & target relation

- **Key idea:** **opinions have targets**, i.e., opinion terms are used to modify aspects and entities.
 - “The pictures are absolutely **amazing**.”
 - “This is an **amazing** software.”
- The syntactic relation is approximated with the **nearest** noun phrases to the opinion word in (Hu and Liu 2004).
- The idea was generalized to
 - **syntactic dependency** in (Zhuang et al 2006)
 - **double propagation** in (Qiu et al 2009). A similar idea also in (Wang and Wang 2008)

Extract aspects using DP (Qiu et al. 2009; 2011)

- *Double propagation (DP)*
 - Based on the definition earlier, **an opinion should have a target**, entity or aspect.
- Use dependency of opinions & aspects to extract both aspects & opinion words.
 - Knowing one helps find the other.
 - E.g., “The **rooms** are **spacious**”
- It extracts both aspects and opinion words.
 - A domain independent method.

The DP method again



Explicit and implicit aspects

(Hu and Liu, 2004)

- **Explicit aspects:** Aspects explicitly mentioned as nouns or noun phrases in a sentence
 - “The **picture quality** is of this phone is great.”
- **Implicit aspects:** Aspects not explicitly mentioned in a sentence but are implied
 - “This car is so **expensive**.”
 - “This phone will not easily **fit in a pocket**.”
 - “Included **16MB** is stingy.”
- Some work has been done (Su et al. 2009; Hai et al 2011)

(3) Using supervised learning

- Using sequence labeling methods such as
 - **Hidden Markov Models** (HMM) (Jin and Ho, 2009)
 - **Conditional Random Fields** (Jakob and Gurevych, 2010).
 - Other supervised or partially supervised learning.
- (Liu, Hu and Cheng 2005; Kobayashi et al., 2007; Li et al., 2010; Choi and Cardie, 2010; Yu et al., 2011; Fang and Huang, 2012).

Identify aspect synonyms (Carenini et al 2005)

- Once aspect expressions are discovered, group them into aspect categories.
 - E.g., power usage and battery life are the same.
- **Method**: based on some similarity metrics, but it needs a taxonomy of aspects.
 - **Mapping**: The system maps each discovered aspect to an aspect node in the taxonomy.
 - **Similarity metrics**: string similarity, synonyms and other distances measured using WordNet.

(4) Topic Modeling

- Aspect extraction has two tasks:
 - (1) extract aspect expressions
 - (2) cluster them (same: “picture,” “photo,” “image”)
- Top models such as pLSA (Hofmann 1999) and LDA (Blei et al 2003) perform both tasks at the same time. A topic is basically an aspect.
 - A document is a distribution over topics
 - A topic is a distribution over terms/words, e.g.,
 - {*price, cost, cheap, expensive, ...*}
 - Ranked based on probabilities (not shown).

Many Related Models and Papers

- Use topic models to model aspects.
- Jointly model both aspects and sentiments
- Knowledge-based modeling: Unsupervised models are often insufficient
 - Not producing coherent topics/aspects
 - To tackle the problem, *knowledge-based topic models* have been proposed
 - Guided by user-specified prior domain knowledge.
 - Seed terms or constraints

Aspect sentiment classification

- For each aspect, identify the sentiment about it
- Work based on sentences, but also consider,
 - A sentence can have multiple aspects with different opinions.
 - E.g., The battery life and picture quality are *great* (+), but the view *finder* is *small* (-).
- Almost all approaches make use of **opinion words and phrases**. But notice:
 - Some opinion words have **context independent orientations**, e.g., “good” and “bad” (almost)
 - Some other words have **context dependent orientations**, e.g., “long,” “quiet,” and “sucks” (+ve for vacuum cleaner)

Roadmap

- Sentiment analysis problem
- Document sentiment classification
- Sentence subjectivity & sentiment classification
- Aspect-based sentiment analysis
- **Mining comparative opinions**
- Opinion lexicon generation
- Some interesting sentences
- Summary

Comparative Opinions

(Jindal and Liu, 2006)

- *Gradable*

- *Non-Equal Gradable*: Relations of the type *greater or less than*
 - “*The sound of phone A is better than that of phone B*”
- *Equative*: Relations of the type *equal to*
 - “*Camera A and camera B both come in 7MP*”
- *Superlative*: Relations of the type *greater or less than all others*
 - “*Camera A is the cheapest in market*”

Analyzing Comparative Opinions

- **Objective:** Given an opinionated document d , **Extract comparative opinions:**

$(E_1, E_2, A, po, h, t),$

E_1 and E_2 ; entity sets being compared

A : their shared aspects - the comparison is based on

po : preferred entity set

h : opinion holder

t : time when the comparative opinion is posted.

- **Note:** not positive or negative opinions.

An example

- Consider the comparative sentence
 - “*Canon’s optics is better than those of Sony and Nikon.*”
 - Written by John in 2010.
- The extracted comparative opinion/relation:
 - ({Canon}, {Sony, Nikon}, {optics}, *preferred*:{Canon}, John, 2010)

Common comparatives

- In English, comparatives are usually formed by adding *-er* and superlatives are formed by adding *-est* to their **base adjectives** and **adverbs**
- Adjectives and adverbs with two syllables or more and not ending in *y* do not form comparatives or superlatives by adding *-er* or *-est*.
 - Instead, *more*, *most*, *less*, and *least* are used before such words, e.g., *more beautiful*.
- Irregular comparatives and superlatives, i.e., *more*, *most*, *less*, *least*, *better*, *best*, *worse*, *worst*, etc

Some techniques (Jindal and Liu, 2006, Ding et al, 2009)

- Identify comparative sentences
 - Supervised learning
- Extraction of different items
 - Label sequential rules
 - Conditional random fields (CRF)
- Determine preferred entities (opinions)
 - Lexicon-based methods: Parsing and opinion lexicon
- (Yang and Ko, 2011) is similar to (Jindal and Liu 2006)

Analysis of comparative opinions

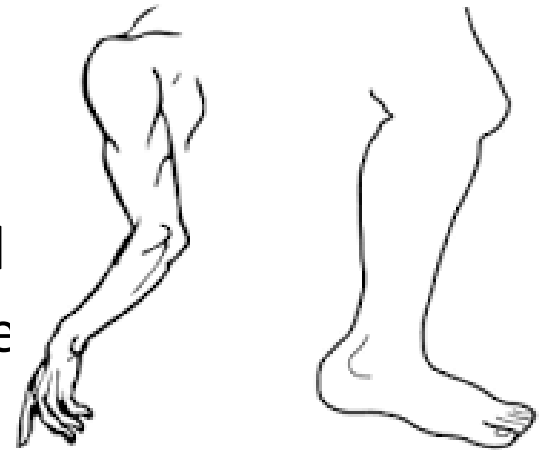
- Gradable comparative sentences can be dealt with *almost* as normal opinion sentences.
 - E.g., “*optics of camera A is better than that of camera B*”
 - **Positive:** (camera A, *optics*)
 - **Negative:** (camera B, *optics*)
- **Difficulty:** recognize non-standard comparatives
 - E.g., “*I am so happy because my new iPhone is *nothing like* my old slow ugly Droid.*”

Roadmap

- Sentiment analysis problem
- Document sentiment classification
- Sentence subjectivity & sentiment classification
- Aspect-based sentiment analysis
- Mining comparative opinions
- **Opinion lexicon generation**
- Some interesting sentences
- Summary

Sentiment (or opinion) lexicon

- **Sentiment lexicon**: lists of words and expressions used to express people's subjective feelings and sentiments/opinions.
 - Not just individual words, but also phrases and idioms, e.g., "cost an arm and a leg"
- They are instrumental for sentiment analysis.
- There seems to be endless variety of bearing expressions.
 - We have compiled more than 6,700 ind
 - There are also a large number of phrase



Sentiment lexicon

- **Sentiment words or phrases** (also called polar words, opinion bearing words, etc). E.g.,
 - **Positive**: beautiful, wonderful, good, amazing,
 - **Negative**: bad, poor, terrible, 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, only an one-time effort
 - **Corpus-based approach**
 - **Dictionary-based approach**

Corpus-based approaches

- **Rely on syntactic patterns in large corpora.** (Hazivassiloglou and McKeown, 1997; Turney, 2002; Yu and Hazivassiloglou, 2003; Kanayama and Nasukawa, 2006; Ding, Liu and Yu, 2008)
 - **Can find domain dependent orientations** (positive, negative, or neutral).
- (Turney, 2002) and (Yu and Hazivassiloglou, 2003) are similar.
 - Assign opinion orientations (polarities) to words/phrases.
 - (Yu and Hazivassiloglou, 2003) is slightly different from (Turney, 2002)
 - use more seed words (rather than two) and use log-likelihood ratio.

Corpus-based approaches (contd)

- **Sentiment consistency:** Use conventions on connectives to identify opinion words (Hazivassiloglou and McKeown, 1997).
E.g.,
 - **Conjunction:** conjoined adjectives usually have the same orientation.
 - E.g., “This car is *beautiful* **and** *spacious*.” (conjunction)
 - AND, OR, BUT, EITHER-OR, and NEITHER-NOR have similar constraints.
- **Learning using**
 - **log-linear model:** determine if two conjoined adjectives are of the same or different orientations.
 - **Clustering:** produce two sets of words: positive and negative

Context dependent opinion

- **Find domain opinion words is insufficient.** A word may indicate different opinions in same domain.
 - “The battery life is *long*” (+) and “It takes a *long* time to focus” (-).
- Ding, Liu and Yu (2008) and Ganapathibhotla and Liu (2008) exploited sentiment consistency (both inter and intra sentence) based on contexts
 - **It finds context dependent opinions.**
 - **Context:** (adjective, aspect), e.g., (long, battery_life)
 - It assigns an opinion orientation to the pair.

The Double Propagation method

(Qiu et al 2009, 2011)

- The same DP method can also use dependency of opinions & aspects to extract new opinion words.
- Based on dependency relations
 - Knowing an aspect can find the opinion word that modifies it
 - E.g., “The **rooms** are **spacious**”
 - Knowing some opinion words can find more opinion words
 - E.g., “The **rooms** are **spacious** and **beautiful**”

Dictionary-based methods

- Typically use WordNet's synsets and hierarchies to acquire opinion words
 - Start with a small seed set of opinion words.
 - Bootstrap the set by searching for synonyms and antonyms in WordNet iteratively (Hu and Liu, 2004; Kim and Hovy, 2004; Kamps et al 2004).

Semi-supervised learning

(Esuti and Sebastiani, 2005)

- Use supervised learning
 - Given two seed sets: positive set P , negative set N
 - The two seed sets are then expanded using synonym and antonymy relations in an online dictionary to generate the expanded sets P' and N' .
- P' and N' form the training sets.
- Using all the glosses in a dictionary for each term in $P' \cup N'$ and converting them to a vector
- Build a binary classifier
 - Tried various learners.

Which approach to use?

- Both corpus and dictionary based approaches are needed.
- Dictionary usually does not give domain or context dependent meaning
 - Corpus is needed for that
- Corpus-based approach is hard to find a very large set of opinion words
 - Dictionary is good for that
- In practice, corpus, dictionary and manual approaches are all needed.

Roadmap

- Sentiment analysis problem
- Document sentiment classification
- Sentence subjectivity & sentiment classification
- Aspect-based sentiment analysis
- Mining comparative opinions
- Opinion lexicon generation
- **Some interesting sentences**
- Summary

Some interesting sentences

- “Trying out Chrome because Firefox keeps crashing.”
 - Firefox - negative; no opinion about chrome.
 - We need to segment the sentence into clauses to decide that “crashing” only applies to Firefox(?).
- But how about these
 - “I changed to Audi because BMW is so expensive.”
 - “I did not buy BWM because of the high price.”
 - “I am so happy that my iPhone is nothing like my old ugly Droid.”

Some interesting sentences (contd)

- The following two sentences are from reviews in the paint domain.
 - “For paintX, one coat can cover the wood color.”
 - “For paintY, we need three coats to cover the wood color.”
- We know that paintX is good and paintY is not, but how, by a system.

Some interesting sentences (contd)

- Conditional sentences are hard to deal with (Narayanan et al. 2009)
 - “If I can find a good camera, I will buy it.”
 - But conditional sentences can have opinions
 - “If you are looking for a good phone, buy Nokia”
- Questions are also hard to handle
 - “Are there any great perks for employees?”
 - “Any idea how to fix this lousy Sony camera?”

Some interesting sentences (contd)

- Sarcastic sentences
 - “What a great car, it stopped working in the second day.”
- Sarcastic sentences are common in political blogs, comments and discussions.
 - They make political opinions difficult to handle
- Some initial work by (Tsur, et al. 2010)

Some more interesting sentences

- “My goal is to get a tv with good picture quality”
- “The top of the picture was brighter than the bottom.”
- “When I first got the airbed a couple of weeks ago it was wonderful as all new things are, however as the weeks progressed I liked it less and less.”
- “Google steals ideas from Bing, Bing steals market shares from Google.”

Opinion mining is hard!

- “This past Saturday, I bought a *Nokia* phone and my girlfriend bought a *Motorola* phone with *Bluetooth*. We called each other when we got home. *The voice on my phone was not so clear, worse than my previous Samsung phone. The battery life was short too. My girlfriend was quite happy with her phone. I wanted a phone with good sound quality. So my purchase was a real disappointment. I returned the phone yesterday.*”

Roadmap

- Sentiment analysis problem
- Document sentiment classification
- Sentence subjectivity & sentiment classification
- Aspect-based sentiment analysis
- Mining comparative opinions
- Opinion lexicon generation
- Some interesting sentences
- **Summary**

Summary

- This chapter presented
 - The problem of sentiment analysis
 - It provides a structure to the unstructured text.
 - Main research directions and their representative techniques.
- Still many problems not attempted or studied.
- None of the subproblems is solved.

Summary (contd)

- It is a fascinating NLP or text mining problem.
 - Every sub-problem is highly challenging.
 - But it is also restricted (semantically).
- Despite the challenges, applications are flourishing!
 - It is useful to every organization and individual.
- The general NLU is probably too hard, but can we solve this highly restricted problem?
 - We have a good chance.

Recommendation System

Purchasing Behavior Model, PBM

- Input:

- User statistics from different sources
- Location/Trajectories
- Users' posts
- Browsing histories
- Time
- Payment history

- Output:

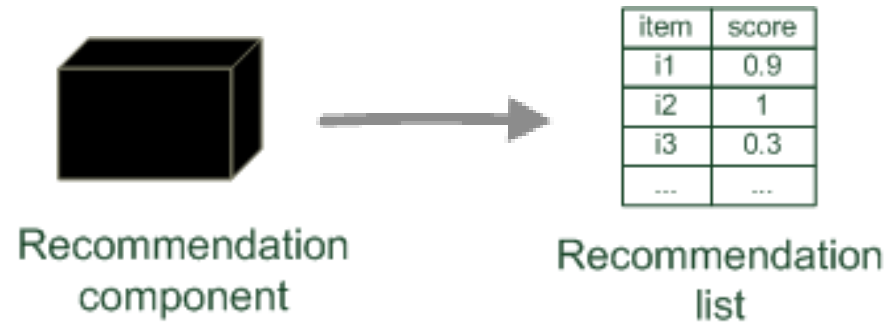
- Purchasing behavior model for each user

Three Challenges

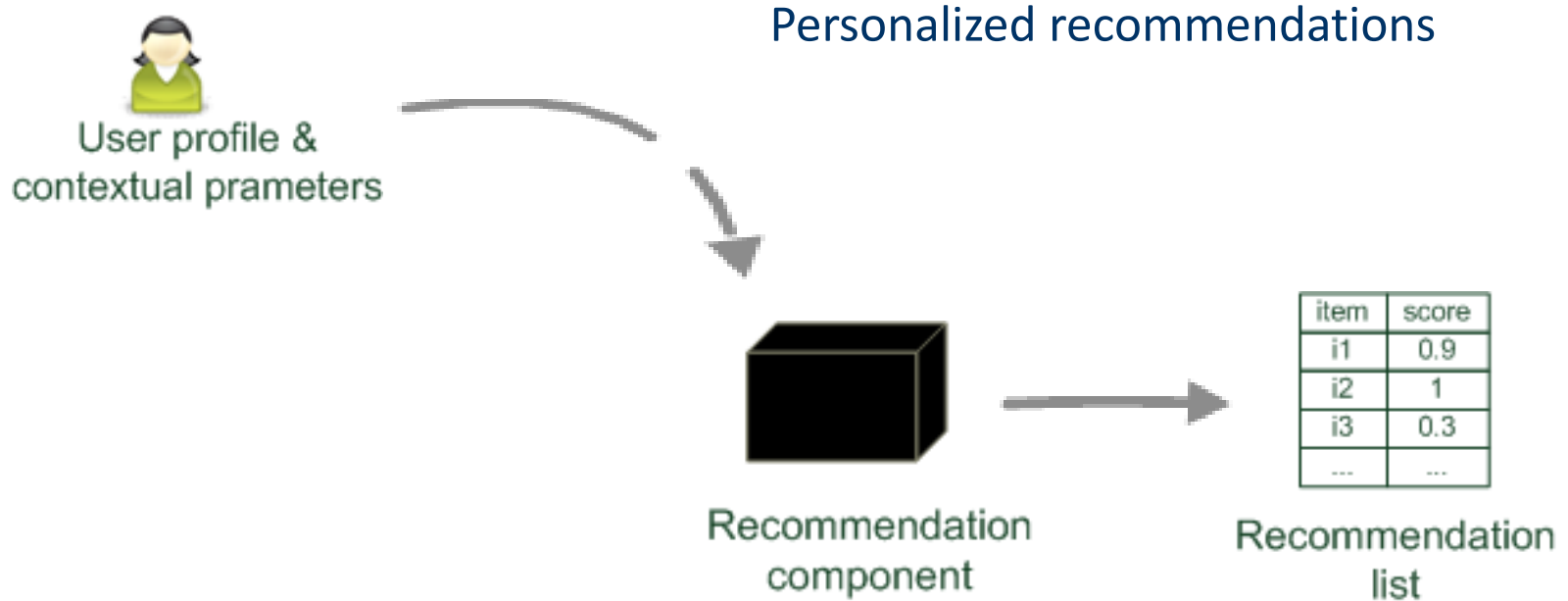
- Cold start problem
 - New users or new shops
- Implicit feedback
 - Has never seen vs. doesn't like
- Multi-Source data integration
 - Heterogeneous data sources
 - Curse of dimensionality

Paradigms of recommender systems

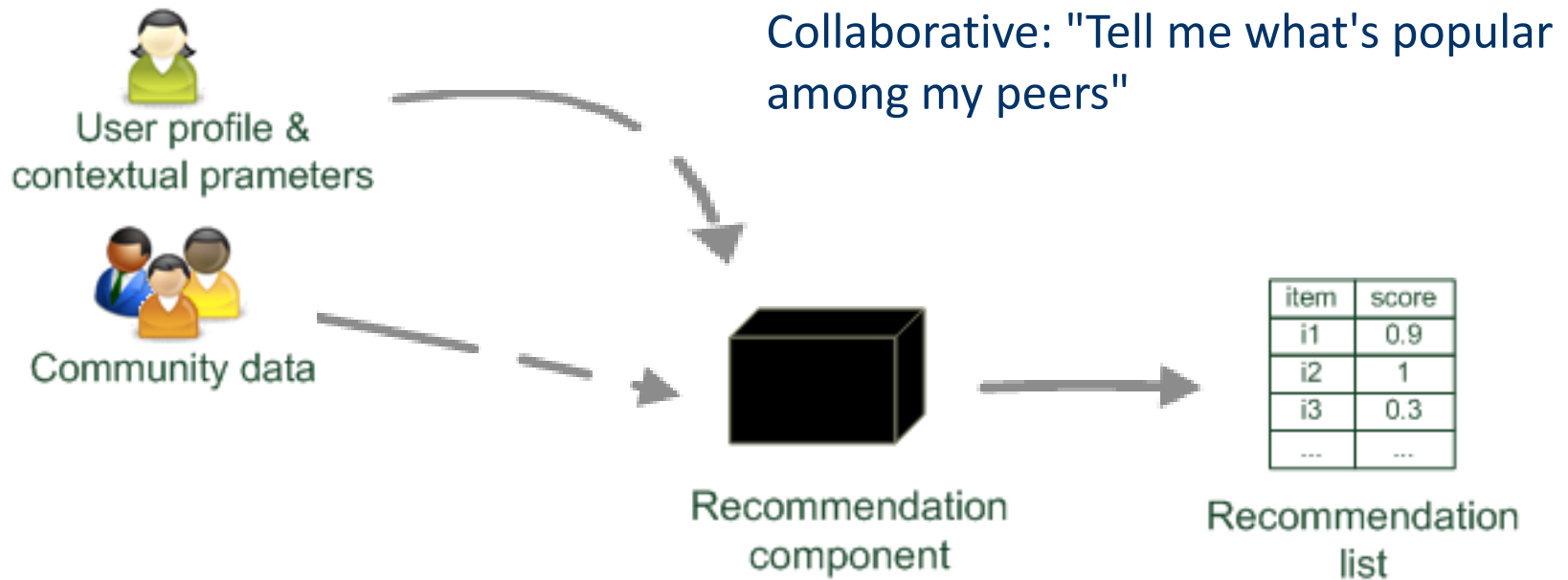
Recommender systems reduce information overload by estimating relevance



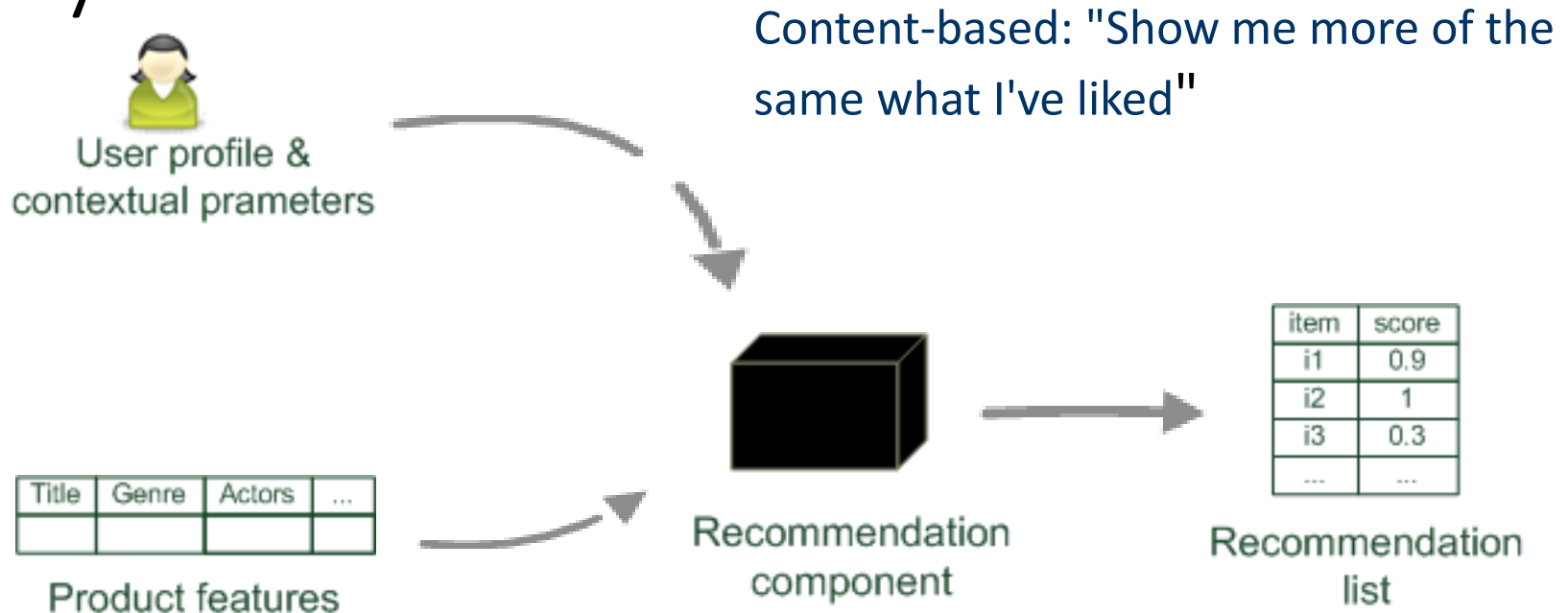
Paradigms of recommender systems



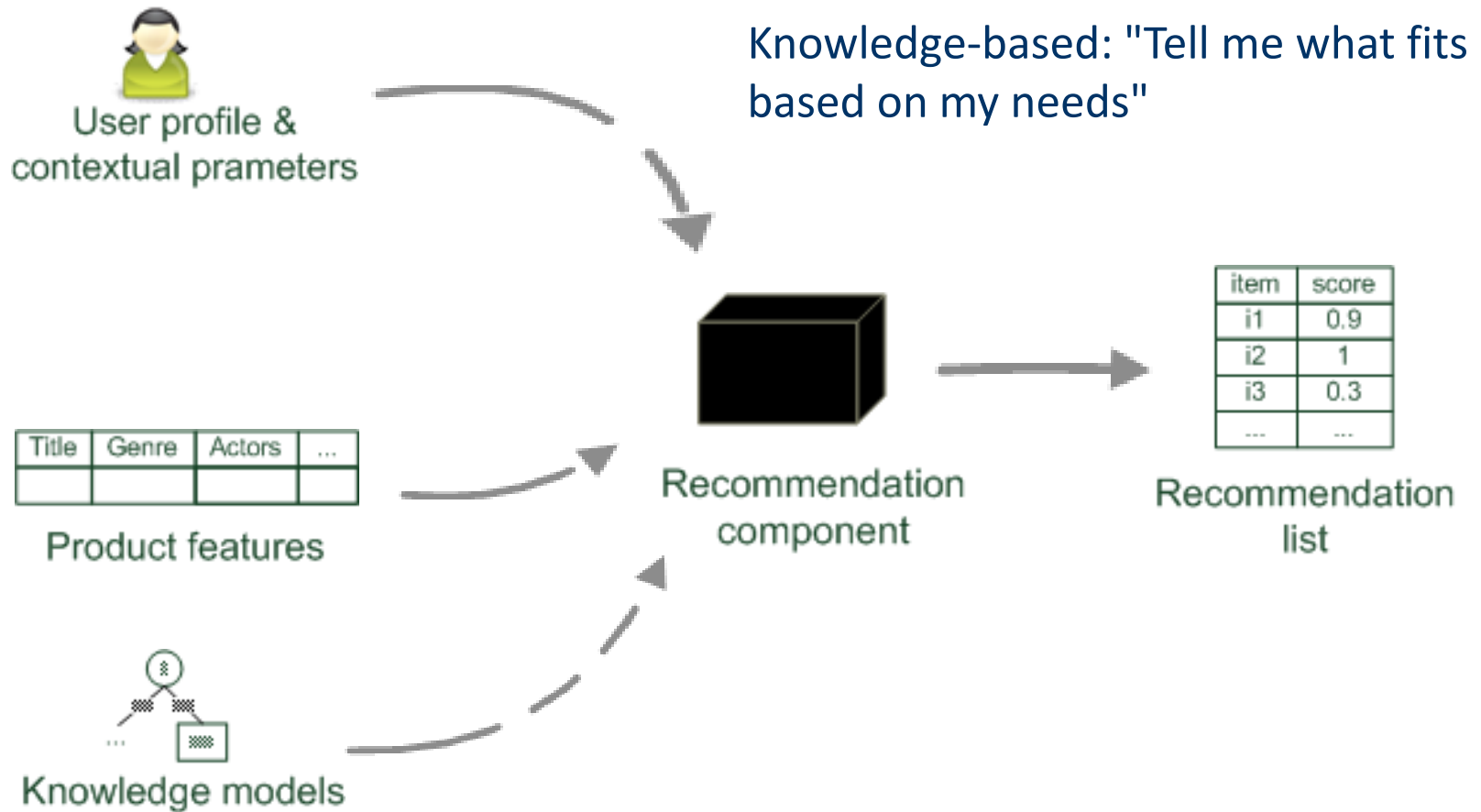
Paradigms of recommender systems



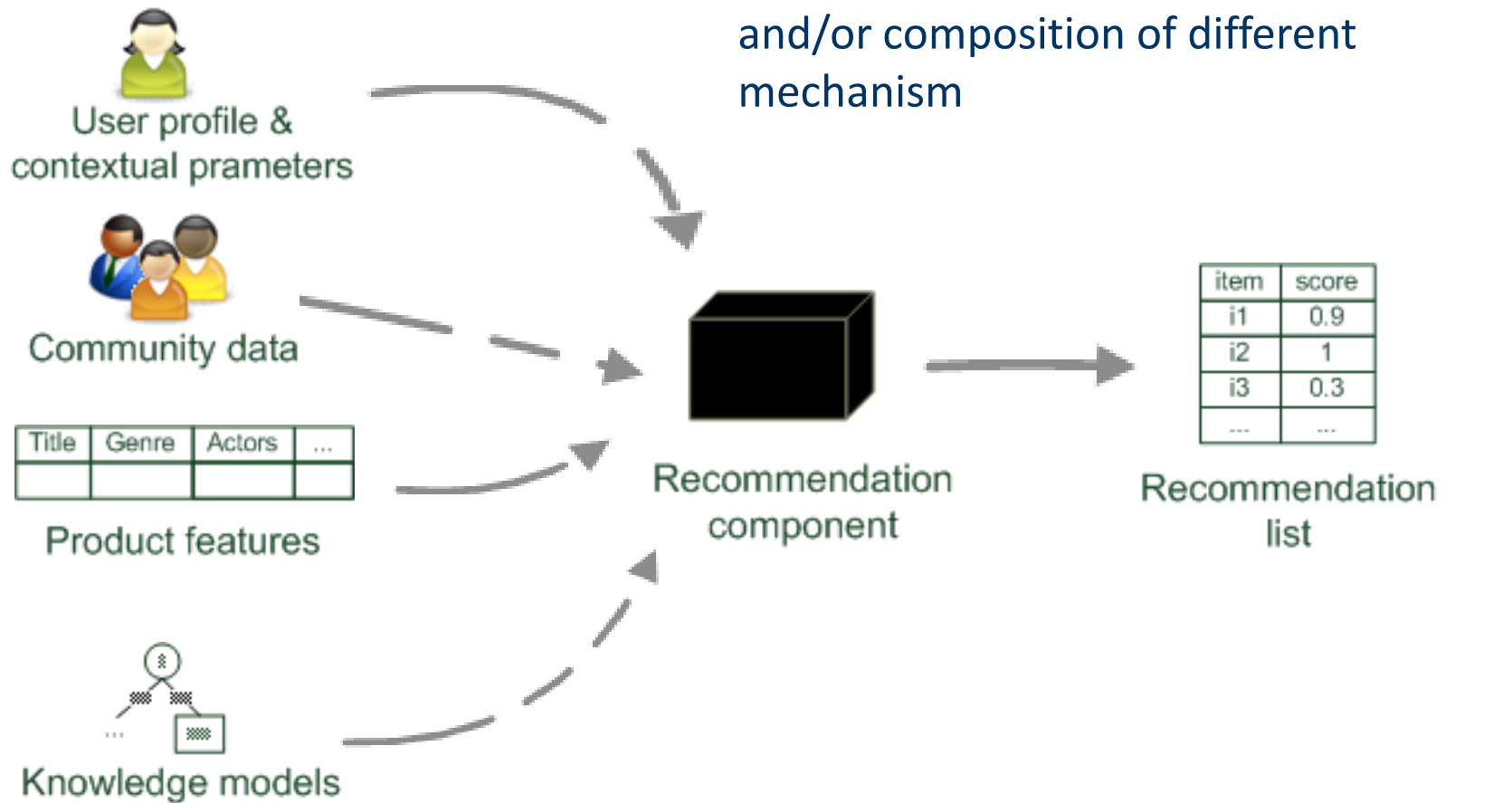
Paradigms of recommender systems



Paradigms of recommender systems



Paradigms of recommender systems





2018

2018 Q1 Youtube網紅影響力排行榜		類型	粉絲熱度
1	小玉	實驗開箱	954298.6
2	放火 Louis	生活搞笑	874356.3
3	人生肥宅 x尊	實驗開箱	854748.6
4	黃阿瑪的後宮生活	寵物影片	695482.5
5	眾量級 CROWD	情侶搞笑	581814
6	蔡阿嘎	搞笑影片	567327.9
7	聖結石Saint	生活輕鬆	559502.8
8	安啾咪	手作開箱	517028.8
9	千千進食中	美食分享	504166.1
10	蔡阿嘎 Life	生活花絮	499762.4

排名	KOL 名稱	FB 粉絲數	FB 平均 互動率	FB 平均觀 看率	IG 粉絲數	IG 平均 互動率	IG 平均 觀看率	YT 粉絲數	YT 平均 互動率	YT 平均 觀看率
1	 <u>蔡阿嘎</u>	1,916,383	1.84%	22.43%	1,497,760	4.86%	16.95%	2,440,000	0.86%	22.14%
2	 <u>這群人</u>	2,044,306	0.37%	19.51%	721,923	2.35%	13.22%	3,230,000	1.51%	75.03%
3	 <u>Nico品筠&Kim京燁【那對夫妻】</u>	2,701,001	1.12%	22.00%	693,922	2.61%	16.70%	592,000	-	-
4	 <u>486先生</u>	907,901	0.28%	6.99%	-	-	-	31,300	0.18%	19.28%
5	 <u>黃阿瑪的後宮生活</u>	1,391,691	1.07%	17.25%	989,424	5.81%	-	1,460,000	1.72%	55.91%
6	 <u>館長</u>	1,279,642	2.09%	18.12%	621,726	4.29%	22.32%	939,000	0.23%	10.70%
7	 <u>蔡桃貴</u>	277,438	8.17%	68.55%	1,340,237	9.23%	35.12%	808,000	1.65%	62.74%

Recommender systems: basic techniques

	Pros 	Cons 
Collaborative	No knowledge-engineering effort, serendipity of results, learns market segments	Requires some form of rating feedback, cold start for new users and new items
Content-based	No community required, comparison between items possible	Content descriptions necessary, cold start for new users, no surprises
Knowledge-based	Deterministic recommendations, assured quality, no cold-start, can resemble sales dialogue	Knowledge engineering effort to bootstrap, basically static, does not react to short-term trends

Related Work

- Collaborative filtering
- Matrix factorization
- Deep learning
- Clustering

Collaborative filtering

The CF Ingredients

- List of **m Users** and a list of **n Items**
- Each user has a **list of items** with associated **opinion**
 - **Explicit opinion** - a rating score
 - Sometime the rating is **implicitly** – purchase records or listen to tracks
- **Active user** for whom the CF prediction task is performed
- **Metric** for measuring **similarity between users**
- Method for selecting a subset of **neighbors**
- Method for **predicting a rating** for items not currently rated by the active user.




Collaborative Filtering

- The basic steps:

1. Identify set of ratings for the **target/active user**
2. Identify set of users most similar to the target/active user according to a similarity function (**neighborhood** formation)
3. Identify the products these similar users liked
4. **Generate a prediction** - rating that would be given by the target user to the product - for each one of these products
5. Based on this predicted rating recommend a set of top N products

User-based CF

Example

	4	5	6	7	8	9	
							$\text{sim}(u,v)$
	2		2	4	5		NA
	5		4			1	0.87
			5		2		
		1		5		4	
			4			2	
	4	5		1			NA

User-based CF

Example



2		2	4	5	
5		4			1
		5		2	
	1		5		4
		4			2
4	5		1		

$\text{sim}(u,v)$

NA

0.87

1

NA

User-based CF

Example

	4	5	6	7	8	9	
							sim(u,v)
	2		2	4	5		NA
	5		4			1	0.87
			5		2		1
		1		5		4	-1
			4			2	
	4	5		1			NA

User-based CF

Example



2		2	4	5	
5		4			1
		5		2	
	1		5		4
3.51*	3.81*	4	2.42*	2.48*	2
4	5		1		

$\text{sim}(u,v)$

NA

0.87

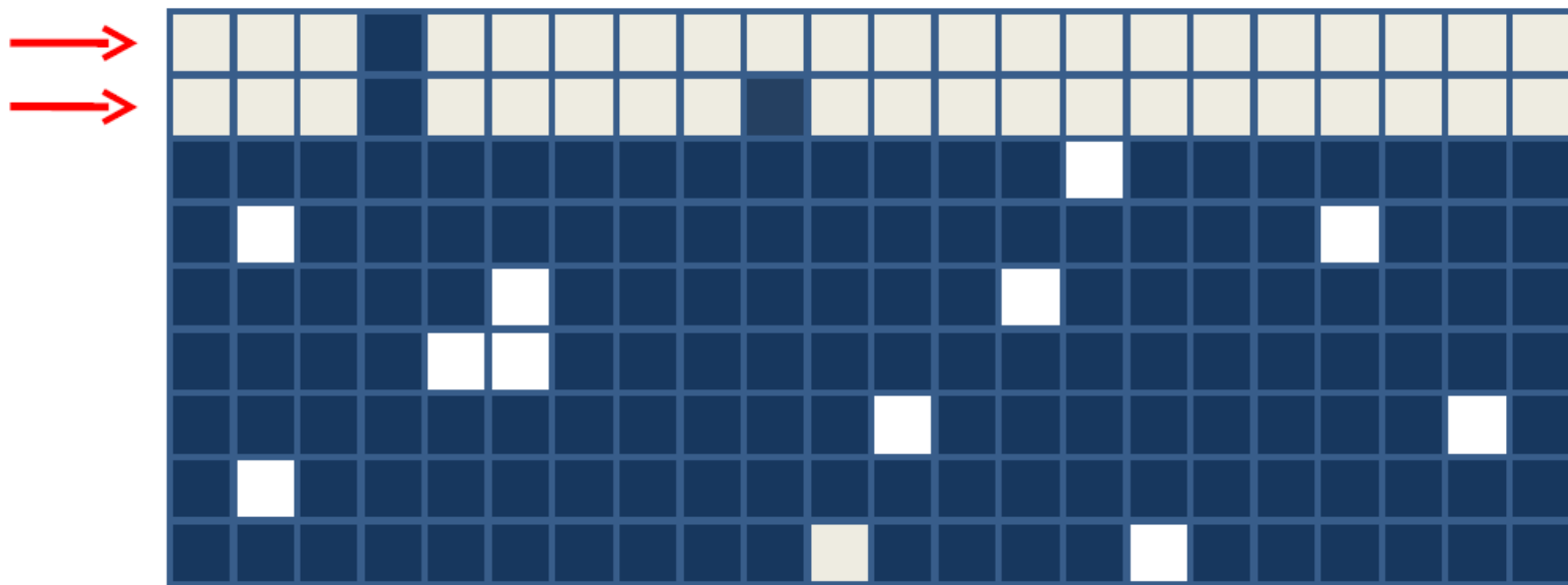
1

-1

NA

Problems with User-based Collaborative Filtering (1)

- User Cold-Start problem
not enough known about new user to decide who is similar (and perhaps no other users yet..)



- Need way to motivate early rater

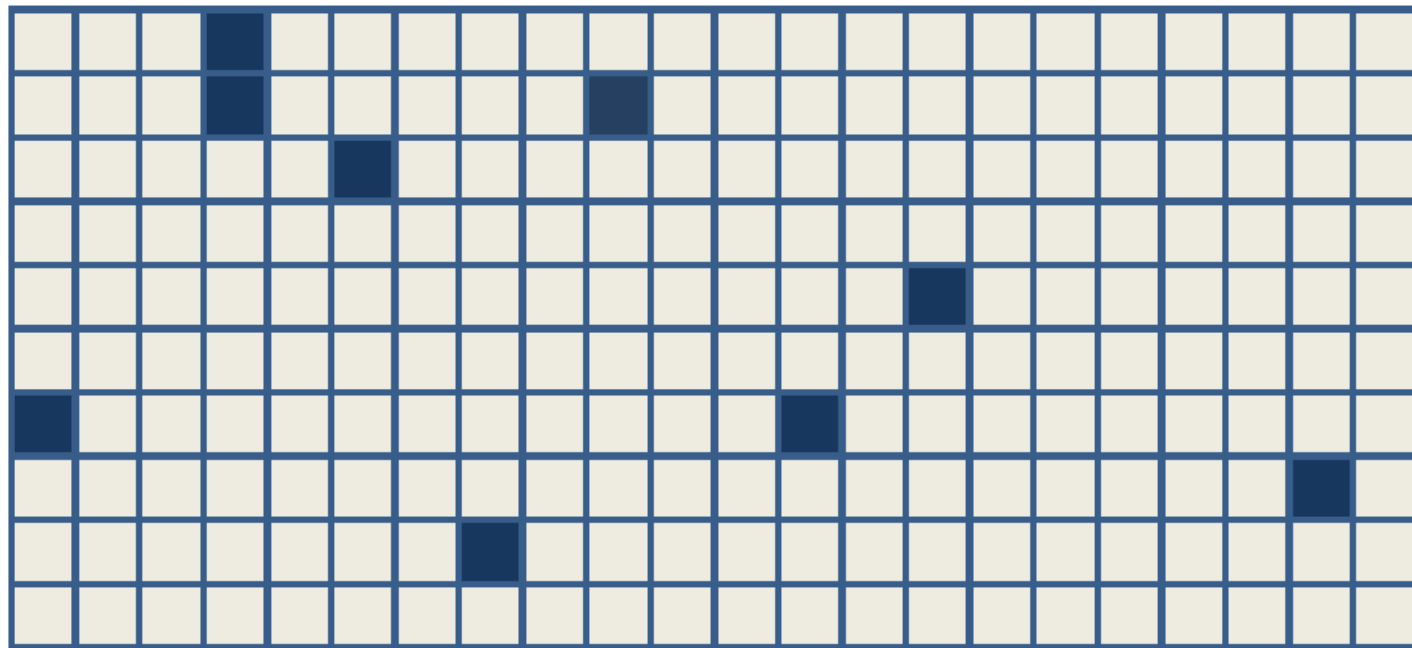
Problems with

User-based Collaborative Filtering (2)

- Sparsity

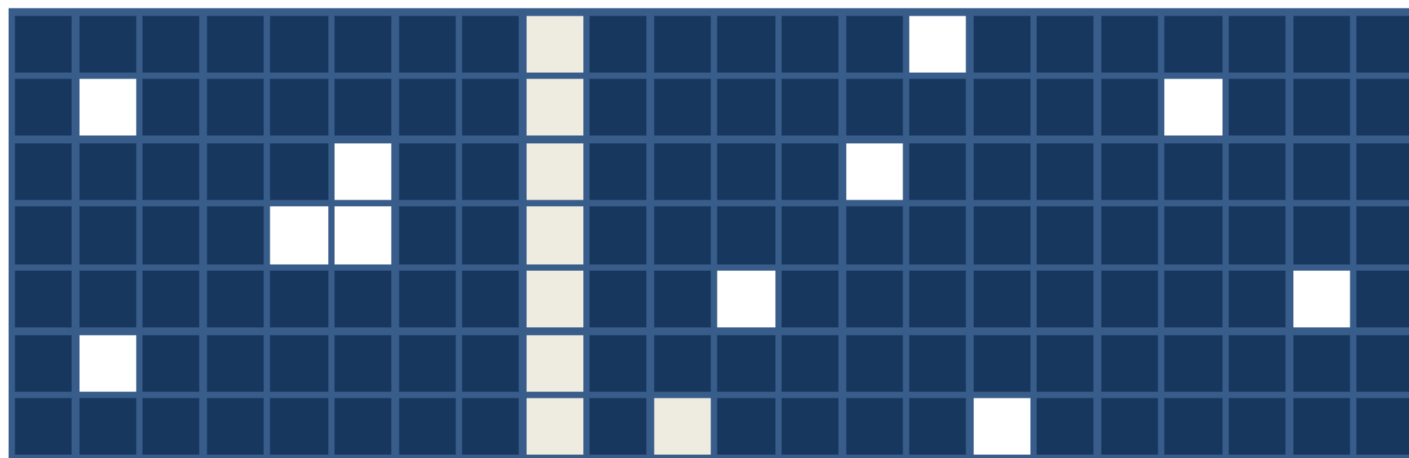
when recommending from a large item set, users will have rated only some of the items

(makes it hard to find similar users)



Problems with User-based Collaborative Filtering (3)

- Scalability
 - with millions of ratings, computations become slow
- Item Cold-Start problem
 - Cannot predict ratings for new item till some similar users have rated it [No problem for content-based]



Item Based CF Algorithm

- Look into the items the target user has rated
- Compute how similar they are to the target item
 - Similarity **only using** past **ratings** from other users!
- Select k most similar items.
- Compute Prediction by taking weighted average on the target user's ratings on the most similar items.

Item Similarity Computation







- Similarity: find users who have rated items and apply a similarity function to their ratings.
 - Cosine-based Similarity (difference in rating scale between users is not taken into account)

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|_2 \|\mathbf{B}\|_2} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}},$$

- Adjusted Cosine Similarity (takes care of difference in rating scale)

$$\text{sim}(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_u)^2}}.$$



	2		2	4	5	
	5		4	?		1
			5	?	2	
		1		5		4
			4	?		2
	4	5		1		



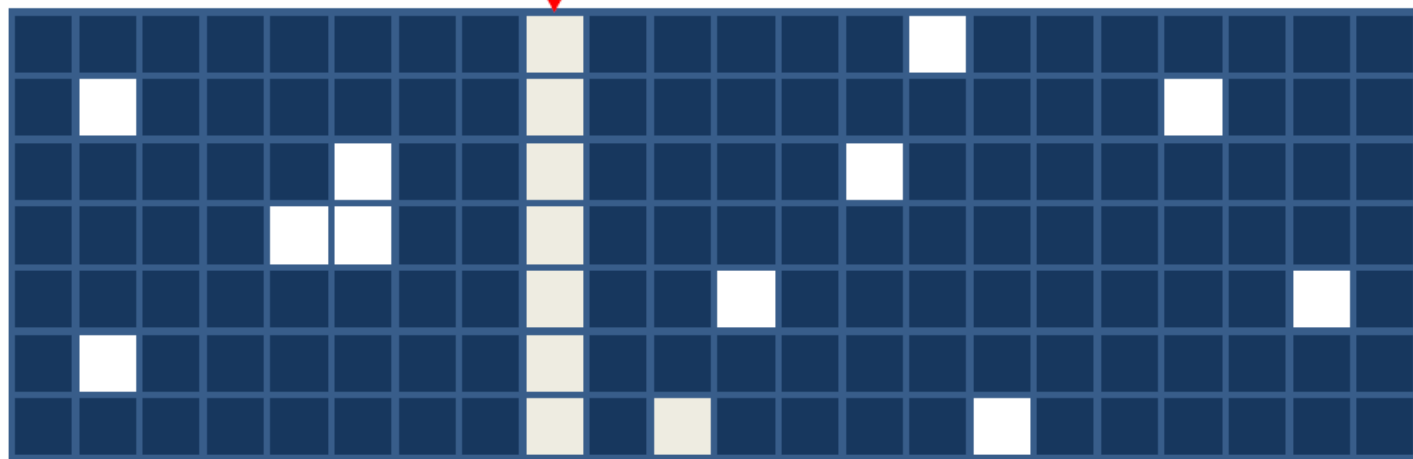
Problems with Item-based Collaborative Filtering (1)

- Item Cold-Start problem

Cannot predict which items are similar till we have ratings for this item

[No problem for content-based.]

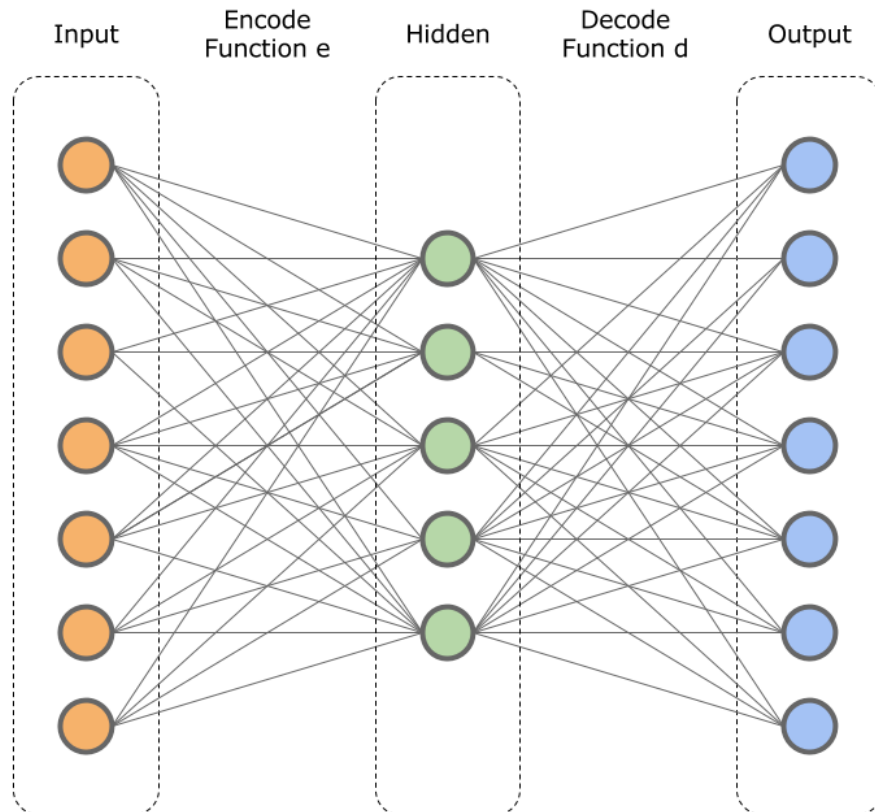
Also a problem for User-based collaborative filtering, but it is a bigger problem here.]



Model Based CF Algorithms

- First develop a model of user
- Type of model:
 - Probabilistic (e.g. Bayesian Network)
 - Clustering
 - Rule-based approaches (e.g. Association Rules)
 - If a visitor has some interest in Book 5, she will be recommended to buy Book 3 as well
 - Classification
 - Regression
 - LDA

Deep Autoencoders For Collaborative Filtering



$$\phi : X \longrightarrow Z : x \mapsto \phi(x) = \sigma(Wx + b) := z$$

$$\varphi : Z \longrightarrow X : z \mapsto \varphi(z) = \sigma(\tilde{W}z + \tilde{b}) := x'$$

$$L(x, x') = \frac{1}{n} \sum_{i=1}^n \|x_i - x'_i\|^2$$

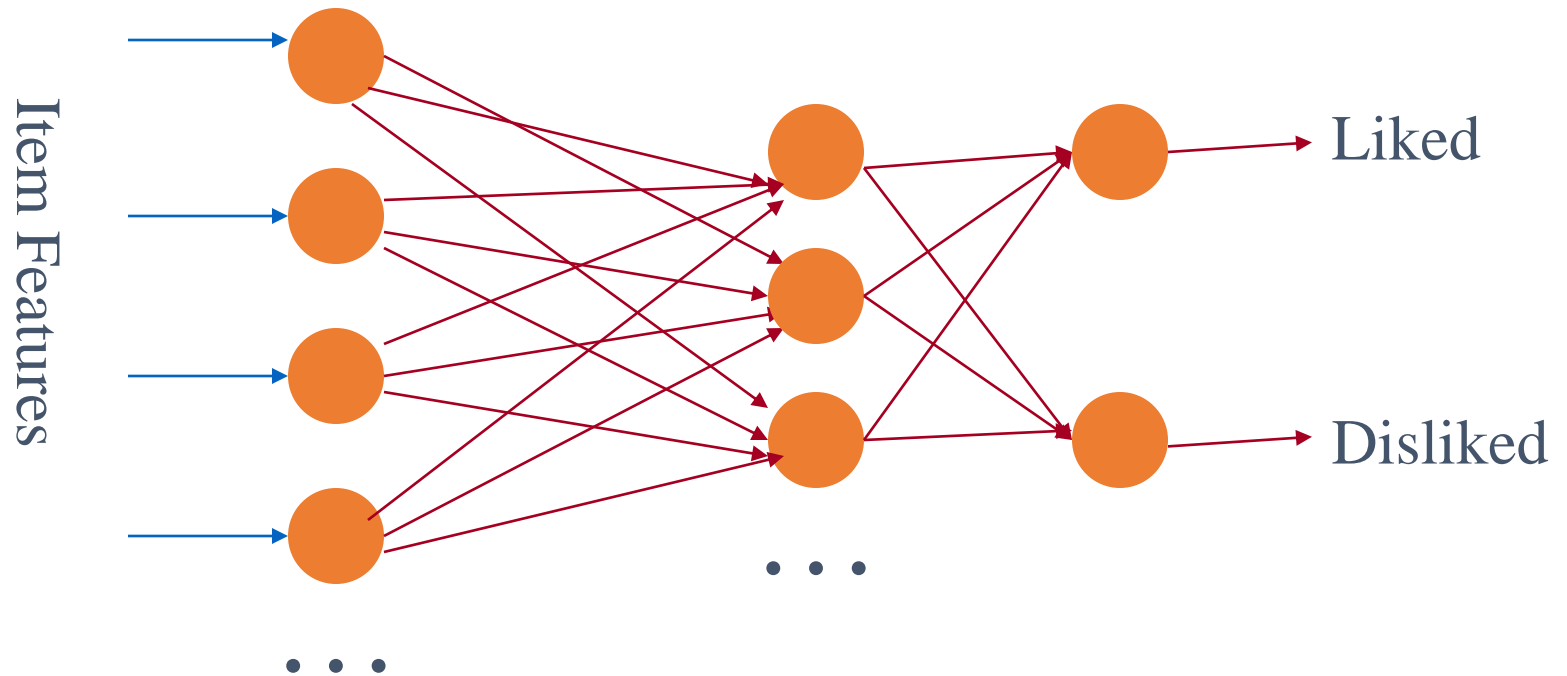
$$= \frac{1}{n} \sum_{i=1}^n \|x_i - \sigma(Wz_i + b)\|^2$$

$$= \frac{1}{n} \sum_{i=1}^n \|x_i - \sigma(W\sigma(\tilde{W}x_i + \tilde{b}) + b)\|^2$$

MovieLens Data Description

- Dataset contains 1,000,209 anonymous ratings of approximately 3,900 movies made by 6,040 *MovieLens* users. The import file we need is *ratings.dat*. This file contains 1,000,209 lines all having the following format:
- user_id::movie_id::rating:time_stamp.
- 1::595::5::978824268

NN Recommender



- Calculate recommendation score as $y_{\text{liked}} - y_{\text{disliked}}$

Matrix Factorization

Matrix Factorization

Number in table: number of figures a person has

A
B
C
D
E

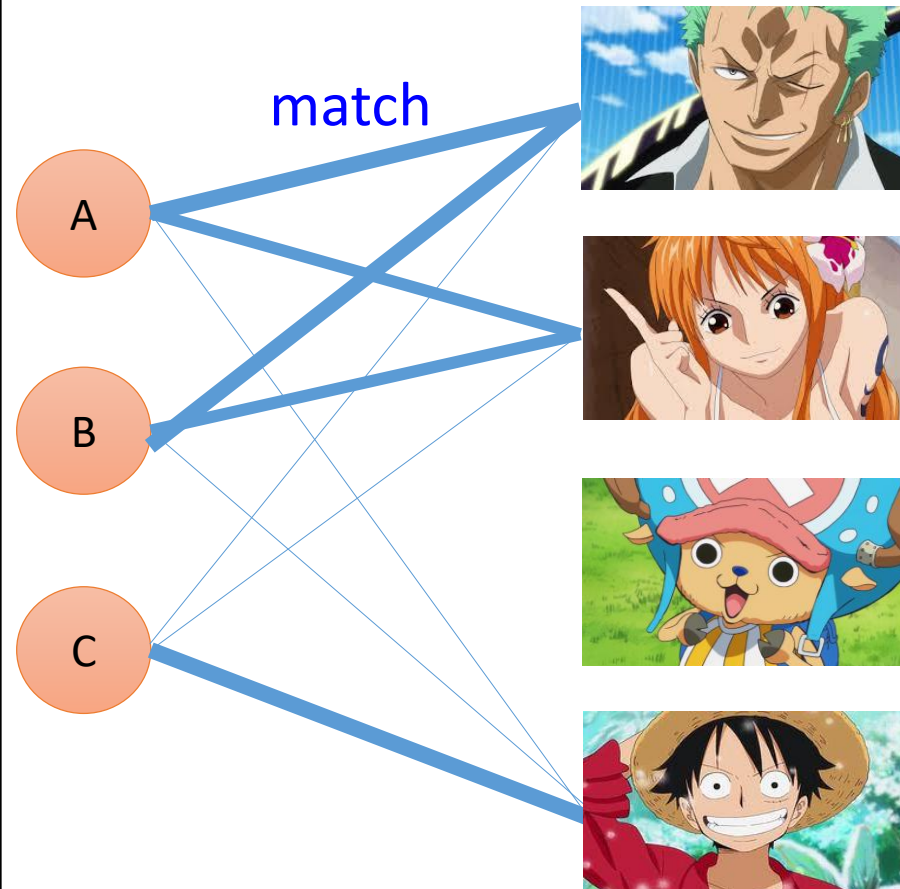
There are some common *factors* behind customers and characters.

<http://www.quuxlabs.com/blog/2010/09/matrix-factorization-a-simple-tutorial-and-implementation-in-python/>

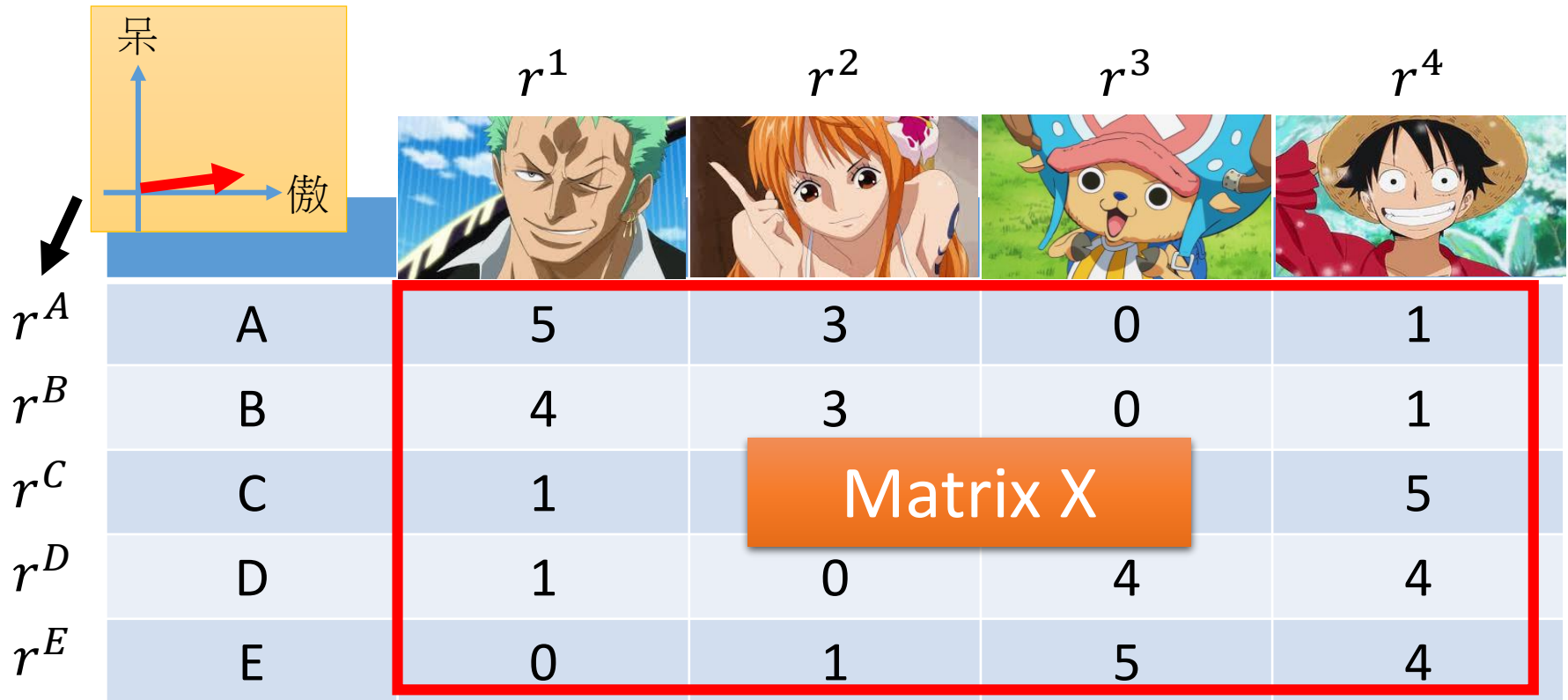
Matrix Factorization

The factors are latent.

Only implicit
feedback



Not directly
observable

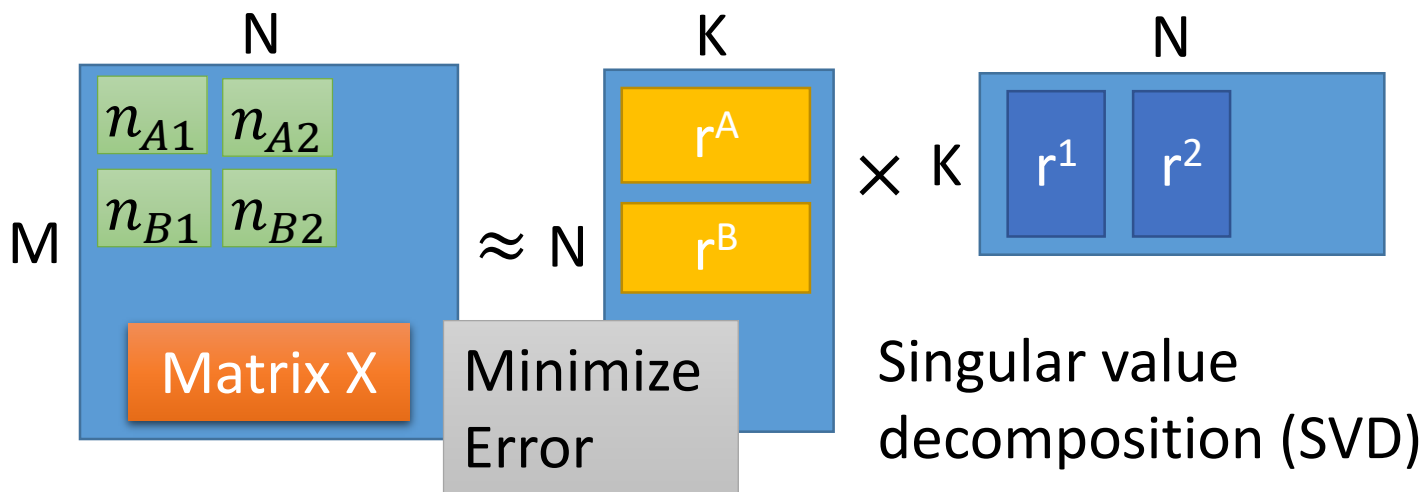






No. of Otaku = M

No. of characters = N

No. of latent factor = K

$$\begin{aligned}
 r^A \cdot r^1 &\approx 5 \\
 r^B \cdot r^1 &\approx 4 \\
 r^C \cdot r^1 &\approx 1 \\
 &\vdots
 \end{aligned}$$



	r^j	r^1	r^2	r^3	r^4
r^i					
r^A	A	5 n_{A1}	3	?	1
r^B	B	4	3	?	1
r^C	C	1	1	?	5
r^D	D	1	?	4	4
r^E	E	?	1	5	4

$$r^A \cdot r^1 \approx 5$$

$$r^B \cdot r^1 \approx 4$$

$$r^C \cdot r^1 \approx 1$$





⋮

Minimizing

$$L = \sum_{(i,j)} (r^i \cdot r^j - n_{ij})^2$$

Find r^i and r^j by gradient descent

Only considering the defined value

		r^1	r^2	r^3	r^4
					
r^A	A	5 n_{A1}	3	-0.4	1
r^B	B	4	3	-0.3	1
r^C	C	1	1	2.2	5
r^D	D	1	0.6	4	4
r^E	E	0.1	1	5	4

Assume the dimensions of r are all 2 (there are two factors)

A	0.2	2.1
B	0.2	1.8
C	1.3	0.7
D	1.9	0.2
E	2.2	0.0

1 (索隆)	0.0	2.2
2 (娜美)	0.1	1.5
3 (喬巴)	1.9	-0.3
4 (魯夫)	2.2	0.5

More about Matrix Factorization

- Considering the individual characteristics

$$r^A \cdot r^1 \approx 5 \quad \longrightarrow \quad r^A \cdot r^1 + b_A + b_1 \approx 5$$

b_A : how customer A likes to buy

b_1 : how popular character 1 is

Minimizing
$$L = \sum_{(i,j)} (r^i \cdot r^j + b_i + b_j - n_{ij})^2$$

Find r^i, r^j, b_i, b_j by gradient descent (can add regularization)

- Ref: Matrix Factorization Techniques For Recommender Systems

Matrix Factorization

$$\min \sum_{i,j} (\langle u_i, v_j \rangle - x_{ij})^2 + \frac{\alpha}{2} \|u\|^2 + \frac{\beta}{2} \|v\|^2$$

- It can also incorporate contents F...

$$\min \sum_{i,j} (\langle u_i, v_j \rangle - x_{ij})^2 + \frac{\alpha}{2} \|u\|^2 + \frac{\beta}{2} \|v\|^2 + \|F - Wv\|^2$$

Compared with Collaborative Filtering

- Predict unobserved entries based on partial observed matrix

movies

		2		1			4				5	
		5		4				?		1		3
			3		5			2				
	4			?			5		3		?	
			4		1	3				5		
				2				1	?			4
		1					5		5		4	
			2		?	5		?		4		
		3		3		1		5		2		1
		3				1			2		3	
		4			5	1			3			
			3				3	?			5	
	2	?		1		1						
			5			2	?		4		4	
		1		3		1	5		4		5	
	1		2			4				5	?	

users

Overfitting

- Most state-of-the-art recommender methods have a large number of model parameters and thus are prone to **overfitting**.
- Low Rank Approximation

The diagram illustrates the Low Rank Approximation concept for matrix factorization. It shows a matrix Y (a 10x10 grid of numbers) being approximated by the product of two matrices, U and V' .

Matrix Y (10x10):

	2		4	5		1	4	2	
3	1			2	2		5		4
4		2		4	1		3	1	
3			3	4		2			4
2	3		1		3		2		
	2	2		1			4		5
	2		4	1	4		2	3	
1		3		1	1			4	3
	4		2	2		5	3	1	

Matrix U (10x5):

Matrix V' (5x10):

Matrix X (10x10, rank k):

Solution to Overfitting

- Typically L2-regularization is applied to prevent overfitting, e.g.:
 - Maximum margin matrix factorization

$$\underset{U, V}{\text{minimize}} \quad \sum_{(u, i, y) \in S} \max(0, 1 - y_{ui} U_u^T V_i) + \frac{\lambda}{2} [\text{tr} U U^T + \text{tr} V V^T]$$

- Probabilistic matrix factorization

$$\underset{U, V}{\text{minimize}} \quad \sum_{(u, i, r) \in S} (U_u^T V_i - r_{ui})^2 + \frac{\lambda}{2} [\text{tr} U U^T + \text{tr} V V^T]$$

Deep Learning

Deep Learning Basic Model

- Restricted Boltzmann Machines (RBM) for recommendation
- Collaborative Deep Learning
- Collaborative Denoising Auto-Encoder
- Collaborative Recurrent Autoencoder
- A Multi-View Deep Learning Approach for Cross Domain User Modeling in Recommendation Systems (MV-DNN)
- Temporal Deep Semantic Structured Model (TDSSM)
- Neural Autoregressive Collaborative Filtering (NF-NADE)

Q and A