

ステップ2： テキストに感情の点数を付ける

佐藤 聖

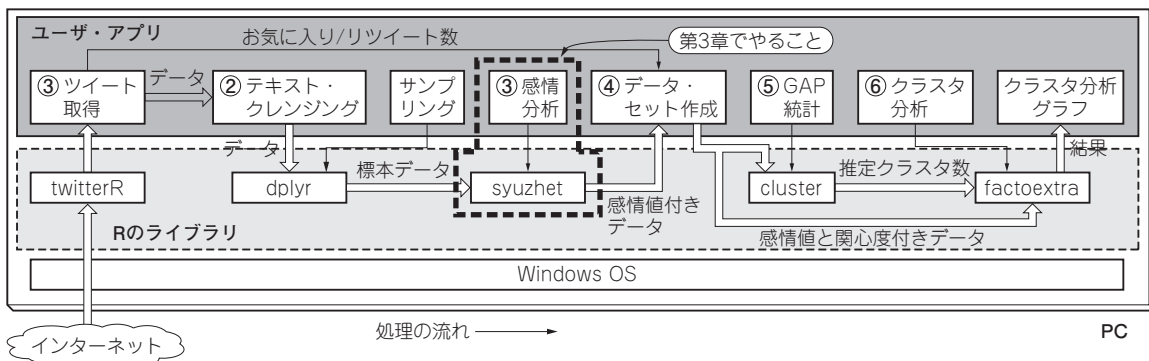


図1 特集でやる「人工知能で市場動向分析」のはじめの一歩…テキストに感情の点数を付ける

ここでは、第2章でテキスト・クレンジング処理済みのデータの感情を査定します(図1)。図2のように、単語に含まれる感情を査定することによって、経済ニュースの各ヘッドラインから内容が肯定的、否定的または中性的の情報を得たいと思います。各ヘッドラインの感情査定結果は、まちまちであることが想定されます。

単語が肯定的か否定的かを 点数にするための「感情辞書」

● ネット上に大量にある

査定には感情モデルに基づいて作成された感情辞書を使います。インターネットを検索すると、さまざまな言語の感情辞書が大量に見つかります。それらは自然言語の研究で作成されたもの、単位取得を目的として論文を書くために作成されたようなもの、商用パッケージまであります。たくさんの感情モデルを理解する時間がなかったので、よく見かけるものを例として表1にまとめました。通常R言語では、感情辞書がパッケージに含まれて配布されているものがあり、すぐに分析が行えるので便利です。

● 辞書ごとに査定結果が異なる

感情辞書の役割は、単語に対して感情値を査定するためのデータベースです。さまざまな感情辞書があ

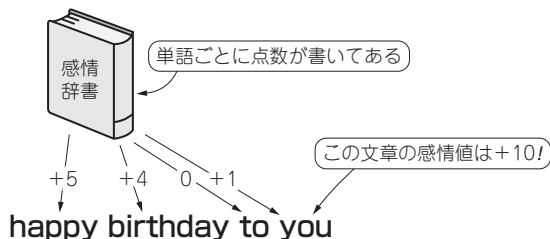


図2 肯定的か否定的かは辞書を引きながら点数で決める

り、同じ単語でも辞書によって割り当てられる感情値が異なります。例えば、ある感情辞書では肯定的であっても、他の感情辞書では否定的とされる場合があります。分析したい内容に合わせて感情辞書を選択します。感情値の表現方法として肯定的/否定的の様な単純な査定方法から(表2)、さまざまな感情表現に細分化して査定する感情辞書があります(表3)。

処理の全体像

● 判定の流れ

基本的に感情分析は図3のプロセスで行います。テキスト・データを入力すると単語ごとに読み取ります。単語は前後を空白文字で囲まれているASCII文字を単語として識別して読み取ります。読み取られた単語