

知識情報工学専攻	学籍番号	033718
申請者氏名	須藤 悠	

指導教員氏名	村越 一支
--------	-------

## 論文要旨 (修士)

論文題目	コサイン距離を用いた自己組織化マップによる換言知識の獲得
------	------------------------------

換言知識は、情報検索や質問応答などの意味的な言語処理において重要な応用横断技術である。しかし、既存のシソーラスだけでは人間が日常的に行っている換言知識を網羅することはできないとされる。

従来の教師あり学習による自然言語学習では教示が必要であるため、正答の作成にかかる手間が大きいという問題があった。しかし、教師なし学習では教師信号の入力が不要であり、人手による正答の教示や報酬の設定を必要としない。そのため、タグ無しコーパスから意味的な知識を獲得するための研究が行われている。

長谷川らは、二つの固有表現の間に成立する静的な関係に注目し、一文の中に現れる二つの固有表現の間の文脈から、教師なし学習と tf-idf 法によって動詞句の換言知識の獲得が可能であることを示した。長谷川らの手法では、教師なし学習に距離測度にコサイン距離を用いた最長距離法が利用されている。しかし、最長距離法では出力がデンドログラム (樹状図) として得られるが、出力となるクラスタ数が与えられない場合、クラスタは学習結果から人手により解決する必要がある。また、デンドログラムは可視化性能が低く、複数のフレーズや換言フレーズ間の関係を把握しにくい。そこで本研究では、こうした問題を解決するために自己組織化マップを用いる。

自己組織化マップ (Self-Organizing Map; SOM、以下 SOM) は、教師なし学習を行う人工ニューラルネットワークの一種である。特徴ベクトルを用いた競合学習によって、高次元のデータであってもデータ間の位相を保ったまま低次元空間に射影するので、高次元のデータとその非線型性に対しても良好なクラスタリング性能を持つ。言語処理に用いられた研究には、大量の記号列のクラスタリングや、3 つ組を用いた推論、単語の意味マップの形成といった例があるが、換言知識の獲得に用いられた例はない。

本研究では、SOM による換言知識の獲得手法を示す。学習対象が tf-idf 法によるベクトル空間であるため、通常 SOM の競合学習で距離測度に用いられるユークリッド距離や内積ではなく、コサイン距離を用いる方法を提案する。学習規則にも通常 SOM のものではなく、内積 SOM のものを用いる。これは、コサイン距離が内積をノルムによって正規化した尺度であるためである。

提案した手法の有効性を検証するため、従来手法であるユークリッド距離型 SOM と内積型 SOM との比較を行う。英語自然文から固有表現を用いて文脈を収集し、得られたフレーズを tf-idf によってベクトル化する。こうして得られたベクトルを用いて比較実験を行った。その結果、提案手法によって従来手法と比べ多くの換言知識の候補が得られることを確認した。