

センター試験『国語』評論読解問題ソルバーの改良の検討

A Machine-Learning Based Solver for Comprehension Questions
in Contemporary Japanese of the National Center Test加納 隼人 佐藤 理史 松崎 拓也
Hayato Kanou Satoshi Sato Takuya Matsuzaki名古屋大学大学院 工学研究科 電子情報システム専攻
Graduate School of Engineering, Nagoya University

This paper describes a new solver for comprehension questions in Contemporary Japanese of the National Center Test. A target question consists of a text body, a question sentence, and five choices. The core of the new solver is an SVM binary classifier that determines which is better between two choices. The solver selects the winner of the round-robin tournament of five choices. Our solver correctly solves 94 questions in 240 comprehension questions, which outperforms previous solvers.

1. はじめに

2011年に、国立情報学研究所で、「ロボットは東大に入れるか(略称:東ロボ)」というプロジェクトが開始された[新井 12]. このプロジェクトは大学入試問題を計算機で解くことに挑戦するものであり、その中間目標は、2016年までに、大学入試センター試験において東京大学の二次試験に進めるレベルの点数を取ることである.

我々は、2013年5月からこのプロジェクトに参加し、大学入試センター試験の『国語』現代文の問題を解くシステムの開発に取り組んでいる. 特に本研究では、『国語』第1問の評論の、読解問題を対象とする. 対象とする読解問題は、センター試験『国語』現代文評論の配点の約2/3を占める.

本稿では、この読解問題の自動解法として、複数の特徴量と機械学習を用いる新しい解法を提案する. 提案する解法は、「2つの選択肢のうちどちらがより本文に合致しているか」を判定する判定器を用いて、各選択肢のスコア計算を行う. この判定器は、機械学習によって構成する.

以下、本稿は次のように構成されている. まず、2節で、大学入試センター試験の『国語』の構成と、本研究で対象とする読解問題について説明する. 次に、3節で、提案する解法について説明し、4節ではその評価実験について述べる. 最後に、5節でまとめを述べる.

2. 対象とする問題

大学入試センター試験の国語では、大問4問が出題される[教学 13]. 大問の第1問が評論、第2問が小説、第3問が古文、第4問が漢文の問題である. 第1問の「評論」は、何らかの評論から抜き出された文章(本文)と、それに対する設問から構成されている. ほとんどの場合、設問は6問あり、表1のような構成になっている.

本研究では、問2から問5の読解問題を対象とする. この問題の具体例を図1に示す. この図に示すように、対象とする問題は、1つの設問文と5つの選択肢が与えられ、正解の選択肢を1つ選ぶ選択式の問題である. 以降では、この問題を単

表1: 『国語』評論の設問構成

問1	漢字問題(小問5問)
問2-問5	本文中の傍線が引かれた部分(傍線部)について、その内容や理由を問う読解問題
問6	本文全体について、内容や論の進め方などを問う問題

に読解問題とよぶ. 読解問題の配点は、第1問「評論」全体の配点の約2/3を占める.

3. 複数の特徴量と機械学習を用いた解法

3.1 基本方針

我々は、評論読解問題の自動解法として、これまでに本文照合法および節境界法という2つの解法を提案し、それらがセンター試験過去問に対して有効に作用することを示した[佐藤 14, 加納 15]. しかし、これら2つの解法は、1つの特徴量のみによってすべての読解問題を解こうとする解法である. たとえば、本文照合法は、本文の一部分(照合領域)と各選択肢とのオーバーラップ率[服部 13]が最も大きい選択肢を出力する. 1つの特徴量のみを用いるのではなく、複数の特徴量を組み合わせて用いれば、さらに性能が向上することが期待できる.

本稿で提案する解法は、この考え方に基づくものである. すなわち、解答器を、複数の特徴量を用いた機械学習により構成する. 具体的には、「2つの選択肢のうちどちらがより本文に合致しているか」を判定する判定器を学習し、その判定器を用いて選択肢の総当たり比較を行った結果、最も多く「本文に合致している」と判定された選択肢を出力する. 本稿で提案する解法を、Binary Classifier-Based Method (BCBM) とよぶこととする.

3.2 ソルバー構成

BCBMの全体像を、図2に示す. この図に示すように、BCBMは、次のような方法で読解問題を解く.

1. **入力:** 本文, 設問, 選択肢集合と, 照合領域抽出用パラメータを入力する.
2. **照合領域の抽出:** パラメータに基づき, 選択肢と本文の表層的類似度の計算に用いる本文の一部(照合領域)を抽

連絡先: 加納隼人, 名古屋大学大学院 工学研究科 電子情報システム専攻, 〒464-8603 愛知県名古屋市千種区不老町 C3-1(631) IB 電子情報館南棟 159号室, 052-789-4435, h.kanou@nuee.nagoya-u.ac.jp

本文:
 ... 成人のあいだで通常いとなまれている言語的交通も、そうした面をつよもつことがあるのではないだろうか。会話はそれ自体としてある種の快感をさそう行為でありえ、緊張を解きほぐすふるまいとなりうる。ひとはまた特定の目的とともに会話を開始するというよりは、むしろ多く会話にまぎこまれ、気づいたときには会話の流れに身をゆだねてしまっている。D ことばによるやりとりには、目的一手段という枠組みではとらえがたい面が存在する。ことばは、たんなる手段あるいは道具ではなく、なによりもまず、交流のかたちそのものである。...

問5 傍線部D「ことばによるやりとりには、目的一手段という枠組みではとらえがたい面が存在する」とあるが、それはどういうことか。その説明として最も適当なものを、次の1~5のうちから一つ選べ。

- 1 ことばによるやりとりは、音そのものの強弱・高低などの音楽性が根本にあるため、伝達すべき情報の価値よりも、声の調子や質によって、どれだけ聞き手の緊張をほぐしたり快感を誘ったりできるかが重要な場合があるということ。
- 2 ことばによるやりとりは、成人間では情報の伝達を主眼とするが、やりとりの途中でいつの間にか会話に流されることもあるから、伝達のしかたや話の展開にも留意しなければ、十分な理解はむずかしい場合があるということ。
- 3 ことばによるやりとりは、情報の伝達を基本としているものの、ひとりごとや意味のないつぶやきなどのように、特定の聞き手を意識せず、自問自答したり漠然としただれかに語りかけたりする場合があるということ。
- 4 ことばによるやりとりは、もともと唱うことやリズムによる共鳴作用から発しているもので、意味の伝達という目的のためよりも、むしろ音声の交換という目的のためにこそ存在している場合があるということ。
- 5 ことばによるやりとりは、必ずしも特定の目的や相手に伝えたい情報があらかじめ存在して、それが言語化され発話されるという手順が踏まれるものとは限らず、相手との交流自体に意義がある場合があるということ。

図 1: 対象とする評論読解問題の例 (2007 年センター試験追試 第 1 問の問 5)

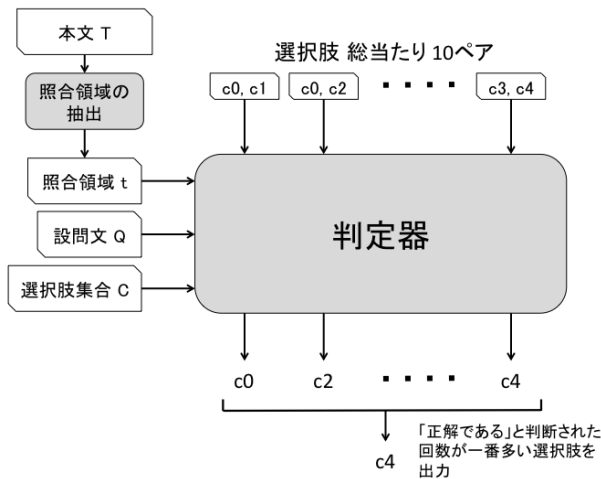


図 2: BCBM の概念図

出する。

3. **選択肢の点数の計算:** 選択肢 5 つについて総当たり 10 ペアを作成し、そのそれぞれについて、以下を実行する。

(*) 選択肢ペアを判定器に入力し、「本文に合致している」と判定された方の選択肢に 1 点を与える。

4. **出力:** 点数の最も高い選択肢を解答として出力する (点数が同じ場合は、より選択肢番号の若い選択肢を優先する)。

照合領域は、本文中の傍線部を中心とした連続領域であり、どの領域を抽出するかをパラメータとしてソルバーに与える。具体的には、傍線部を含む段落から前後どれだけの段落までを対象とするかによって指定し、P-m-n のように表す。たとえば、P-1-0 は、「傍線部を含む段落と、その 1 つ前の段落」を照合領域とすることを意味する。また、整数で指定する以外に特別な位置として、a (本文冒頭)、b (1 つ前の設問の傍線部を含む段落)、c (1 つ後ろの設問の傍線部を含む段落) を指定できる。

3.3 判定器

判定器は、ある 2 つの選択肢について、どちらがより「本文に合致している」かを判定する。図 3 に、判定器の全体像

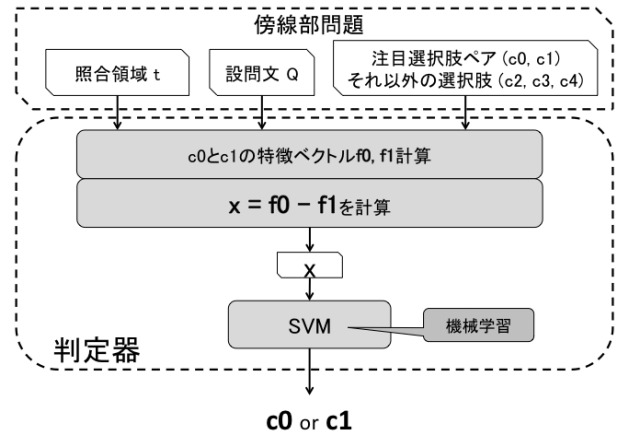


図 3: 判定器の構成

を示す。この図に示すように、判定器は、次のような 2 値分類問題を解く。

入力 注目選択肢ペア (c_0, c_1) 、他の選択肢 (c_2, c_3, c_4) 、照合領域 t 、設問文 Q

出力 $\{c_0, c_1\}$ のどちらか

2 値分類器の学習アルゴリズムは、サポートベクターマシン (SVM) を用いる。SVM による学習の実現には、WEKA [Hall 09] の classifiers/functions/SMO パッケージを使用する。カーネルは線形カーネルを用い、他のパラメータはすべてデフォルト値を用いる。

判定器は、具体的には次のような処理を行う。

1. 入力から、選択肢 c_0 の特徴ベクトル f_0 、選択肢 c_1 の特徴ベクトル f_1 を計算する。
2. 2 つのベクトルの差分 $x = f_0 - f_1$ を計算して、SVM に入力し^{*1}、判定結果を得る。

判定器の学習は、正解選択肢と間違い選択肢のペアを用いて行う。具体的には、学習セット中の各読解問題 1 問について、

*1 WEKA の前処理によって、 x の各要素は、すべて 0 から 1 の範囲の実数に正規化される。

表 2: 各特微量
説明

名前	式	説明
cor	$f_1 = \text{overlap_ratio}_D(C^1; t, c_i)$	照合領域と選択肢との文字オーバーラップ率
kkor	$f_2 = \text{overlap_ratio}_D(K^1; t, c_i)$	照合領域と選択肢との漢字・カタカナオーバーラップ率
bor	$f_3 = \text{overlap_ratio}_D(C^2; t, c_i)$	照合領域と選択肢との文字 bigram オーバーラップ率
lor	$f_4 = \text{overlap_ratio}_D(L; t, c_i)$	照合領域と選択肢との形態素原形オーバーラップ率
wor	$f_5 = \text{overlap_ratio}_D(W; t, c_i)$	照合領域と選択肢との形態素表層形オーバーラップ率
comp	$f_6 = \frac{1}{4} \sum_{j \neq i} \text{overlap_ratio}_B(C^1 c_j, c_i)$	選択肢 c_i の、他の選択肢との文字オーバーラップ率の平均値
lack	$f_7 = \text{lack_words}(C, i)$	選択肢 c_i には出現しないが、他の全ての選択肢に出現する名詞の数
uniq	$f_8 = \text{unique_words}(C, i)$	選択肢 c_i には出現するが、他の全ての選択肢に出現しない名詞の数
cn-c	$f_9 = \text{cn_score}(C^1; t, c_i)$	照合領域と選択肢との CN スコア
cn-b	$f_{10} = \text{cn_score}(C^2; t, c_i)$	照合領域と選択肢との CN スコア
cn-l	$f_{11} = \text{cn_score}(L; t, c_i)$	照合領域と選択肢との CN スコア
cn-w	$f_{12} = \text{cn_score}(W; t, c_i)$	照合領域と選択肢との CN スコア

正解選択肢と間違い選択肢のペアを 8 個 (4 個 × 2 方向) 作成し、学習データとする。

3.4 選択肢の特徴ベクトル

判定器は、注目選択肢 c_i ($i = 0, 1$) をそれぞれ特徴ベクトル $\mathbf{f}_i = [f_1, \dots, f_{12}]$ に変換する。各特微量の詳細を表 2 に示す。 f_1 から f_5 までの特微量は、照合領域と選択肢とのオーバーラップ率 [服部 13] であり、本文照合法ソルバー [佐藤 14] において用いられている特微量である。 $\text{overlap_ratio}_D(E; s_1, s_2)$ は、文字列 s_1 と s_2 との間の共通する要素の割合を表す*2。要素としてどのような単位を用いるかは、集合 E に何を選ぶかによって指定する。 C^n は文字 n -gram 集合、 K^n は漢字・カタカナ n -gram 集合、 L は形態素原形集合、 W は形態素表層形集合である。

f_6 から f_8 までの特微量は、他の選択肢との差異に注目した特微量である。センター試験『国語』の読解問題の解答においては、選択肢同士の比較が重要なステップであるということが、文献 [津田 13] などの受験参考書において述べられている。この文献によると、他の 4 つの選択肢に比べて話題の欠けている選択肢や、表層的に大きな差異のある選択肢は、正解である可能性が低いという。 f_6 から f_8 までの特微量は、この観点をソルバーに導入するためのものである。なお、 f_7, f_8 の計算においては、形式的でない名詞 (IPAdic の品詞細分類が“一般”、“サ変接続”、“形容動詞語幹”、“固有名詞”のいずれかである名詞) のみを対象としている。

f_9 から f_{12} までの特微量は、照合領域と選択肢の CN スコア [加納 15] であり、節境界法ソルバーにおいて用いられている特微量である。 $\text{cn_score}(E; s_1, s_2)$ は、以下のような方法で計算する。

Step1 文字列 s_1 と s_2 に節境界検出に基づいた節分割を行い、それぞれを節の集合 S_1, S_2 に変換する。節境界検出には、節境界検出プログラム Rainbow [加納 14] を用いる。

Step2 S_2 内の各節 $x_2 \in S_2$ のスコアを計算し、その平均値を、 $\text{cn_score}(E; s_1, s_2)$ とする。節 x_2 のスコアは、 x_2 と、 S_1 内の各節 $x_1 \in S_1$ との類似度の最大値とする。

節同士の類似度は、節同士のオーバーラップ率と、2 つの節の節ラベルが一致する場合のボーナスの和とする。CN スコアの詳細は文献 [加納 15] を参照されたい。

*2 添字が D となっているものは方向性があり、 B となっているものは方向性がない。

表 3: データセットの内訳

問題	年代	試験数	設問数	選択肢数
センター過去問	2001 - 2015	13	52	260
代ゼミ問題集	2005, 2014	14	56	280
駿台問題集	2006, 2014	14	56	280
河合塾問題集	2014, 2015	10	40	200
旺文社問題集	2014	4	16	80
代ゼミ模試	2012 - 2014	5	20	100
計		60	240	1200

4. 評価実験

4.1 読解問題データセット

実験には、センター試験の過去問題、および各予備校のセンター試験対策問題集の問題を使用した。使用したデータセットの構成を表 3 に示す。この表に示すように、データセットには全 240 問の評論読解問題が含まれる。

4.2 判定器の性能評価

ソルバーの性能評価の前に、判定器の 2 値分類器としての性能評価を行った。データセットの各設問につき、正解選択肢と間違い選択肢のペアを 8 個 (4 個 × 2 方向) 作成し、そのデータ内で、10 分割交差検証を行った。したがって、データ内には、正例と負例が同数含まれる。

データセット全 240 問すべて (1920 インスタンス) を用いた場合、および、センター過去問 52 問のみ (416 インスタンス) を用いた場合の 2 通りについて、照合領域を変更し、判定器の精度 (正しく分類されたインスタンスの割合) を測定した。評価結果を表 4 に示す。この表では、データセットすべてを用いた場合、センターのみを用いた場合の精度を、この順に斜線で区切って示している。表の P-m-n は、本文の照合領域を表す。SVM3 は、データセット全体内で SVM2 に Best First 法による特徴選択を行った判定器、SVM4 は、センター過去問内で SVM2 に Best First 法による特徴選択を行った判定器である。

2 つの場合のいずれにおいても、SVM2 が最も優れた性能であり、データセット全体に対しては最大 69.8%、センター過去問に対しては最大 79.3% の精度を示した。また、CN スコアを使用していない SVM1, SVM3 は、使用している SVM2, SVM4 に比べて、データセット全体内での精度は同程度であるが、センター過去問内での精度が低い。これより、CN スコアは、センター過去問に対して特に有効な特微量であることがわかる。

表 4: 判定器の精度 (単位%, データセット全体/センターのみ)

	SVM1	SVM2	SVM3	SVM4
cor	✓	✓	✓	✓
kkor	✓	✓	✓	✓
bor	✓	✓		
lor	✓	✓		
wor	✓	✓	✓	
comp	✓	✓	✓	✓
lack	✓	✓	✓	
uniq	✓	✓	✓	✓
cn-c		✓		
cn-b		✓		✓
cn-l		✓		✓
cn-w		✓		✓
P-0-0	64.5/68.5	64.5/74.8	64.5/68.5	65.4/73.3
P-b-0	65.9/72.4	66.5/ 79.3	65.2/72.6	64.9/77.6
P-1-1	66.9/70.2	67.1/74.8	66.7/71.4	65.9/74.8
P-1-0	65.8/67.8	65.9/77.2	65.9/70.7	66.1/76.7
P-a-0	66.2/72.1	67.2/77.6	66.5/72.4	65.8/75.0
P-b-c	68.9/73.6	67.9/77.2	68.4/74.3	67.6/76.0
P-b-1	67.6/74.3	68.1/77.4	66.4/73.3	66.4/75.7
P-a-c	68.4/74.8	69.8 /76.0	68.8/74.5	68.9/75.5

表 5: ソルバーの正解数と内訳

	本文照合法	節境界法	BCBM
データセット全体 (240)	87	79	94
センター過去問 (52)	24	30	23
代ゼミ問題集 (56)	22	18	20
駿台問題集 (56)	16	17	22
河合塾問題集 (40)	8	7	12
旺文社問題集 (16)	8	1	9
代ゼミ模試 (20)	9	6	8

4.3 BCBM ソルバーの性能評価

交差検証^{*3}により, BCBM ソルバーの評価を行った. 実験用データにはデータセット全 240 問すべてを用い, 正解数を調査した. 判定器は, 最も精度の良かった SVM2 を用い, 照合領域は, データセット全体において SVM2 の精度が最も良かった P-a-c を採用した.

比較のため, 本文照合法ソルバー, 節境界法 (CN スコア採用) ソルバーの正解数も調査した. これらのソルバーのパラメータは, データセット全体において最も成績の良いものを採用した. 具体的には, 本文照合法ソルバーは照合領域 P-a-0, オーバーラップ単位 C^1 , 事前選択 ps, 節境界法ソルバーは照合領域 P-b-c, オーバーラップ単位 W , 事前選択 ps を採用した. なお, 本文照合法ソルバーは第一著者による実装であり, 文献 [佐藤 14] の実装とはスコアのタイプブレーク法など細かな仕様が異なる. 各ソルバーの正解数と, データの出典ごとの内訳を表 5 に示す. 左列の括弧内の数字は, 読解問題の総数を示す.

データセット全体において, BCBM ソルバーの正解数は 240 問中 94 問であり, 本文照合法ソルバー, 節境界法ソルバーの正解数を上回った.

本文照合法および節境界法と異なる BCBM の傾向として, センター以外の問題でも正解率があまり落ちないということがある. 表 5 を見ると, 既存の 2 つの解法, とくに節境界法は, センター過去問の正解数に比べ他の問題の正解率が全体

的に低いことがわかる. それに対し BCBM は, センター過去問の正解数では既存の 2 つの解法を下回るが, 他の問題においても正解率が 4 割前後からあまり変化せず, 安定している. BCBM では機械学習を導入し, かつセンター過去問以外の問題も学習に用いたことで, 既存の解法よりも正解率が安定するようになったと考えられる.

5. まとめ

本研究では, センター試験『国語』評論読解問題を, 複数の特徴量と機械学習を用いて解く新たな解法を提案し, 実装した. 提案する解法である BCBM は, 「2 つの選択肢のうちどちらがより本文に合致しているか」を判定する判定器を機械学習により構成し, その判定器を用いて選択肢の総当たり比較を行った結果, 最も多く「本文に合致している」と判定された選択肢を出力するという解法である. BCBM は, 読解問題 240 問に対し, 既存の解法を上回る 94 問という正解数を示した.

今後, BCBM ソルバーの性能を上げるための方法の 1 つとして, 判定器の多値分類への拡張がある. 選択肢 5 つの総当たり 10 ペアのうち 6 ペアは「どちらも本文に合致しない」ペアである. しかし, 現在の判定器は, そのようなペアに対してもどちらかの選択肢を「本文に合致する」と判定し, 点を与えてしまう. そこで, 判定器を「 c_0 が本文に合致する」「 c_1 が本文に合致する」「どちらも本文に合致しない」の 3 値分類を行えるように拡張することで, ソルバーの性能向上が期待できる. このような分類を正確に行えるような新たなフレームワーク, 特徴セットを考案する必要がある.

謝辞

本研究は, 国立情報学研究所のプロジェクト「ロボットは東大に入れるか」から, データの提供を受けて実施した. 本研究の一部は, JSPS 科研費 24300052 の助成を受けて実施した.

参考文献

- [新井 12] 新井 紀子, 松崎 拓也: ロボットは東大に入れるか?—国立情報学研究所「人工頭脳」プロジェクト—, 人工知能学会誌, Vol. 27, No. 5, pp. 463–469 (2012)
- [Hall 09] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H.: The WEKA Data Mining Software: An Update, *SIGKDD Explorations*, Vol. 11, No. 1, pp. 10–18 (2009)
- [服部 13] 服部 昇平, 佐藤 理史: 多段階戦略に基づくテキストの意味関係認識: RITE2 タスクへの適用, 情報処理学会研究報告 2013-NL-211 No.4/2013-SLP-96 No.4, 情報処理学会 (2013)
- [加納 14] 加納 隼人, 佐藤理史: 日本語節境界検出プログラム Rainbow の作成と評価, FIT2014 講演論文集 第 2 分冊 pp.215-216 (2014)
- [加納 15] 加納 隼人, 佐藤理史, 松崎拓也: 節境界検出を用いたセンター試験『国語』評論傍線部問題ソルバー, 情報処理学会研究報告 2015-NLP-220, 情報処理学会 (2015)
- [教学 13] 教学社編集部: センター試験過去問研究 国語 (2014 年版センター赤本シリーズ), 教学社 (2013)
- [佐藤 14] 佐藤理史, 加納隼人, 西村翔平, 駒谷和範: 表層類似度に基づくセンター試験『国語』現代文傍線部問題ソルバー, 自然言語処理 vol.21 No.3 pp.465-483, 言語処理学会 (2014)
- [津田 13] 津田 秀樹: センター試験マル秘裏ワザ大全 国語 2014 年度版, 洋泉社 (2013)

*3 解こうとしている問題と, その本文と同一の本文を参照する問題を除いたすべての読解問題を学習セットとする.