The State of Sentiment

Bing Liu
University Of Illinois at Chicago
liub@cs.uic.edu

Introduction

- Sentiment analysis (SA) or opinion mining
  - computational study of opinion, sentiment, appraisal, evaluation, attitude, and emotion.
- Very active research area
  - Intellectually very challenging
    - It touches every aspect of NLP
    - NLP has no easy problems …
  - Unlimited applications
  - Spread from CS to management & social sciences.
    - E.g., politics, business, financial market, health, …

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

- Id: John on 5-1-2008 -- "*I bought an iPhone yesterday. It is a nice phone. **The touch screen is really cool**. The voice quality is clear too. …*"
- **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 summary

(Hu and Liu, 2004)

(Liu et al. 2005)

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

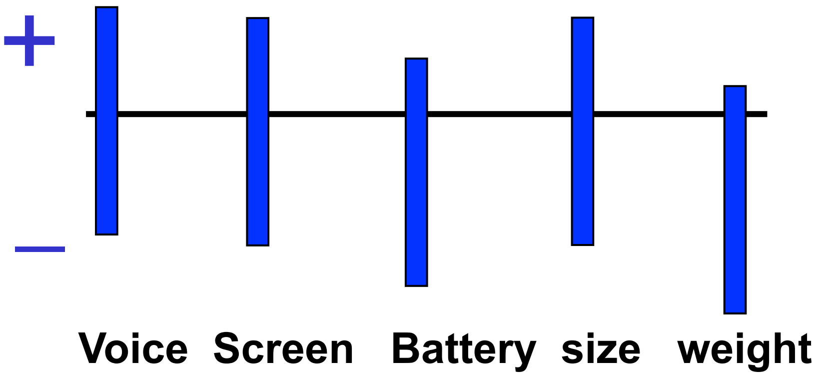
- The screen is easily scratched.
- I have a lot of difficulty in removing finger marks from the touch screen.

…

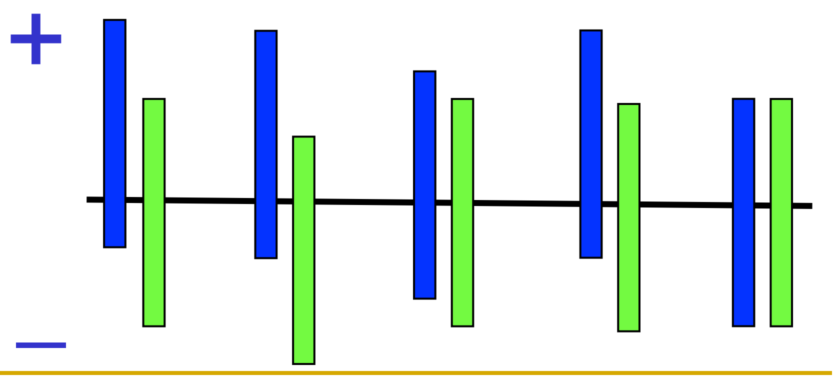
Aspect: **voice quality**

…

- Opinion Summary of 1 phone



- Opinion comparison of 2 phones



SA is a very rich problem

- (***entity***, ***aspect***, ***sentiment***, ***holder***, ***time***)
  - target entity: Entity extraction & resolution
  - aspect of *entity*: Aspect extraction & resolution
  - sentiment: Aspect sentiment classification
  - *opinion holder:* *Information/data extraction*
  - *time:* *Information/data extraction*
- About all NLP problems
  - Synonym grouping (voice = sound quality)
  - Lexical semantics
  - Coreference resolution
  - …..

The State of Sentiment

- Deep learning
  - You must have heard a lot about it.
- Lifelong learning: It works.
  - Lifelong sentiment classification
  - Lifelong aspect extraction
- …

Classic ML paradigm

- Learning in isolation: Given a dataset,
  - run a machine learning (ML) algorithm on the dataset to learn a function
    - without considering any related information or past learned knowledge.
- Existing ML algorithms such as
  - Deep NN, SVM, NB, CRF, and topic models
  - have been very successful.
- Let's call this learning paradigm: Machine Learning (ML) 1.0.

Limitations of the classic paradigm

- But such “isolated learning” has weaknesses.
  - No memory: Knowledge learned is not retained.
    - Knowledge is not cumulative.
    - Cannot learn by leveraging past learned knowledge
  - Needs a large number of training examples.
    - Humans can learn effectively from a few examples.
- Humans never learn in isolation.
  - Always use prior knowledge to help learning

Lifelong learning (LL)

- Learn as humans do.
  - *lifelong (machine) learning*
    - Retain learned knowledge from previous tasks and use it to help future learning tasks
- Let us call this paradigm Machine Learning 2.0
  - Ultimate goal of machine learning
- Big data provides an excellent opportunity for LL
  - Especially, big text data, of all domains/topics.
  - Extensive sharing of concepts across tasks/domains due to the nature of the natural language

(1) Lifelong sentiment classification
(Chen, Ma and Liu, 2015)

- *"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. ...."*
- Goal: Classify docs or sentences as + or -.
  - Need to manually label a lot of training data for each domain, which is very labor-intensive
  - Can we not label for every domain or at least not so many docs/sentences?

Exploiting past information/data

- But it is “well known” that the sentiment classifier (SC) built for domain A will not work for domain B.
  - E.g., SC built for “camera” will not work for “earphone”
- Classic solution: transfer learning
  - Using labeled data in a past domain S (camera) to help learning in the target domain T (earphone). T may have only unlabeled data.
  - If S and T are very similar, S can help.
- May not be a great solution!

Lifelong learning

Imagining - we have worked on a *large number of past domains* with their training data P.

- do we need any data from T?

- No (in many cases) –
  - A naive *lifelong learning* method works wonders.
    - Improve accuracy by as much as 19% (= 80%-61%)
- Yes (in some others): e.g., we build a SC using P, but it works poorly for toy reviews.
  - Why? Because of "toy."

Exploiting Knowledge via Penalties

- Domain dependent sentiment words

$$\frac{1}{2}\alpha \sum_{w \in V_T} \left(\left(X_{+,w} - N^t_{+,w} \right)^2 + \left(X_{-,w} - N^t_{-,w} \right)^2 \right)$$

- Domain-level knowledge: If a word appears in one/two past domains, the knowledge associated with it is probably not reliable or general.

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

One result

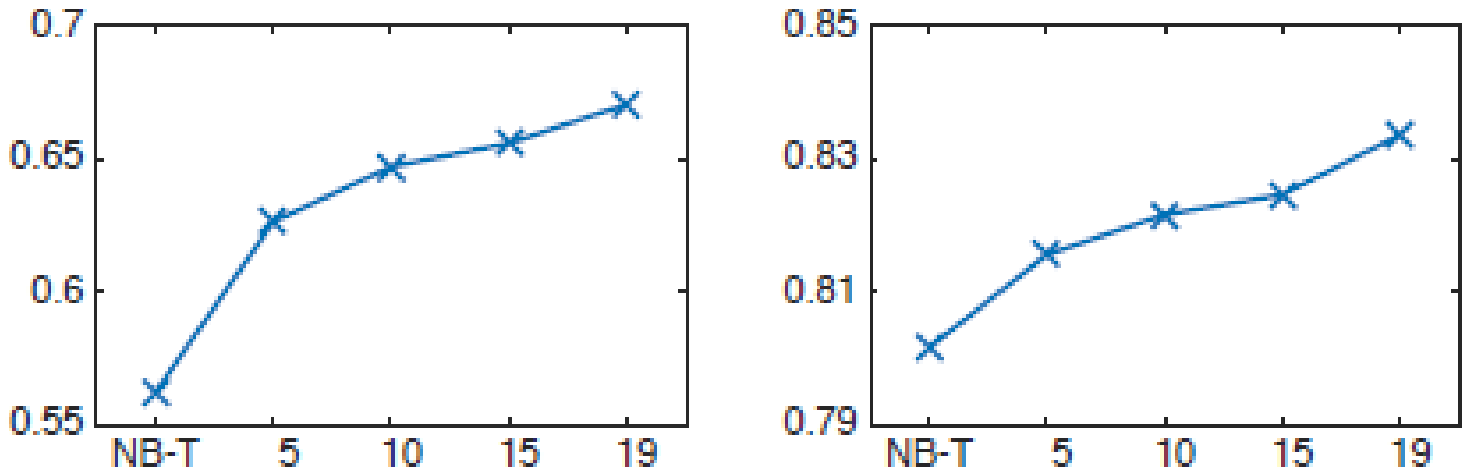


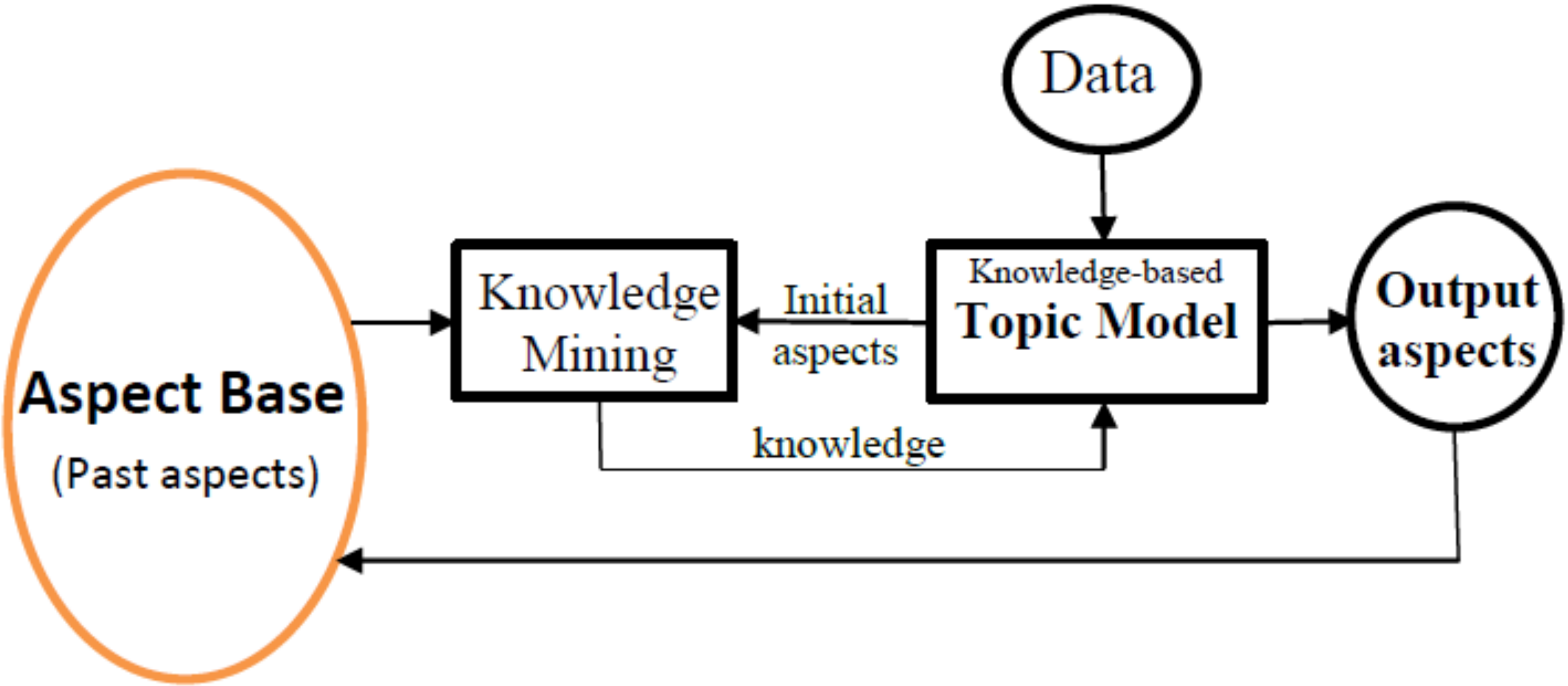
Figure 1: Figure 1. (Left): Negative class F1-score of LSC with #past domains in natural class distribution. (Right): Accuracy of LSC with #past domains in balanced class distribution.

(2) Lifelong aspect extraction
(Chen and Liu, ICML-2014, KDD-2014)

- "*The battery life is long, but pictures are poor*."
  - Aspects: battery life, picture
- Observation:
  - 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*.
  - It is rather "silly" not to exploit such sharing in learning or extraction.

Lifelong Topic Modeling (LTM)
(Chen and Liu, ICML-2014)

- for aspect extraction



Experiment results

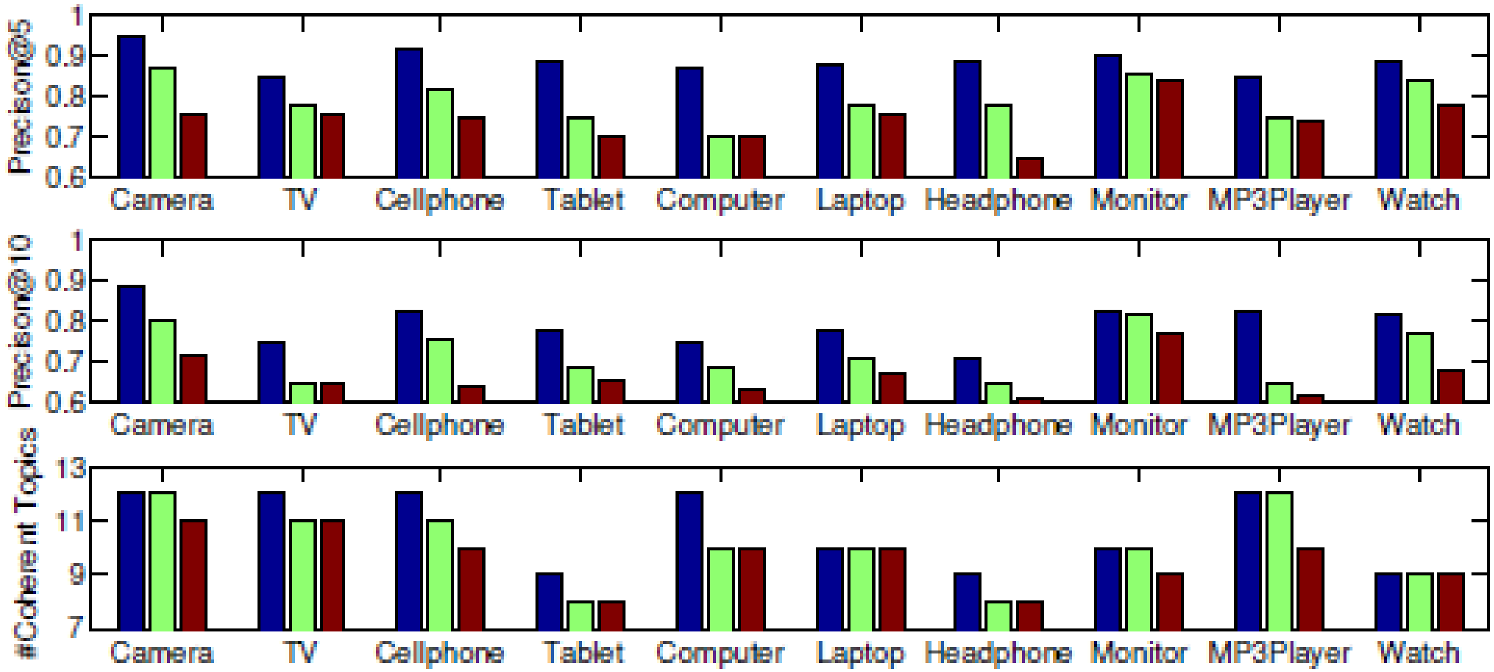


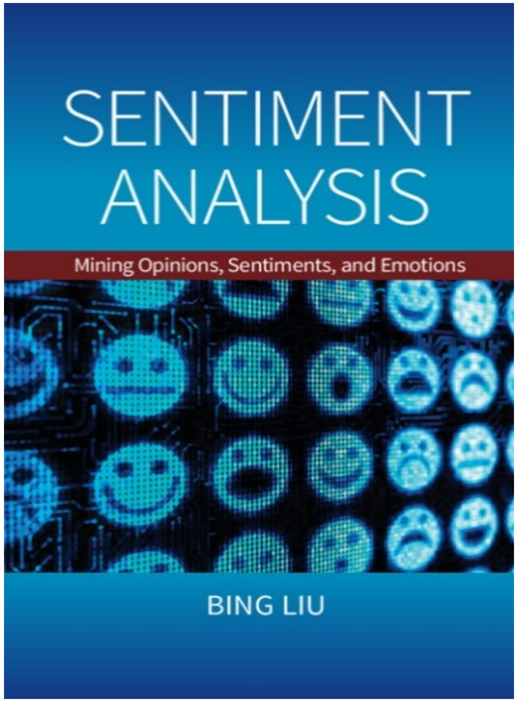
Figure 2. Top & Middle: Topical words *Precision*@5 & *Presicion*@10 of coherent topics of each model respectively; Bottom: number of coherent (#Coherent) topics discovered by each model. The bars from left to right in each group are for LTM, LDA, and DF-LDA. On average, for *Precision*@5 and

Summary

- SA is a well-defined semantic analy. problem
  - Two key concepts form its core
    - (1) sentiment and (2) sentiment target or aspect
- Observation: Due to highly focused nature, SA tasks have a significant amount of concept sharing across application domains
  - which makes *lifelong learning* very promising
  - Exploit such sharing and the big data to push SA to a new level.

More information

- (New book) B. Liu. *Sentiment Analysis*: *Mining Opinions, Sentiments, and Emotions*. Cambridge University Press, 2015.



- My sentiment analysis page
  - http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html