week2

May 17, 2018

1 Week 2 Overview

During this week's lessons, you will learn more about word association mining with a particular focus on mining the other basic form of word association (i.e., syntagmatic relations), and start learning topic analysis with a focus on techniques for mining one topic from text.

1.1 Goals and Objectives

After you actively engage in the learning experiences in this module, you should be able to:

- Explain how to discover syntagmatic relations from text data
- Explain the computation task of mining and analyzing topics in text data, particularly its input and the expected output.
- Explain the problems with defining a topic as just one term when mining and analyzing topics in text data.
- Explain the limitations of using one term to represent a topic and how they can be addressed by representing a topic as a distribution over words.
- Explain basic concepts in statistical language models such as "language model", "unigram language model", "likelihood", Maximum Likelihood estimate.
- Explain how to mine one topic from a text document, i.e., estimate a unigram language model

1.2 Key Concepts

- Entropy
- Conditional entropy
- Mutual information
- Topic and coverage of topic
- Language model
- Generative model
- Unigram language model
- Word distribution
- Background language model
- Parameters of a probabilistic model
- Likelihood
- Bayes rule
- Maximum likelihood estimation
- Prior and posterior distributions