

Sentiment Analysis and The Fourth Paradigm

MSE 2400 EaLiCaRA
Dr. Tom Way

Necessary Background

- Fourth Paradigm
- Natural Language Processing
- Machine Learning
- Sentiment
- Sentiment Analysis
- Sentiment Tracking
- Word Cloud
- Various Tools

Fourth Paradigm

- Scientific breakthroughs powered by advanced computing capabilities that help researchers manipulate and explore massive datasets.

Nth paradigm?

The historian of science Thomas Kuhn gave paradigm its contemporary meaning when he adopted the word to refer to **the set of practices that define a scientific discipline at any particular period of time**. Kuhn himself came to prefer the terms **exemplar** and **normal science**, which have more precise philosophical meanings. However in his book *The Structure of Scientific Revolutions* Kuhn defines a **scientific paradigm** as:

- * what is to be observed and scrutinized
- * the kind of questions that are supposed to be asked and probed for answers in relation to this subject
- * how these questions are to be structured
- * how the results of scientific investigations should be interpreted

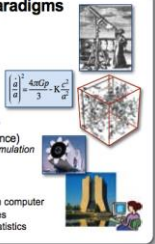
Alternatively, the Oxford English Dictionary defines paradigm as "a pattern or model, an exemplar." Thus an additional component of Kuhn's definition of paradigm is:

- * how is an experiment to be conducted, and what equipment is available to conduct the experiment.

Jim Gray's four scientific paradigms

Science Paradigms

- Thousand years ago: science was **empirical**
describing natural phenomena
- Last few hundred years: **theoretical** branch
using models, generalizations
- Last few decades: a **computational** branch
simulating complex phenomena
- Today: **data exploration** (eScience)
unify theory, experiment, and simulation
 - Data captured by instruments or generated by simulator
 - Processed by software
 - Information/knowledge stored in computer
 - Scientist analyzes database files using data management and statistics



- 1. empiricism**
observe phenomenon and attempt to classify
Ptolemy's universe of concentric spheres
- 2. theory**
describe above classifications with mathematical models
Newtonian/Einsteinian gravity
- 3. computation**
build "virtual" physical systems via solution of math models
Cosmic structure formation
- 4. data-driven synthesis (?)**
unite empirical, theoretical and computational branches with data (X-info and Comp-X)
Matter/energy content of the universe

Natural Language Processing

- Natural language processing (NLP) is a field of computer science, artificial intelligence, and linguistics concerned with the interactions between computers and human (natural) languages.

NLP Examples

- Automatic summarization – generate a summary from original
- Machine translation – Google Translate
- Morphological segmentation – break up words into phonemes
- Natural language generation – generate text automatically
- Natural language understanding – analyze and learn meaning
- Optical character recognition (OCR) – “read” the text
- Parts of speech tagging – identify nouns, verbs, adjectives, etc.
- Question answering – parse a question and generate a response
- Speech recognition – “hear” what is spoken, convert to text
- Word sense disambiguation – pick the correct definition of a word
- Readability – at what grade level is text written?
- And more... such as?

Machine Learning

- Machine Learning (ML), a branch of artificial intelligence, is about the construction and study of systems that can learn from data.

ML Examples

- Decision tree – maps observations to conclusions
- Association rule – discovering relationships between things
- Neural networks – learn to recognize patterns and apply that learning
- Support vector machines – given training data it creates categories, can classify new data into the correct category
- Bayesian networks – fine-tunes probabilities of classifying data as it gets more data
- More...

Sentiment

- What is “sentiment”?
- an attitude, thought, or judgment prompted by feeling
- a specific view or notion
- an idea colored by emotion

Sentiment Analysis

- Sentiment analysis or opinion mining refers to the application of natural language processing, computational linguistics, and text analytics to identify and extract subjective information in source materials.
- It is an example of a Computational Agent

Sentiment Analysis Example

- A basic task in sentiment analysis is classifying the *polarity* of a given text - is it positive, negative, or neutral for some desired measure?
- Advanced, “beyond polarity” sentiment classification looks, for instance, at emotional states such as “angry,” “sad,” and “happy.”

Sentiment Tracking

- Performing Sentiment Analysis over time
- It is a Computational Agent that learns
- Watching the sentiment (positive, neutral, negative) as it changes over time

Word Cloud

- A word cloud (or tag cloud) is a visual representation for text data, typically used to depict keyword metadata (tags) on websites, or to visualize free form text.
- How does it fit into the Fourth Paradigm?

Word Cloud example



Example of tools

- Machine translation: <http://translate.google.com>
- Readability: <http://www.readability-score.com>
- Sentiment analysis: <http://www.sentimentanalyzer.appspot.com>
- Word cloud: <http://www.tagxedo.com>