

平成24年度 情報工学コース卒業研究報告要旨

石川 研究室	氏 名	中 村 晃
卒業研究題目	Wikipediaの著者のバイアスに基づくクラスタリング	
<p>WikipediaはWeb上の百科事典であり、利用者は誰でも著者となって編集を行うことができる。そのため、記事には虚偽の記述などの信頼性に欠ける記述が含まれる。また、著者は独自の判断で編集を行うために、編集にはその著者の考え方や価値観が反映される。思い込みや思想などから考え方や価値観が偏っている著者は、同様に偏った内容の記述を行うといったことが考えられる。このように考え方や価値観に偏りを持った著者をバイアスのかかった著者と呼び、偏った考え方や価値観が反映された記述をバイアスのかかった記述と呼ぶ。著者のバイアスが記事に反映されている場合、その記事の百科事典としての質と信頼性は低下する。これらの理由から、Wikipediaの信頼性は問題視されている。</p> <p>信頼性の問題を解決するために、自動的に記述の信頼度を測定する手法が提案されている。この手法では、記述が多くの編集を経て残存している場合、その記述と記述した著者の信頼性が高いと仮定することにより、信頼度を算出している。ところが、宗教や政治問題などの様々な意見が存在するような記事においては、信頼度の算出を適切に行うことは困難である。一つの記事で二つの対立する意見がある場合、片方の意見を持つ著者グループの著者は、もう片方の意見を持つ著者グループによって書かれた記述を削除するといったことが考えられる。この場合、少数派の著者グループよりも多数派の著者グループの方が信頼性が高いと評価される。このように、いくつかの観点が存在するような情報に対して、多数決によってそれらの正誤を判定することは適切であるとはいえない。Wikipediaには「中立的な観点」という方針があり、これは、すべての記事は特定の観点到に偏らずあらゆる観点からの描写を平等に扱い、中立的な観点到に沿って書かれていなければならないというものである。著者はこの方針に従って記述や削除を行う。しかし、中立的な観点到は人や文化によって異なる。ある観点到に基づいた著者グループと、それとは別の対立する観点到に基づいた著者グループの間で編集合戦が生じる原因は、それぞれの著者グループの中立的な観点到が異なっているからである。どちらの著者グループも自分の観点到や記述が正しいと思っており、片方の観点到から見れば、もう片方の記述は誤っているように見える。</p> <p>そこで、本研究では記述の信頼度算出は行わずに、バイアスによる記述の分類を行うことを目標とする。そのために、バイアスに基づいた著者のクラスタリング手法を提案する。著者をバイアスによって区別することができれば、著者の記述もバイアスによって区別することができる。利用者が記事を閲覧する際に、各記述がバイアスによって色分けなどで区別されていれば、記述の信頼性判定に役立てることができると考えられる。</p> <p>バイアスに基づいた著者のクラスタリングには、著者間の賛同・反対関係を用いる。著者間の賛同・反対関係は、編集における残存・削除行為から特定する。著者は、自分と同様の価値観や考え方によって書かれた記述や、正確な事実に基づいた記述を残存させ、対立する価値観や考え方によって書かれた記述や、誤りを含む記述を削除する。このとき、著者は賛同する著者の記述を残存させ、反対する著者の記述を削除すると考えることができる。この仮定に基づけば、記事の編集における残存・削除行為を用いて著者間の賛同・反対関係を特定することができる。また、このことから、賛同関係にある著者は同様のバイアスを持ち、反対関係にある著者は異なるバイアスを持つと考えられる。したがって、著者間の関係を用いることにより、バイアスに基づいた著者のクラスタリングを行うことができる。</p>		