別記様式第４号

平成 30 年度 独創的研究助成費 実績報告書

平成　31 年　3 月　28 日

<table>
<tr><td>報　告　者</td><td>学科名</td><td>情報システム工学科</td><td>職 名</td><td>教授</td><td>氏 名</td><td colspan="2">菊井　玄一郎</td></tr>
<tr><td>研 究 課 題</td><td colspan="7">エンティティの属性推定を用いた固有表現の意味理解に関する研究開発</td></tr>
<tr><td rowspan="3">研 究 組 織</td><td colspan="2">氏　　名</td><td colspan="2">所 属 ・ 職</td><td>専 門 分 野</td><td colspan="2">役　割　分　担</td></tr>
<tr><td>代表</td><td>菊井　玄一郎</td><td colspan="2">情報システム工学科
教授</td><td>自然言語処理</td><td colspan="2">全体統括・方式検討</td></tr>
<tr><td>分担者</td><td>石井　颯人

河田　尚孝</td><td colspan="2">システム工学専攻・
博士前期・１年
システム工学専攻・
博士前期・１年</td><td>同上

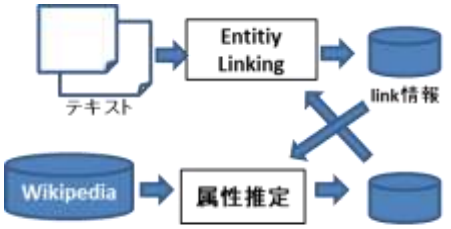
同上</td><td colspan="2">エンティティ属性推定法
の検討・実験
エンティティリンキング
手法の検討・実験</td></tr>
<tr><td>研究実績
の概要</td><td colspan="7">本研究計画は次の二つのサブテーマからなる.
サブテーマ１：エンティティリンキング
サブテーマ２：固有表現属性推定
以下，実績の概要について述べる.
 

図1</td></tr>
</table>

※ 次ページに続く

研究実績の概要	**１．エンティティ・リンキング** エンティティリンキングは,①入力テキストから固有物に対する言語表現(mention:言及)を抽出する「言及抽出」と，②得られた言及と実世界の事物（ここではwikipediaのエントリ）を紐づける「曖昧性解消」という２つのステップに分けられる．日本語においては①の性能が低いことがまず問題となっている．言及抽出は，通常，文中における言及とその（意味的）カテゴリを同時に推定する「固有表現抽出」が利用されるが，後段の処理②においてカテゴリ情報はさほど有用でないことから，カテゴリの粒度は自由に調整できる．そこでカテゴリ粒度の調整で言及抽出精度（再現率）が改善できるかを調査した．また，系列ラベリングによる固有表現抽出の再現率向上を目的として，一度通常の系列ラベリングを行ったあと，複数のトークン（単語）からなる固有表現を一つのトークンにまとめ，学習データも同様にまとめることにより再ラベリングを行う「再ラベリング法」を新たに考案した. 評価実験の結果，カテゴリ粒度は「関根の拡張固有表現」の最も細かい階層(数量表現等を除く 154 カテゴリ)，中間層(同 11 カテゴリ)，全てを１つにまとめたもののうち 11 カテゴリが最もよかった．また再ラベリングにより精度(適合率)が 2.8 ポイント下がったものの，再現率で 5.3 ポイント，F 値で１ポイント向上することが分かった．これらの結果を国内大会（目録 1，採録），国際会議（目録 2,審査中）に投稿した. **２．属性推定** 本サブテーマについては理化学研究所革新知能融合センター(AIP)のプロジェクト「森羅」[1]に参加して，このプロジェクトの提供するデータ（課題）を利用することにより，研究を進めた. 今回はプロジェクトの初年度ということで人名，企業名など 5 つのカテゴリについてwikipediaからランダムに抽出された 600 項目（例：夏目漱石）に対して人手で属性（生没年月日）が付与された「正解データ」が配布された. 我々は配布された正解を元に，wikipediaから属性を抽出するパターンを機械学習する方法を提案した．結果の提出形式の誤解があり参考結果にはなったが，パターン照合のアプローチとして妥当な結果となった. 提案手法は 2018 年 10 月 18 日開催の「森羅プロジェクト 2018 結果報告会」において報告した（資料目録 3）. [1]関根ほか：“Wikipedia 構造化プロジェクト「森羅 2018」，言語処理学会第 25 回年次大会，pp.69-72，2019.
成果資料目録	学会発表： １）河田尚孝，菊井玄一郎：“エンティティリンキングのための言及抽出手法”，　2019 年度人工知能学会全国大会，1N-3-J-9-02 (2019). ２）N. Kawata and G. Kikui:”A Mention Detection Method for Entity Linking” submitted to XXXX(ブラインド査読のため伏せる). ３)石井颯人，菊井玄一郎：Infobox を利用した属性抽出，森羅 2018 成果報告会 (2018).