

クラスター分析 Cluster Analysis

科目コード(Course Number) 20DIFe05

複合科学研究科 School of Multidisciplinary Sciences 情報学専攻
Department of Informatics 知能システム科学 Intelligent Systems Science

学年(Recommended Grade) 1年 2年 3年 4年 5年

2単位(credit) 後学期 2nd semester

Michael Houle

〔授業の概要 Outline〕

本講義では、知識発見のためのクラスター分析のテーマを取り巻く理論上、実用上の問題を扱う。アプリケーションやサポートに必要なデータ構造とともに、クラスタ化戦略を比較再検討する。特に、検討するさまざまな手法のスケラビリティと適用可能性に関するデータ表現の含意とアルゴリズム的設計選択に注目する。

This course deals with the theoretical and practical issues surrounding the topic of cluster analysis for knowledge discovery. A comparative review of clustering strategies will be presented, as well as their applications, and the data structures needed to support them. Particular attention will be given to the implications of data representations and algorithmic design choices on the scalability and applicability of the various approaches studied.

〔教育目標・目的 Aim〕

クラスター分析を取り巻く主要問題、特に、実装に関する問題を理解すること。最新の手法に精通すること。クラスタ化戦略のスケラビリティと効果を評価する能力を身につけること。

An understanding of the main issues surrounding cluster analysis, particularly as regards implementation. Familiarity with state-of-the-art methods. The ability to assess the scalability and effectiveness of clustering strategies.

〔成績評価 Grading criteria〕

授業参加、レポート、プレゼンテーションを基に評価する。

Assessment based on class participation, reports, and presentations.

〔授業計画 Lecture plan〕

担当教員：Michael Houle

開講日：木曜4限（14:45-16:15）

授業計画：

- (1) クラスターリングvs 分類
- (2) 特徴、類似、および検索
- (3) スケラビリティと次元の呪い
- (4) クラスター化戦略の概要
- (5) 分割型クラスターリング
- (6) 階層型手法
- (7) 格子に基づく手法
- (8) モデルに基づく手法
- (9) 密度に基づく手法
- (10) 共有近傍手法

- (11) サブスペース クラスターリング
- (12) その他のクラスターリング戦略
- (13) アプリケーション：外れ値検出
- (14) アプリケーション：特徴選択
- (15) その他のアプリケーション

Lecturers: Michael Houle

Schedule: Thursday, 4th slot (14:45-16:15)

Contents:

- (1) Clustering versus classification.
- (2) Features, similarity and search.
- (3) Scalability and the curse of dimensionality.
- (4) Overview of clustering strategies.
- (5) Partitional clustering.
- (6) Hierarchical methods.
- (7) Grid-based methods.
- (8) Model-based methods.
- (9) Density-based methods.
- (10) Shared-neighbor methods.
- (11) Subspace clustering.
- (12) Other clustering strategies.
- (13) Application: outlier detection.
- (14) Application: feature selection.
- (15) Other applications.

〔実施場所 Location〕

国立情報学研究所(NII)：講義室1（12階1212号室）

National Institute of Informatics (NII): Lecture room 1 (12F, room1212)

〔使用言語 Language〕

英語

English

〔教科書・参考図書 Textbooks and references〕

教科書:なし

参考書:論文

Textbooks:None.

References:Research papers.

〔関連URL Related URL〕

URL:

〔上記URLの説明 Explanatory Note on above URL〕

〔備考・キーワード Others/Keyword〕

履修の条件:人工知能とコンピュータサイエンスについての基礎知識.

Prerequisites:Basic knowledge of artificial intelligence and computer science.