一般口演 | 知識工学
一般口演9
自然言語処理・テキストマイニング
2021年11月20日(土) 09:10 〜 11:10 E会場 (2号館2階222+223)

[3-E-1-08] 学部生の Web試験解答におけるコピー＆ペーストが成績に与える影響のテキストマイニングによる推定

*田中 武志[1]（1. 広島大学　病院　医療情報部）
*Takeshi Tanaka[1]（1. 広島大学　病院　医療情報部）

キーワード：Undergraduate Education, Medical Informatics Education, Text Mining

[背景]A大学の歯学部（歯科医師、歯科衛生士、歯科技工士に対応した3つの専攻を有する）では2年次生を対象に医療情報学の講義（医療情報倫理、患者の権利、情報セキュリティ、医療情報システム、標準化、医療安全、等）を行っている。成績の90%を決める筆記試験を毎年実施しているが、2020年度は COVID-19の流行により試験を Web上で実施した。 Web試験はコピー＆ペーストによるカンニングを防ぐことが困難であるが幸い明確な同一解答は学生間で見られなかった。しかし語句や言い回しの類似が疑われる解答が散見された。[目的]テキストマイニングの手法を用いて学生間の解答の類似性を調べ、それが成績にどのように影響しているのかを推測する。[方法] 特にカンニングによる不正な利益を得る高得点者と悪影響を受ける低得点者の解答に着目し、解答毎の名詞、動詞、形容詞の出現頻度を元に解答をクラスタリングした（ R ver.4.1.0 Windows64bit版を使用）。同じような得点の学生が集まるクラスタを抽出したあと、そのクラスタとは異なる得点帯の解答をサンプリングし、名詞の出現度などに違いがあるかを調べた。[結果]3専攻の本試験受験者90名を対象とした。満点の8割以上を得点している高得点者が集まるクラスタを見つけることはできなかった。しかし特定の低得点者の一群はクラスタリングの方法（距離とグループ分けの定義）を変更しても同一クラスタが形成されるため、このクラスタに着目した。名詞の出現頻度を調べると高得点者および平均程度の得点者の解答と比べて語句「患者」の出現頻度が少ない傾向が見られた。[考察]低得点者のクラスタの学生は試験において効果的でない解答案を共有し試験に用いていた可能性が高いと考える。またその語句の出現頻度は患者の居る実際の場面を想定した具体性の欠如の傾向を示す可能性があると考える。

学部生の Web 試験解答におけるコピー＆ペーストが成績に与える影響のテキストマイニングによる推定

田中 武志[*1]
*1 広島大学病院医療情報部

An Estimation of Impact of Copy-and-Paste Operation by Undergraduate Students in Answering a Web Examination on Their Grades by Text Mining

Takeshi Tanaka[*1]
*1 Dept. Medical Informatics, Hiroshima University Hospital, Hiroshima, Japan

Abstract: [Background] The Faculty of Dentistry of University A has lectures on medical informatics for second-year students. In 2020, the examination of the lectures was conducted on the Web due to the COVID-19 epidemic. Although no clear identical answers were found among students, there were some answers that were suspected to be similar in terms and phrases. [Purpose] Using text mining techniques, to find cheats by students in the examination, the author examined the similarity of answers between students and inferred its impact on grades of student. [Method] Based on the frequency of appearance of nouns, verbs, and adjectives for each answer, answers were classified in several methods. The author found a cluster of answers with similar scores (with focusing on the answers of high-scoring and low-scoring students) that is robust against changing methods of classification. After classifying the answers in score bands that were different from that cluster, the author investigated whether there were differences of word-appearance-frequency between the cluster and score bands. [Result] The author found a suspicious cluster of answers, in which many answers were in the lowest-score band. Answers in the cluster had different word-appearance-frequencies from those of other score bands. [Discussion] The author argues that the result is useful to prevent cheats by students in the examination in the future.

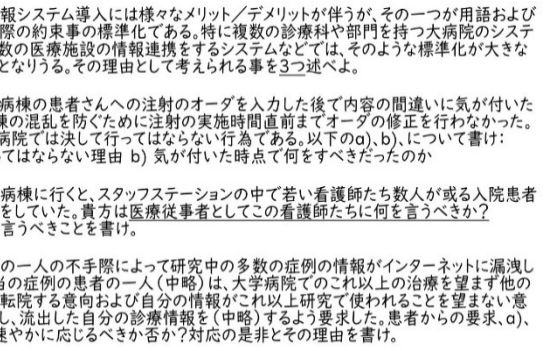
Keywords: Undergraduate Education, Medical Informatics Education, Text Mining.

1. 背景

国立大学 A 大学の歯学部（歯科医師、歯科衛生士、歯科技工士に対応した 3 つの専攻を有する;以下、順に S 専攻、E 専攻、G 専攻と称する）では 2 年次生を対象に医療情報学の講義（医療情報倫理、患者の権利、情報セキュリティ、医療情報システム、標準化、医療安全、等）をオムニバス形式で行っており、成績の 90%を決める筆記試験を毎年実施している。問題の多くは記述式であり、学生に基本的な用語と考え方を問うものである（図1）。また試験に際しテキストやノート、資料の持ち込みは無しとしている。複数の講師の都合で試験問題を毎年大きく変更することは実質的に困難な状況であり、毎年若干の問題修正を行いつつ、毎回試験問題を回収していた。しかしに年が経つにつれて学生が似たような内容の回答をするようになり、学生が過去の出題問題に対して「模範解答」を作成していることが推察される状況になった。また学生の「模範解答」と思しき答案の内容が必ずしも優良な解答とはいえないことも多々あった。しかしながら筆記試験の形態であれば、どのような形であれ、学生が一度内容を咀嚼したものをアウトプットする形になるため、それなりに学習した成果が解答となっていると考えて、不合格点を取らない限りは大目に見ていた状況であった。

しかし2020年度はCOVID-19の流行により、留学生を含めて 100 名ほどの受講学生を試験会場となる教室に集めることができなかったため、学生は在宅のまま大学が保有する Web 上の学習支援システムで試験を実施せざるを得なかった。解答時間に制限をかけ、Web ブラウザの動作を制限したものの、本質的に Web 試験は学生間の解答のコピー＆ペーストによる不正行為を防ぐことが困難である。また Web 試験で解答しやすい形式に問題文を改変したものの、前述の理由で問題内容の抜本的な変更は行わなかった。

試験実施後、採点の際に同一解答の有無を確認したところ、幸いにして学生間で全く同一の解答は見られなかった。しかし特に日本語の試験解答において語句や文章の類似が疑われる解答が散見された。



- 医療情報システム導入には様々なメリット／デメリットが伴うが、その一つが用語および連絡の際の約束事の標準化である。特に複数の診療科や部門を持つ大病院のシステムや複数の医療施設の情報連携をするシステムなどでは、そのような標準化が大きなメリットとなりうる。その理由として考えられる事を3つ述べよ。
- 貴方は病棟の患者さんへの注射のオーダを入力した後で内容の間違いに気が付いたが、病棟の混乱を防ぐために注射の実施時間直前までオーダの修正を行わなかった。これは病院では決して行ってはならない行為である。以下のa)、b)、について書け:
  a) 行ってはならない理由 b) 気が付いた時点で何をすべきだったのか
- 貴方が病棟に行くと、スタッフステーションの中で若い看護師たち数人が或る入院患者の噂話をしていた。貴方は医療従事者としてこの看護師たちに何を言うべきか?
  貴方が言うべきことを書け。
- 医局員の一人の不手際によって研究中の多数の症例の情報がインターネットに漏洩した。該当の症例の患者の一人（中略）は、大学病院でのこれ以上の治療を望まず他の病院へ転院する意向および自分の情報がこれ以上研究で使われることを望まない意向を示し、流出した自分の診療情報を（中略）するよう要求した。患者からの要求、a)、b)、に速やかに応じるべきか否か?対応の是非とその理由を書け。

図 1 2020 年度の問題文（抜粋）
試験解答を明示しないため一部省略あり。

2. 目的

コピー＆ペーストによる試験の不正行為の有無を調べるため、テキストマイニングの手法を用いて日本語解答の学生間の類似性を調べ、それが学生の成績にどのように影響しているのかを推測する。

3. 方法

3.1 解析の方針

テキストマイニング的手法で日本語の文章から著者を判別

するためには二万字程度以上の字数が必要との議論[1]があるが、試験問題の模範解答は千字弱程度であるため完全に著者を判別することは困難である。またコピー元の文章との単純一致を避けつつ元の文章の意味を保つように変更する場合、文体や言い回しの変更が最もあり得る方法であるが、その場合主要な意味を担う名詞、動詞、形容詞の変更は少ないと考えられる。よって解答の中の名詞、動詞、形容詞の出現頻度に着目し、それを元に学生の解答間の類似度を探った。特にコピー＆ペーストを用いた不正行為による利益を得る高得点者と悪影響を受ける低得点者の解答に着目した。

3.2 テキストデータの処理

学生の解答データのうち、選択問題の解答を他の設問の解答と重複しないキーワードに置き換え、残りの問題(医療情報倫理についての国際ガイドライン(記述)、医療情報の標準化の意義(記述)、患者情報保護(記述二問)、医療安全と医療情報システム(記述二問)、コンピュータウィルス対策(記述)、電子カルテの三要件(単語解答)、医療情報の開示と訂正(単語解答および記述)、著作権保護(記述))の解答を学生毎に一つのテキストファイルにまとめた。

3.3 解析方法

統計解析用ソフトウェアとして R ver.4.1.0 Windows 64bit 版および R の日本語形態素解析ライブラリ RMeCab Ver. 1.07 を用いた。[2](以下、登場する関数名は RMeCab の関数名である。)

テキストファイル毎の各名詞、各動詞、各形容詞の頻度を測定してテキストファイルをベクトル空間の座標で表し(docMatrix 関数を使用)、各テキストの座標間の距離を測定し階層型クラスタリングの手法を用いて(dist 関数を使用し hclust 関数を用いてクラスタ間の距離を示すデンドログラムを描いて)各テキストファイルを分類した。テキスト間の距離の定義方法として、ユークリッド距離、キャンベラ距離、マンハッタン距離、最長距離法の4つ、各クラスタ間の距離の定義方法として、最小分散(ウォード)法、群平均法、最短距離法、完全連結(最長距離)法、McQuitty 法、重心法の 6 つ、の合計 24 通りの組み合わせを試しながら特定のテキストファイル群によるクラスタが共通して形成されるか否かを試行した。特に高得点者(得点率 80%以上)と低得点者(得点率 60%以下)の学生のテキストファイルに着目し、同じような得点の学生が集まるクラスタを抽出した。クラスタを抽出したあと、そのクラスタのテキストとそれ以外得点群のテキストで、言葉の出現頻度などに違いがあるかを調べた。

4. 結果

4.1 解答データの基本属性

Web 上での期末本試験(日本語)には 90 名の学生が解答した。3 つ専攻の内訳は、S 専攻 53 名(内、高得点者 2 名、低得点者 7 名)、E 専攻 18 名(内、高得点者 1 名、低得点者 2 名)、G 専攻 19 名(内、高得点者 4 名、低得点者 0 名)であった。各解答のテキストデータは、解答者一人あたり数百字から千数百字程度であった。

4.2 解答データのクラスタリング

クラスタリングを行った結果、S 専攻の低得点者を含むクラスタが 24 通り中 23 の定義の組み合わせにおいて見られた。最短距離法と最長距離法の組み合わせのみ該当のクラスタが明確に出現しなかったが、S 専攻の低得点者が近い距離で固まる傾向が見られた。24 の組み合わせ中 18 の場合で 7, 8 人程度の、ほぼメンバーが同じクラスタを形成した。(表 1)

表 1 で S 専攻低得点者を含むクラスタの人数が 7 名になっている場合では、クラスタの生成順序の違いはあるものの、全てメンバーは同じであった。人数が 8 人(および 9 人以上)になっている組み合わせでは、デンドログラムからクラスタ形成の順序を読み取ると、前述の 7 名に加えて 8 人目(および 9 人目講)が加わる形になっていた(図 2、3)。8 人目以降のメンバーは、多少共通性がみられるものの、完全に同じメンバーではなかった。

表 1　距離の定義法の組み合わせと S 専攻低得点者を含むクラスタの人数

＼テキスト間 クラスタ間＼	最小 分散	群 平均	最短 距離	完全 連結	McQ- Uitty	重心
ユークリッド	7	8	8	7	8	3*
最長距離	7	8	不明	7	8	5*
マンハッタン	7	8	8	8	9	7
キャンベラ	7	8	8	8	11	4*

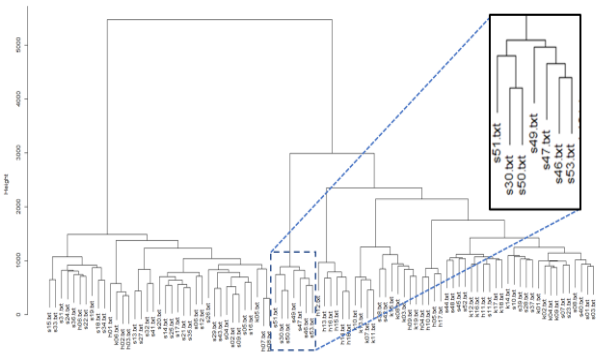


図 2　デンドログラム(最小分散法／キャンベラ距離)
右上は S 専攻低得点者含む 7 名のクラスタ部分の拡大図。

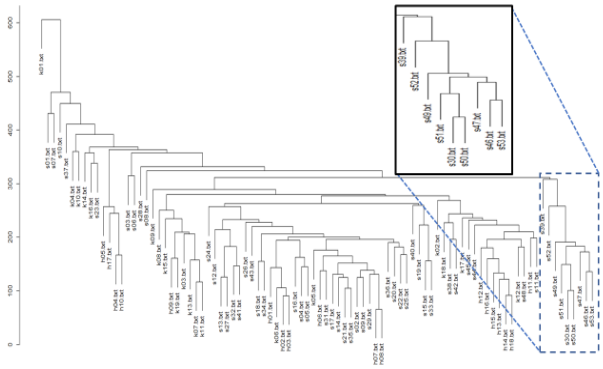


図 3　デンドログラム(McQuitty 法とマンハッタン距離)
右上は S 専攻低得点者含む 9 名のクラスタ部分の拡大図。図 2 の 7 名のデータに加えて「s52」、次に「s39」のデータが最後に加わる形になっている。

またクラスタの人数が 7 名未満の場合(表 1 でアスタリスクを付けた場合)のメンバーは全て 7 名の場合のサブセットであり、それ以外のメンバーが入ることは無かった。しかし、1 名の中位得点者の解答は必ずこのサブセットに含まれていた。

よってこの 7 名の解答を S 専攻低得点者クラスタのメンバー（以下、クラスタ A）と定義し、以降の分析で用いた。7 名の内訳は、低得点者 5 名、比較的低位の得点者 1 名、中位の得点者 1 名であり、全て S 専攻である。

クラスタ A のようにほとんどの距離の定義の組み合わせにおいて形成されるクラスタは他には発見されなかった。しかし幾つかの距離の定義方法の組み合わせに於いては、例えばE専攻の比較的成績下位のメンバー（低得点者を含む）のクラスタなどが散見された。しかしながら、それらのクラスタについてはメンバーがクラスタ A ほど定かでないため、今回の解析には含めていない。

表 2　クラスタ A とその他の群における一般名詞の出現頻度（上位 10 語）

クラスタA（7 名）		成績上位者（7 名）		左記以外（76 名）	
一般名詞	頻度/人	一般名詞	頻度/人	一般名詞	頻度/人
情報	9.71	情報	13.00	情報	8.87
医療	5.57	患者	10.14	患者	7.32
個人	3.86	医療	5.71	医療	3.89
患者	3.57	ウイルス	3.71	ウイルス	2.74
原則	3.57	個人	3.57	個人	2.38
ウイルス	2.29	リスク	2.43	最新	2.00
本人	2.29	義務	2.14	人	1.43
最新	1.86	最新	2.14	リスク	1.42
条件	1.86	ID	2.00	状態	1.39
データ	1.57	カルテ	2.00	ソフト	1.38

表 3　クラスタ A とその他の群における動詞の出現頻度（上位 10 語；同率 10 位含む）

クラスタA（7 名）		成績上位者（7 名）		左記以外（76 名）	
動詞	頻度/人	動詞	頻度/人	動詞	頻度/人
する	21.00	する	28.71	する	22.13
ある	4.29	れる	7.14	れる	5.05
れる	3.00	ある	4.57	ある	3.99
できる	2.57	いる	4.43	できる	3.46
反する	2.29	できる	4.00	なる	2.83
おく	2.14	おく	3.86	おく	2.78
通す	2.00	なる	3.57	いる	2.50
なる	1.86	得る	3.14	行う	2.18
伝える	1.86	行う	2.29	得る	1.43
取る	1.57	られる	1.86	しまう	1.43
－	－	しまう	1.86	－	－

4.3 分類したクラスタにおける名詞出現頻度

クラスタ A における一般名詞の出現頻度を調べると、頻度の高い順に「情報」「医療」「個人」「患者」「原則」となったが、7 名の高得点者および高得点者と上記クラスタのメンバー以外の 76 名では「情報」「患者」「医療」「ウイルス」「個人」となり、クラスタ A における「患者」の出現頻度が他の群よりも 1/2 から 1/3 程度に低くなる傾向が見られた。また「原則」もクラスタ A 以外では上位には現れない（表 2）。

4.4 分類したクラスタにおける動詞出現頻度

クラスタ A における動詞の出現頻度を調べると、概ね頻度の高い動詞は他の群と共通性が見られるが、五番目の「反する」は他の群では上位には出現しなかった（表 3）。

4.5 分類したクラスタにおける形容詞出現頻度

全体的に形容詞の出現頻度は名詞、動詞と比べると低いが、クラスタ A においては「多い」の出現頻度が他の群と比べると高く、一人一回と76名の群の5倍以上の頻度であった（表 4）。

表 4　クラスタ A とその他の群における形容詞の出現頻度（上位 6 語）

クラスタA（7 名）		成績上位者（7 名）		左記以外（76 名）	
形容詞	頻度/人	形容詞	頻度/人	形容詞	頻度/人
ない	1.00	ない	1.86	ない	1.38
多い	1.00	正しい	1.00	正しい	0.33
よい	0.43	やすい	0.43	多い	0.17
いい	0.14	いい	0.29	よい	0.13
新しい	0.14	怪しい	0.29	やすい	0.12
良い	0.14	よい	0.29	いい	0.11

5. 考察

5.1 クラスタリングの妥当性

クラスタ A の抽出については、用語などの特徴を調べるためにある程度の人数が必要であることを考慮し、且つクラスタに不確かなメンバーが含まれないよう配慮した。形容詞については元々の出現数が少ないことから明確なことは言いにくいが、名詞と動詞については表 2、3 のようにクラスタ A とそれ以外の傾向の違いを得られたことから、概ね妥当であったと思われる。

今回はコピー＆ペーストによる不正行為の有無に関して試験解答のデータ以外の情報が一切無く、また元々試験問題の解答という語彙が重なり易いデータのため、様々なクラスタリングの方法を試しながら方法に因らず共通に出現する頑強なクラスタを探索する方法をとらざるを得なかった。

クラスタ A とは別の解答の類似したクラスタを探索する際に、クラスタ A が効果的に分離できることをクラスタリングの方法を決める基準とすれば、より効果的なクラスタの探索が可能になると考える。

5.2 解答の類似が成績に与えた影響とその原因

クラスタ A は、用語の頻度について他の群とは異なる特徴が見られるだけでなく、クラスタリング方法の変更に対して非

常に頑強であり、且つクラスタの中に必ず 1 名の中位得点者が含まれていた。従ってクラスタ A の特徴は単なる低得点者の解答傾向を反映した結果とは考え難い。

クラスタ A のメンバーの多くが低得点者であり上位の成績の者が居ないことから、彼らが試験で高得点を取るために効果的ではない「模範解答」を共有していた可能性は十分にあると考える。

用語の傾向(「原則」「反する」)から「何故クラスタ A のメンバーの解答が効果的でないか」を考察すると、「個人情報保護」や「OECDの(個人情報保護)8原則」などの用語を乱用して解答していることが考えられる。

多くの医療情報関係の教育と同様、本講義の内容においても個人情報保護法およびその元になっている OECD の個人情報保護の 8 原則は重要な地位を占めており、講義の中で多くの時間を割いて解説している。しかしながら試験問題では、それらの原則および他の法律やガイドラインを考慮して、患者の存在を想定し具体的にとるべき行動を問う問題が多くなるため、具体的なことを解答に書こうとすると単語「患者」が多くなる。

しかしながら講義内容をよく理解してない学生が、具体的な内容に触れることなく解答する為に、安易に重要と思われるキーワードを解答して少しでも点数を取ろうとすること(例えば「OECD の 8 原則に反するから」云々と解答すること)は充分に考えられる。また著者がクラスタ A の解答を直接読んだ際に、そのような傾向が実際にあると感じられた。

そのような高得点を取れない解答でも、試験に合格すれば「模範解答」として次年度の学生に伝えられる可能性少なくない。実際、数年前のS専攻の試験解答には上記のような傾向の解答が多く、次年度の講義中に学生に注意を促したことがあった。不正行為があったかどうかは別にして、そのような出来の悪い「模範解答」が今回共有されていた可能性は充分にあると考える。

5.4 コピー＆ペーストによる不正行為の有無

「背景」で書いたように、試験において学生間で完全に一致する解答はなく、また過去の出題を元に作成された「模範解答」が学生の間に幅広く出回っていることが窺える状況を考えると、この解析結果はあくまでも状況証拠でしかなく、実際に不正行為が行われたかどうかまでは判断できない。クラスタ A のメンバーが試験時間内に同じ「模範解答」を見ていなくても、試験勉強の際に共通の「模範解答」を使っていた所為でこの結果になった可能性もある。

したがって、本解析の結果を用いて試験における不正の証明を行おうとすることは得策ではない。

5.4 改善のための方策

教育においては何よりも講義内容に対する学生の習熟と理解を高めることが重要である。学生の考え方を変えてより講義のテーマに沿った質の高い解答ができるよう指導することの方が重要である。

すでに成績判定を終えてしまった学年ではなく、次年度の学生たちに今回の解析結果を提示し、安易なキーワードの乱用では高得点に繋がらないこと、そしてWeb試験の場合は類似した解答があった場合に不正受験の疑いがかかる可能性があることを伝え、講義内容に真剣に取り組むことを促す必要があるであろう。その際、解析結果の提示が新たな「模範解答」を生まぬよう、学生への具体的な語句の提示には慎重になるべきと考える。

その一方で、単純にキーワードを並べただけの解答にならないような問題の出題方法も検討すべきであろう。

2021年8月現在、大学関係者のワクチン摂取率が高くなったとはいえ、COVID-19の流行は未だ収束の見えぬ状況である。そのため今後数年間は Web 試験への対策とそれに基づいた学生への指導を考え続ける必要があると考える。

6. 結語

自然言語処理の手法を用いて医療情報学の Web 試験の解答間の類似性を調べた結果、低得点者の解答を多く含むクラスタが抽出され、他の群と異なる語句の頻度の傾向を持つことが明らかになった。コピー＆ペーストによる不正行為が実際に行われたかどうかは定かではないが、このクラスタの学生の間で試験において高得点を取る上で効果的でない解答案を共有し試験に用いていた可能性が高いと考えられる。

学生の解答の質を向上させるために、この解析結果の一部を次年度の学生に提示して啓発する一方で、安易なキーワード乱用による解答を避けるような問題作成が必要であると考える。

参考文献

1) 松浦司, 金田康正. n-gram の分布を利用した近代日本語文の著者推定. 計量国語学 2000;22(6):225-38.
2) 石田基広. テキストの自動分類. R によるテキストマイニング入門. 森北出版株式会社, 2008:131-42.