

## 平成 21 年度 情報工学コース卒業研究報告要旨

村瀬 研究室	氏 名	鈴木規之
卒業研究題目	Wikipedia 人名エントリによるニュース中の人物 同定	
<p>本研究では、Wikipedia を利用し、ニュース放送中に出現した人物名称に該当する人物を識別することを目的とする。これは例えば、ニュース放送中に出現した、「鳩山総理大臣」という人物名称から、Wikipedia の「鳩山由紀夫」のエントリを抽出することである。</p> <p>我々が普段目にするニュース放送中には、多くの人物が登場する。従来、そのようなニュース映像中の人物を、ニュース映像の解析に用いる研究が行われている。人物名はニュース中では音声や字幕で提示される。しかし、人物名がフルネームで表されることもあるが、苗字と役職名で表されることも多い。そのため、ニュース中の言語情報以外の情報で人物の同定を行う必要がある。</p> <p>視聴者の多くは、苗字と役職名で示された人物の同定が可能である。これは、視聴者が、人物名称という情報のほかに、ニュース映像中の人物の顔、役職名で示された役職に就任している人物、または、そのニュース放送の話題に関する事前知識を利用して識別しているからである。一方、ニュース中の映像を用いた人物の同定手法も提案されている。しかし、映像からの人物切り出し精度の問題や、顔照合の精度の問題から、対応付け精度に限界があることが報告されている。そこで、提案手法では、放送中の言語情報以外の情報獲得の手段として、Wikipedia を利用する。</p> <p>ニュース放送中の 1 つの話題をトピックという。1 つのトピックの中に人物名称が出現する場合、ニュースの話題は人物名称が持つ役職情報に関連した話題であり、関連性のある単語が出現する場合が多い。本研究では、このようなニュース放送の文章の特徴に着目し、クローズドキャプションと Wikipedia の各エントリとの類似度比較を行うことで、該当する人物を抽出する。</p> <p>提案手法は、以下の手順によって行う。</p> <ol style="list-style-type: none"><li>1. ニュース放送文 ( Closed-Caption ) から人物名称 ( 苗字と役職名 ) を抽出する</li><li>2. 人物名称から苗字と役職情報を分離し、役職情報から Wikipedia のカテゴリ情報を抽出する</li><li>3. 苗字とカテゴリ情報を利用し、Wikipedia から同定候補となるエントリを抽出する</li><li>4. 人物名称が出現するニューストピックと、3. で抽出した Wikipedia エントリから、それぞれの文章に出現する単語を要素とした文書特徴ベクトルを作成する</li><li>5. ニューストピックより作成したベクトルに、人物名称の単語を用いて重み付けをする</li><li>6. Wikipedia の各エントリより作成したベクトルに、各エントリのタイトルの単語による重み付けをする</li><li>7. ニュースの単語ベクトルと、Wikipedia エントリの単語ベクトルの類似度比較を行い、類似度最大のものを同定結果とする</li></ol> <p>150 日分のニュース映像を用いた実験より、71.1% の正解率を得ることができ、本手法の有効性を確認した。</p>		