

情報理論を用いて受講者の自由記述から理解度を推測する授業改善支援システム A Method for Evaluating Student Comprehension through Descriptive Analysis Utilizing Information Theory

篠田 彪河¹⁾ 池田 了哉¹⁾ 辻本 了柁¹⁾ 金道 敏樹¹⁾
Hyuga Shinoda Ryoya Ikeda Ryosei Tsujimoto Toshiki Kindo

1 はじめに

受講者の力をよりよく伸ばす教育を実現するためには、授業の質を評価し、それを改善する仕組みが必要である。授業の質を評価する方法としてよく用いられるのは、授業後の受講者アンケートによる方法である [1][2]。このアンケートによる方法は簡便で優れた方法であるが、受講者の主観評価を前提とするために、受講者が授業内容を誤って「理解している」場合には「授業を理解できた」という回答になることを避けられない。この意味で、教師が伝えようとした知識（数学の概念など）が正しく受講者に伝わっているかまでを把握したい場合には、アンケートは必ずしも十分ではない。

ところで、教師が受講者の理解度を把握するために、授業のポイントを問にした自由記述のレポートを課すことは、よくある。そして、大学で教鞭をとる著者の一人は、その教育活動を通じて、こうした自由記述文は受講者が授業のポイントをどれだけ理解しているかを、非常によく表していると感じている。

我々は、本論文で、この自由記述文を情報量に注目して解析することで、教師が伝えようとした知識を受講者が正しく理解しているかを把握することが可能かを検討する。具体的には、受講者全員の自由記述文の集合を全体事象とし、これを高得点の受講者達（高得点群）の自由記述文の部分集合を上位事象とし、その余事象である受講者（低得点群）の自由記述文の部分集合を下位事象とする。この二つの事象を統計的に調べて、両者を区別する言葉とその構造が存在するか明らかにする。

この調査・検討においては、

- 対象とする自由記述文が特定的话题
- 多様な受講者を対象とするがそのクラス分類が得点の高得点群と低得点群と単純

という、自然言語処理としては小規模な問題であることを踏まえ、自由記述文がどのような意味を持つのかといった意味論への深入りを避け、単語及び2つの単語の対の出現パターンが形作る文の構造に集中する。

そして、この簡便な方法でも課題が問うている内容についての低得点群の理解の範囲と高得点群のそれとの違いを可視化できることを示す。

以下、第2章で、自由記述文を統計処理できる形式に変換する前処理（Word Vectorに変換するまで）を述べ、第3章で高得点群の自由記述文と低得点群の自由記述文とが与えられた際に高得点群と低得点群を分ける特徴を抽出する方法について述べる。第4章で実際の授業で提出された自由記述文において、高得点群と低得点群が使用する単語の違いについて実験結果を示す。第5章で高得点群と低得点の自由記述文の構造解析し、その違いを可視化する。第6章で、単語及び2つの単語の対の出現

パターンの構造までを見ることで、教師が伝えようとした知識を受講者が正しく理解できているかを把握することができるかについて議論する。

2 自由記述文のモデル

我々が取り扱う課題 a に対する答えである自由記述文を含むデータ $D(a)$ は、

$$D(a, p) = (d(a, p), s(a, p)), \quad p = 1, 2, \dots, n \quad (1)$$

書ける。ここで、 $d(a, p)$ は受講者 p が課題 a に対して書いた自由記述文であり、 $s(a, p)$ は $d(a, p)$ に対する評価者の採点結果である。

自由記述文 $d(a, p)$ は、単語 w の列

$$d(a, p) = (w_1(a, p), w_2(a, p), \dots, w_{n(a, p)}(a, p)) \quad (2)$$

で表現される。課題が同じであれば、書き手によらず自由記述文の中で同じ単語が使われる可能性が高いが、語順までが同じであることはほとんど期待できない。すなわち、二人の書き手 p_1, p_2 が書いた自由記述文中の i 番目の単語は、

$$w_i(a, p_1) \neq w_i(a, p_2)$$

であり、そのままでは簡単な統計解析には馴染まない。

今、語順を無視することになると、単語の順番は解析する側が自由に決められるから、十分多くの単語数 n_d を含み、その辞書中の語順を定めた辞書

$$\mathcal{D} = \{w_\alpha, \alpha = 1, 2, \dots, n_d\} \quad (3)$$

を用意すれば、自由記述文 $d(a, p)$ は、語順が共通の列

$$d(a, p) = ((w_\alpha, n_\alpha(a, p)), \alpha = 1, 2, \dots, n_d) \quad (4)$$

と書ける。ここで、 $n_\alpha(a, p)$ は自由記述文 $d(a, p)$ 中での出現回数である。

なお、本論文の中では、自由記述文中の語順はアルファベット (i, j など)、辞書中の語順はギリシャ文字 (α, β など) と区別して表現する。

この自由記述文 $d(a, p)$ はさらに簡略化でき、ワードベクトル

$$\mathbf{v}(a, p) = (n_1(a, p), n_2(a, p), \dots, n_{n_d}(a, p)) \quad (5)$$

でモデル化できる。

本論文では、自由記述文のワードベクトル $\mathbf{v}(a, p)$ が高得点群の受講者と低得点群のそれとの違いに興味があるので、一つの自由記述文の中での単語出現回数はそれほど重要ではないと考え、One-Hot ベクトル

$$n_\alpha(a, p) = \begin{cases} 1 & \text{単語 } w_\alpha \text{ の出現回数が } 1 \text{ 以上の場合} \\ 0 & \text{単語 } w_\alpha \text{ が出現回数が } 0 \text{ の場合} \end{cases}$$

に簡略化する。

1) 金沢工業大学 工学部 情報学科

3 高得点群/低得点群の違いの抽出方法

3.1 積算ワードベクトル

前節の議論から、我々が取り扱うデータは、One-Hot ベクトルと得点の組

$$D(a) = \{(v(a, p), s(a, p)), p = 1, 2, \dots, n\}$$

とモデル化された。

この One-Hot ベクトルと得点の組は、得点の閾値 θ_s を導入すると、

$$D^+(a) = \{(v(a, p), s(a, p)) | s(a, p) \geq \theta_s\} \quad (6)$$

$$D^-(a) = \{(v(a, p), s(a, p)) | s(a, p) < \theta_s\} \quad (7)$$

のように、高得点群のデータ $D^+(a)$ と低得点群のデータ $D^-(a)$ に分けることができる。

そして、それぞれの群から、課題 a について次のような高得点群の積算ワードベクトル $V^+(a)$ と低得点群のそれ $V^-(a)$ を

$$V^+(a) = \sum_{p \in D^+(a)} v(a, p) \quad (8)$$

$$= (N_1^+(a), N_2^+(a), \dots, N_{n_d}^+(a)) \quad (9)$$

$$N^+(a) = \sum_{p \in D^+(a)} 1 \quad (10)$$

と計算することができる。ここで、 $N_\alpha^+(a)$ は課題 a の自由記述回答で単語 w_α を使った高得点群の人数であり、 $N^-(a)$ は低得点群の人数である。

この積算ワードベクトルは、高得点群のデータ $D^+(a)$ と低得点群のデータ $D^-(a)$ に現れる単語の使用者数のヒストグラム (図 1) と等価である。

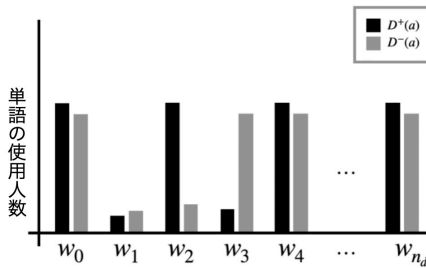


図 1: 単語の出現数が高得点群と低得点群で違うことを示すヒストグラムの概念図

3.2 成績の違いを特徴づける単語の抽出

辞書 \mathcal{D} に含まれる単語の範囲で成績の違いを特徴づける単語を取り出すには、前述の上位事象と下位事象を区別する上で情報量が多い単語 ($N_\alpha^+(a) \gg N_\alpha^-(a)$ または $N_\alpha^+(a) \ll N_\alpha^-(a)$ である単語) を選べば良い。我々は、高得点群がよく使う単語と低得点群がよく使う単語とを区別したいので、文献 [4] で提案された識別情報量

$$CIC(a, w_\alpha) = q(a, w_\alpha) \log \frac{q(a, w_\alpha)}{p(a)} - (1 - q(a, w_\alpha)) \log \frac{1 - q(a, w_\alpha)}{1 - p(a)} \quad (11)$$

$$q(a, w_\alpha) = \frac{N_\alpha^+(a, w_\alpha)}{N_\alpha^+(a, w_\alpha) + N_\alpha^-(a, w_\alpha)} \quad (12)$$

$$p(a) = \frac{N^+(a)}{N^+(a) + N^-(a)} \quad (13)$$

を用いる。

この識別情報量を使うと、辞書 \mathcal{D} に含まれる単語を

$$CIC(a, w_\alpha) \gg 0 \quad \text{肯定的単語}$$

$$CIC(a, w_\alpha) \sim 0 \quad \text{中立的単語}$$

$$CIC(a, w_\alpha) \ll 0 \quad \text{否定的単語}$$

(14)

と分類できる。ここで、肯定的単語は高得点群が特徴的に使う単語であり、否定的単語は低得点群が特徴的に使う単語である。中立的単語は、上位事象と下位事象を区別する上で役に立たない単語である。

4 受講者の理解度を可視化する実験

我々が取り扱う課題に対する解答文は一般の自由記述解答文と比較して表現の幅が非常に狭い。このことを明示するために、以下では対象の文章を自由記述解答文と呼ぶ。

4.1 対象とする自由記述解答文

今回の実験では、金沢工業大学工学部情報工学科 2 年生向けの「確率と統計」の授業で第 6 週目に課された課題「平均値、中央値、最頻値がどのような値か、200 文字程度で説明しなさい。」という問に対する自由記述解答文 (105 件) を、筆者らがテキストデータ化したものを用いる。

形態素解析には、MeCab をエンジンとして、mecab-ipadic-NEologd を辞書として用いる [3]。また、表 1 に示す固有名詞や、表記揺れの候補を辞書に追加し、統計用語が意図せず分割されることを防いだ。

表 1: 辞書に追加した単語

最頻値	基本統計量	左図	登場回数
中央値	頻度分布	右図	データ解析
平均値	データ全体	異常値	データ分布
正規分布	階級値	観測数	出現頻度
小さい	データ間	データ値	偏差値
データ数	代表値	要素数	あたい
左右対称	合計値	山型	いちばん
左右非対称	中心傾向	推定値	うける
右隣	偶数個	データ量	わかる
左隣	奇数個	一般分布	すべて
データ内	データセット内		

4.2 自由記述解答文の採点

自由記述解答文の得点は、

- 平均値の導出方法が正しく述べられている
- 最頻値の導出方法が正しく述べられている
- 中央値の導出方法が正しく述べられている
- 平均値は外れ値の影響あることが述べられている
- データが正規分布であるとき、平均値、最頻値、中央値が一致することが述べられている。

という 5 つの要素がいくつ含まれているかで与えた。

採点の結果は図 2 のような得点分布であったので、

- 4 点以上を高得点群
- 3 点以下を低得点群

とした。

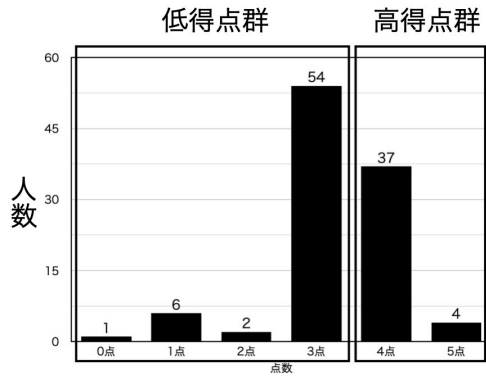


図 2: 自由記述解答文の得点分布

4.3 使用単語の違い

まず、高得点群と低得点群との間で、単語の使用人数に違いがあるかを確認する。

結果を分かり易くするために、出現頻度が多い単語について単語を識別情報量 $CIC(a, w_\alpha)$ について降べき順に並べた上で、各単語が上位事象と下位事象に出現する確率を、図 3 示す。

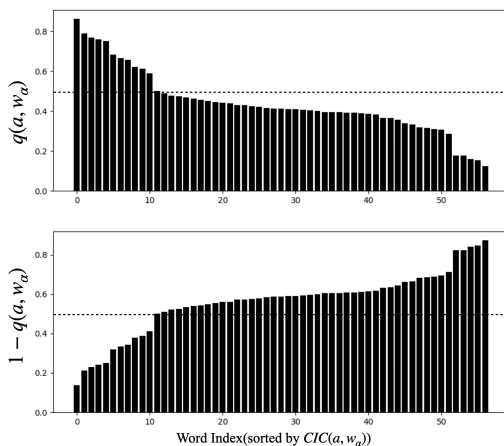


図 3: 単語の出現頻度が高得点群と低得点群との間には違いがあることを示す図。

図 3 から明らかなように、単語の出現頻度レベルにおいて高得点群と低得点群との間には違いがあることが分かる。

今解析している問において、出現確率が大きく異なる単語を書き出すと、高得点群では、「にくい、影響、特徴、受ける、外れ値」であり、低得点群では、「2つ、グラフ、平均、求める、偶数」である。このリストアップされた単語を見ると、低得点群の単語リストからは何らかの意味を汲み取ることは難しいが、高得点群のものからは「低得点群が論じていない、外れ値の影響を論じている」と解釈できそうである。これから、単語の出現確率を見るだけでも、高得点群が何を理解しており、低得点群が何を理解できていないかをある程度推定できることが分かる。

もう一点興味深いところは、この結果だけから結論を下すことは慎重にすべきであるが、高得点群に特徴的に現れる単語 $\{w_\alpha | q(a, w_\alpha) > 0.5\}$ と低得点群に特徴的に現れる単語 $\{w_\alpha | (1 - q(a, w_\alpha)) > 0.5\}$ の数を比較すると、

低得点群に特徴的に現れる単語数の方がかなり多いことがある。これは分かっていることを文で表現する際には必要な単語数は少ないが、分かっていないことをその場で解答しようとする并使用する単語数が増えてしまう現象を捉えているのかもしれない。

5 自由記述解答文の構造解析

高得点群の理解と低得点群の理解（もしくは理解不足）をより詳細に推定するために、我々は単語順序対を導入し、自由記述解答文の簡単な構造解析を行う。

5.1 単語順序対の導入

我々が導入する単語順序対とは、自由記述解答文中の名詞、動詞、形容詞のごく局所的な順序関係を取り出すものである。具体的には、まず、自由記述解答文を形態素解析して得られる m 個の単語からなる分かち書きから、例えば w_2 と w_{i+1} から w_{j-1} が助詞などであった場合には

$$w_1, w_2, w_3, \dots, w_i, \dots, w_j, \dots, w_m \\ \rightarrow w_1, w_3, w_4, \dots, w_i, w_j, \dots, w_m$$

とそれを除いて、名詞、動詞、形容詞だけの単語列を順序を保って抽出する。そして、この単語列において隣接する単語の順序対 $(w_1, w_3), (w_3, w_4), \dots, (w_i, w_j), \dots$ が、ここで言う単語順序対である。各単語順序対は、辞書 \mathcal{D} の語順を使って、

$$\{(w_\alpha, w_\beta), (w_\gamma, w_\delta), \dots, (w_\xi, w_\eta), \dots\}$$

と自由記述解答文中の語順によらない単語順序対の集合へ変換する。この単語順序対を導入することで、単語間の方向を持ったつながりの強さ、おそらくそれは自由記述解答文の論理を弱いながらも反映させることができると、我々は考えている。

この二つの単語、単語 w_α の次に単語 w_β という単語順序対 $o_{\alpha\beta}$ に対しても、我々は単語と同様に単語順序対の識別情報量 $CIC(a, o_{\alpha\beta})$ を考えることができる。さらに、識別情報量が $CIC(a, o_{\alpha\beta}) \gg 0$ である単語順序対に注目することで高得点群特有の、識別情報量が $CIC(a, o_{\alpha\beta}) \ll 0$ である単語順序対に注目することで低得点群特有の局所的な単語の繋がりを取り出すことができる。

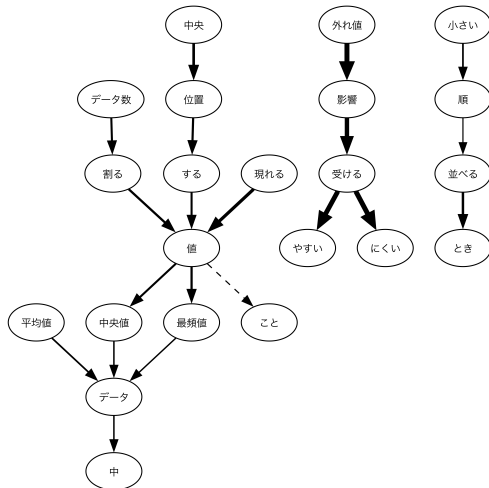
5.2 自由記述解答文の単語ネットワーク表現

ところで、我々が取り扱っているのは、制限のない自由記述解答文ではなく、数学の課題に対する自由記述解答文である。当然、表現の幅は狭く、重要な単語は高得点群と低得点群に関係なく当然使われていると考えられる。つまり、重要な単語は、中立的な単語になっている可能性が高い。このことを考えると、自由記述解答文の構造解析においては、重要な単語を洩らす事なく、かつそれがどのように他の単語と繋がっているかを明らかにする工夫が求められる。

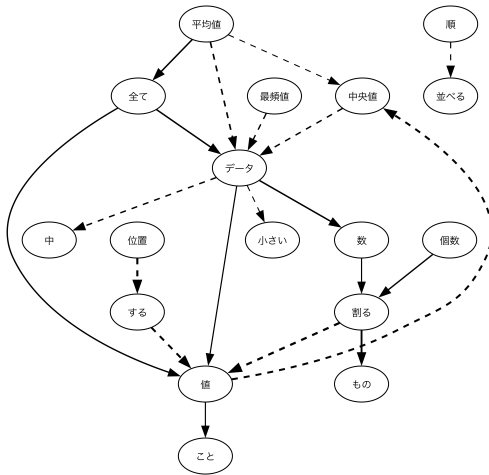
この要請に対する解として、高得点群と低得点群の自由記述解答文を単語ネットワークで可視化する際に、

- 重要な単語を残すために、ノードとなる単語は使用者の人数 $N_{\alpha\beta}^+(a)$ が大きい単語順序対 $o_{\alpha\beta}$ から選ぶ
- ノード間（単語間）のつながりの強さは、単語順序対の識別情報量 $CIC(a, o_{\alpha\beta})$ で表現する

とする方法を採用した。



(a) 高得点群の自由記述解答文の単語ネットワーク表現



(b) 低得点群の自由記述解答文の単語ネットワーク表現

図 4: 重要な単語を残すためにノードとなる単語は使用者の人数が多いものから順に選び、ノード間(単語間)のリンクの強さは単語順序対の識別情報量で表現した自由記述解答文の単語ネットワーク表現。

高得点群と低得点群の自由記述解答文の単語ネットワークを、それぞれ単語順序対を使用者の人数 $N_{\alpha\beta}^+(a)$, $N_{\alpha\beta}^-(a)$ が大きい方から 20 個選んで作成した結果を図 4 に示す。

図中の矢印は単語順序対で表現されている語順を示している。実線リンクは、注目している得点群の単語ネットワークにおいて、当該群中に顕著に現れる単語間リンクを表現している。破線リンクは、注目している得点群の単語ネットワークにおいて、当該群中にほとんど現れない(逆の群に顕著に現れる)単語間リンクを表現している。

5.3 自由記述解答文の単語ネットワークの解釈

前節の実線リンクと破線リンクの意味を踏まえて、高得点群のネットワーク表現を解釈すると、そこには次のような記載があることが分かる：

- 中央 (に) 位置する値 (は) 中央値
- データ数 (で) 割る値
- 外れ値 (の) 影響 (を) 受ける, やすい or にくい

• 小さい順 (に) 並べるとき。
「データ数 (で) 割る値」から「平均値」へのリンク、「小さい」「順 (に)」「小さい順 (に) 並べるとき」から「中央値」へのリンクがない点に不満があるものの、妥当な記載ができていたことが読み取れる。

一方、低得点群のネットワーク表現を解釈すると、実線リンクを辿ることで

- 平均値 (は) 全て (の) データ数 (で) 割るもの

という記載があるものの、「中央値」「最頻値」に繋がる記載はほとんどないことが読み取れる。

これらの結果は、「低得点群は、平均値は理解しているものの、高得点群が理解している中央値を十分には理解できていない」ことを示唆している。

6 まとめ

教師が伝えようとした知識を受講者が正しく理解できているかを課題に対する自由記述解答文から把握することができるかを、情報量に注目して検討した。

この検討に当たって、自由記述解答文中の局所的な語順を表す単語順序対を新たに導入し、自由記述解答文を方向を持った単語ネットワークで表現する文構造解析方法を提案した。

単語と単語順序対の中でも、特に高得点群と低得点群の表現の違いを強く表すものを文献 [4] で提案された識別情報量を用いて抽出して、高得点群と低得点群の文構造の違いを単語ネットワークの違いとして可視化した。提案した文構造解析方法は意味解析に立ち入らない簡略な方法であるが、新たに導入した単語順序対によって、低得点群の理解の範囲と高得点群のそれとの間に違いがある程度読み取れるようになった。

今回取り上げた問題は比較的簡単なものであり、自由記述解答文も 200 文字程度と短いものなので結論は急ぐべきではないが、可視化された文構造の違いを詳細に見ることで、低得点群が何を理解できていないかまでも単語ネットワークから読み取れる可能性がある。

今後は、自由記述解答文が 2 倍程長くなる専門性の高い問題を取り上げ、提案した文構造解析方法の有効性と限界を見極めるとともに、提案手法を実装した理解度を推測する授業改善支援システムを試作し性能評価に取り組む予定である。

倫理審査

本研究は、金沢工業大学倫理委員会の承認(承認番号 23-09-010)を得て、個人が特定されないよう配慮した。

参考文献

- [1] 堤 厚博, 西岡 圭太, 谷口 進一: 数学教育における重要単元・項目抽出の試み, 日本工学会教育協会, 工学教育研究講演会講演論文集, pp.12-13, 2018.
- [2] 石崎 龍二, 佐藤 繁美: 統計演習科目における学生の自己評価に基づいた教育効果の検証, 福岡県立大学人間社会学部紀要 2019, vol.27, No.1, pp.125-142, 2019.
- [3] 工藤 拓: MeCab: Yet Another Part-of-Speech and Morphological Analyzer, 入手先 <<https://taku910.github.io/mecab/>> (参照 2024-06-14).
- [4] Toshiaki KINDO, Hideyuki Yoshida, Tetsuro Morimoto, and Taisuke Watanabe: Adaptive Personal Information Filtering System that organizes personal profiles automatically, Proceeding of IJCAI97, pp.716-721,1997.