

ユーザを意識した大規模データマイニングと活用に関する研究



情報知能工学部門 高度ソフトウェア工学講座 准教授 峯 恒憲

ユーザが求める情報を、必要な時に提供する仕組みの実現を目指し、小規模から大規模まで、様々なデータのマイニングと活用について、企業や学術研究機関と共同で研究開発を進めています。

<https://www.m.ait.kyushu-u.ac.jp/>

データマイニングと情報推薦

質問、学業状況振り返り文
レポート内容
問題の解答文(短文)

質問、要望
自動回答
お勧め提示
レビュー評価
気になる商品・レビュー

質問回答、アドバイス提示

チャットボット(推薦)システム
(感情、状況、要望、意図の推定)

NLP, Feature Engineering
Graph Convolutional Network
Machine / Deep Learning, Big Data Mining

- 質問応答(チャットボット)システム
 - 非質問文や領域外質問の識別
 - 少ない教師データからの質問・応答対学習
- ユーザの感情推定
 - 色と感情の対応
- 有用レビュー文推定 (Amazonなど)
- 有用な間接アクセス情報取得グラフマイニング
 - ユーザのアイテムアクセス履歴(暗黙フィードバック)隣接行列に基づく協調フィルタリング
- 固有表現抽出
 - 特許文や文献などからの特定表現抽出
- 教育データマイニング

学生(生徒)の授業後コメント文
学習記録、テスト結果、操作ログ

Deep Learning

学生(生徒)の授業後コメント文

アドバイス指示

あなたの学習状況は良好ですが、友達との協力は今一歩、協力しながら学習を進めることで、理解も深まると考えられます。

自然言語処理
(文章解析, 生成)

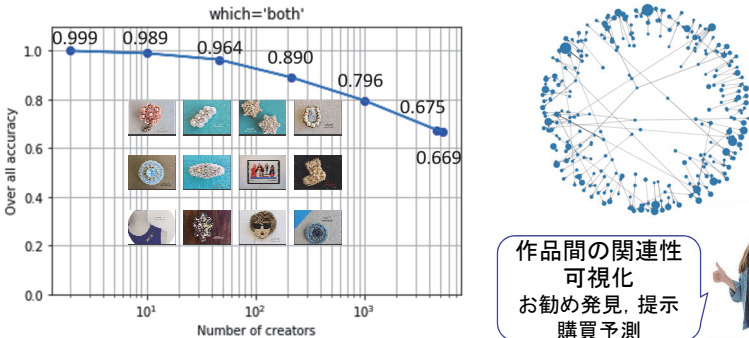
成績推定, アドバイス文・報告文生成

小学生から大学生まで

短文記述式解答文の自動採点

入試解答の自動採点に向けて

- 成績推定, アドバイス文・報告文生成
 - 小学生から大学生まで
- 短文記述式解答文の自動採点
 - 入試解答の自動採点に向けて
- 手作り作品データマイニング
 - 作品群間の影響



- 作家識別, お勧め作品推薦, 売上予測
- 作品傾向分析
- 手作り作品画像や説明文などの情報を利用

交通データマイニング

地方の公共交通(路線バス)

課題 データ収集・分析に基づく運行最適化

路線バスデータ取得・分析の仕組み, マルチモーダルデータの分析・活用

バスプローブ (1年半以上継続獲得中)
ドライブレコーダ データ(3年分)
周辺交通流 (ETC2.0データ)
天候情報 (気象台, AMeDAS)
人流 (バス停センサ)

マルチモーダルデータの取得・分析・活用

様々な種類の不完全(粗く, ズれを含む)データから, 下記の目的に沿う特徴導出

- 変動を考慮した移動・到着・遅延時間予測, 時刻表とのずれ算出, 乗車人数予測
- 事故防止に向けた危険運転挙動・危険状況推定, 運転特性推定

持続可能な地方公共交通の維持・発展に貢献

昭和バス48台にETC2.0車載器を設置

九州大学 (分析, 評価)

昭和自動車 (時刻表改訂) (利用者への通知)

ETC2.0簡易路側機設置箇所
ETC2.0簡易路側機
5.8GHz無線通信
AC100V (50/60Hz)



移動時間・到着時間・遅延時間推定

時刻表上の時間 3 min.

実際の時間 3 min. 50 sec.

遅延時間 +50 sec.

ターゲットバス停までの

- 移動時間
- 停車回数・停車時間

事故抑止のための危険(急ブレーキ)状況推定

場所以外の情報(時間, 曜日, 方向, 速度など)の影響大

気象情報なしでも高確率で識別

気象台(MO)やAMeDASから取得した気象関連データを加えると精度向上

Measurement	MO	AMeDAS
Atmospheric pressure (気圧)		MO
Temperature (気温)		MO and AMeDAS
Humidity (湿度)		MO
Rainfall amount (雨量)		AMeDAS
Weather (天気)		MO
Wind speed (風速)		MO and AMeDAS
Wind direction (風向)		MO and AMeDAS
Visibility (視野)		MO

同じ経路上の多くのデータ

K	Precision	Recall	F1	Volume
ALL	0.9212	0.9431	0.9320	263M
10K	0.9211	0.9429	0.9319	263M