

分散表現を利用したタグ集合の階層化

鈴木 宏明

要旨

近年、SNSの台頭などによりWeb上へ投稿されるコンテンツが増加している。これらのコンテンツにはタグを用いて特徴づけされているものも多いが、一般的にコンテンツ同士の関係性は明示的ではない。本研究では、どのコンテンツ同士に関係があり、どのコンテンツ同士に関係がないかを明示的にするために、コンテンツに付与されたタグの集合を構造化することを目的とする。

これまでに既存手法として、分散表現を利用した語の is-a 関係のモデル化手法として、下位概念を入力して上位概念を予測する手法が提案されている。本研究ではこれをタグ集合へと拡張するとともに、複数の上位概念予測手法を提案する。また提案手法の有効性を評価するために、ニコニコ動画におけるアニメに関連するタグを持つ動画約32万件を対象に評価実験を行った。既存研究との比較という意味では、一部の提案手法にその有効性が確認できた。その一方で、上位タグ集合の予測精度は必ずしも十分ではなく、今後取り組むべきいくつかの課題が示唆された。