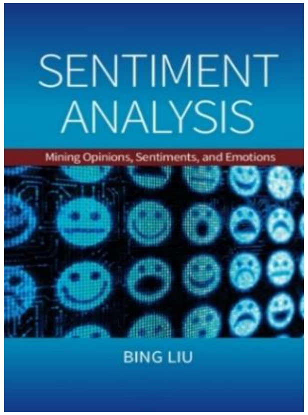
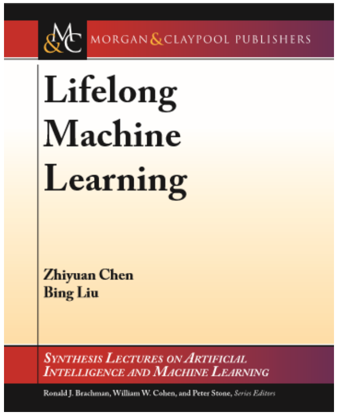
Opinions, Deceptions, and Lifelong Machine Learning





Bing Liu
University of Illinois at Chicago
liub@uic.edu

Outline

- Opinion Mining and Sentiment Analysis
  - Problem definition
  - Sentiment classification (document & sentence)
  - Aspect-based sentiment analysis
- Lifelong Machine Learning
  - Lifelong learning for aspect extraction
  - Lifelong learning for sentiment classification
- Deceptive Opinion Detection
- Summary

Introduction

- Sentiment analysis 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.
  - Spread from CS to management, health, finance, medicine, political and social sciences
- Because of the Web and the social media

Sentiment Analysis (SA) problem
(Hu and Liu 2004; Liu, 2010; 2012)

- Id: John 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 old Blackberry. …*"
- **Definition**: An *opinion* is a quadruple,
  (*target*, *sentiment*, *holder*, *time*)
- A more practical definition:
  (*entity*, *aspect*, *sentiment*, *holder*, *time*)
  - E.g., (iPhone, touch_screen, +, John, 5-1-2008)
- SA goal: Given an opinion doc, mine all quintuples

Opinion summarization
(Hu and Liu, 2004)

- Classic text summarization is not suitable.
  - Opinion summary can be defined conceptually,
  - not dependent on how the summary is produced.
- Opinion summary needs to be quantitative
  - 60% positive about X is very different from 90% positive about X.
- One main form of opinion summary is
  - 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.

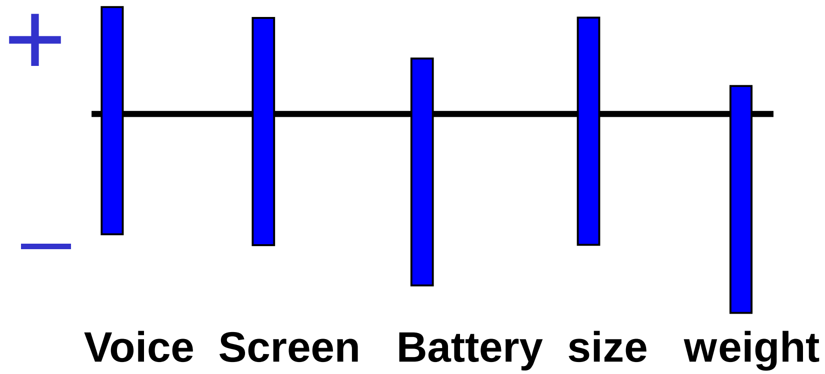
…

Aspect: **voice quality**

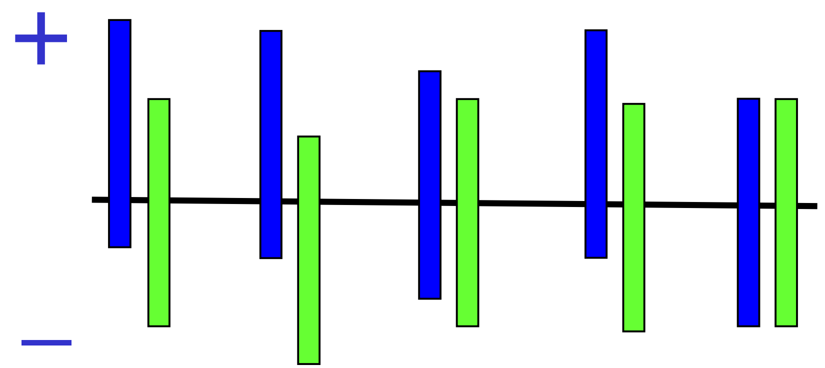
…

(Liu et al. 2005)

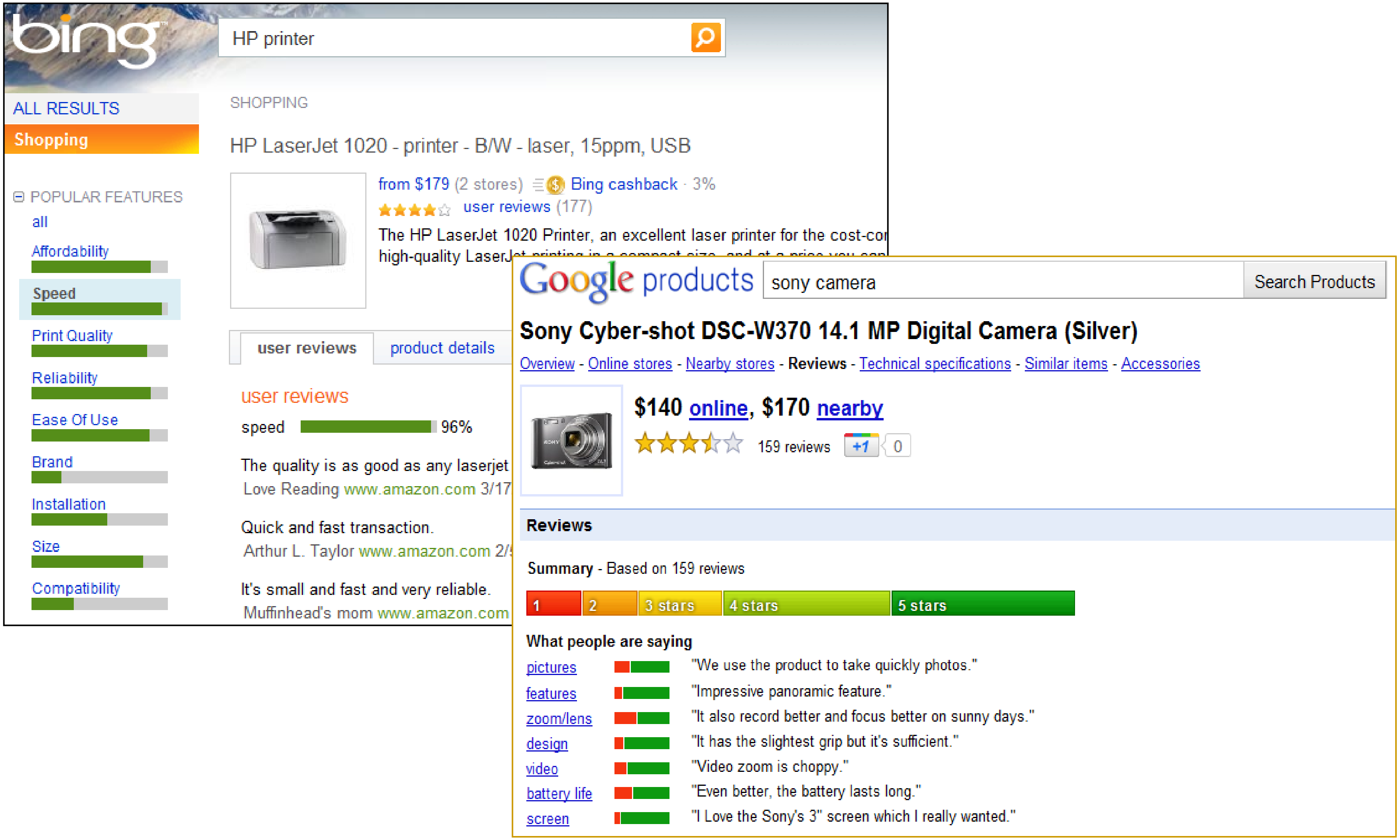
- Opinion Summary of 1 phone



- Opinion comparison of 2 phones



Aspect-based opinion summary



Summarization for BestBuy (Samsung) (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**."

ADDSTRUCTURE

Document sentiment classification

- Classify a whole opinion doc (e.g., a review) based on the overall sentiment (Pang & Lee, 2008)
    - Classes: Positive, Negative (and possibly neutral)
- Assumption: The doc contains opinions about a single entity.
- Reviews usually satisfy the assumption
    - Positive: 4/5 stars, negative: 1/2 stars
- But forum discussions often do not

Solution methods

- Supervised learning: applied all kinds of supervised learning methods, NB, SVM, DNN (Pang et al, 2002; Dave et al 2003; Gamon, 2004; Li et al 2010; Paltoglou & Thelwall, 2010; Xia et al. 2013; Socher et al 2013; etc)
  - Features: n-grams, sentiment words/phrases, POS tags, negation, position. dependency, word embedding
  - IR weighting schemes
- Unsupervised methods
  - Based on predefined patterns (Turney, 2002)
  - Lexicon-based methods (Taboada et al. 2011)
    - A list of positive and negative words with weighting and combination rules.

Sentence sentiment classification

- Classify each sentence: 3 classes
  - Positive, negative, neutral
  - Ignore mixed sentences, e.g., “*Apple is going well in this poor economy*”
- Supervised learning (Wiebe et al., 2004; Wilson et al 2004, etc)
  - Using similar features as for documents
- Lexicon based methods (Hu and Liu 2004; Kim and Hovy, 2004)

Aspect-based sentiment analysis

- Document/sentence sentiment classification does not give details (Hu and Liu, 2004).
- They help but do not solve the problem of

  (*entity*, *aspect*, *sentiment*, *holder*, *time*)
  - Do not identify entity, aspect, holder or time.
  - Do not assign sentiment to entity/aspect.
- For applications, we often need to solve the full problem, i.e., *aspect-based analysis*.

Aspect extraction

- "*The battery life is long, but pictures are poor*."
  - Aspects: battery life, picture
- Many approaches
  - Frequency-based: frequent noun phases (Hu & Liu, 2004)
  - Syntactic dependency: opinion and target relation (Hu & Liu 2004; Zhuang, Jin & Zhu 2006; Wang & Wang, 2008; Wu et al. 2009; Blair-Goldensohn *et al.*, 2008; Qiu et al. 2009, Kessler & Nicolov, 2009; etc).
  - Supervised sequent labeling (e.g., CRF) (Jin and Ho 2009; Jakob and Gurevych, 2010, etc)
  - Topic modeling (Mei et al, 07; Titov et al 08; Li, Huang & Zhu, 10, …)
  - Many others (Kobayashi et al. 2006; Fang & Huang 2012; Liu, Xu & Zhao. 2013; Zhu, Wan & Xiao 2013, etc)

Aspect sentiment classification

"*Apple is doing very well in this poor economy*"

- Lexicon-based approach:
  - Opinion words/phrases, *good*, *bad*, *cost an arm and leg*
  - Parsing and orientation shifters: simple sentences, compound sentences, conditional sentences, questions, modality, verb tenses, negations and other sentiment orientation shifters, etc.
- Supervised learning (tricky):
  - SVM, deep learning, and many other methods have been applied using distance feature weighting, parse tree, attentions, etc.

Some interesting/hard sentences

- "We brought the mattress yesterday, and a body impression has formed."
- "Are there any great perks for employees?"
- "Great for insomniacs" (book)
- "The laptop next to Lenovo looks so ugly."
- "I am so happy that my iPhone is nothing like my old ugly Droid."
- "After taking the drug, I got severe stomach pain"
- "The top of the picture is brighter than the bottom."

Outline

- Opinion Mining and Sentiment Analysis
  - Problem definition
  - Sentiment classification (document & sentence)
  - Aspect-based sentiment analysis
- Lifelong Machine Learning
  - Lifelong learning for aspect extraction
  - Lifelong learning for sentiment classification
- Deceptive Opinion Detection
- Summary

Classic Learning Paradigm (ML 1.0)
(Chen and Liu, 2016-book)

- Isolated single-task learning
  - Given a dataset, run an ML algo. to build a model
    - No consideration any previously learned knowledge
- Weaknesses of “isolated learning”
  - Knowledge learned is not retained or accumulated
    - Needs a large number of training examples
    - Suitable for well-defined & narrow tasks in restricted env.
    - Cannot learn by itself automatically

Machine learning: ML 2.0

(Thrun, 1996b; Silver et al 2013; Chen and Liu, 2014, 2016-book)

- Human beings never learn in isolation
  - We learn continuously:
    - Accumulate knowledge learned in the past and use it to learn more knowledge, and we are self-motivated
    - Learn effectively from a few or no examples
- *Lifelong Machine Learning* (LML) mimics this human learning capability
    - Without LML, an AI system is unlikely to be intelligent
    - **EU:** Lifelong Learning for Intelligent Systems (LLIS) (2017)
    - **DARPA:** Lifelong learning machine (L2M) (2017)

Humans Don't Learn in Isolation

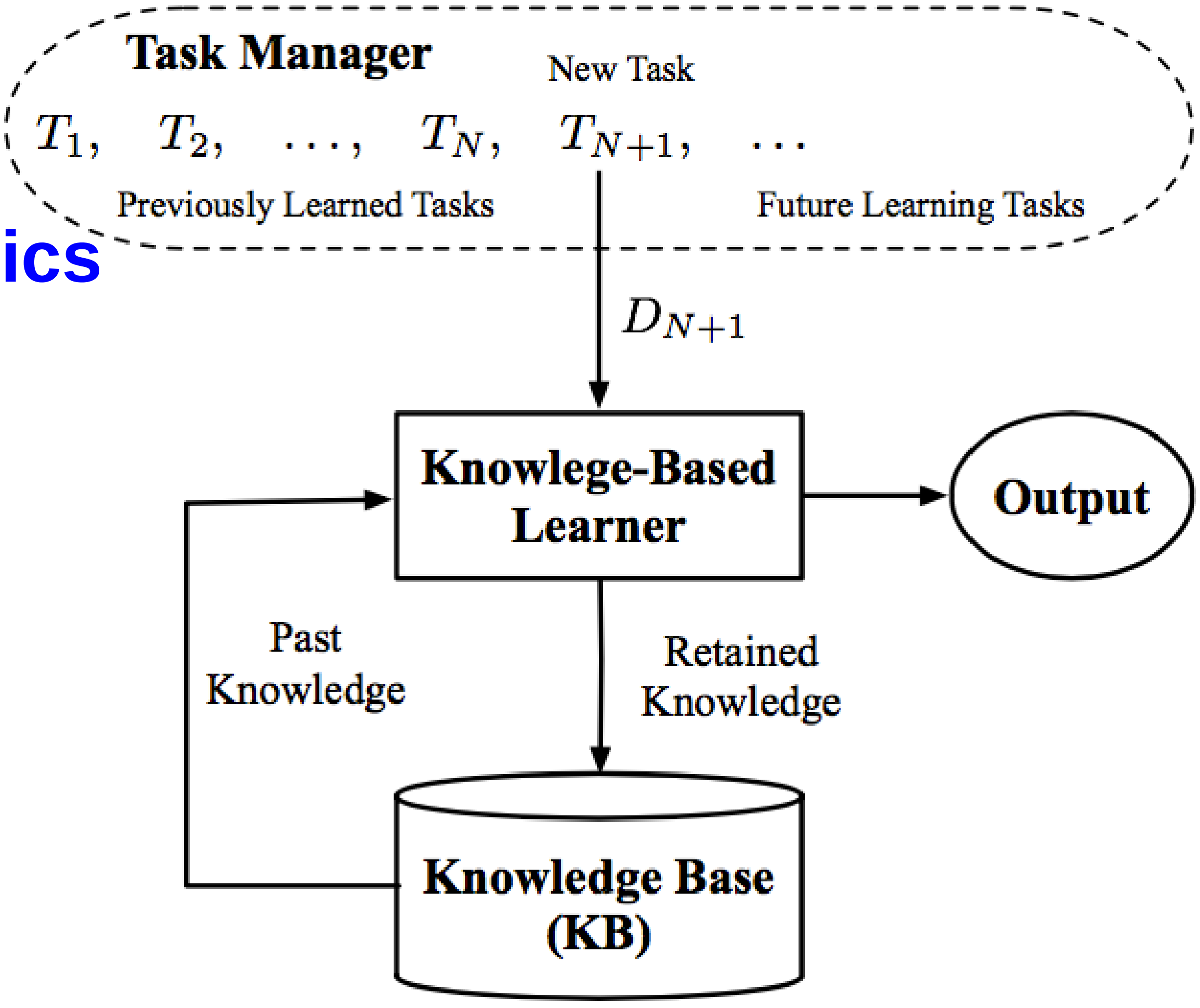
- Nobody has ever given me 1000 positive and 1000 negative car reviews and ask me
  - to build a classifier to classify reviews of cars
- I can do it without any training reviews as
  - I have learned how people praise and criticize things
- If I don't have the accumulated knowledge, NO
  - E.g., I don't know Arabic and if someone gives me 2000 training +/- reviews in Arabic, I cannot learn.

Definition of LML

(Chen and Liu, 2016 – book)

- The learner has performed learning on a sequence of tasks, from 1 to *N*.
- When faced with the (*N*+1)th task, it uses the relevant knowledge in its *knowledge base* (KB) to help learning for the (*N*+1)th task.
- After learning (*N*+1)th task, KB is updated with learned results from (*N*+1)th task.

Lifelong Machine Learning Systems



Key Characteristics

- Continuous learning and adaptation
- Knowledge accumulation (KB)
- Use the past knowledge to help new learning (KBL)

Transfer, Multitask → Lifelong

- Transfer learning vs. LML
  - Transfer learning is not continuous
  - No retention or accumulation of knowledge
  - Only one directional: help target domain
- Multitask learning vs. LML
  - Multitask learning retains no knowledge except data
  - Hard to re-learn all when tasks are numerous
- Online multi-task learning is LML

Aspect Extraction: Topic Modeling

- *"The battery life is long, but pictures are poor."*
  - Aspect terms: battery life, picture
- Aspect extraction actually has two tasks:

  (1) extract aspect terms
  - "picture," "photo," "battery," "power"

  (2) cluster them (synonym grouping).
  - Same aspects: {"picture," "photo"}, {"battery," "power"}
- Top modeling (Blei et al 2003) performs both tasks at the same time. A topic is an aspect.
  - E.g., {*price*, *cost*, *cheap*, *expensive*, …}

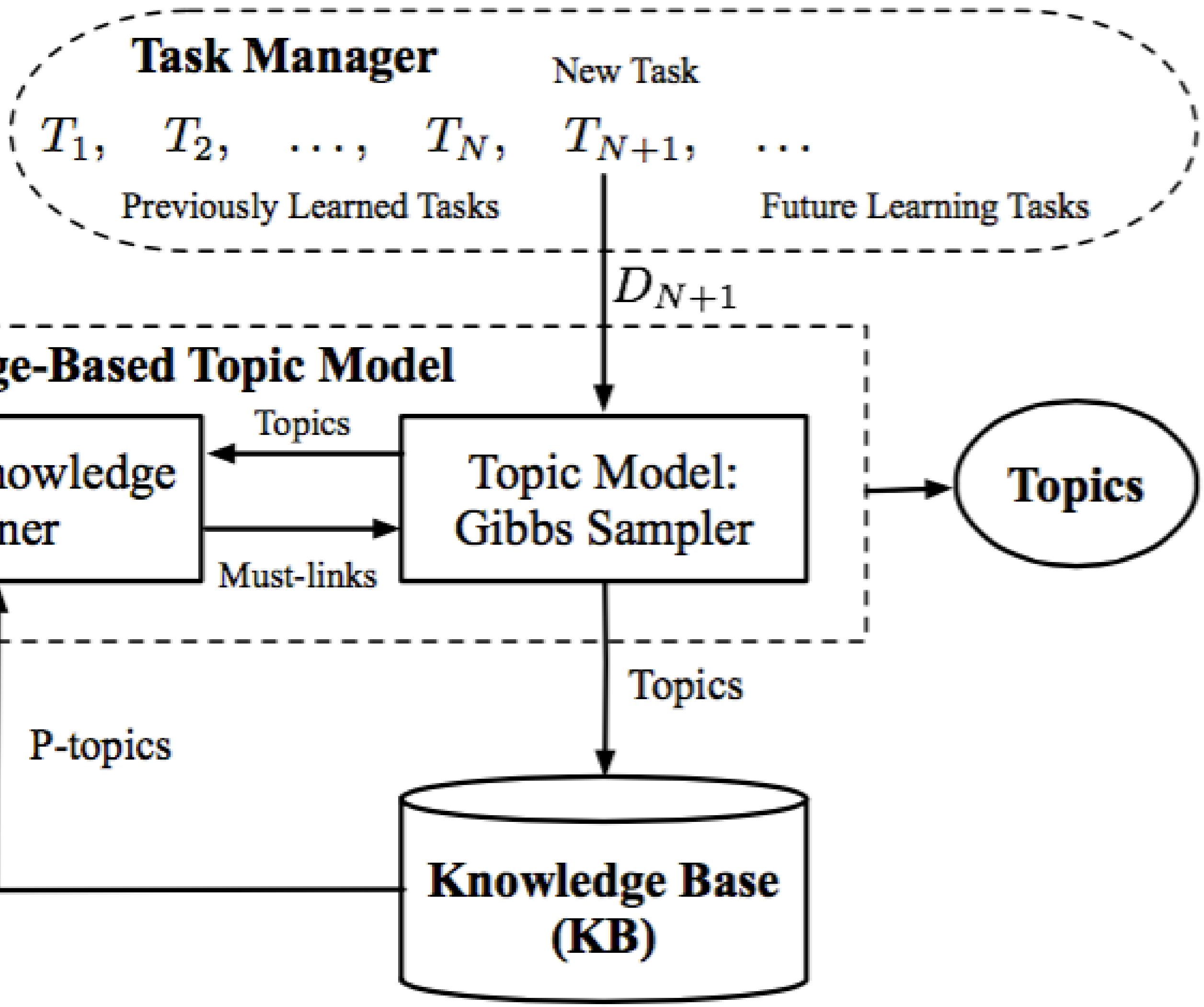
Key observation in practice

(Chen and Liu, ICML-2014)

- A fair amount of aspect overlapping across reviews of different products or domains
  - Every product review domain has the aspect *price*,
  - Most electronic products share the aspect *battery*
  - Many also share the aspect of *screen*.
- This sharing of concepts / knowledge across domains is true in general, not just for SA.
  - It is “silly” not to exploit such sharing in learning

LTM: Lifelong Topic Modeling

(Chen and Liu, 2014a)



What Knowledge?

(Chen and Liu, 2014a, 2014b; Wang et al. 2016)

- Should be in the same aspect/topic

  => Must-Links

  e.g., {picture, photo}

- Should not be in the same aspect/topic

  => Cannot-Links

  e.g., {battery, picture}

Approach: shared local knowledge

- Some pieces of knowledge from previous tasks/domains are useful to new task, e.g.,
  - {*price*, *cost*} and {*price*, *expensive*} should belong to the same topics.

$$P(z_i = t | \boldsymbol{z}^{-i}, \boldsymbol{w}, \alpha, \beta, \mathbf{A}') \propto$$
$$\frac{n_{d,t}^{-i} + \alpha}{\sum_{t'=1}^{T} (n_{d,t'}^{-i} + \alpha)} \times \frac{\sum_{w'=1}^{V} \mathbf{A}'_{t,w',w_i} \times n_{t,w'}^{-i} + \beta}{\sum_{v=1}^{V} (\sum_{w'=1}^{V} \mathbf{A}'_{t,w',v} \times n_{t,w'}^{-i} + \beta)}$$

Lifelong Sentiment Classification
(Chen, Ma, and Liu 2015)

- "*I bought a cellphone a few days ago. It is such a nice phone. The touch screen is really cool. The voice quality is great too. ....*"
- Goal: classify docs or sentences as + or -.
  - Need to manually label a lot of training data for each domain, which is highly labor-intensive
- Can we not label for every domain or at least not label so many docs/sentences?

A Simple LML Approach

Assuming we have worked on a *large number of past domains* with all their training data *D*

- Build a classifier using *D*, test on new domain
  - Note - using only one past/source domain as in *transfer learning* does not work well.
- In many cases – improve accuracy by as much as 19% (= 80%-61%). Why?
- In some others cases – not so good, e.g., it works poorly for toy reviews. Why? "toy"

Objective Function

- Maximize the probability difference

$$\sum_{i=1}^{|D^t|} \left(P\left(c_j|d_i\right) - P\left(c_f|d_i\right)\right)$$

- c_j: labeled class in ground truth
- c_f: all classes other than c_j

Exploiting Knowledge via Penalties

- Penalty terms for two types of knowledge
  - Document-level knowledge
  - Domain-level knowledge

$$\frac{1}{2}\alpha \sum_{w \in V_S} \left(X_{+,w} - R_w \times X^0_{+,w}\right)^2 + \frac{1}{2}\alpha \sum_{w \in V_S} \left(X_{-,w} - (1 - R_w) \times X^0_{-,w}\right)^2$$

  - R_W: ratio of #tasks where *w* is positive / #all tasks
  - $X^0_{+,w} = N^t_{+,w} + N^{KB}_{+,w}$ and $X^0_{-,w} = N^t_{-,w} + N^{KB}_{-,w}$

One Result

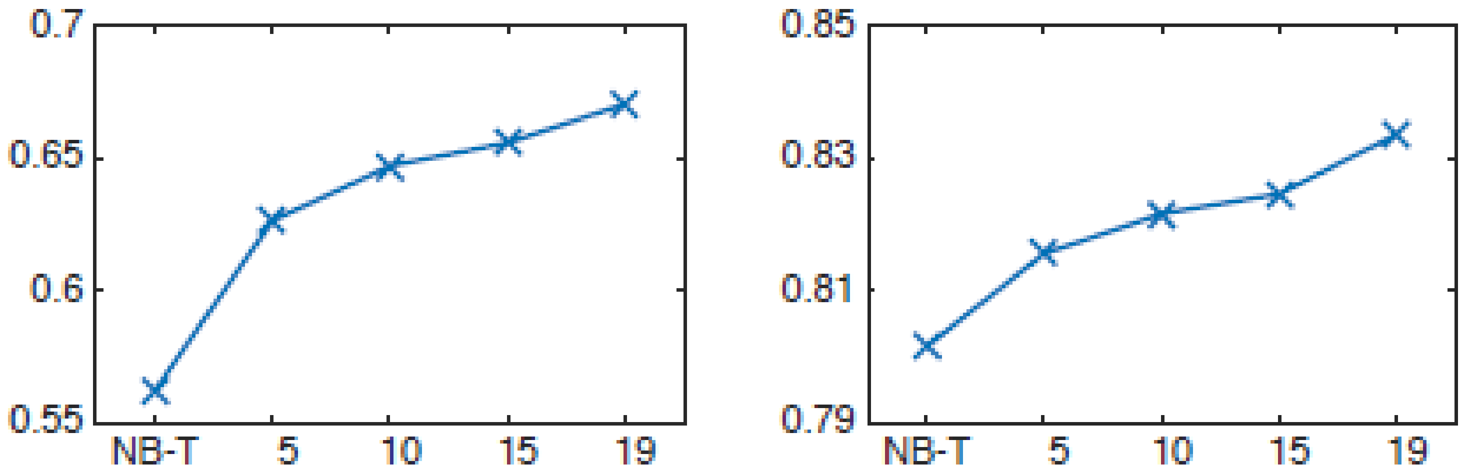


Figure 1: Figure 1. (Left): Negative class F1-score of LSC with #past domains in natural class distribution. (Right): Accuracy of LSC with #past domains in balanced class distribution.

Separating entities and aspects
(Shu et al 2016)

- Sentiment target can be an entity or aspect. Previous algorithms cannot separate them
  - "This camera is great" and "Pictures from Sony cameras are great"
- Use relaxation labeling
  - unsupervised graph-based label propagation.
    - Unsupervised classification
  - It is augmented with lifelong learning (*Lifelong-RL*) to exploit previously learned knowledge.

Shared knowledge in lifelong learning

Lifelong-RL uses two forms of knowledge

- Prior edges: graphs are usually not given or fixed but are built based on text data.
  - If the data is small, many edges may be missing
  - But such edges may exist in the graphs of some previous tasks
- Prior labels: initial $P^0(L(n_i))$ is quite hard to set, but results from previous tasks can be used to set it more accurately.

Improving model in execution
(Shu et al 2017)

- Can a model’s performance be improved after training without manual labeling?
- This paper proposes a technique that does this in the context of CRF for info extraction.
- It exploits dependency features
  - As the model sees more data, more features are identified
  - These features help produce better results in the new domain using the same model.

Outline

- Opinion Mining and Sentiment Analysis
  - Problem definition
  - Sentiment classification (document & sentence)
  - Aspect-based sentiment analysis
- Lifelong Machine Learning
  - Lifelong learning for aspect extraction
  - Lifelong learning for sentiment classification
- Deceptive Opinion Detection
- Summary

Opinion spamming (fake reviews)
(Jindal and Liu, 2007, 2008)

- Opinion spamming refers to people giving **fake or deceptive** reviews/opinions, e.g.,
  - Write undeserving positive reviews for some target entities in order to promote them.
  - Write unfair or malicious negative reviews for some target entities in order to damage their reputations.
- Motivation: positive opinions mean profits and fame for businesses and individuals
- Writing fake reviews has become a business
  - e.g., reputation management firms

Is this review fake or not?

I want to make this review in order to comment on the excellent service that my mother and I received on the Serenade of the Seas, a cruise line for Royal Caribbean. There was a lot of things to do in the morning and afternoon portion for the 7 days that we were on the ship. We went to 6 different islands and saw some amazing sites! It was definitely worth the effort of planning beforehand. The dinner service was 5 star for sure. One of our main waiters, Muhammad was one of the nicest people I have ever met. However, I am not one for clubbing, drinking, or gambling, so the nights were pretty slow for me because there was not much else to do. Either than that, I recommend the Serenade to anyone who is looking for excellent service, excellent food, and a week full of amazing day-activities!

What about this?

The restaurant is located inside of a hotel, but do not let that keep you from going! The main chef, Chef Chad, is absolutely amazing! The other waiters and waitresses are very nice and treat their guests very respectfully with their service (i.e. napkins to match the clothing colors you are wearing). We went to Aria twice in one weekend because the food was so fantastic. There are so many wonderful Asian flavors. From the plating of the food, to the unique food options, to the fresh and amazing nan bread and the tandoori oven that you can watch as the food is being cooked, all is spectacular. The atmosphere and the space are great as well. I just wished we lived closer and could dine there more frequently because it is quite expensive.

A Study of Amazon Reviews

- June 2006 Crawl: 5.8mil reviews, 1.2mil products and 2.1mil reviewers.

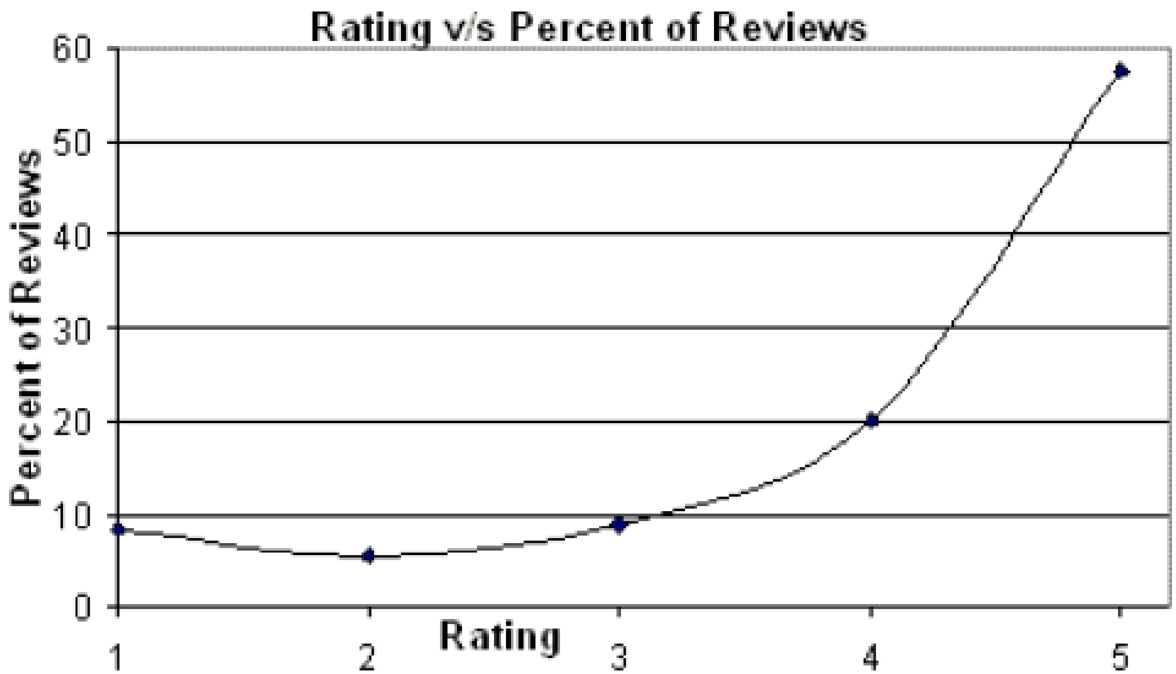


Figure 4. Rating vs. percent of reviews

Supervised Learning Approaches

- AUC = 0.78 assuming duplicate reviews as fake [Jindal & Liu, 2008]. Duplicate reviews as fake is incomplete and a naïve assumption
- F1 = 0.63 using manually labeled fake reviews [Li et al., 2011]. Manual labeling of fake reviews is unreliable.
- Accuracy = 90% using n-grams on Amazon Mechanical Turk (AMT) crowdsourced fake reviews [Ott et al., 2011].

N-gram Features on Yelp Fake Reviews (Mukherjee et al. 2013)

n	P	R	F1	A
Uni	86.7	89.4	88.0	87.8
Bi	88.9	89.9	89.3	88.8

(a) Ott et al., (2011)

P	R	F1	A
65.1	78.1	71.0	67.6
61.1	82.4	70.2	64.9

(b) Hotel

P	R	F1	A
65.1	77.4	70.7	67.9
64.9	81.2	72.1	68.5

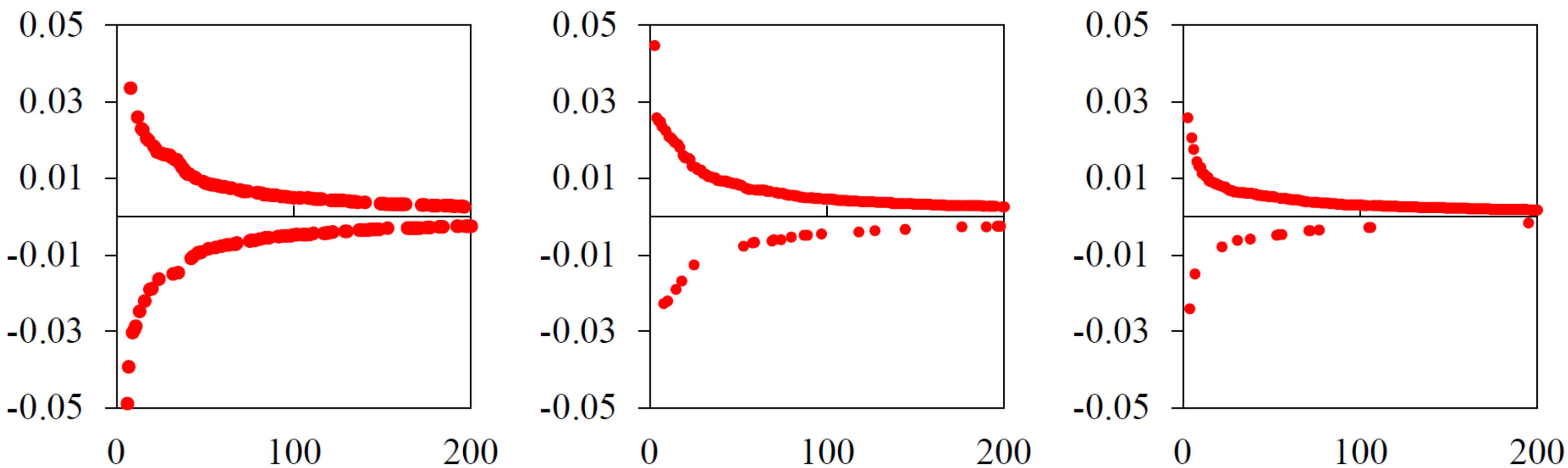
(c) Restaurant

Table 4: Comparison with Ott et al. (2011) based on SVM 5-fold CV results. Feature Set: Uni: Word unigrams, Bi: Word bigrams.

Why? Information theoretic analysis

- A deep probe using KL-Divergence
  - $KL(F||N) = \sum_i F(i) \log_2 \left(\frac{F(i)}{N(i)}\right)$ to measure:
  - "How much fake reviews differ from non-fake linguistically?"
- Word-wise contribution $\Delta\ KL$
  - $\Delta KL_{Word}^{i} = KL_{Word}(F_i||N_i) - KL_{Word}(N_i||F_i)$

Plots of Δ *KL* results



(a) Ott et al. (b) Hotel (c) Restaurant

Word-wise difference (ΔKL_{Word}) across top 200 words

1. Turkers did not do a good job at Faking
2. Yelp Spammers are smart but overdid Faking!

Using behavioral features

Feature Setting	P	R	F1	A
Unigrams	62.9	76.6	68.9	65.6
Bigrams	61.1	79.9	69.2	64.4
Behavior Feat.(BF)	81.9	84.6	**83.2**	**83.2**
Unigrams + BF	83.2	80.6	81.9	83.6
Bigrams + BF	86.7	82.5	**84.5**	**84.8**

(a): Hotel

P	R	F1	A
64.3	76.3	69.7	66.9
64.5	79.3	71.1	67.8
82.1	87.9	**84.9**	**82.8**
83.4	87.1	85.2	84.1
84.1	87.3	**85.7**	**86.1**

(b): Restaurant

Classification results: behavioral (BF) and *n*-gram features

Outline

- Opinion Mining and Sentiment Analysis
  - Problem definition
  - Sentiment classification (document & sentence)
  - Aspect-based sentiment analysis
- Lifelong Machine Learning
  - Lifelong learning for aspect extraction
  - Lifelong learning for sentiment classification
- Deceptive Opinion Detection
- Summary

Summary

- SA is a fascinating NLP problem
  - Every sub-problem is highly challenging.
  - But it is also restricted (semantically).
  - Applications are flourishing!
- Deception/fake is a huge problem on the Web
  - Detection is needed
- To solve these or any problems that require intelligence,
  - lifelong learning is probably necessary.

Thank you

Q&A