Overview Text Mining and Analytics

ChengXiang "Cheng" Zhai
Department of Computer Science
University of Illinois at Urbana-Champaign
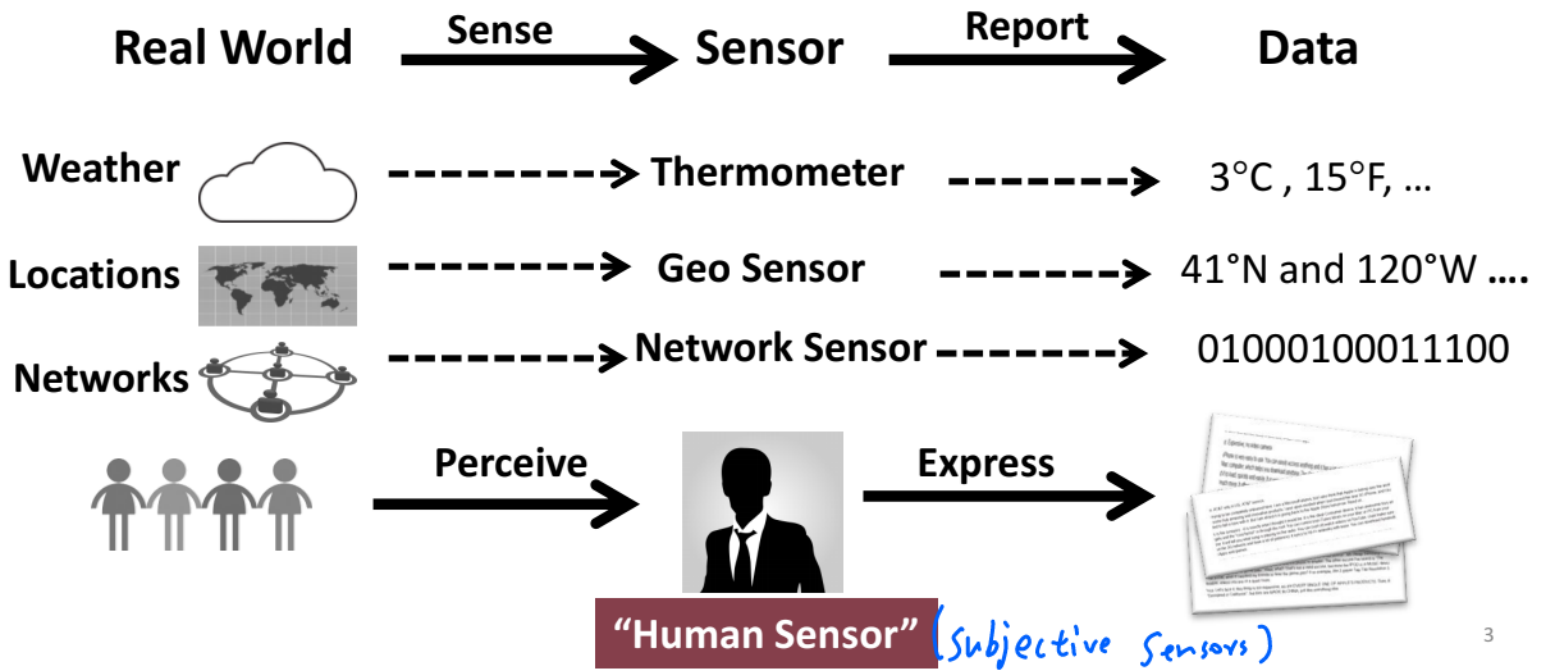
Text Mining and Analytics

process   *result*
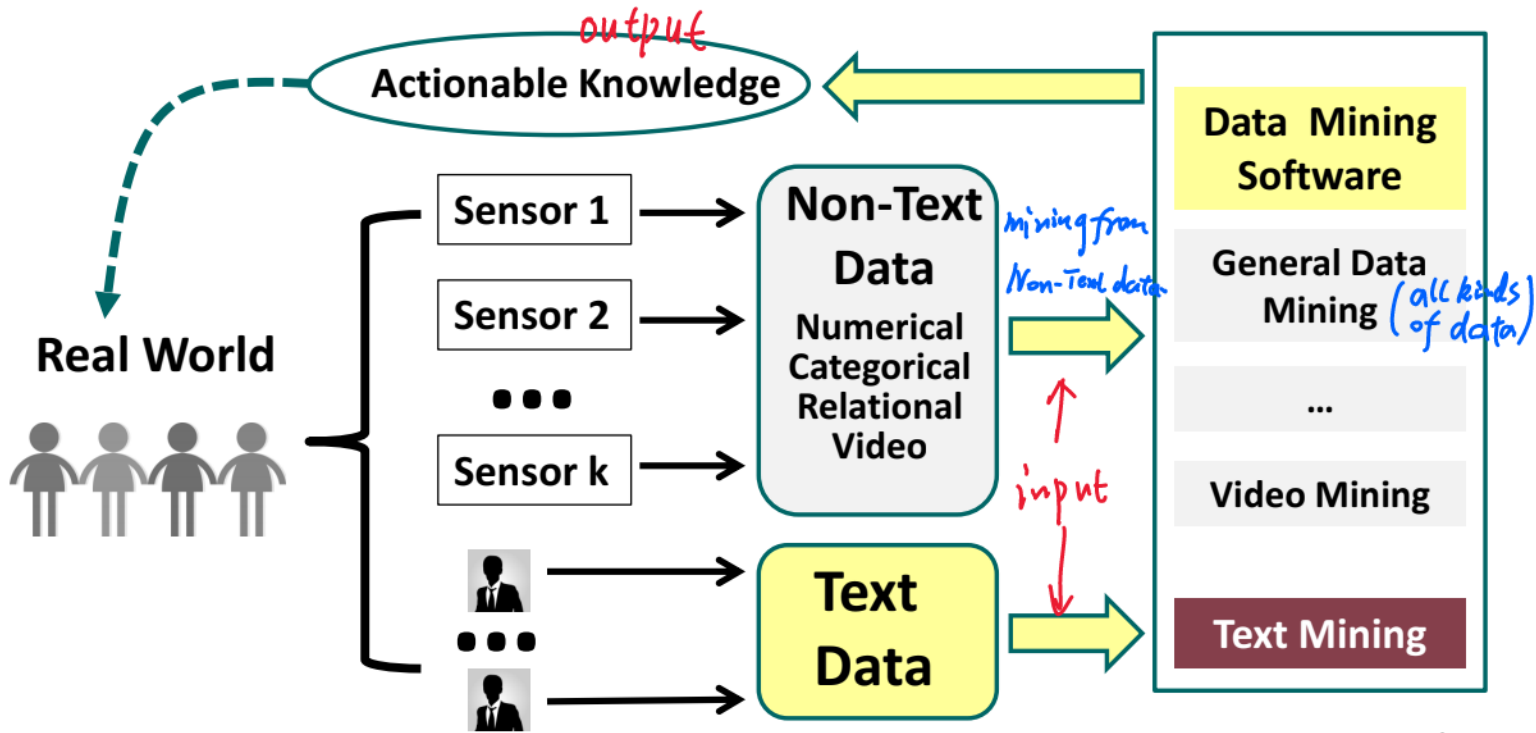
- Text mining ≈ Text analytics
- Turn text data into **high-quality information** or **actionable knowledge**   *goal*
  - **Minimizes human effort** (on consuming text data)
  - Supplies knowledge for **optimal decision making**
- Related to **text retrieval**, which is an essential component in any text mining system
  - Text retrieval can be a preprocessor for text mining
  - Text retrieval is needed for knowledge provenance

Text vs. Non-Text Data:
Humans as Subjective "Sensors"

Real World		Sense →	Sensor	Report →	Data
Weather	☁	- - - →	Thermometer	- - - →	3°C , 15°F, …
Locations	🗺	- - - →	Geo Sensor	- - - →	41°N and 120°W ….
Networks		- - - →	Network Sensor	- - - →	01000100011100

Perceive → "Human Sensor" (Subjective Sensors) → Express →

3

The General Problem of Data Mining

Actionable Knowledge — *output*

Real World

Sensor 1 →

Sensor 2 →

••• → **Non-Text Data**
Numerical
Categorical
Relational
Video

Sensor k →

→ **Text Data**

•••

Data Mining Software

General Data Mining *(all kinds of data)*

...

Video Mining

Text Mining

mining from Non-Text data

input

The Problem of Text Mining

Actionable Knowledge

Real World

Non-Text Data

Text Data

Joint Mining of Text + Non-Text
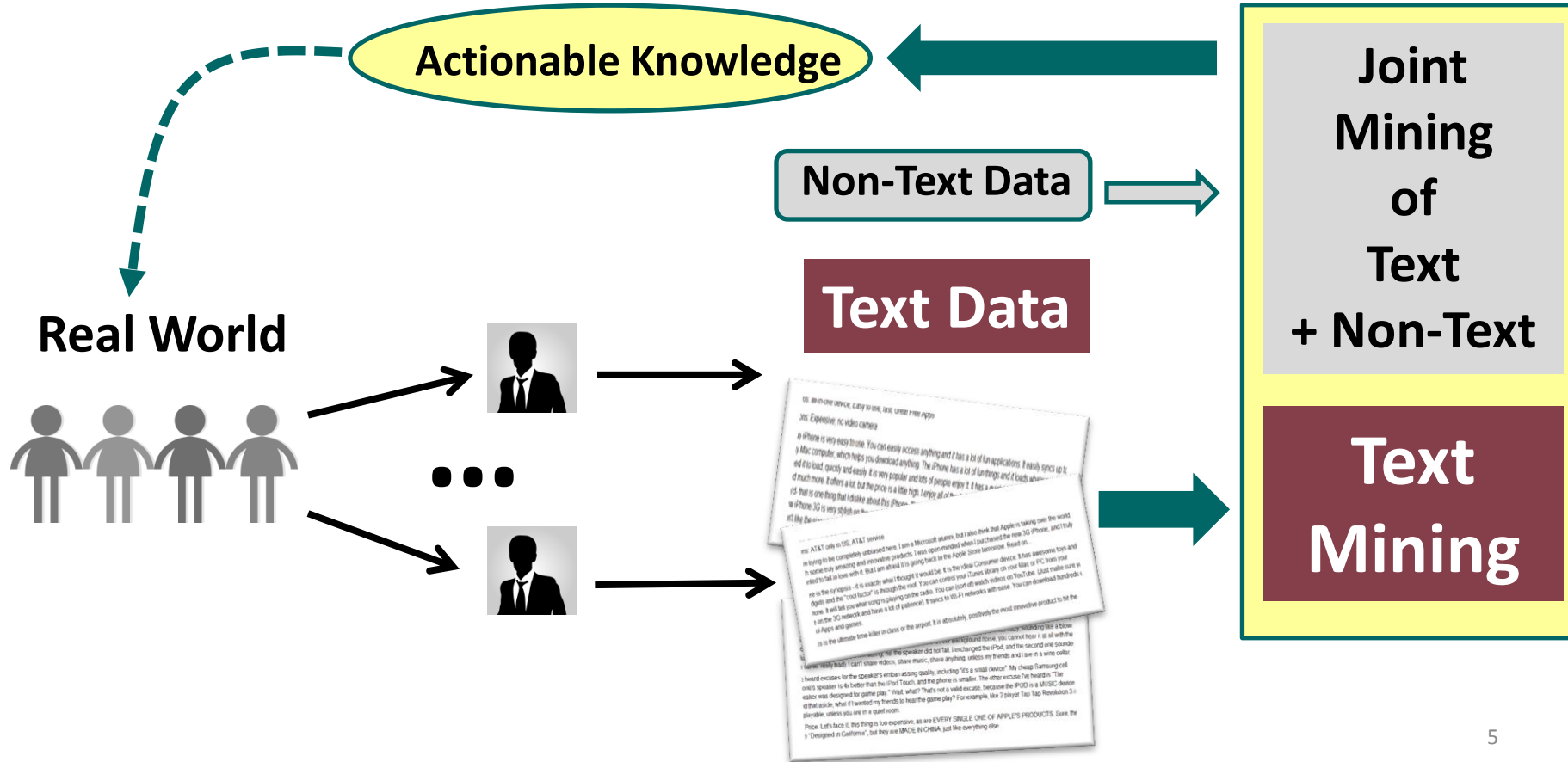
Text Mining

Landscape of Text Mining and Analytics

+ Non-Text Data

4. Infer other real-world variables (predictive analytics)

2. Mining content of text data

Real World

Observed World   *can be biased*   Text Data   **+ Context**

Perceive
(Perspective)

Express
(English)

3. Mining knowledge about the observer

1. Mining knowledge about language

6
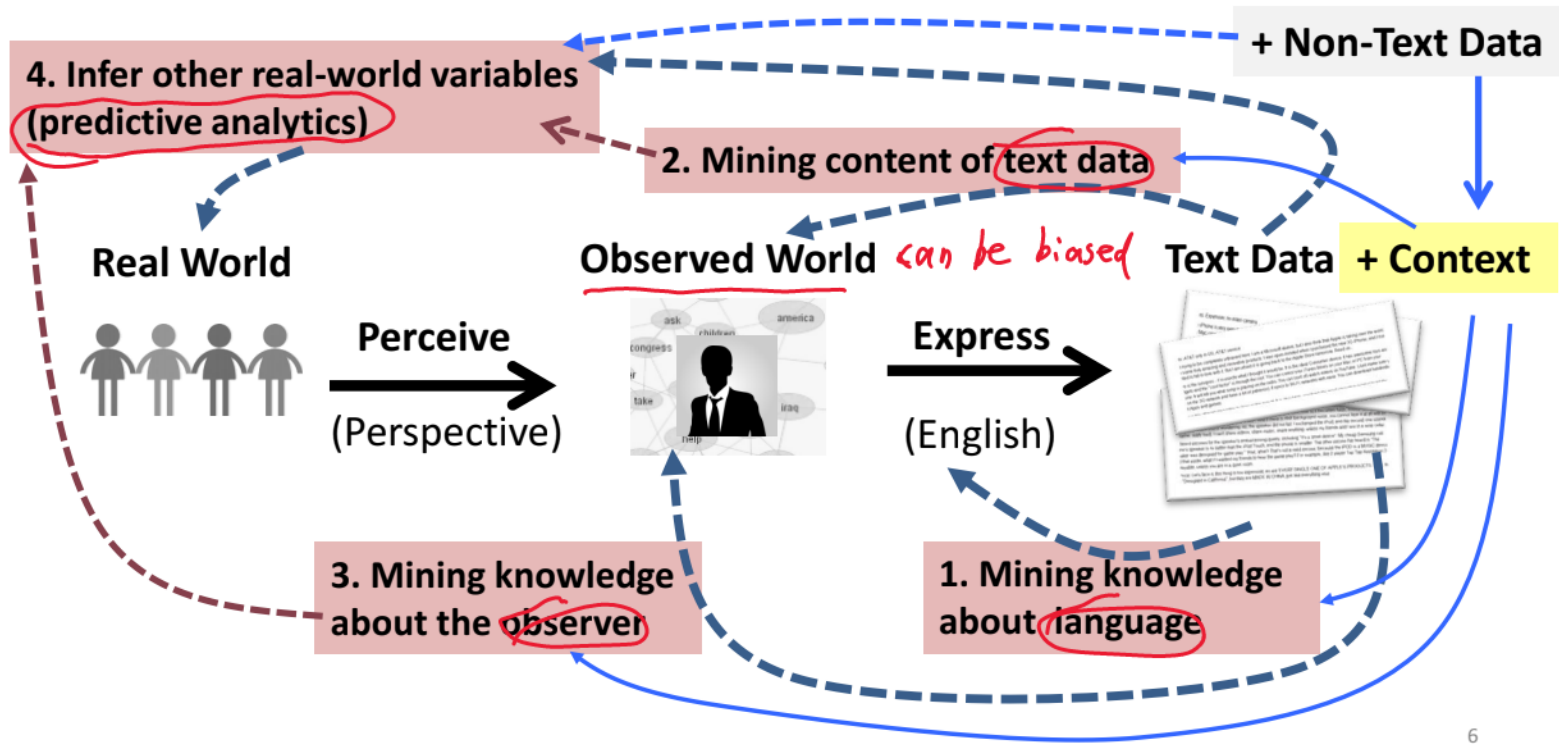
Topics Covered in This Course

5. Text-based prediction

3. Topic mining & analysis

Real World　　　　**Observed World**　　　　**Text Data**

Perceive

(Perspective)

Express

(English)

1. **Natural language processing & text representation**

4. Opinion mining & sentiment analysis

2. Word association mining & analysis