

論文審査の結果の要旨および担当者

報告番号	※ 甲 第 12800 号
------	---------------

氏 名 廖 宸一

論 文 題 目

A Study on Composite Data Mining Methods for Linking Real-World Information with Web Resources
(実世界情報とWeb資源を関連付けする統合的データマイニング手法における研究)

論文審査担当者

主査	名古屋大学	教授	河口 信夫
委員	名古屋大学	教授	古橋 武
委員	名古屋大学	准教授	岩田 哲
委員	名古屋大学	准教授	井手 一郎

論文審査の結果の要旨

廖宸一君提出の論文「A Study on Composite Data Mining Methods for Linking Real-World Information with Web Resources(実世界情報とWeb資源を関連付けする統合的データマイニング手法における研究)」は、Web上に存在している様々な情報を、実世界情報に関連付けするための様々な情報処理手法を明らかにしている。本論文は7章からなり、各章の概要は以下の通りである。

第1章では、本論文の導入として、研究の背景、動機と目的、本論文の構成、主な貢献について述べている。特にインターネット上に大量に存在している情報と実世界情報の関連付けが不十分であり、これを実現する技術の必要性について述べている。

第2章では、本研究に関連する既存研究について述べている。まず、Webページからのイベント情報の抽出手法において、Webページの構造化手法やイベントデータの分類手法について述べている。次に、サイネージ識別手法において、文字認識を用いた手法と、画像認識を用いた手法に分類して説明している。

第3章では、Webページからイベントデータを抽出するための基本的な手法として、Webページのセグメンテーション手法を提案している。WebページはHTMLによって半構造化されているが、これは視覚的な構造であって、意味的な構造ではない。これを構造の類似度に基づくマッチングと、バックトラッキング手法を用いて、98%を超えるセグメント化の精度を実現している。

第4章では、セグメント化されたWebデータから、適切なイベントデータを抽出する手法を提案している。本手法では、イベントデータかどうかを判別するクラス分類機を、重み付きのWord2Vecを用いて構成し、これにより、91%を超えるF1スコアを達成している。また、セグメント化とイベントデータ抽出を用いて、Event.Lockyと呼ぶ実世界のイベント検索アプリケーションを開発した。

第5章では、実世界のサイネージをWebページへ関連付けるための基礎的なデータ収集手法として、Webイメージデータセットの構成手法を提案している。この手法では、店舗名とURLを入力とするだけで、その店舗に関係するサイネージの画像を収集できる。店舗名を用いてインターネット検索を通じて得られた画像には、様々な画像が存在するが、本手法では、店舗前画像を判別するための判別器を構築し、適切な画像選択を行っている。さらに、サイネージ画像を抽出するためのオブジェクト識別器を構築し、画像の抽出を行っている。また、店舗ロゴについても店舗のWebページを用いて抽出する手法を提案している。

第6章では、Webから収集された店舗サイネージ画像と、実世界のサイネージ画像に対するハイブリッドマッチング手法を提案している。マッチング手法としてはDeepMachingを用いるが計算負荷が高いため、不要なマッチングを行わない工夫を行っている。また、マッチング結果の適切さをRANSACを通じて評価する枠組みも実現した。これにより、92%の店舗の関連付けが可能になった。

第7章では、実世界情報とWeb資源の関連付けに関し、本研究の成果と今後の課題を述べている。

以上のように本論文では、実世界情報とWeb資源を関連付けるための様々な情報処理手法を提案し、実用上十分な精度で関連付けを実現している。これらの提案手法と得られた結果は、実世界情報とWeb資源の関連付けやその周辺技術において重要であり、工学の発展に寄与するところが大きいと判断できる。よって、本論文の提出者である廖宸一君は博士(工学)の学位を受けるに十分な資格があると判断した。