

## レファレンス記録に対する NDC 自動付与の研究

佐藤 綾香

レファレンスサービスとは、図書館利用者に対し、その必要とする情報や情報源を効率よく入手できる様に図書館員が援助するサービスである。レファレンス業務は図書館員の経験や知識に依存するものである。近年これらのレファレンス記録を「蓄積、共有」することが求められている。主な理由として、原田らは以下の4点を挙げている。①レファレンス記録は自身がレファレンスツールとなる。②レファレンスマニュアルの作成に役立てることができる。③司書の教育目的に使える。④外部から図書館を評価するための判断材料としても用いることができる。国立国会図書館が提供しているレファレンス協同データベース事業はこれらを実現したものである。現在、全国の公立図書館、大学図書館、専門図書館等計 561 館が参加している。今後、レファレンス協同データベースを活用することによって、レファレンスサービスの向上に繋がることが期待される。

レファレンス記録は、検索のほかに、日本十進分類法に基づく分類番号である「NDC」、「地名」、「人物名」によるカテゴリ別に関連することが可能である。「NDC」に関しては、全てのレファレンス記録に付与することが可能であるが、現在 NDC の付与率は 100%ではない。原因として、NDC の付与は図書館員にとって負担となることが考えられる。図書館員の負担を減らすこと、また全ての記録に対して NDC を付与することで利用者の NDC 別閲覧の機会を増やすことが可能となる。米国では「QuestionPoint」と呼ばれる協働型のレファレンスサービスが行われている。日本での実現する際は、利用者から寄せられたレファレンス質問を、そのテーマを得意とする図書館に自動転送するシステムが有効である。

本研究では、フリーのデータマイニングソフト Weka を用い、機械学習による自動分類を試みた。原田らの先行研究では、レファレンス記録の 9 割を学習用、1 割を評価用とし実験を行っていた。しかし、学習用データとして、レファレンス記録約 3 万件では不足であると考えた。今回、年数万件の書誌事項が登録されている日本全国書誌に着目し学習用データとして採用し、レファレンス質問文を NDC 別に分類した。先行研究と同様、SVM、ナイーブベイズ、決定木の 3 手法による機械学習により、再現率の比較をした。先行研究の再現と、本研究手法を比較し評価を行った。

実験の結果、全体の再現率は SVM で最も高い結果となったが、先行研究の再現結果にわずかに及ばなかった。類ごとに見ると、3 類(社会科学)において再現率 80.5%、類によっては再現率 10%以下と言う結果となった。NDC と関連性の高い単語の選び方として、今後、図書タイトルとレファレンス質問文の性質の違いを考慮することが求められる。類による再現率に差が見られたが、3 類は日本全国書誌を使った本研究の方が良かったことから類ごとに学習用に用いるデータを変えることが有効であると考えられる。

(指導教員 辻慶太)