データ拡張による動機づけ面接のための MITI コード推定手法の改良

Improved MITI Code Estimation Method for Motivational Interviewing with Data Extension

比嘉 和樹[*1]，當間 愛晃[*2]
Kazuki HIGA[*1], Naruaki TOMA[*2]
[*1] 琉球大学工学部工学科知能情報コース
[*1]Presently with Computer Science & Intelligent Systems Program, School of Engineering, Faculty of Engineering, University of the Ryukyus
[*2] 琉球大学
[*2]University of the Ryukyus
Email: e205759@ie.u-ryukyu.ac.jp, tnal@ie.u-ryukyu.ac.jp

あらまし：人が現在の自分を変えたいと考え，行動するとき，多くの場合それには多大な努力を要する．動機づけ面接はそのような人をサポートするためのカウンセリングの一種である．本研究では動機づけ面接の評価指標である MITI (Motivational Interviewing Treatment Integrity)で評価のために面接中の各発話に付与されるコードを MITI コードと呼称し，動機づけ面接コーパスを用いたデータ拡張と MITI コード推定実験を行った．

キーワード：動機づけ面接，データ拡張

1. はじめに

禁煙やダイエットのように，自分をより良くするために習慣を変える努力をするとき，人は「変えたい」という思いと「変えたくない」という思いが拮抗し，一人の力で努力を継続することは非常に困難である．そういった人に対して有効な支援法の１つに、動機づけ面接[(1)]が挙げられる．動機づけ面接は相談者と面接者の１対１で行われ，相談者は自身の持つ両価性について語り，面接者は相談者の持つ問題の理解に努め，変化に前向きな気持ちを持たせる目的で対話を行う．

MITI (Motivational Interviewing Treatment Integrity)[(2)]は動機づけ面接を客観的に評価する指標の名称であり，MITI には面接を全体的に評価する過程と個別に発話の意味を示す MITI コードを付与して集計する過程がある．後者の過程は専門家による手作業で行われるため，大きな手間がかかる．そのため日本語の動機づけ面接コーパスが非常に少ないのが現状である．

本研究の目的は，手作業による MITI コード付与作業のコスト軽減のために，精度の高い MITI コード推定を行う機械学習モデルを行うことである．同時に小規模なコーパスの欠点を補うために有効なデータ拡張の手法の探索を行う．

2. 動機づけ面接コーパスの概要

本研究で用いる動機づけ面接コーパスは，過労死等防止調査研究センター・木内により構築されたものであり，9 件の動機づけ面接(総発話数 373 件)が収録されている．コーパス中で用いられている，簡略化された MITI コードの一覧を表 1 に示す．

3. データ拡張手法

本研究では動機づけ面接コーパスに対して 2 つの手法でデータ拡張を行った．

表 1 MITI コードの種類

コード	意味	説明	発話数
SR	単純な聞き返し	意味を変えずに繰り返す	31
CR	複雑な聞き返し	意訳、意味を付加して繰り返す	71
QUO	開かれた質問	自由に答えられる質問	35
QUC	閉じた質問	選択肢が限定された質問	20
AF	是認	理解を示す、肯定	5
-	要約	相談内容を要約する	1
Not Coded	該当なし	他のコードに該当しない発話	24

是認と要約のコードに対する ChatGPT-4 を用いた手法と，BERT(Bidirectional Encoder Representations from Transformers)[(3)]を用いた類義語置換による手法の 2 種類である．この 2 種について本節で説明を行う．

3.1 ChatGPT-4 を用いたデータ拡張

発話数が少ない是認と要約のコードについて、そのまま類義語置換によるデータ拡張を行った場合，それらのラベルについて極端に多様性に乏しいコーパスになり、学習を困難にすることが予測された．そこで，BERT を用いたデータ拡張を行う前に ChatGPT-4 を用いたデータ拡張を行った．ChatGPT に予め本研究で用いる MITI コードの種類と意味の情報を与え、動機づけ面接の面接者としての発話を生成する指示を行った．それぞれのコードについて生成されたものを確認し，主観的に自然と判断できたデータをコーパスに追加した．

3.2 BERT を用いたデータ拡張

TF-IDF または BM25 を用いて ChatGPT-4 による拡張後のコーパス中に存在する単語の重要度算出を行った．TF-IDF を用いた場合は対象の単語を名詞と形容詞のみに絞り，BM25 の場合は全単語を対象にした．その後，コーパスを「相談者の発話」，「直後の面接者の発話」，「直後の相談者の発話」を[SEP]トークンで区切った文章に成形し，面接者発話の MITI コードと同一のコードをラベルとして付与した．そのデータ中の重要度の高い 1 つ以上の単語に対して[MASK]トークンとの置換を行った．続いて BERT モデルにその文章を入力して全ての[MASK]トークンに当てはまる単語をそれぞれ推測して再度置換し，ラベルが同一の新しい文章を[MASK]トークンと同じ数だけ作成してコーパスに追加した．ただし，元の文章に存在していた単語と同じ単語を推測した場合はコーパスへの追加を行わず．生成された文章同士で同じ候補を用いないようにすることで可能な限りコーパスの多様性を確保した．この推定と拡張の手順を面接者発話全てに適用した．

4. 評価実験の手法と結果

評価実験の手法として，BERT を用いた分類実験を行った．[MASK]トークンの数が 1 個から 9 個の場合でそれぞれ 5 分割交差検証法を用いて学習モデルを評価した．図 1 に結果を示す．

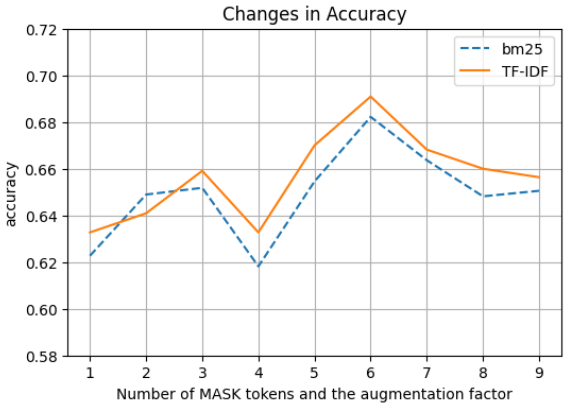

図 1 重要度算出に TF-IDF を用いた場合と BM25 を用いた場合の正答率推移の比較

TF-IDF を用いた場合と BM25 を用いた場合で，グラフには共通した傾向が現れた．いずれも[MASK]トークン 6 個，水増し文章 6 個追加の場合が最も性能が良く，最も良い条件で 69.1%の正答率になった。

図 2 に，図 1 中の最も良い条件で学習したモデルへのテストデータによる評価結果となる混同行列を示す．　図 2 から，意味の近い SR と CR，QUO と QUC の間の正答率が低いことが示された．　また，表 2 より，データ数の多い CR と QUO については Recall がより高く，データ数の少ない SR と QUC については Precision がより高いことから，データ数の偏りの影響を受けている可能性が考えられる．AF と要約については非常に高い精度を得られた．これらは他のラベルの中に意味的に近いものがなく，特徴を捉えやすいためと考えられる．

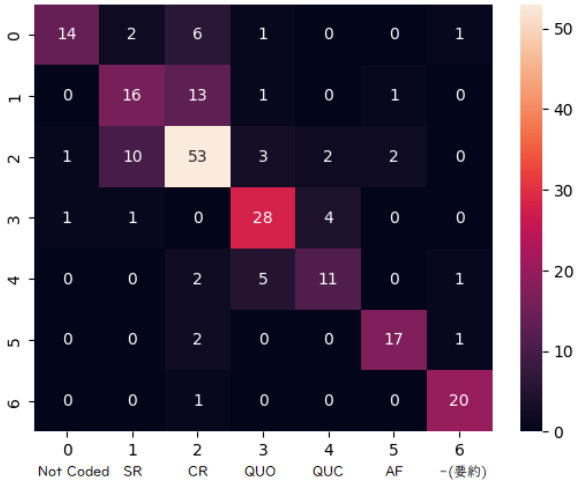

図 2 テストデータの混同行列

表 2 混同行列から算出した Precision-Recall

MITI コード	Precision	Recall
Not Coded	0.875	0.583
SR	0.551	0.516
CR	0.688	0.823
QUO	0.736	0.764
QUC	0.647	0.578
AF	0.850	0.850
-(要約)	0.869	0.952

5. まとめ

本研究では，小規模な動機づけ面接コーパスに対して ChatGPT-4 と BERT による類義語置換を主としたデータ拡張手法を実践し，MITI コード推定実験の正答率を比較することで評価を行った．

提案した手法では意味が近いラベルを混同する可能性が高く、データ拡張の規模や[MASK]を施す単語の数の調整で改善はできたが，更に性能の良いモデルの作成は難しい．今後の研究では，ラベルごとのデータ量の不均等性の是正や，類義語置換以外の手法による多様性に富んだ拡張手法を探求する必要がある．

参考文献

(1) W.R.Miller, S.Rollnick：”Motivational Interviewing, Third Edition: Helping People Change (Applications of Motivational Interviewing)”, 2012

(2) T.B.Moyers, J.K.Manuek, D.Ernst ： ” Motivational Interview□ing Treatment Integrity Coding Manual 4.2.1”, https://f6ee60c9-a097-4186-b1b8-5e89d6bb4f51.filesusr.com/ugd/d6c3d4_f15ca5cae476449a9f811cdd9c9fbda0.pdf

(3) Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova ： ” BERT: Pretraining of Deep Bidirectional Transformers for Language Understanding”, Proceedings of the 2019 Conference of the North American Chapter of Association for Computational Linguistics, pp.4171-4186(2019)