

システム情報工学研究科修士論文概要

| | | | |
|--|----------|----------|-------------|
| 年 度 | 平成 24 年度 | 学位名 | 修士(工学) |
| 専 攻 | 知能機能システム | 専攻 | 著者氏名 胡 碩 |
| 指導教員氏名 宇津呂 武仁 | | | |
| 論文題目 時系列ニュースのトピックモデルにおける二言語間対応の同定 | | | |
| 論文概要 <p>本論文では、日本語および中国語の二言語の時系列ニュースを対象として、教師なしの時系列トピックモデルを適用し、時系列のトピックの推移を推定することにより、日中二言語間のトピックの対応を同定する手法を提案する。本論文の枠組みにおいては、まず、Wikipedia の言語間リンクを用いて、エントリの日中対応の抽出を行う。Wikipedia の言語間リンクを用いることにより、あるエントリに対して、日本語と中国語の間で、翻訳を行うことができる。次に、ニュース記事の日中対応の推定を行う。日付が同一の日本語ニュース記事と中国語ニュース記事の間で、日中対訳語組数を求め、10 組以上の日中対訳語組数を持つ日本語ニュース記事と中国語ニュース記事の組に対して、ニュース記事が日中間で対応するとみなす。次に、日中間で対応する日本語および中国語ニュース記事集合において、時系列トピックモデルである DTM(dynamic topic model)によって、各言語での時系列トピックの推定を行う。最後に、時系列トピックの日中対応の推定においては、ある連続期間における日本語と中国語の時系列ニュース記事を対象として、DTM モデルで生成した日本語トピック集合と中国語トピック集合の間で、2 組以上の日中対応記事組を持つ日本語トピックと中国語トピックについて、その日付において日中間で対応するトピックであるとみなす。</p> <p>以上の手法により、2010 年 2 月 25 日から、3 月 23 日までの一ヶ月の日中ニュース記事を対象として、評価実験を行った。その結果、日本語から中国語へのニュース記事対応付け精度は 63.2%、中国語から日本語へのニュース記事対応付け精度は 53.3%となった。また、日中方向のトピック対応付け精度は 80%、中日方向のトピック対応付け精度は 66.7%となった。この結果において、大きく精度を下げている要因は、日中両国における自国経済についてのニュースであった。これらの自国経済についてのニュースを除外した場合には、日本語から中国語へのニュース記事対応付け精度が 88.1%に、中国語から日本語へのニュース記事対応付け精度が 86.6%に、それぞれ改善した。そして、日中方向のトピック対応付け精度は 100%に、中日方向のトピック対応付け精度は 80%に、それぞれ改善した。</p> | | | |
| 審査日 平成 25 年 1 月 31 日 | | | |
| 審査員 | (大学名 職名) | (学位) | (氏名) |
| 主査 | 筑波大学 教授 | 博士(工学) | 宇津呂 武仁 |
| 副査 | 筑波大学 准教授 | 博士(工学) | 中内 靖 |
| 副査 | 筑波大学 准教授 | 博士(情報科学) | 星野 准一 |