

機械翻訳済み文章からの質問生成を用いた SCDV による質問応答

深田明希[†] 関谷一真[‡] 濱上知樹[§]
 横浜国立大学[†] 横浜国立大学[‡] 横浜国立大学[§]

1 はじめに

様々な分野において自然言語技術による質問応答システムの需要が増えている。ドメイン知識を持たないユーザーのあいまいな質問から、求められる回答を対話的に導き出す技術は、業務の自動化やナレッジ検索など、多くの応用が期待される分野である。

質問応答システムは、事実や名称などを答えとする factoid 型と理由や方法を答えとする non-factoid 型に大別される。しかし、多くの応用ではこれらが混在したニーズがある。そこで本研究では、回答として事実や名称、理由や方法を区別することなく、回答に相当するソース文章を直接提示する方法を用いる。そのために、予めソース文章から推測される質問文を複数作成し、ユーザーの質問文との意味的類似性をもとにこれをキーとした検索によって回答ソースを提示する方法を検討する。このとき、ユーザーの質問文とキーとして予め生成する質問文が正しくマッチされることが重要である。そこで本稿では、質問文同士の表現として機械翻訳による英語表現を使うことで、意味的曖昧さを補完する方法を提案する。

2 従来手法の課題

質問応答システムの回答精度はユーザーの質問文の明確さに依存する。そのため、質問文の意味が一意に限定されることが望まれる。また、意味解析では構文解析が用いられるため、各言語の文構造が意味解析の正確性に寄与する。英語文はすでに分かち書きされており、構文が鮮明な状態での解析が可能である。一方、日本語文は分かち書きされておらず、表記や語順等によって意味解釈が多様になりやすい。質問文の意味に多様性があると、意図した回答が得られにくいという問題が生じる。

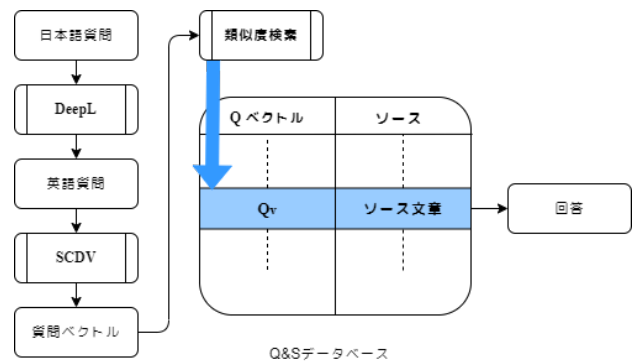


図1 提案システム全体図

3 提案手法

上記の問題を解決するために、機械翻訳を仲介した質問応答を提案する。まず、対象文書を分割し、各ソース文章を英語に機械翻訳することにより曖昧性を解消する。次に、自動質問生成機を用いて各ソース文章に関する質問を生成し、質問とソース文章をペアにする。そして、入力質問と類似度の高い質問に対応するソース文章を回答する質問応答システムを構築する。

3.1 機械翻訳による構文の曖昧性の解消

機械翻訳機の構文解析エンジンを通して、日本語の曖昧な構文を整理し、曖昧性が解除された英文へ変換する。本研究では、文脈や慣用句にも対応した DeepL^{*1}の API を用いた。

3.2 質問の自動生成

平叙文から質問文を生成するためには、その言語によらず、各形態素の品詞特定と正確な構文解析が必要である。本研究では、英文から機械学習によって質問を生成する Python ライブラリ (QuestGen^{*2}) を用いて質問と回答のペアをバッチで作成した。

4 実験・結果・考察

4.1 実験環境

4.1.1 SCDV モデル [1]

- 単語ベクトル: 事前学習済み FastText[2][3]
- 単語セット: 機械翻訳前処理済み本学履修案内中の英単語 1057 個
- クラスタ数:60

Question answering with SCDV using question generation from machine-translated sentences

[†] Hiroki Fukada, Yokohama National University

[‡] Kazuma Sekiya, Yokohama National University

[§] Tomoki Hamagami, Yokohama National University

^{*1} <https://www.deepl.com/translator>

^{*2} <https://questgen.ai/>

- 疎率:0.04

4.1.2 質問集

QuestGen を用いて機械翻訳済みのソース文章 137 個から 309 個の質問を生成し、各質問を言い換えて計 1204 個の質問を作成した。なお前処理として

1. 英語の小文字統一化
2. 数字の 0 統一
3. ストップワードおよび一文字の英単語, 数字, 記号の除去

を行った。ストップワードは Python 用の自然言語処理ライブラリ "NLTK"^{*3} 内の 127 語とし、類似度評価はコサイン類似度を用いた。

4.2 評価方法

履修案内に関する質問を人手で 88 個作成し、各質問と類似度が高い回答 3 つをペアにした各データに対し、{5. 十分満足できる, 4. 大体満足できる, 3. 不満はない, 2. 不満がある, 1. かけ離れている} のオピニオン評価を行った。

4.3 結果・考察

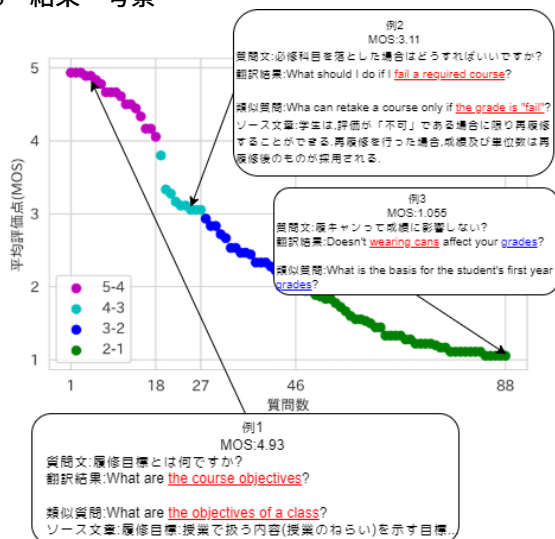


図2 平均評価点の評価者数平均の降べき順分布

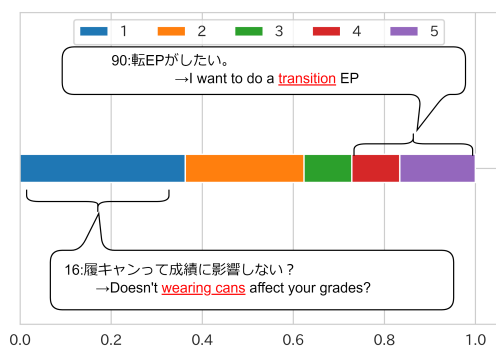


図3 全評価に対する各評価点の回数の割合

図 2 に平均評価点の分布, 図 3 に全評価に対して各評価点を与えられた回数の割合を示す。

図 2 において, MOS が 4 ~ 5 となった質問の多くは, 例 1 のように質問文内と履修案内中の表現が一致していた。例 2 では, 「科目を落とす」と「評価が不可」という異なる表現が「fail a grade」と統一化された一方, 例 3 では「履キャン」という略語は「wearing cans」と誤訳されていた。よって MOS の度合いは, 機械翻訳による意味補完で, 質問文内の表現が履修案内中と同義なものに変換されるかによると考えられる。

また図 3 より, この質問応答に満足していないと評価された割合が半分を超えていた。評価の改善にはモデル構築の単語セットの最適化, 略語表現等に対する翻訳機の精度向上が必要である。

5 おわりに

本研究では, 機械翻訳を通して意味的あいまいさが補完された文章を使用した質問応答システムを提案した。省略語でも履修案内上に存在する単語, 一般的な省略語俗語, 日本語独特の表現は, 機械翻訳を通すことで同義の表現が統一化される場合があり, 回答精度を向上させることが明らかとなった。今後は, 質問文とソース文章間の類似度を使用した場合も評価して, 本研究との比較を行う。

参考文献

- [1] Dheeraj M., Vivek G., Bhargavi P. and Harish K. "SCDV: Sparse Composite Document Vectors using soft clustering over distributional representations", Proc. of the 2017 Conference on Empirical Methods in Natural Language Processing, pp.659-669, 2017.
- [2] Armand Joulin; Edouard Grave; Piotr Bojanowski; Tomas Mikolov, "Bag of Tricks for Efficient Text Classification", Proc. of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Vol.2, Short Papers, pp. 427-431, 2017.
- [3] Piotr Bojanowski; Edouard Grave; Armand Joulin; Tomas Mikolov, "Enriching Word Vectors with Subword Information", Transactions of the Association for Computational Linguistics, Vol.5 pp.135-146, 2017

*3 <https://www.nltk.org/>