

地域科学特別演習 I

8 単位 (必修) 1 年 (通年), 2 年 (通年)

石田 基広・准教授 / 地域科学専攻 (博士前期課程) 地域創生

【授業目的】 データマイニングあるいはテキストマイニングを応用した修士論文の指導を行う

【授業概要】 大量のテキストデータを対象に統計解析環境 R や, C, Java などの言語を用いて各種自然言語処理を行う。また自然言語処理を施したテキスト群を対象に, 各種の統計的指標を検討する。さらには文書・ターム行列を作成し, 各種のクラスタリング手法を適用するなどを行う。

【キーワード】 統計学, データマイニング, テキストマイニング

【履修上の注意】 コンピュータに慣れておくこと

【到達目標】 自然言語処理の基本技術を習得する。また構造分割されたテキストから各種の統計量を抽出し, 統計的に解析あるいは検定する手法と, その結果の解釈方法を学ぶ。

【授業計画】

1. テキストとは何か
2. 日本語テキストを解析するための予備知識
3. 形態素分割と, 係り受け解析
4. 文書・ターム行列とは何か
5. 各種クラスタリング手法の理論
6. テキスト分析のための統計的手法の理論
7. テキスト分析のための統計的手法の応用
8. ブログや WEB データの解析事例
9. アンケート分析への応用事例
10. 実践的分析 1
11. 実践的分析 2
12. 成果の検討 1
13. 成果の検討 2
14. 発展的解析手法の紹介
15. まとめ

【教科書】 『R によるテキストマイニング入門』 森北書店

【参考書】 『R の基礎とプログラミング技法』 シュプリンガー・ジャパン

【WEB 頁】 <http://cms.ias.tokushima-u.ac.jp>

【授業コンテンツ】 <http://cms.db.tokushima-u.ac.jp/cgi-bin/toURL?EID=218109>

【連絡先】

⇒ 石田 (ishida-m@ias.tokushima-u.ac.jp) MAIL