

BA4 自然資源の管理・利用についての伝統的知識に関する 情報検索システムの開発

Development of an Information Retrieval System for Indigenous and Local Knowledge of Management of Nature

地球循環共生工学領域 08E15056 花田拓実 (Takumi HANADA)

Abstract: Traditional knowledge, known as Indigenous and Local Knowledge (ILK), is critical to conserving biodiversity. However, the inheritance of ILK is getting difficult all over the world. Against this trend, machine learning is expected to have the potential to facilitate the challenge of inheriting ILK. The purpose of this study is to develop an information retrieval system to support the succession of Japanese ILK for managing nature. First, 529 articles from "Kikigaki Koushien" were collected, which is a knowledge base related to nature's contributions and inheritance and revival of tradition. And the articles were preprocessed and the natural language data frame with 10472 records was built. Next, traditional knowledge was learned from the data frame using Sentence-BERT, which was pre-trained on Japanese Wikipedia. Finally, we conducted an experiment with 19 subjects to evaluate the naturalness and usefulness of the 10 generated conversations and discussed future development perspectives.

Keywords: natural language processing, inheritance of traditional knowledge, Kikigaki Koushien, Sentence-BERT

1. 背景と目的

生物多様性保全の分野では先住民族・地域住民の伝統的知識 (Indigenous and Local Knowledge, ILK) が重視されている。ILK は生物多様性の評価の基準の一つであるが¹⁾、日本では継承が困難になりつつある²⁾。人工知能分野では、ユーザーに仏教的観点からアドバイスをする仏教の経典を学習した「ブッダボット」³⁾が開発されるなど、伝統的知識の継承を支援する試みが行われている。これらの背景から、本研究では自然言語処理技術と機械学習技術を用いて、自然資源の利用と管理の伝統的知識のデータベースから、対話的かつ自然に情報を検索できるシステムを開発することを目的とした。

2. 方法

2. 1 伝統的知識データフレームの構築と自然言語処理モデルの学習

「聞き書き甲子園」は、高校生が自然とかかわる名人にインタビューをして伝統的知識を記録するという取り組みである⁴⁾。過去の 1753 本の記事のうち、自然資源の管理・利用の「自然のめぐみ」と、知識の継承や復活の「伝承と再生」の 2 つのカテゴリで、合計 529 本の記事を情報検索システムの学習用に収集した。これらの記事に含まれる 6,226 個の文章に元の記事の識別 ID と、記事のジャンル ID を付与し、512 文字でスライシングすることで計 10,472 レコードの学習用データフレームを作成した。

学習には、機械学習モデル BERT⁵⁾ を文章ベクトルの埋め込み性能に特化させた Sentence-BERT⁶⁾ を用いた。学習用データフレームから基準となる文章を選び、記事の識別 ID が同じ文章を類似文章、ジャンル ID が異なる文章を非類似文章として、Triplet Objective Function を指定してモデルを追加学習した。

2. 2 情報検索システムの設計と性能評価

追加学習済み Sentence-BERT でユーザーの質問文と伝統的知識データフレームのコサイン類似度を計算し、類似度の上位 3 個の文章をランダムに結合して情報検索システムからの回答文を生成した。この回答に対して、情報検索システムの性能の評価実験を実施した。評価対象は自然資源の管理と利用に関する 10 の質問と情報検索システムの回答の対とした会話である。被験者 (N = 19) はその会話の自然度とシステムの回答の有用性を [1:very negative, 2:negative, 3:positive, 4:very positive] の 4 段階で評価した。

表1 情報検索システムの応答の例

| | | |
|--------|----|--|
| 高評価の会話 | 質問 | 都市に住んでいる私たちは自然管理とは無縁でしょうか？ |
| | 回答 | 人が自然に依存していることは間違いがなくて、とても共存なんておこがましい表現はよくないかもしれないけれど、出来る範囲で自然の恩恵を受けながらそこの自然を保てる形の活動ができると一番いいなと思うし。だから、共存できる森にするために私たちも何かできることがあるのではないでしょうか。それが中学生になると”私たちの出来る社会貢献とは何か？”って答えをこちらが言わなくとも自分たちで考えるようになる。 |
| 低評価の会話 | 質問 | 茶碗を紙で拭いて捨てるのとそのまま下水に流すのではどちらが環境負荷が低いですか |
| | 回答 | そうなった時のことを考えた時、今の山をこのままにしておかないと手入れをして、いい山を残していきたいなど考えています。昔は、全部手作業でやってたから軽いドリルでも一日使っていると夜には肩とか手があがらなくなって茶碗も持てなくなっている日、はくろう病になってしまったものだから、もっと簡単に作業できないか毎日、毎日考えたんです。やっぱり江戸時代から続いているから、これから若い人たちにも受け継いでいきたいですね。 |

3. 結果と考察

情報検索システムの応答の例と、全10問の会話に対する評価実験の結果を表1と図1に示す。全10問の会話に対する評価値の平均は自然度2.0(S.D.0.8)、有用性2.0(S.D.0.9)となった。このうち最も評価が高い会話では自然度2.9(S.D.0.5)、有用性2.9(S.D.0.6)となった。これは、”都市に居住していて普段自然の管理に携わることが少ない住民が管理に参加すべきか”という質問に対して、”人は自然に依存しているので、自然を守るために活動に参加するべきである”という、妥当と思われる回答をした。一方、評価が低い会話では自然度1.3(S.D.0.6)、有用性1.3(S.D.0.5)となった。この会話は水と紙によって食器を洗浄した際に生じる環境負荷の大小比較を問う質問に対し、茶碗という単語は入っているものの、”山を次の世代に継承するためには管理が必要であるが作業が過酷である”といった質問内容に無関係と思われる回答を返した。これは聞き書き甲子園の記事に食器の洗浄や下水への負荷にまつわる知識がないことによると考えられる。情報検索システムは、質問の内容によっては自然で有用な会話が生成できる可能性がある一方、さらなる性能向上が必要である。

4. 今後の課題

まず、聞き書き甲子園には全部で1,753本の記事が存在するため、更にデータフレームを増やすことが必要である。また、伝統地の有用性については、多角的な視点から伝統的知識の価値を評価する必要があるため、幅広いステークホルダーに評価を依頼することも重要である。特に伝統的知識の保有者や研究者など、より伝統的知識の継承の実践者の意見を反映させていくことが必須である。

参考文献

- 1) Ichikawa K. Indigenous and Local Knowledge in the Context of IPBES. JOURNAL OF RURAL PLANNING ASSOCIATION 2017, 36 (1), 34–37. <https://doi.org/10.2750/arp.36.34>.
- 2) 香坂玲; 内山倫太; 田代藍. 過疎化・人口減の縮小社会における伝統的生態学的知識の喪失とイノベーション. 日本健康学会誌 2018, 84 (6), 214-223. https://doi.org/10.3861/kenko.84.6_214.
- 3) 熊谷誠慈; 三浦典之; 栗野皓光; 上田祥行. Psyche Navigation System構想. 人工知能 2021, 36 (6), 684-694. https://doi.org/10.11517/jjsai.36.6_684.
- 4) 聞き書き甲子園 <https://www.kikigaki.net/> (accessed 2022-01-29).
- 5) Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805 [cs] 2019.
- 6) Reimers, N., Gurevych, I. Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks. arXiv:1908.10084 [cs] 2019.

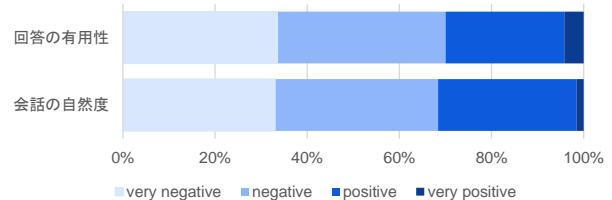


図1 情報検索システムの評価実験の結果