



HOKKAIDO UNIVERSITY

Title	A Study on Advisory Online Comments for a Motivational Dialogue System [an abstract of dissertation and a summary of dissertation review]
Author(s)	Patrycja Swieczkowska EWA
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第14135号
Issue Date	2020-03-25
Doc URL	https://hdl.handle.net/2115/78233
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
Type	doctoral thesis
File Information	Patrycja_Swieczkowska_abstract.pdf, 論文内容の要旨



学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称 博士（情報科学） 氏名 Patrycja Swieczkowska

学 位 論 文 題 名

A Study on Advisory Online Comments for a Motivational Dialogue System

（動機付け対話システムのためのオンラインコメントにおけるアドバイスに関する研究）

日常生活における動機付けの欠如は、現代の世界における差し迫った問題である。ますます多くの人々が、学校や職場で、また運動などの日々の努力において、やる気を失っている。したがって、この研究の最終的な目的は、ユーザにやる気を起こさせるアドバイスを与えるための動機付け対話システムを構築することである。これまでの研究で私は、このタスクが単純ではないということを確認し、日常的なタスクのための動機付け対話システムの開発を行った。この対話システムは気軽な対話や懇切な態度でユーザにモチベーションを与える予定であったが、評価を行った結果、5 ポイントのスケールで、動機付け機能が平均 3.9、対話機能が平均 3.1 と評価された。すなわち、効果的にやる気を起こさせる対話システムは、対話の気軽さ・親切さだけでは不十分で、ユーザの動機付けの欠如に対して動機付けを起こさせる会話を行うという機能が不可欠ということが明らかとなった。そこで、このような機能を開発するために本研究を行った。

以上の目的を達成するためには、まず、助言および動機付けのテキストが持つ性質を明らかにする必要がある。これは以下に述べるいくつかの点で有効である。まず、助言的または動機付けが存在するテキストのみを収集することにより、このような目的を持つ対話システムの学習データのコーパスを作成できる。できるだけ膨大なデータセットを収集するために、多くのデータソースを利用することが望ましい。「教えて!goo」のようなデータであればラベル（アドバイスかどうか）や評価（どのくらい良いアドバイスであるか）が付与されているが、ブログや一般的なディスカッションサイトの場合にはラベル無しデータが存在している。それらのラベル無しのデータを利用するためには、まずアドバイスではないテキスト、つまりノイズを除去する必要がある。その後、アドバイスの中から性質の良いもののみを抽出する必要がある。本論文で提案する分類アルゴリズムとランク付けアルゴリズムを用いて、前述したような目的を達成することができるものと考えられる。第二に、助言的または動機付けのテキストがどのようなものかを明らかにした上で、システムがどのような発話を生成すべきかについて重要な資料となる。本論文では、動機付けと助言的なテキストの性質に関する研究について述べる。以上より、本研究の主な成果は、助言的テキストの特徴の発見、その特徴の詳細な分析、あるいは助言的データを分類・ランク付けできる 2 つのアルゴリズムを提案したという 3 点である。自然言語処理分野において、動機付けをするアドバイスの分析は先行研究がほとんど存在しないため、その分析と助言的データを処理できるアルゴリズムを提案したことが本研究の新規性である。

分析のために使用したのは、ディスカッションプラットフォーム Reddit からダウンロードしたオンラインコメントである。ユーザがアドバイスを求めた投稿のみを使用したため、コメントは動機付けまたは助言を含んでいる。コメントのレベルは、他のユーザが付与したスコアで知ることができる。したがって、最高の助言と動機付けを行ったコメントのみを収集することとした。

データを収集した後、コメントの内容を分析し、その性質について考察を行った。分析した上で、

助言テキストの特徴な素性のセットを作成した。一般的に、最高のスコアが付与されたコメントには、多くの命令型およびアドバイス表現が含まれており、与えられたアドバイスは非常に具体的であることが確認された。さらに、アドバイスを与えた人は、過去に同じ問題を抱えていたためその問題をよく理解することができた場合が多いことが明らかになった。これらの性質をいくつかの素性として作成した上でセンチメント分析も行った。結果的に 14 のアドバイス素性が存在した。さらに、word2vec の分散表現の利用も行った。Word2vec は自然言語処理において重要なツールで、単語の意味を分散表現で適切に表現するものである。この分散表現を素性として利用することで単語の意味的特徴も利用することができる。

素性の有効性は、ニューラルネットワークを用いて 2 つのタスクにより確認を行った。最初のタスクは、データからノイズを除去するための分類である。2 つの浅いニューラルネットワークを組み合わせ、アドバイスを含むテキストと通常のテキストの分類を行った。2 つ目のタスクは、性質の良いデータを選択するためのアドバイスのランク付けである。畳み込みニューラルネットワークを使用して、テキストが 3 つというグループ内で動機付けおよび助言テキストのランク付けを行った。分類タスクでは、順伝播型ネットワークが相応しかったが、ランク付けタスクでは別のアプローチが必要となった。順伝播型ネットワークでは、各素性がわずかに異なる計算を行うのに対し、畳み込みニューラルネットワークでは、フィルターで素性のグループの計算を同様に行うことができる。ランク付けタスクの場合、入力順に関わらず各コメントがランキングで同等の評価を受けるように、コメントグループ内の各コメントの素性が最初のレイヤーで同じ計算を行う必要がある。そのため、各コメントの素性グループに同じサイズのフィルターを使用した。

前述したように、動機付けアドバイスの分析は、自然言語処理分野でこれまで研究されていなかったため、今回の実験でベースラインは存在しない。しかし、私の実験は段階的に実験を行い、徐々に改善を行っている。そのため最初の実験をベースラインと考えることができる。分類実験では、最初に SVM と順伝播型ネットワークを使用し、最終的に提案手法で実験を行った。提案手法としては、14 のアドバイス素性セットと word2vec の素性セットを個別に使用している 2 つの順伝播型ネットワークを連鎖させた。最初にデータを一つのネットワークで分類し、アドバイスでないとラベル付けられたものをデータから削除し、その残りのデータを次のネットワークで分類するという流れである。なお、ネットワークを適用する順序は任意である。最初の分類実験では、F 値が 0.760 であったが、最終実験では、ネットワークの順番により 0.943 ~ 0.971 に向上した。同様に、適合率は 0.844 から 0.965 ~ 0.977 に向上した。すなわち、提案手法がデータからノイズを効果的に除去できることが確認された。また、ランキング実験では 0.971 の精度が得られた。この実験では 14 のアドバイス素性と word2vec の素性を合わせて使用したので、ベースラインとしては word2vec の素性のみを使用する畳み込みネットワークが考えられる。そのネットワークの精度は 0.885 であった。これは、提案手法よりかなり低いものである。

本研究により、14 のアドバイス素性の有効性及び 2 つの提案手法の有効性が確認された。さらに、モチベーションに関する他の研究プロジェクトで使えるような助言と動機付けのテキストのコーパスの開発を行った。また、動機付けおよび助言テキストの性質について多くの知見を得ることができた。したがって、本研究では動機付けの対話システムを開発するための最初の段階を遂行したと考えられる。