

定常的な出現を考慮した地域の特徴の抽出に関する研究

舛井 秀成

まちづくりやマーケティングを行う際には、地域の特徴を考慮して戦略を立てることが重要である。地域の特徴を抽出する方法の1つに、ソーシャルメディア上の投稿を分析する方法が考えられる。ソーシャルメディア上の投稿に現れるキーワードは、地域の特徴を多様な表現で反映していることから、同じ特徴を反映したキーワードをまとめあげてラベルを付与する必要がある。そこで、ソーシャルメディアにおいて、特定の地域の市民の投稿に現れるキーワードに対して、DBpediaのカテゴリを用いて地域の特徴を付与する手法を提案する。

地域の特徴を抽出するためには、特定のカテゴリに関して言及する投稿の量に着目することが考えられるが、ユーザによって投稿数は異なる。したがって本研究では、特定のカテゴリに関して言及するユーザ数に着目する。

一方で、人気のある地域のイベントによって、特定の時期に特定のカテゴリについての投稿が多くなることもあるが、特定の時期にだけ多く投稿がみられるものを、その地域の特徴と考えることは難しい。そこで、提案手法では、ある地域において定常的に表れる特徴を抽出する。具体的には、期間を複数に分割し、各期間においてあるカテゴリについて言及するユーザの割合を算出し、全期間に渡って、ある地域において一定以上、他の地域では一定未満の値を示すものを、その地域の特徴とする。

定常的な出現を考慮した特徴抽出の有効性を示すために、実験を行った。実験データは、2016年8月、9月、10月における水戸市民、つくば市民のツイートである。8月は水戸市が2,127ユーザ、156,638ツイート、つくば市が2,707ユーザ、230,765ツイートであり、9月は水戸市が2,042ユーザ、132,917ツイート、つくば市が2,657ユーザ、205,343ツイートであり、10月は水戸市が2,036ユーザ、112,829ツイート、つくば市が2,601ユーザ、178,674ツイートであった。評価にあたっては、提案手法により抽出された特徴のうち、地域の特徴として適切なものを人手で選択して確認を行った。水戸市では、「水戸黄門」、「時代劇」という特徴が確認されたことから、水戸市民は歴史的な事柄に関心が深いと考えられる。つくば市では、「漫画作品」、「アニメ作品」、「自転車競技」という特徴が確認されたことから、学生や若者が多く、つくば市民の特徴を反映していると考えられる。また、両地域の結果を踏まえると、DBpediaを用いることにより、作品に関する特徴や、自分の活動に関する特徴が得られると考える。

今後の課題としては、抽出した地域の特徴中の単語を手がかりにして、時期に依存した特徴や、様々なカテゴリに属する単語を含む特徴を削除する手法の検討が挙げられる。

(指導教員 関 洋平)