平成22年度 情報工学コース卒業研究報告要旨

石川 研究室	氏　　名	岩 井　一 晃
卒業研究題目	Twitter 検索のためのキーワード推薦	

現在、Twitter に代表されるマイクロブログは多くの利用者によって利用されている．マイクロブログの利点として，身近な人とのコミュニケーションツールとなる点や，日常の生活では知り合えない人の発見，交流ができるという点が挙げられる．日常では知り合えない人を発見する方法として，利用者が投稿文を検索する方法がある．利用者が自ら検索を行うとき，公式サイトにある検索フォームよりキーワードを入力する方法がある．この場合，多くの投稿文が存在し，また検索中にも多くの投稿文が投稿されると考えられる．この中から新しい話題を発見し，新しい話題のキーワードも含め，複数のキーワードを検索フォームより入力し，投稿文を見ると仮定する．この時，大量に投稿される投稿文を利用者は見る必要があり，利用者の負担が大きくなると考えられる．

この問題を解決するために，本研究では利用者の入力したキーワードにより投稿文を取得し，利用者の入力したキーワードに関連性の高いと考えられる文字列を新しいキーワードとして推薦するシステムを提案する．新しいキーワードを推薦する方法において，利用者のキーワードにより取得した投稿文における頻出文字列を新しいキーワードとする方法が考えられる．この方法において頻出単語を抽出する際，方法の一つとして，形態素解析の利用が考えられる．しかしマイクロブログでは文がくだけた表現になる傾向がある．投稿文のくだけた表現は形態素解析で利用される辞書に登録されていない単語である可能性が高い．よってマイクロブログ特有のくだけた文は形態素解析で解析できない．これより，形態素解析によって取得できる新しい単語はあまり良い精度が得られないと考えられる．

そこで，本研究ではキーワード抽出の支援を n-gram と発言者の情報を用いた関連語判別を用いて行う．本研究のシステムは図 1 のようになっている．n-gram を用いて頻出文字列を取得することにより，形態素解析で判別することのできなかった文字列を取得することができる．n-gram は文字列を文字の並びだけで認識している．そのため n-gram によって投稿文を解析した場合，形態素解析によって出現回数を算出する場合よりも誤判断が発生する可能性が低いと考えられる．n-gram により頻出文字列を解析した場合，頻出文字列の部分文字列も取得してしまうが，これらはノイズとなるため，削除する．

また本研究では発言者の情報を用いてノイズ除去を行う．発言者の情報を用いることによって，少数の発言者が何度も投稿している文字列を取り除くことができる．これによって高い精度で新しいキーワードを提示できると考えられる．

本論文では，提案した手法のキーワード推薦の精度を再現率精度曲線を用い，形態素解析を用いてキーワード推薦を行ったシステムとの比較実験を行い，本手法の考察を行った．

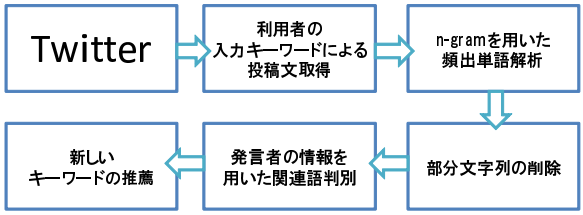

図 1 本研究システムの流れ