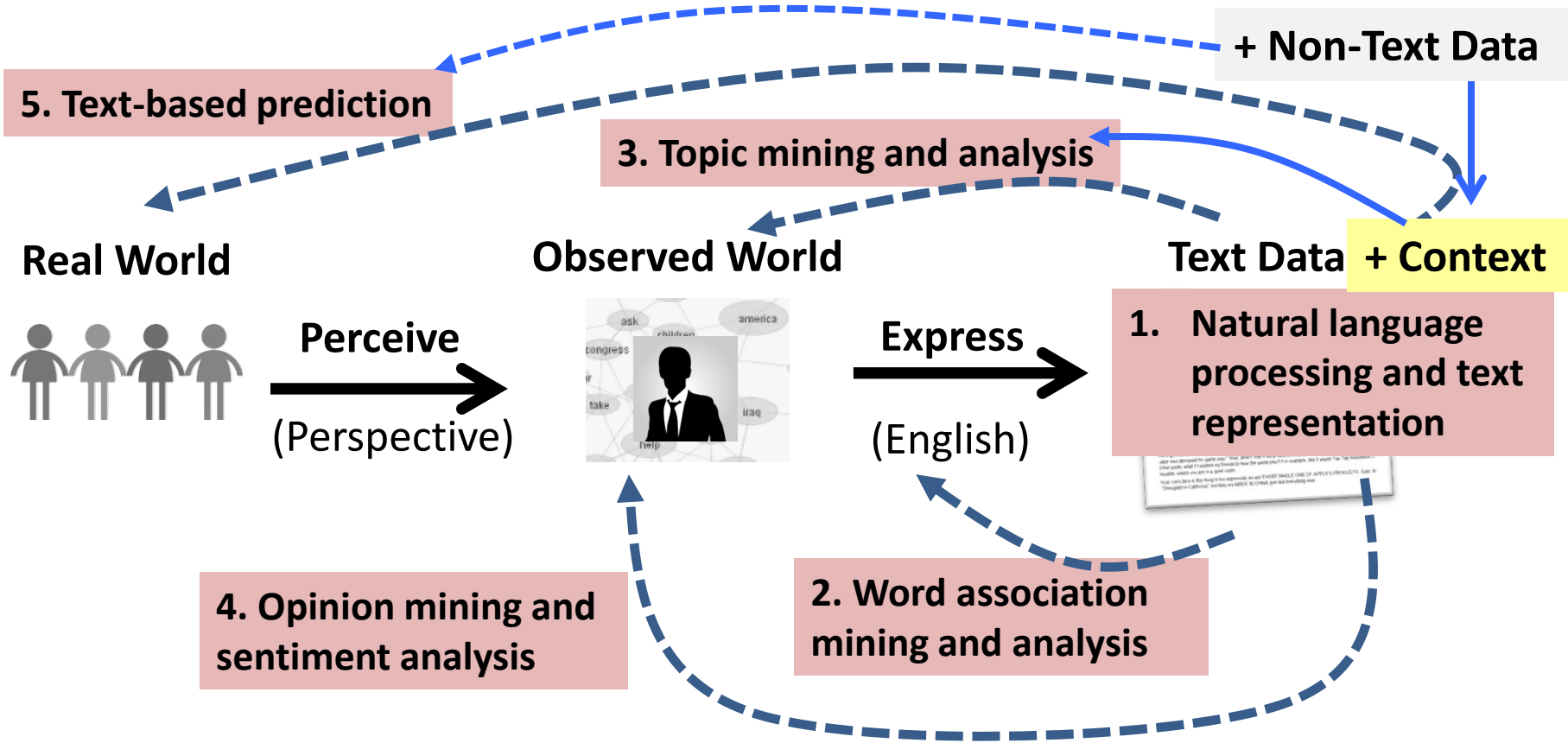
Course Summary

ChengXiang “Cheng” Zhai
Department of Computer Science
University of Illinois at Urbana-Champaign

Topics Covered in This Course



Key High-Level Take-Away Messages

13. Joint mining of text and non-text
14. Contextual PLSA
15. NetPLSA
16. Causal topic mining

6. Probabilistic Topic Model (PLSA, LDA)
7. Generative model; ML estimate; EM
8. Text clustering: model vs. similarity-based
9. Text categorization: generative vs. discriminative
10. Evaluation of clustering and categorization

Perceive

(Perspective)

1. NLP → Text representation → Knowledge discovery
2. Robust TM = Word-based rep + Statistical analysis

(English)

11. Sentiment classification: ordinal regression
12. Latent Aspect Rating Analysis

3. Paradigmatic and syntagmatic relations
4. Text similarity: Vector space, BM25
5. Co-occurrence analysis: Entropy, MI

What to Learn Next

- **Natural Language Processing**
  - Foundation for all text-based applications
  - More NLP ➔ Deeper knowledge discovery
- **Statistical Machine Learning**
  - Backbone techniques for NLP and text analysis
  - Key to predictive modeling and "big data" applications
- **Data Mining**
  - General data mining algorithms can always be applied to text
- **Text/Information Retrieval**
  - Essential system component in any text-based application (human in the loop)
  - Some techniques useful for text data mining

Main Techniques for Harnessing Big Text Data: Text Retrieval + Text Mining