

# 小論文の自動採点

## A Method of Essay Scoring based on Word Vectors

笠原 要\*1\*2

Kaname Kasahara

天野 成昭\*1

Shigeaki Amano

佐藤 哲司\*1

Tetsuji Satoh

\*1 日本電信電話 (株) NTT コミュニケーション科学基礎研究所

NTT Communication Science Laboratories, NTT Corporation

Previous methods of essay scoring cannot grade an essay if there are several model opinions and reasons to support them, because they regard an essay as a bag of words. We propose a new method of essay scoring based on relationships between opinions and reasons and word vectors acquired from documents.

### 1. はじめに

コンピュータを用いて試験を行うデジタルテストにおいて、答案データの採点を自動化できれば、試験実施に要するコストの削減が期待できる。また、e-Learningでの復習テストの採点が自動化できれば、採点結果を生徒に直ちに伝えられるので、生徒の習熟度が向上すると期待される。問題に対して複数の選択肢から受験者に回答を選ばせる選択式の問題ならば自動採点は容易であるため、実際に様々な自動採点が行われている。一方小論文では、1つの正解に対して様々な言語表現が可能であり、さらに複数の正解が考えられる場合もあるため自動評価は難しい。しかし、小論文の自動採点は、受験生の作文能力や論理的思考、保有する分野知識などの多側面より評価できる点で重要な試験形式である。さらに、言語理解や知識獲得等の様々な人工知能研究成果を評価するためのターゲットとしても有用であり、様々な検討が行われている。

小論文を自動採点する方法としては、答案から抽出可能な表層的な特徴の最適値を採点者が採点した答案より学習し、それに基づいて答案を評価する方法がある [Page 66, Burstein 98, 石岡 02]。これらでは、模範解答を予め作成せずに採点できる長所があるが、しかし、答案集合の一部を手で採点した結果を事例とする場合には、それを作成する時間が必要となり時間やコストを要する点が問題である。また、過去に行った同一の問題の採点結果を利用する場合には、新たな設問の試験には適用できない点が問題である。

一方、答案のテキストを予め作成した模範答案や設問のテキストを比較し、関連性の高さに基づいて内容を採点することを特徴とする方法も提案されている [Landauer 97, 長坂 01]。同一の内容の模範答案に対して様々な言語表現が考えられるため、テキスト同士を文字や単語表記の出現傾向で比較するのは不十分である。そのため、これらの方法では、設問に関連する文書集合やテキストコーパスより単語を多次元空間に配置し、さらにテキストも多次元空間に配置し、テキスト間の意味的な関連性の高さで内容の近さを評価している。例えば [Landauer 97] では、複数の文書を対象としそれぞれに含まれる単語の出現頻度を要素とした行列を作成し、それを特異値分解して文章や単語の多次元空間を作成して検索を行う Latent Semantic Analysis, LSA [Deerwester 90] を利用している。しかし、これらの方法では、テキスト全体を単語の集合 (bag of

words) と単純化しているため、文内の単語の構文や意味関係や、さらには内容に関する文間の結びつき等は無視している。従って、設問に対して複数の妥当な意見が存在し、さらに、意見のそれぞれを説明する理由が複数考えられるような複雑な知識構造を必要とする小論文については適用が困難である。

そこで本稿では、上記の問題点に対応し、採点済みの答案を学習事例としてなくても適用可能で、複数の内容の正解が可能な設問においても自動採点可能な方法を提案する。

### 2. 提案方法

本章では、答案の内容を評価する新しい自動採点方法を提案する。まず、対象とする小論文の設問形式を規定し、そのための知識の表現および獲得方法について述べる。そして、これらを用いて答案を自動採点する方法を説明する。

#### 2.1 設問形式

小論文の設問では、ただ1つの正解が存在するというよりも複数の意見に分かれるトピックを扱うことが多い。例えば、「日本政府や地方自治体は、積極的にダムを建設すべきか否かについて述べよ」という設問では、立場や見方によって建設する /しないのどちらも適切となる。このような複数の意見のいずれもが妥当となる設問は、提示された文章を要約し、その要約した内容に関する受験生の考えを問う設問や、指定された複数のキーワードを用いて文章を作成する設問でも同様である。

小論文の設問で複数の意見に分かれるトピックを扱うことが多い理由として、予めただ1つの正解の意見があるような設問では、選択形式の設問と傾向が似通ってしまい、受験生の柔軟な思考能力を評価することができなくなるためと考えられる。そこで、複数の内容の正解が存在する設問に対する答案を自動評価する方法を本稿では検討する。ここでは検討の第一歩として、ある事柄や考え方の是非を論ずる最も単純な設問形式を取り上げる。是非のいずれも言語表現は様々考えうる場合が多いが、形式を単純化するために、答案では、是非を選択させることとする。

このような2択を問う小論文の設問について、どちらの選択肢についても、それを説明する理由はただ1つとは限らない。例えば上記のダムに関する設問の場合、積極的にダムを建設すべきであるという意見に対して、電力を確保するために必要という理由や、水害を防ぐために必要という理由など様々に考えられる。勿論、これらの理由が相互に関係しあう場合もあり得るが、明快な答案を作成するためには、主要な理由をまず述べるのが受験生に望まれると予想される。そこで、ある事柄や考えの是非を問う設問に対して、いずれかの選択とその主な理

連絡先: 天野 成昭, 日本電信電話 (株) NTT コミュニケーション科学基礎研究所, 〒 619-0237 京都府相楽郡精華町光台 2-4 email: amano@cslab.kecl.ntt.co.jp

\*2 現在 西日本電信電話 (株) に所属

由を述べた答案を自動採点の対象とする。このような設問の設定および、それに対する答案の一例を図 1 に挙げる。

**設問**  
日本政府や地方自治体は、積極的にダムを建設すべきか否か回答し、それを選んだ理由を述べよ。

**答案例**  
Yes/No 建設すべきである (建設すべきではない)  
理由 ダムの建設によって、周囲の自然が破壊され、貴重な動植物が絶滅する恐れがあるため。

図 1: 小論文の設問 / 答案例

## 2.2 選択理由の知識表現と獲得

前節で述べた設問と答案形式に対する模範答案としては、是非それぞれに対する適切な理由が考えられる。さらに、個々の理由に対する言語表現は様々であり、受験生が答案で是非のどちらを選択しても、それに対する適切な理由が記述されていれば高い採点ができるような理由の表現が必要である。そこで小論文の答案を内容において自動採点するために、図 2 のような 3 層構造の理由の知識表現を行う。最上位は選択肢であり、是非の 2 通りを考える。次に、それぞれの選択肢を選ぶ際に妥当な理由を分類と見なす。そして、個々の理由を説明する複数の言語表現を対応づける。このような理由を設定することやその言語表現を記述するためには、設問に関する分野知識や専門用語を必要とする場合があるため、専門家による作成を行うことが必要である。これについては、分野に関するテキストコーパスを利用した概念検索によって自動化が可能と予想される。

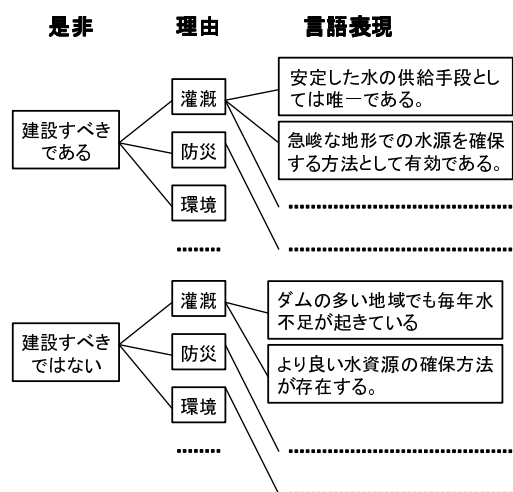


図 2: 理由の知識構造

概念検索とは、キーワードどうしの意味的な関連性を考慮した検索技術であり [日経 98]、先に挙げた LSA [Deerwester 90] もその 1 つと考えられている。また、国語辞典を用いて見出し語を多次元空間に配置する方法 [笠原 97] や、テキストコーパス中の単語同士の共起傾向に基づいて単語を空間配置する方法 [Schutze 92] を用いた情報検索手法 [Schutze 94, 熊本 99] も提案されている。本稿では、テキストコーパスを用いた概念検索 [Schutze 94] を利用する。国語辞典を用いて単語を多次元空間に配置する方法 [笠原 97] では、一般常識的な観点で単語

が配置されるのに対して、[Schutze 92] では、設問に関する分野のテキストコーパスを用意できれば、分野に応じた単語の配置が可能となるためである。

理由知識を自動獲得する手順を以下に述べる。

- (1) 設問に関する文書集合を用意する。
- (2) 設問に含まれるキーワード、あるいは採点者が重要と考えられるキーワードを含む文書を選択する。
- (3) (2) で得られた文書集合より単語間の共起傾向を収集し、単語の多次元空間を作成する。
- (4) (2) で得られた文書それぞれについて含まれる単語の空間配置の重心を求める。
- (5) 文書の多次元空間中の配置を自動分類し、空間中で各分類の重心に近くに配置される単語を理由、分類に含まれる文書を理由の言語表現とする。

なお単語の多次元空間作成および文書の分類の方法の詳細については文献 [熊本 99] を参考されたい。また、後述の通りこの概念検索手法は、答案と理由の比較においても利用される。

## 2.3 答案の採点

設問に関する専門家が作成、あるいは、分野コーパスより自動作成される選択理由知識を用いて、受験生が作成する答案を評価し自動採点する方法について述べる。

まず、答案の選択理由が用意した理由の分類のいずれに含まれるかを判定し、その理由に対応した是非と答案で選択された是非を比較する。その際に、答案における理由と構築された理由の分類やそれに含まれる言語表現との関連性の高さは、前節でも述べた概念検索手法を利用する。あらかじめ獲得しておいた理由に対応する言語表現の多次元空間中での配置は、どのような分布であるかは予め明らかではない。そのために、分布を仮定した分類手法よりも、k 近傍法、すなわち、分類中の k 個の言語表現と答案の選択理由を比較し、最も関連が高い言語表現を含む分類に答案の理由を識別するような手法が汎用であるとも考えられる。例えば、図 1 の答案に対して図 2 の理由知識を適用し、答案の理由が“建設すべきである”下の理由に分類された場合には答案の選択と一致しないので、得点は 0 とする。一方、“建設すべきではない”下の理由に分類された場合には、部分点を与える。この方法では、選択理由にいずれかの選択理由とは無関係な文章を書いた場合、適切に評価できない。そのため、理由の知識表現としては、2 択とは無関係な文書を用意して対応する必要がある。例えば、「XX ダムで昨日、殺人事件が発生し、犯人が逮捕された。」という文書を無関係な理由として分類の言語表現として用意しておく。答案の理由がこれに分類される場合には、得点は 0 とする。

答案の選択とその理由が予め用意した理由の分類と対応する選択と一致する場合であっても、複数存在する理由の言語表現の中で例外的なものに近しい場合や、複数の言語表現が集まる多次元空間中の配置に近い場合があり、これらを同一に評価することはできない。そこで、代表的な言語表現と答案の理由が近い場合に高い得点を与える。理由の分類中の言語表現の分布は予め予測はできないため、ここでは、理由に含まれる言語表現の配置の重心に対する各言語表現の関連性の度合いを概念検索の手法により計算し、その統計情報に基づいて採点する方法を用いる。例えば、答案の理由と分類の重心の関連性の度合いを計算し、それが関連性の度合いの平均から標準偏差を差し引いた値以上である場合には、満点とし、それ以下の場合

には、満点の半分とする方法などが考えられる。これについては、答案の採点者の採点結果より各種パラメータを最適化する必要があるが、これは分野知識というよりは、採点に関する一般的な感覚と考えられるので、一度獲得すれば、他の分野の問題においても同様に適用可能な数値となると期待される。

### 3. 実験

小論文の自動採点の基本的な方式検討を行うための評価実験を行った。設問としては、図1に示した例と同じ設問を用いた。まず、毎日新聞1991年から2000年までの記事より、単語「ダム」を含む2279記事を抽出し、これを用いて [Schutze 92] の方法で単語の多次元空間を作成した。

次に、理由知識の作成を行った。先に示したように理由知識を自動作成する方法もあるが、ここでは採点の結果の検討を重視するために、主として人手で理由知識の作成を行った。まず、www を利用し、ダム建設の是非それぞれの立場を代表する国土交通省のダム建設に関する基本的考え方 (URL [http://www.mlit.go.jp/river/dam/thinking\\_index.html](http://www.mlit.go.jp/river/dam/thinking_index.html)) と、と世界銀行と世界自然保護連合が主導で開催された世界ダム委員会の最終報告書の要旨 (URL [http://www.mekongwatch.org/documents/pr\\_20001116\\_01.html](http://www.mekongwatch.org/documents/pr_20001116_01.html)) に基づき、是非それぞれに対応する6個の理由、合計12個を設定し、それを補足する説明を人手で作成した (表3)。

そして、2279記事中の「ダム」を含む953文について、これらの12の理由あるいはどれにも分類されないかを人手で評価した。その結果、640文は12の理由のいずれかに分類された。これを用いて答案を評価する実験を現在行っている。

表 1: 選択の理由

是非	理由	説明
是	採算	恩恵に見合うコストで建設, 管理されている
	受益	ダムによる恩恵を利用者が享受している
	退去	退去者や周辺住民の生活が保障されている
	環境	周囲の自然環境や生態系が保全されている
否	灌漑	水足の地域に適切に水を供給できている
	防災	台風等による水害を防ぐことができている
	採算	計画額を大幅に超える費用が出費されている
	受益	ダムによる恩恵を利用者が享受していない
否	退去	退去者らが生活環境の変化で不具合が生じる
	環境	周囲の自然環境や生態系が損なわれている
	灌漑	水不足の地域に適切に水を供給できていない
	防災	台風等による水害を防げない

### 4. おわりに

本稿では、小論文を自動採点する方法として内容に注目し、是非を問う設問に対して、その是非に対応する理由が適切であるかを評価する方法を提案した。今後は実験の評価および、それを通して方式のパラメータの最適化を試みる予定である。また、実験では理由知識を半自動で作成したが、これを自動化する方法もあわせて評価する予定である。さらに文中の「～ではない」のような否定表現を考慮した評価の高度化も検討する予定である。

### 参考文献

- [Burstein 98] Burstein, J., Kukich, K., Wolff, S., Lu, C., Chodorow, M., Braden-Harder, L., and Harris, M. D.: Automated Scoring Using a Hybrid Feature Identification Technique, in *Proceedings of the Thirty-Sixth Annual Meeting of the Association for Computational Linguistics and Seventeenth International Conference on Computational Linguistics*, pp. 206–210 (1998).
- [Deerwester 90] Deerwester, S., Dumais, S., Furnas, G., Landauer, T., and Harshman, R.: Indexing by latent semantic analysis, *Journal of the American Society for Information Science*, Vol. 41, pp. 391–407 (1990).
- [Landauer 97] Landauer, T. K. and Dumais, S. T.: A Solution to Plato's Problem: The Latent Semantic Analysis Theory of the Acquisition, Induction, and Representation of Knowledge, *Psychological Review*, Vol. 104, pp. 211–240 (1997).
- [Page 66] Page, E.: Grading essays by computer: Progress report, in *Notes from the 1966 Invitational Conference on Testing Problems*, pp. 87–100 (1966).
- [Schutze 92] Schutze, H.: Dimensions of Meaning, in *Proceedings of Supercomputing 92*, pp. 787–796 (1992).
- [Schutze 94] Schutze, H.: Cooccurrence-based vector expansion, Technical report istl-qca-1994, Xerox PARC (1994).
- [笠原 97] 笠原, 松澤, 石川: 国語辞書を利用した日常語の類似性判別, *情報処理学会論文誌*, Vol. 38, No. 7, pp. 1272–1284 (1997).
- [熊本 99] 熊本, 島田, 加藤: 概念ベースの情報検索への適用 概念ベースを用いた検索特性の評価, *情処研報*, 第 SIG-ICS 115 巻, pp. 9–16 (1999).
- [石岡 02] 石岡, 亀田: コンピュータによる日本語小論文の自動採点システム, *電子情報通信学会技術研究報告*, No. TL2002-40, pp. 43–48 (2002).
- [長坂 01] 長坂, 古瀬, 石井, 西川: Web 環境における記述問題自動採点エンジンの開発, *高度情報化支援ソフトウェアシーズ育成事業* (2001).
- [日経 98] 日経 BP: 概念検索が可能な検索エンジン, *日経オープンシステム*, No. 67, pp. 108–111 (1998).