

非タスク指向型対話システムでの話題同定
のためのコーパス構築に関する研究
Study on Corpus Construction for Topic Identification
in Non-task-oriented Dialogue System

学籍番号：201621616

氏名：柴田 尚樹

Shibata Hisaki

近年、スマートスピーカー、スマートフォンなどの普及に伴い、様々な対話システムが研究されている。このうち、雑談を行う非タスク指向型対話システムには、幅広い話題への対応という課題がある。この課題を解決する方法として、幅広い話題の記事を扱い、利用者によって日々記事が更新される Wikipedia から有益な記述を抽出して発話に利用するという方法が考えられる。そこで、本研究では、会話中の発話を、話題を表す記事として“会話している人らにとって有益な記述”が存在する Wikipedia 記事に結びつけることで、非タスク指向型対話システムでの話題同定に Wikipedia を利用するためのコーパスを構築する。

このコーパスは3つのステップによって構築され、そのうち2つのステップは、2段階に分けたクラウドソーシングである。この2段階は、クラウドソーシングのワーカーに有益な記述の抽出を行ってもらう段階と、前段階でワーカーが抽出した有益な記述についてどれだけの合意が得られるかを確認する段階である。何が有益であるかの判断が人により異なることを考慮し、どちらの段階でも、タスク1つにつきワーカー5人の回答を得る。そして、クラウドソーシングで獲得した有益な記述および各段階のワーカーの回答の集計を、コーパスとして保存する。

コーパス構築後、同じ記事中の同じ記述でも有益であるか有益ではないかの判断が人によって分かれることがあるなど、人によって有益とする判断基準が異なることを確認した。また、同じ発話から抽出された同じ記事内の記述であっても、記述によって有益であるとするワーカーの数が大きく異なる場合があるため、本研究で構築したコーパスを利用し、非タスク指向型対話システムの話者を同定する際には、記述の利用が重要であると言える。

本研究で構築したコーパスが、発話中の話題を同定する機械学習に加え、会話の話題の変化を追う機械学習、挨拶のみの会話など Wikipedia を参照する必要がない会話の判別を行う機械学習や Wikipedia 記事中の有益な記述を判別する機械学習に利用され、非タスク指向型対話システムがより多くの話題を扱う上での足がかりになることが期待される。

研究指導教員：手塚 太郎

副研究指導教員：若林 啓