

2019年度 地域貢献研究助成費 実績報告書

2020年 3月 31日

報告者	学科名	情報システム工学科	職名	教授	氏名	菊井 玄一郎
研究課題	岡山弁による地域情報提供システムに関する研究					
研究組織	氏名	所属・職		専門分野	役割分担	
	代表	菊井 玄一郎	情報システム工学科 ・ 教授	自然言語処理	研究総括, 全体方式検討, 岡山弁対応の検討	
	分担者	石井 颯人	大学院システム工学 専攻・博士前期2年	同上	エンティティ属性抽出の 方式検討, 実験	
		河田 尚孝	同上	同上	エンティティ・リンキング の方式検討・実験	
保延 渉		大学院システム工学 専攻・博士前期1年	同上	質問応答処理の方式検討・ 実験		
研究実績 の概要	<p>本研究課題全体は3つのサブテーマから構成される。 サブテーマ1：地域特有の用語（固有表現）に対する辞書データベースの自動構築 このサブテーマはさらに次の2つに分けて実施した (A) エンティティリンキング (B) エンティティ属性抽出 サブテーマ2：自治体等におけるFAQ データベースの自動構築 サブテーマ3：音声対話処理の岡山弁適応 サブテーマ1, 3については既存手法に比べて精度向上などの成果があり, サブテーマ2 については新たなアプローチを提案し有効性を検証した。また国際会議1件, 国内大会2 件, 支部大会1件の外部発表を行った。以下, それぞれについて説明する。</p> <p>サブテーマ1の実績 (A) エンティティリンキング 昨年度に考案した「再ラベリング法による言及抽出」が今年度の国際会議(IEEE, BCD)に 採択され, 国内大会とともに発表を行った(資料目録1, 2)。計画に従って, サブワー ドモデルを組み合わせる方法を試みたが精度向上には至らなかった。本年度後半では rankingSVM による曖昧性解消を組み合わせてエンティティリンキング全体の精度を評価 した。その結果, 本研究で提案した再ラベリング法の有効性が確認できた。 (B) エンティティ属性抽出 昨年度に引き続き, 理化学研究所(理研) AIPの「森羅プロジェクト」に参加して属性 抽出手法の検討を行った。同プロジェクトで昨年度に最高精度を達成した「機械読解 (DrQA)に基づく手法」の追試を行った結果, 箇条書きなどのリスト構造の扱いに問題があ ることを突き止め, リスト構造をテンプレート生成によって自然言語文に変換する前処理 を追加した。その結果, F値21.7から21.4に向上した。これらは森羅2019最終報告会 (理研AIP), および, IEEE HISSで発表した(資料目録3, 優秀賞)。</p>					

※ 次ページに続く

<p>研究実績 の概要</p>	<p>サブテーマ2の実績</p> <p>本サブテーマでは特に市役所等での市民からの問い合わせ対応の効率化, 高度化を目標として, FAQ 集の構築支援手法の確立を目指している. 今年度は, まず, 岡山県内の基礎自治体 (の役所) における FAQ の整備状況を調査した. その結果, 岡山市, 倉敷市は例外的に充実しているものの, 大半の自治体で FAQ 自体が存在しない, あるいは, 存在しても一部の分野のみである, ということが分かった. FAQ 構築においては特に質問部分の作成にノウハウが必要なことから, 「充実している複数の自治体の FAQ で多数出現する質問を見つけて提示する手法」を提案した. この手法はデータマイニングの方法に基づくものである. 実験の結果, 適合率98%, 再現率36%となった. 再現率の向上も必要であるが, この成果は十分実用可能である. (資料目録4)</p> <p>サブテーマ3の実績</p> <p>(A) 岡山弁の音声認識</p> <p>既存研究では一部しか扱えなかった「連母音融合」(長い→ナゲー), 「助詞の縮約」(栗を→クリョー)を新たに考案した「拗音分割方式」により, 想定される全ての表現に対応可能とした. その結果, 岡山弁音声に対する認識精度が74.1から76.8に向上した.</p> <p>(B) 岡山弁の音声合成</p> <p>音声合成のための発音辞書の再構成, 格助詞が連続する場合の縮約規則, 文末のイントネーションの調整を行い, 自然性の向上を図った. 音素レベルでの性能向上は確認したが, 合成音声を聞かせての評価は今後の課題である.</p>
<p>成果資料目録</p>	<ol style="list-style-type: none"> 1. Naotaka Kawata, Genichiro Kikui, "Mention Detection Method for Entity Linking", IEEE International Conference on Big Data, Cloud Computing, Data Science & Engineering (BCD), 2019. 2. 河田尚孝, 菊井玄一郎, "エンティティリンクングのための言及抽出手法", 第33回人工知能学会全国大会, 1N3-J-9-02, 2019. 3. 石井颯人, 菊井玄一郎, "機械読解によるWikipediaからの属性抽出における表, 箇条書きの変換の効果", 第21回IEEE広島学生シンポジウム, B2-17, 2019. 4. 保延渉, 菊井玄一郎, "共起単語に着目した自治体FAQのための質問生成", 情報処理学会第82回全国大会, 7T-2, 2020.