

研究情報

1

耳より情報

2

研究情報

「言葉の山」から価値ある情報を探し出す

情報利用研究領域 竹崎あかね

商品の改良・新商品開発につながるヒントを発見するために、最近、商品に関する感想や質問などの「言葉」に関するデータ（テキストデータとよびます）が活用されるようになってきました。しかし、大量のテキストデータを人力で解析すると膨大な時間が必要になるだけでなく、その作業を複数人で分担すると解析方針に一貫性が無くなります。そこで、コンピューターによって大量のテキストデータから効率的に知見を見つけ出すテキストマイニングと呼ばれる技術の導入が進んできています。

テキストマイニングは、1) テキストデータの内容を抜き出し（言語処理）、2) 統計的手法により必要な知見を見つけ出す（マイニング）ものであり、新聞記事等の推敲されたテキストデータの高速な解析を可能にできました。これまでに多くの言語処理ツールが公開され広く利用されていますが、農産物に関するテキストデータへ利用した事例はまだ少ない状況にあります。

インターネット通販の野菜商品レビュー（購入者の感想）は農産物に関するテキストデータの一つであり、それらを解析することで利用者の嗜好など販売促進に関する多くのヒントを得られると期待できます。そのため私たちは、既存の言語処理ツールで農産物に関するテキストデータの内容が高精度で抜き出せるかを、ネット通販の野菜商品レビュー（購入者の感想）に用いて確認してみました。

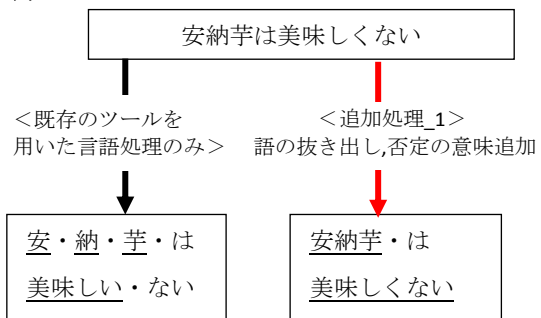
商品を購入した人が実際にどのような感想を持ったかは、この「購入者の感想」を言語処理して抽出したキーワード（名詞・動詞・形容詞に限定）で表現できると仮定しました。例えば、「安納芋は美味しくない」という文章の場合は「安納芋、美味しくない」が記述内容を表現するキーワードと判断できます。しかし、既存の言語処理ツールをそのまま使うと、「安納芋」は「安」・「納」・「芋」と、「美味しくない」は「美味しい」・「ない」と分割されてしまうため、キーワードは「安、納、芋、美味しい」となり、書いてある内容を正確に表現できません（例1）。したがって、このような問題を解決するためには、「安納芋」のような一つの言葉を抜き出したり、「美味しくない」の「ない」のような否定の意味を落とさない処理が必要となります。また、このような処理をした場合でも、例えば、「安納芋は美味しくない」と同じ意味である、「安納芋は旨くない」はキーワードが「安納芋、旨くない」となるため、2つの文章の内容が同じと判定されません（例2）。このため、「美味しい」と「旨い」を同じ意味として抜き出す処理が必要となります。

農産物に関するテキストデータには、品種名など新聞記事にあまり出てこない単語が多く出てきたり（サツマイモの品種の一つである「安納芋」もその一例です）、品種名などの表記が複数あること（例；「安納芋」は「安のう芋」や「あんのういも」と表されます）がありました。

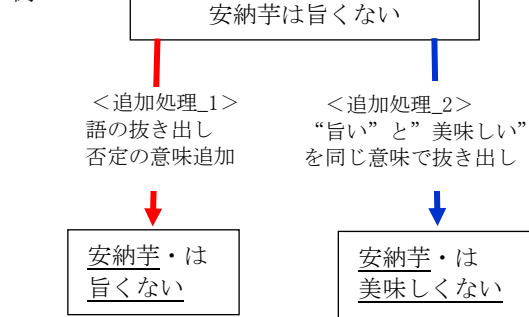
以上のように、農産物に関するテキストデータは独特であることから、既存の言語処理ツールだけでは「購入者の感想」の内容を正しく抜き出すことができませんでした。したがって既存の言語処理ツールを利用する際には、少なくとも農産物の用語や用語間の関係の情報（例；「安納芋」がサツマイモの品種を示す一つの用語であるということ、「安納芋」と「あんのういも」が同じ意味であること）が必要と判断しました。

今後は、これらの結果をもとに農産物に関するテキストデータの解析精度を向上させ、必要な情報を抜き出す方法を確立する予定です。

例1



例2



遺伝子情報を活用した効率的な品種改良

ウシの乳量などを高めるための形質改良にはこれまで長い年月がかかってきました。たとえば、乳量の多い雌牛をつくるための優良な種牛（雄親）を選ぶ場合、種牛候補の雄牛は乳を出さないで、雌牛と交配して産まれた多数の娘牛の乳量を調べないとどの雄牛が優良なのかわかりません。そのため、これまでの方法では、新たに優秀な種牛ができるまでに5年以上の時間を要していました。また、この方法ではその間種牛になれなかった雄牛についても飼育しておく必要があり、餌代等の多大な経費がかかるという問題がありました。

一方、最近では家畜の遺伝子情報が容易に得られる状況になっています。それにより候補の雄牛から

DNA検査によって優良な雄牛を選ぶことができるようになったことから、時間もコストも大幅に削減できるようになりました。実際この方法はすでに米国やカナダなどの乳牛の形質改良において利用されており、新しい種牛の育成期間が2年程度と大幅に短縮され、種牛の候補個体の飼育コストも不要となり、効率的な方法であることが実証されています。このような遺伝子情報を用いた効率的な形質改良法は、イネ、野菜、果樹等の作物への展開も期待されています。

(情報利用研究領域 林 武司)

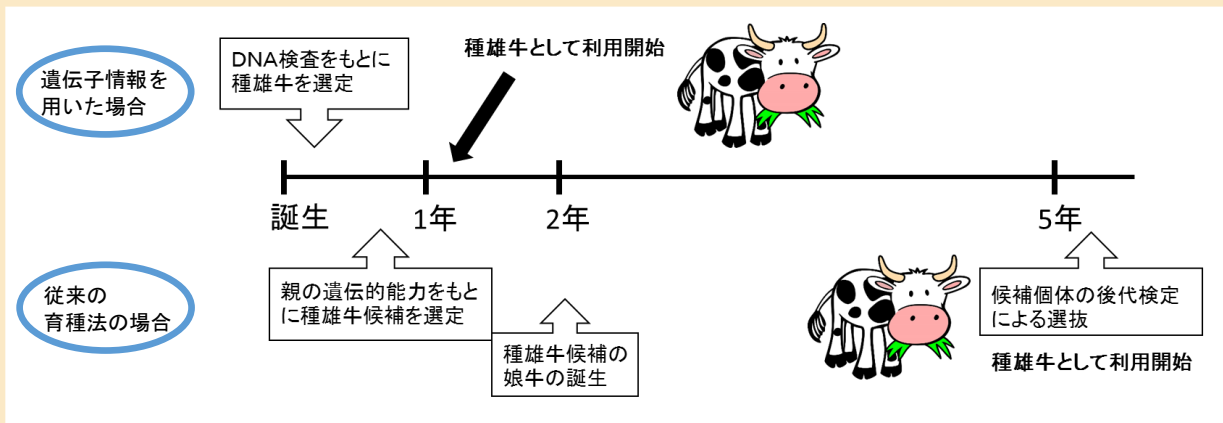


図. DNA情報の活用による育種年限の短縮.

上側：DNA情報を用いた場合の種牛育成期間、下側：従来の育種法を用いた場合の種牛育成期間