

先端科学技術研究科 修士論文要旨

所属研究室 (主指導教員)	ソーシャル・コンピューティング (荒牧 英治 (教授))		
学籍番号	2111191	提出日	令和 5年 1月 17日
学生氏名	西山 智弘		
論文題目	ソーシャルメディアにおける服薬ノンコンプライアンス: 医薬品構造の類似性に基づく転移学習		
要旨			
<p>近年、インターネット上での医薬品の流通・販売の増加に伴い、服薬ノンコンプライアンスと呼ばれる不適切な医薬品使用が問題となっている。そこで、ソーシャルメディア上での不適切な医薬品使用をモニタリングすることを目指した。しかし、モニタリングのためのコーパス構築にはコストがかかるため、化学構造の類似した医薬品に関するコーパスの転移学習を試みた。具体的な手法として、ソーシャルメディア上のテキストを服薬に基づきマルチラベル分類することを実施し、さらに医薬品の化学的類似性を用いて、コーパスにおける転移学習の可能性を確認した。使用したコーパスは、医薬品のコンプライアンス違反に関するMediAコーパスであり、これは20種類の医薬品に言及したツイートにNoncompliant use/mention, Noncompliant sale, General use, General mentionからなるラベルを割り当てたものである。特定の医薬品に関するツイートの分類モデルは、他の1つの医薬品に関するツイート(単一サブコーパス転送学習)と他の薬に関するツイート(マルチサブコーパス漸増学習)の2つのサブコーパスでの転移学習によりトレーニングした。そして、医薬品構造の類似性に基づいて、転移学習に用いるべき有効な医薬品のサブコーパスが存在するかどうかを評価した。医薬品の構造類似度と分類性能の間には、0.278のわずかな相関が見られた。また、サブコーパスの数が少ない場合には、構造的類似性の高い医薬品コーパスを転移学習させたモデルの方が、ランダムにサブコーパスを追加して学習させたモデルよりも良い性能を示した。この結果は、学習コーパスに含まれる医薬品が少ない場合、構造的類似性が未知の医薬品に関するメッセージの分類性能を向上させることを示唆するものである。一方、医薬品の種類が十分に確保されている場合には、Tanimoto構造類似度の影響を考慮する必要はあまりないことが示された。</p>			