

語および感情の結束度を考慮したオンライン小説の段落分割手法の提案

伊藤 志暢

現在、オンライン小説のコンテンツ数は増加しており、メディアミックスなど様々な方法で受容されている。オンライン小説の長さは多様であり、100万文字を超える長さの小説も少なくない。しかし、このような長大な小説の中には章分けがなされていないものも多く、限られた時間で読むときにはどこまで読むべきかの判断が難しい。

そこで本研究ではオンライン小説を複数回に分けて読むための適切な範囲に区切る 2 つの手法を提案する。第一は語の結束度に基づく手法である。この手法は、小説の話中の名詞に対して、同一語句が以前かつどれくらい近傍に出ているのかで語の結束度を求める。話に含まれるすべての名詞についての結束度の平均を話の結束度として話の区切りを判定する。第二は感情に基づく手法である。この手法は、8つの感情に対応した感情語を分類語彙表によって拡張し、小説の話中の感情語の割合について話と話の差で求められる結束度により話の区切りを判定する。

本手法の有効性を検証するため、まず実際のオンライン小説の章分けを正解データとして精度と再現率の指標を使った評価を行った。さらに、感情の結束度による話の区切りについて、利用者実験を行った。実験参加者 18 人に対し小説の区切りを読んでもらい、区切りの 8 つの感情と 5 段階で読後感を尋ねた。対象とした小説区切りは、感情によって区切られた小説区切りから感情語の属性が高いものを 8 個の感情につき 2 個ずつ計 16 個を選んだ。16 区切りのそれぞれに対して、計 121 件の回答を得た。

区切りの評価について、語の結束度の再現率は 0.45、精度が 0.12 だった。いずれもランダムに区切った場合よりも有意に高く、区切りとして一定の適切さがあることが示せた。また語の結束度と感情の結束度の比較では、全体的には語の結束度の方が良い結果が出たものの、いくつかの小説で感情の結束度の方が適切な区切りになっていたこともわかった。

利用者実験による感情付与の評価では、本手法とランダムな選択とであまり変わらないことがわかった。しかし、利用者全員が選んだ感情と本手法で導出した感情が一致している区切りもあり、区切りによっては本手法が適切に感情を付与できる可能性を示した。また、読後感は平均 3.6 であり、感情による区切りはある程度心地よい区切りであったことが示された。

本研究では、語および感情の結束関係からオンライン小説の区切りを作成する 2 つの手法を提案し、評価を行った。今後の課題は、より適切な長さにおける区切りの作成とより実際の読書に近い区切りを提案することである。

(指導教員 松村敦)