Language Modeling

2 units (selection)

Kenji Kita · Professor / Applied Information Media Engineering, Information Science and Intelligent Systems, Systems Innovation Engineering
Masami Shishibori · Associate Professor / Information Science, Information Science and Intelligent Systems, Systems Innovation Engineering

Target〉 The course introduces some probabilistic models of natural language.

Outline〉 Due to the availability of large text corpora, probabilistic approaches to modeling natural language text have become dominant in recent years. This lecture gives an overview of probabilistic modeling of natural language, including n-gram models, hidden Markov models, probabilistic grammars, and maximum entropy models, as well as their applications to natural language processing and information retrieval.

Style〉 Lecture in combination with Portfolio

Keyword〉 ***natural language, probabilistic model, n-gram model, probabilistic grammar***

Fundamental Lecture〉 "**Automata and Formal Languages**"(1.0)

Relational Lecture〉 "**Natural Language Understanding**"(0.5), "**Advanced Machine Translation**"(0.5)

Goal〉 To acquire effective techniques for modeling natural language texts using probabilistic models.

Schedule〉

1. Overview of the course
2. Modeling natural language
3. Estimation and evaluation of probabilistic models
4. N-gram model
5. Hidden Markov model
6. Maximum entropy model
7. Probabilistic grammar
8. Partitions of numbers and Young diagram
9. Symmetric group and its action on polynomials
10. Symmetric form and Young diagram
11. Bumping game
12. Sliding game
13. Product operations on Young tableaux
14. Word problem
15. Recent topics
16. Assignment

Evaluation Criteria〉 Assignment count 100%.

Textbook〉 To be introduced in the class.

Contents〉 http://cms.db.tokushima-u.ac.jp/cgi-bin/toURL?EID=216636

Student〉 Able to be taken by only specified class(es)

Contact〉

⇒ Kita (Dr503, +81-88-656-7496, kita@is.tokushima-u.ac.jp) Mail (Office Hour: Tuesday 12:50 - 14:20)

Note〉 Invited talk by a part-time lecturer will be given.