# Sentiment Analysis and Lifelong Machine Learning

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

# Introduction: my goal

- Briefly introduce sentiment analysis (SA)
- Using SA as a platform for *lifelong learning*
    - Statistical "one-shot" learning methods such as deep NN, SVM, NB, and CRF, have their limits
        - How to go beyond these algorithms?
    - Lifelong learning: retain knowledge learned in the past to help future learning & problem solving
        - Without it, an AI system will never be intelligent
        - Solving a SA problem using *lifelong learning* and *big data* (Chen & Liu, ICML, KDD, 2014).

# Outline

- What is sentiment analysis (SA)?
- SA: a very rich NLP problem
  - SA is not a single problem
  - Aspect sentiment classification
  - Aspect extraction
  - …
- Aspect extraction: lifelong learning & big data
- Summary

# SA: A fascinating problem

- Sentiment analysis (SA) or opinion mining
  - computational study of opinion, sentiment, appraisal, evaluation, attitude, and emotion.
- Popular research & unlimited applications.
  - popular research area in NLP, DM, etc.
  - Spread from CS to management & social sciences
  - More than >300 companies in USA alone.
- Intellectually challenging.
  - CEO 1: "*Anyone who claims >70% accuracy is lying*"
  - CEO 2: "*Our system is as bad as everyone else's*"

# Definition of an 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 **clear** 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*)
- 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)

- An *opinion* is a quintuple

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

  - ***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.

- Called *Aspect-based sentiment analysis*
  - Used in industry

# 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 clear too. It is much better than my old Blackberry. However, my mom was mad with me as I did not tell her before I bought the phone. She thought the phone was too expensive, …*"
- In quintuples

  (iPhone, GENERAL, +, Abc123, 5-1-2008)

  (iPhone, touch_screen, +, Abc123, 5-1-2008)

  (iPhone, voice_quality, +, Abc123, 5-1-2008

  (Abc123, GENERAL, -, momOfAbc123, unknown)
  (iPhone, price, -, momOfAbc123, unknown)

# Outline

- What is sentiment analysis (SA)?
- SA: a very rich NLP problem
  - SA is not a single problem
  - Aspect sentiment classification
  - Aspect extraction
  - …
- Aspect extraction: lifelong learning & big data
- Summary

# 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
- Other NLP problems
  - Synonym grouping: voice = sound quality
  - Lexical semantics: aspect(expensive) = price
  - Coreference resolution:
    - "The sound of this phone is great. ***It*** is cheap too."
    - "The sound of this phone is great. ***It*** is crystal clear."
  - …..

# Aspect sentiment classification

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

- Lexicon-based approach: Opinion words/phrases
  - Parsing: simple sentences, compound sentences, conditional sentences, questions, modality verb tenses, etc (Hu and Liu, 2004; Ding et al. 2008; Narayanan et al. 2009).
- Supervised learning is tricky:
  - Feature weighting: consider distance between word and target entity/aspect (e.g., Boiy and Moens, 2009)
  - Use a parse tree to generate a set of target dependent features (e.g., Jiang et al. 2011)

# Aspect extraction

- "*The battery life is long, but pictures are poor*."
  - Aspect terms: battery life, picture
- Main approaches
  - **Frequency-based approach**: frequent noun phrases (Hu and Liu, 2004).
  - **Syntactic dependency**: opinion and target relation (Hu and Liu 2004; Zhuang et al 2006; Qiu et al., 2009, etc).
    - E.g., "*The pictures are great*"
  - **Supervised sequence labeling** (e.g., HMM, CRF) (Liu et al. 2005; Jin and Ho, 2009; Jakob and Gurevych, 2010, etc)
  - **Topic modeling** (Mei et al., 2007; Titov et al., 2008; etc)

# Aspect extraction: topic modeling

- 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*, …}
    - Ranked based on probabilities (not shown).

# Current extraction paradigm

- Existing methods basically work as follows:
  - Given a collection of opinion documents D,
    - Run an extraction algorithm (supervised or unsupervised learning) on *D* to extract aspects
      - "One-shot" approach
- Results are still not great
  - No matter what you do with the current "one-shot" learning algorithms, I think it is hard to make a major progress.
  - How to improve further?

# Outline

- What is sentiment analysis (SA)?
- SA: a very rich NLP problem
  - SA is not a single problem
  - Aspect sentiment classification
  - Aspect extraction
  - …
- Aspect extraction: lifelong learning & big data
- Summary

# Machine learning (ML) for SA & NLP

- SA and NLP apply many ML algorithms.
- Although ML researchers can still improve them, statistical learning algorithms have their limits, which are fast approaching.
- Then what? How can we progress beyond such "one shot" algorithmic ML algorithms?
  - I believe that answer is: lifelong learning
    - Learn as humans do.
  - Exploiting diverse big data

# Learn as humans do

- We learn and retain the learned knowledge and use the knowledge to help future learning
- Without this lifelong learning (LL) capability,
  - we will never solve the NLP problem.
  - we will never have an intelligent system
- SA/NLP provides an excellent platform for LL
  - Same words/phrases have similar meanings in different domains, which enables easy transfer of knowledge

# 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 rather "silly" not to exploit such sharing in learning

# Big data and aspect sharing

- Why using SA for lifelong learning?
  - Online reviews: Excellent data with extensive sharing of aspect/concepts across domains
  - Hard to find suitable data in other appl. areas
- Why big (and diverse) data?
  - Learn a broad range of reliable knowledge. More knowledge makes future learning easier.

# Lifelong topic modeling (LTM)
(Chen and Liu, ICML-2014)

- For aspect extraction
- Top modeling (Blei et al 2003) find topics from a collection of documents. Topics are aspects.
  - A document is a distribution over topics
  - A topic is a distribution over terms/words, e.g.,
    - {*price*, *cost*, *cheap*, *expensive*, …}
- Questions:
  - How to find good past knowledge and how to use them to help future extraction?

# What knowledge?

- 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}

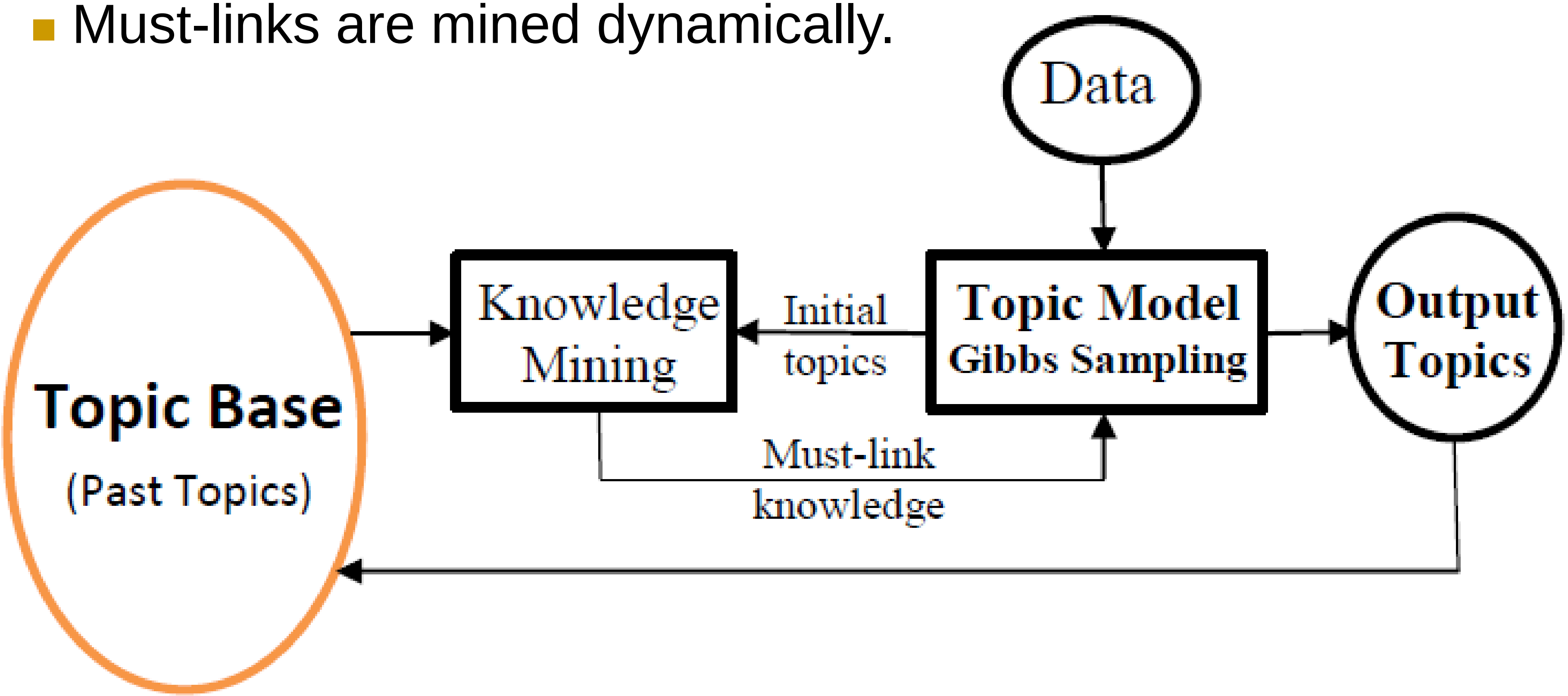
# Problem statement

- Given **a large set of** document collections (big data), $D = \{D_1, \ldots, D_n\}$, learn from each $D_i$ to produce the result $S_i$. Let $S = \cup S_i$
  - $S$ is called the *topic base*
- **Goal**: Given a test/new collection $D^t$, learn from $D^t$ with the help of $S$ (and possibly $D$).
  - $D^t \in D$ or $D^t \notin D$.
  - The results learned this way should be better than without the guidance of $S$ (and $D$).

# Lifelong Topic Modeling (LTM)

(Chen and Liu, ICML-2014)

- Must-links are mined dynamically.



# LTM topic model

- Step 1: Runs a topic model (e.g., LDA) on each $D_i \in D$ to produce a set of topics $S_i$ called *p-topics*.
- Step 2: (1) Mine prior knowledge (*must-links*) (2) use prior knowledge to guide modeling.

---

**Algorithm 2** LTM($D^t$, $S$)

---

1: $A^t \leftarrow$ GibbsSampling($D^t$, $\emptyset$, N); // Run $N$ Gibbs iterations with no knowledge (equivalent to LDA).
2: **for** $i = 1$ **to** $N$ **do**
3: $K^t \leftarrow$ KnowledgeMining($A^t$, $S$);
4: $A^t \leftarrow$ GibbsSampling($D^t$, $K^t$, 1); // Run with knowledge $K^t$.
5: **end for**

---

# Knowledge mining function

- Topic match: find similar topics ($M^t_{j^*}$) from p-topics for each current topic
- Pattern mining: find frequent itemsets from $M^t_{j^*}$

---

**Algorithm 3** KnowledgeMining($A^t$, $S$)

---

**for** each p-topic $s_k \in S$ **do**
    $j^* = \min_j$ KL-Divergence($a_j, s_k$) for $a_j \in A^t$;
    **if** KL-Divergence($a_{j^*}, s_k$) $\leq \pi$ **then**
        $M^t_{j^*} \leftarrow M^t_{j^*} \cup s_k$;
    **end if**
**end for**
$K^t \leftarrow \cup_{j^*}$ FIM($M^t_{j^*}$); // Frequent Itermset Mining.

---

# An example

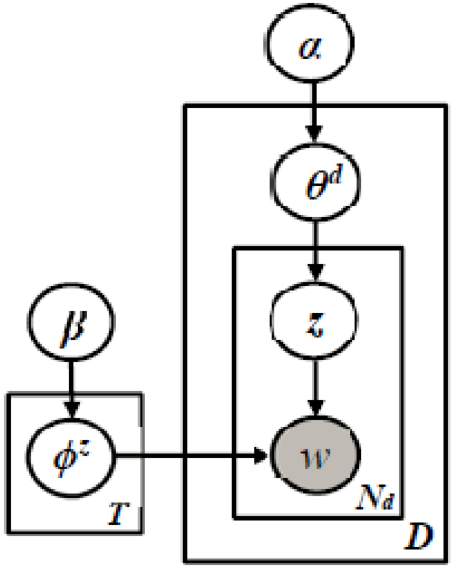
- Given a newly discovered topic:

  {*price*, *book*, *cost*, *seller*, *money*},

  - We find 3 matching topics from topic base *S*
    - Domain 1: {*price*, *color*, *cost*, *life*, *picture*}
    - Domain 2: {*cost*, *screen*, *price*, *expensive*, *voice*}
    - Domain 3: {*price*, *money*, *customer*, *service*, *expensive*}
- If we require words appear in at least two domains, we get two must-links (knowledge):
  - {*price*, *cost*} and {*price*, *expensive*}.
  - Each set is likely to belong to the same aspect/topic.

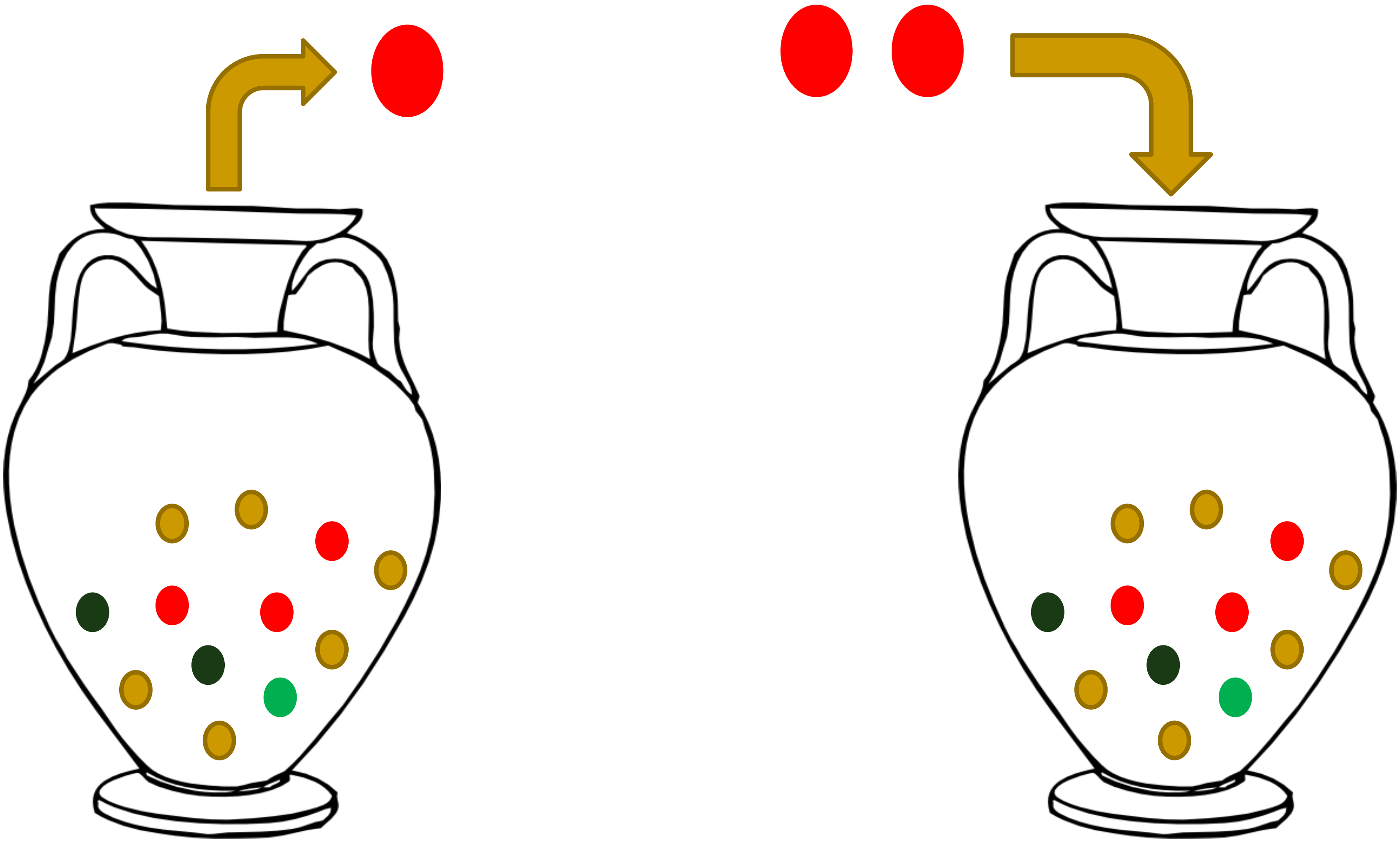
# Model inference: Gibbs sampling

- How to use the *must-links* knowledge?
  - e.g., {*price*, *cost*} & {*price*, *expensive*}

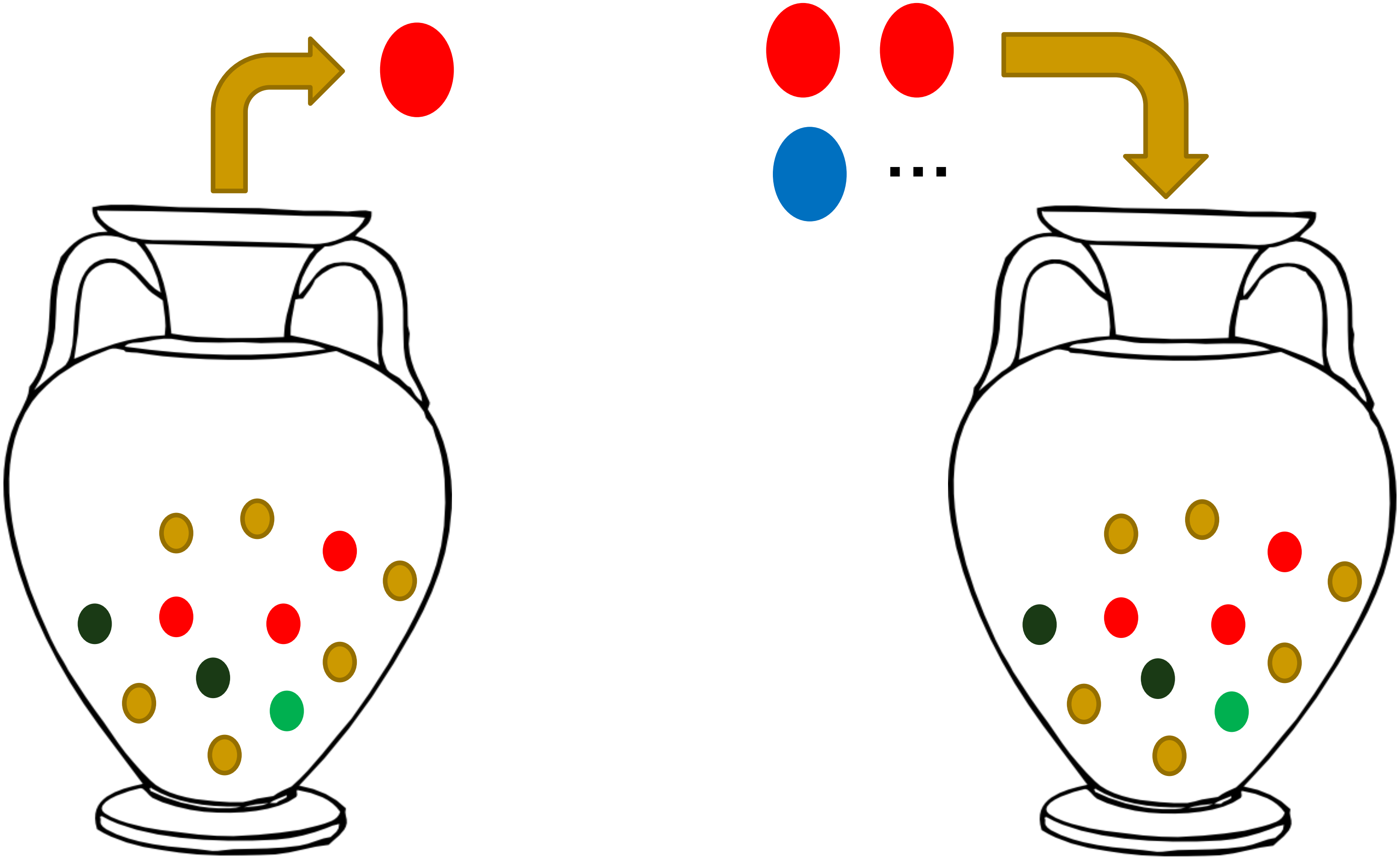



- Graphical model: same as LDA
- But the model inference is very different
  - Generalized Pólya Urn Model (GPU)
- Idea: When assigning a topic *t* to a word *w*, also assign *a fraction of t* to words in must-links sharing with *w*.

# Simple Pólya Urn model (SPU)



# Generalized Pólya Urn model (GPU)



# Gibbs sampler for GPU

- $P\left(z_i = t \,\middle|\, \mathbf{z}^{-i}, \mathbf{w}, \alpha, \beta\right) \propto$

$$\frac{n_{m,t}^{-i} + \alpha}{\sum_{t'=1}^{T}\left(n_{m,t'}^{-i} + \alpha\right)} \times \frac{\sum_{w'=1}^{V} \mathbb{A}_{w',w_i} \times n_{t,w'}^{-i} + \beta}{\sum_{v=1}^{V}\left(\sum_{w'=1}^{V} \mathbb{A}_{w',v} \times n_{t,w'}^{-i} + \beta\right)}$$

# Experiment results

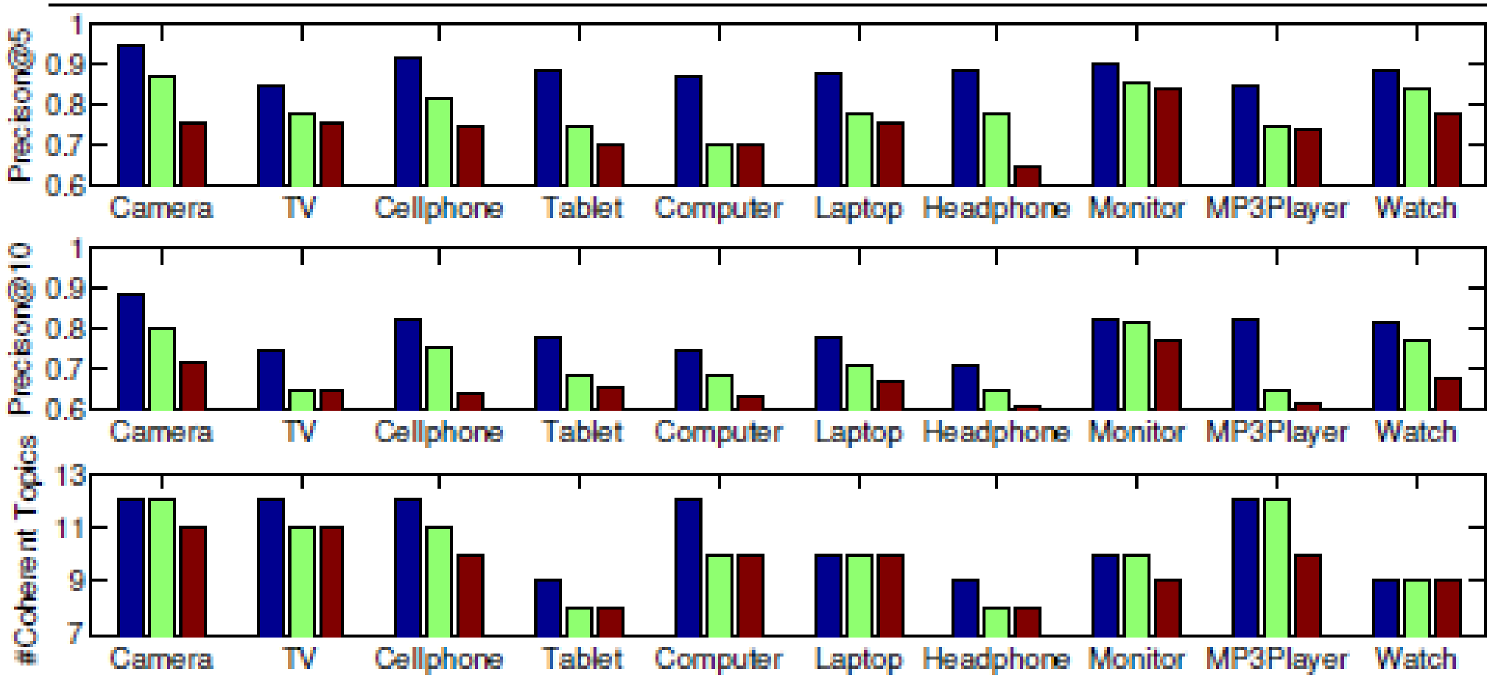



*Figure 2.* Top & Middle: Topical words $Precision@5$ & $Presicion@10$ of coherent topics of each model respectively; Bottom: number of coherent (#Coherent) topics discovered by each model. The bars from left to right in each group are for LTM, LDA, and DF-LDA. On average, for $Precision@5$ and

# Outline

- What is sentiment analysis (SA)?
- SA: a very rich NLP problem
  - SA is not a single problem
  - Aspect sentiment classification
  - Aspect extraction
  - …
- Aspect extraction: lifelong learning & big data
- Summary

# Summary

- SA is a well-defined semantic analysis problem
  - Two key concepts: sentiment and target
- Due to extensive concept/aspect sharing across domains of SA data,
  - SA offers an excellent platform for intelligent and continuous learning – lifelong learning
- Using lifelong learning with big data
  - We can make NLP & AI systems more intelligent.

# More information

- Z. Chen & B. Liu. Topic Modeling using Topics from Many Domains, Lifelong Learning, and Big Data. *ICML-2014*.
- Z. Chen & B. Liu. Mining Topics in Documents: Standing on the Shoulders of Big Data. *KDD-2014.*
- (New book) B. Liu. *Sentiment Analysis: Mining Opinions, Sentiments and Emotions*. Cambridge Univ. Press, March 2015 (about 365 pages).