表層的特徴を用いたセンター試験『国語』評論読解問題の自動解法

Solvers of Japanese Text Comprehension Questions Using Surface Clues

加納 隼人
Hayato Kanou

名古屋大学大学院工学研究科
Graduate School of Engineering, Nagoya University
h_kanou@nuee.nagoya-u.ac.jp

佐藤 理史
Satoshi Sato

(同 上)
ssato@nuee.nagoya-u.ac.jp

松崎 拓也
Takuya Matsuzaki

(同 上)
matuzaki@nuee.nagoya-u.ac.jp

keywords: machine comprehension, university entrance exam

Summary

This paper reports progress from 2014 to 2015 on development of solvers of Japanese comprehension questions in university entrance exam. Target questions are the multiple-choice questions in the *essay* section (Question No.1) in Japanese Language (Kokugo) of National Center Test. In 2014, we introduced a new scoring function using clause boundaries, which are automatically detected by our newly developed tool. The score of a choice is calculated as the average clause-similarity between the choice and a selected part of text body. In 2015, we developed a machine-learning based method, which uses seventeen features to determine the answer. They includes surface-similarity based features, clause-similarity based features, and choice-discriminative features. In addition to the first formal run of Torobo Project in 2013, we participated in the two formal runs in 2014 and 2015; We were only a participant who submitted the result in Contemporary Japanese Language until now. After the 2015 formal run, we conducted an experiment using 276 questions to compare all developed solvers with various parameters. The best performance was obtained by a 2015 solver, which produced 117 (42%) correct answers. For the subset of 56 previous official questions in National Center Test, a 2014 solver was the best, which produced 32 (57%) correct answers. However, there is no statistical significance between the best 2015 solver and our first solver developed in 2013.

1. はじめに

文章を読んでその内容を理解することは、人間の知能を構成する重要な能力の一つであることは疑いない。その能力の機械的実現は、人工知能における大きな目標でもある。ここでは、この能力を**文章の読解力**と呼ぶことにする。

日本の大学入学試験問題では、いずれの科目においても、程度の差はあれ、日本語の文章の読解力が問われる。なぜならば、問題文を読んで何が問われているのかを理解できなければ、問題に答えることができないからである。しかし、直接的に現代日本語文章の読解力が問われる科目は『国語』現代文のみである。『国語』現代文のいわゆる読解問題では、与えられた文章に対して、その内容に関する設問に答えることが求められる。これは、読解力そのものを測るための問題である。

2011年に開始された「ロボットは東大に入れるか(東ロボ)」は、人工知能の各種の要素技術を統合し、大学入

試問題を計算機で解くことに挑戦するプロジェクトである[新井 12]。このプロジェクトでは、2013年秋より年1回、予備校の模試を利用した公開性能評価を実施している。このうち、2013年の公開性能評価で使用した解法とその結果の詳細に関しては、すでに研究報告[佐藤 14a]という形で公表している。

本論文では、センター試験『国語』現代文の問題のうち、評論の読解問題に焦点を絞り、2014年から2015年度末までの進展について総括する。まず、2章で、センター試験『国語』の評論問題の具体例を示し、2013年の解法である**本文照合法**[佐藤 14b]の概略を述べる。3章では、2014年の新解法である**節境界法**[加納 15a]について述べる。4章では、2015年の新解法である**機械学習を用いた解法**[加納 15b]について述べる。5章では、2015年末に実施した、最新の評価実験について述べる。6章では、2014年と2015年の公開性能評価の結果について述べる。7章では、関連研究について述べ、最後の8章では、2015年度末時点における到達点についてまとめる。

本文(一部のみ)

... 成人のあいだで通常いとなまれている言語的交通も、そうした面をつよく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りまといなりうる。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—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—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だとはいえず、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 評論読解問題の例(2007年センター試験追試第1問の問5)

2. 評論読解問題と本文照合法

2.1 センター試験『国語』評論読解問題

大学入試センター試験の『国語』では、例年、大問4問が出題され、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、評論あるいは随筆というジャンルに分類される文章から抜き出された3500字程度の本文と、それに対する設問から構成されている。設問は6問あり、そのうちの4問(問2から問5)が、本論文で対象とする評論読解問題である。それぞれの設問は、通常、本文中の傍線部を参照し、その内容や理由を問う問題であり、与えられた5つの選択肢から1つ選ぶ選択式の問題である。その一例を図1に示す。なお、読解問題の配点は32点(各8点)で、評論全体の配点50点の約2/3を占める。一問当たりの配点が高いことが、『国語』読解問題の特徴である。

2.2 本文照合法

2013年の解法である本文照合法[佐藤14b]は、

- 正解選択肢を選ぶ根拠は、本文中に存在する[船口97, 板野10]
- 意味的に似ているテキストは、表層的にも似ていることが多い

という2つの観察事実に基づく解法である。この解法の骨子は、次のような2ステップの手順である。

- (1) 本文から、その一部を**照合領域**として切り出す。
- (2) 照合領域とそれぞれの選択肢を比較し、照合領域と最もよく合致する選択肢を選ぶ。

照合領域と選択肢が合致するかどうかは、後述する**オーバーラップ率**という指標で計算する。さらに、上記のステップ(2)に先立ち、

- 5つの選択肢の1つをあらかじめ除外する

ためのステップが追加されている。これらのそれぞれのステップを制御するための3つのパラメータがある。

§1 入力

本文照合法の入力は、本文 T 、設問文 Q 、選択肢集合 C 、および、各ステップを制御する3つのパラメータ $\{\alpha, \beta, \gamma\}$ である。これらを、それぞれ、照合領域パラメータ、照合単位パラメータ、事前選択パラメータと呼ぶ。

§2 照合領域の抽出

このステップでは、照合領域パラメータ α の指定に従い、本文から照合領域 t を抽出する。パラメータ α は、傍線部を含む段落の前後何段落(前 m 段落、後 n 段落)を切り出すかを指定するもので、 $P-m-n$ のような形式で指定する。たとえば、 $P-1-0$ の場合は、「傍線部を含む段落と、その1つ前の段落」を照合領域として切り出すことを意味する。整数で段落数を指定する以外に、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)を指定することができる。いずれの場合も、抽出される照合領域は、段落単位の連続領域となる。

§3 選択肢の事前選抜

事前選択パラメータ γ がオンの場合、5つの選択肢の中で「最も他の選択肢と似ていない選択肢」を式(1)で決定し、除外する。オーバーラップ率の一つである関数 bor については、後述する。

$$c_{excluded} = \operatorname{argmin}_{c_i \in C} \sum_{c_j \in C, j \neq i} bor(A; c_i, c_j) \quad (1)$$

§4 照合

5つ(事前選択を行った場合は4つ)の選択肢 c_i のそれぞれに対し、照合領域 t との合致の度合いを表すスコア $score(\beta; t, c_i)$ を式(2)で求め、このスコアの最も大きい選択肢を解答として出力する。パラメータ β と関数 dor は、次で述べる。

$$score(\beta; t, c_i) = dor(\beta; t, c_i) \quad (2)$$

§5 オーバーラップ率

照合と選択肢の事前選抜では、オーバーラップ率 [服部 13] という指標を用いる。この指標は、2つの文字列 x と y の間の共通して出現する要素 (文字, 文字 n -gram, 形態素など) の割合であり, 以下のように定義される。

$$dor(E; x, y) \equiv \frac{l_{overlap}(E; x, y)}{l(E, y)} \quad (3)$$

$$bor(E; x, y) \equiv \frac{2l_{overlap}(E; x, y)}{l(E, x) + l(E, y)} \quad (4)$$

$$l_{overlap}(E; x, y) \equiv \sum_{e \in E} \min(f(e, x), f(e, y)) \quad (5)$$

$$l(E; z) \equiv \sum_{e \in E} f(e, z) \quad (6)$$

これらの式に現れる $f(e, z)$ は, ある集合 E の要素 e が文字列 z に何回現れるのかを表す。照合単位パラメータ β はこの集合 E を規定する (要素のクラスを指定する) もので, 何を単位としてオーバーラップ率を計算するかを制御する。 β の選択肢としては, 以下のようなものがある。

$\mathcal{A}$ (文字集合), $\mathcal{A}^2$ (文字 bigram の集合), $\mathcal{K}$ (漢字・カタカナ文字の集合), $\mathcal{W}$ (形態素表層形の集合), $\mathcal{L}$ (形態素原形の集合)

たとえば, $\beta (= E)$ として文字集合 $\mathcal{A}$ を指定した場合は, 文字を単位としたオーバーラップ率を計算する。

上記の式のうち, 式 (5) は, オーバーラップ, すなわち, 共通要素の数の定義式であり, 式 (6) は, 集合 E で規定される単位で測った文字列 z の長さである。共通要素数 $l_{overlap}$ を y の長さで正規化した式 (3) は, 方向性をもったオーバーラップ率 dor となり, x と y の長さの和で正規化した式 (4) は, 方向性をもたないオーバーラップ率 bor となる。

3. 節 境 界 法

3.1 節 の 利 用

センター試験の評論読解問題の選択肢には, 比較的長い文が現れることが多い。たとえば, 図 1 に示した設問の選択肢②は 1 文で構成され, その長さは 102 文字である。多くの受験参考書は, このような長い選択肢の正誤を判定する際には, まず選択肢をいくつかの部分に分解し, 部分単位で正誤を吟味することを推奨している (たとえば [板野 10])。

この戦略を自動解法に取り入れるためには, 長い文をいくつかの部分に分割することが必要となる。分割によって得られる部分は, 当然のことながら, 意味的にまとまった単位であることが望ましい。しかし, 文を意味的なまとまりに自動分割する技術は未成熟なので, 現実的には, 何らかの形式的 (文法的) まとまりに分割するしかない。

図2 節境界法の構成

我々は, 文の分割単位として「節」という文法単位を採用する。節は, 「述語を中心としたまとまり」[益岡 92] であり, 意味的には, おおよそ「何が何をどうした」ということを表す。

3.2 節を単位とする照合

節境界法は, 「節単位で照合したとき, 最もよく本文と合致する選択肢を選ぶ」という考え方に基づく解法である。節境界法の全体像を図2に示す。本文照合法とは, 照合で用いるスコアのみが異なる。解法を制御する3つのパラメータ $\{\alpha, \beta, \gamma\}$ は, 本文照合法と同様に存在する。

節境界法の照合ステップでは,

- (1) まず, 照合領域と選択肢の両方を, それぞれ節集合に変換し,
- (2) 節単位の類似度に基づいて選択肢のスコアを計算し, 最もスコアの高い選択肢を出力する。

照合領域および選択肢を節集合への変換には, 節境界検出プログラム Rainbow[加納 14] を利用する。Rainbow は, 与えられた文に含まれる節末境界を検出し, その節のタイプを表すラベルを付与するプログラムである。上記の (1) では, Rainbow で検出された節末境界で文を分割し, 節集合を得る^{*1}。以降では, 選択肢 c から得られた節集合を S_c , 照合領域 t から得られた節集合を S_t と表記する。

選択肢のスコアを, これら2つの節集合から計算する。このスコアとして, **CN スコア**と **LN スコア**という2種類のスコアを定義する。

§1 節の類似度

いずれのスコアも, 2つの節の類似度が基本となっている。2つの節 x と y の類似度を以下のように定義する。

$$sim(\beta; x, y) =$$

$$bor(\beta; x, y) + 0.5 \cdot label\_match(x, y) \quad (7)$$

^{*1} 節は必ずしも連続領域とは限らないため, この方法は厳密には節への分割とはならないが, 便宜的に分割された単位を節と呼ぶこととする。

表 1 $label\_match(x, y) = 1$ の条件

ラベル大分類	下位ラベルの一致
ともに補足節 ともに副詞節 ともに連体節 ともに並列節	小分類 (形式名詞, 疑問表現, 引用) まで一致 (19 種類の下位ラベルが) 完全に一致 - 小分類 (順接的, 逆接的) まで一致

この式の第 1 項は, 方向性を持たないオーバーラップ率 (式 (4)) である. 第 2 項は, 2 つの節が同じタイプの節である場合に与えるボーナスである. 具体的には, Rainbow が付与した x と y の節ラベルが表 1 に示す条件を満たす場合は $label\_match(x, y) = 1$, 満たさない場合は 0 とする.

§2 CN スコア

CN スコア $score_{CN}$ は, 選択肢に含まれる節 x のスコアの平均値として定義する.

$$score_{CN}(\beta; S_c, S_t) = \frac{1}{|S_c|} \sum_{x \in S_c} cs_{CN}(\beta; x, S_t) \quad (8)$$

$$cs_{CN}(\beta; x, S_t) = \max_{y \in S_t} sim(\beta; x, y) \quad (9)$$

式 (9) は, 節 x と最もよく似た節を, 照合領域 (の節集合 S_t) から探すことを意味する.

§3 LN スコア

これに対して, LN スコア $score_{LN}$ は, 節の長さを考慮したスコアで, 以下の式で定義する.

$$score_{LN}(\beta; S_t, S_c) = \frac{\sum_{x \in S_c} cs_{LN}(\beta; x, S_t)}{\sum_{x \in S_c} l(\beta; x)} \quad (10)$$

$$cs_{LN}(\beta; x, S_t) = \max_{y \in S_t} [sim(\beta; x, y) \cdot l(\beta; x)] \quad (11)$$

ここで, $l(\beta; x)$ は, 式 (6) で定義される文字列 x の長さである. 式 (11) は, 節 x の長さで重み付けしているため, 節 x が長いときに値がより大きくなる. すなわち, LN スコアは CN スコアに比べて, 選択肢に含まれる長い節が本文と一致することを重視したスコアである.

4. 機械学習を用いた解法

4.1 複数の特徴量の利用

本文照合法とその一部を変更した節境界法は, いずれも単一のスコア (1 つの特徴量) で問題を解く解法である. しかし, 単一の特徴量で解答を決める必然性はない. もし, 複数の特徴量が利用できるならば, それらを組み合わせることによって, 解法の性能を向上させることができると思われる. これは, 以下のような 2 つのステップの手順により実現できる.

- (1) 各選択肢に対して, 複数の特徴量を計算する. すなわち, 各選択肢を, 特徴ベクトル f に変換する.
- (2) 5 つの選択肢に対応する 5 つの特徴ベクトルから, 1 つを選ぶ.

なお, これらのステップは, それぞれ独立に設計できる.

4.2 選択肢の特徴ベクトル表現

これまでの解法を踏襲し, それぞれの選択肢の特徴ベクトルは, 次の要素から計算する.

本文 T , 設問文 Q , 選択肢集合 C ,
注目選択肢番号 k , 照合領域パラメータ α

特徴ベクトル f を構成する 17 個の特徴量を表 2 に示す. これらの特徴量は, 4 グループに大別される.

§1 グループ 1: オーバーラップ率

このグループの特徴量は, 照合領域パラメータ α で定められる照合領域と選択肢とのオーバーラップ率に基づく特徴量である. 特徴量 f_1 から f_5 は, どのような単位でオーバーラップ率を計算するかのみが異なる.

§2 グループ 2: 本文中のよく似た 10 文

このグループの特徴量は, 以下で定義される文集合から計算される特徴量である.

本文 T の中で, 注目選択肢との文字オーバーラップ率 (dor) が最も高い文 10 文.

文を選ぶ対象は照合領域ではなく本文全体であり, 得られる文集合は本文中の連続領域とは限らない.

特徴量 f_6 は, この文集合を仮想的な照合領域とする場合の, 文字オーバーラップ率 (dor) である. 特徴量 f_7 は, この文集合の各文の文番号 (本文先頭から何文目にあるか) の平均と, 注目する傍線部の文番号の差の絶対値である*2. この値が大きい場合, その選択肢は傍線部とは離れた位置の内容を述べている可能性が高いことになる.

§3 グループ 3: 他の選択肢との比較

このグループに属する特徴量は, 主に, 注目選択肢とそれ以外の選択肢との比較に基づく特徴量である.

特徴量 f_8 は, 注目選択肢がそれ以外の選択肢と平均的にどれくらい似ているかを, f_9 は最もよく似た選択肢とどれくらいよく似ているかを表す. いずれの場合も, 似ているかどうかは, 方向性を持たない文字オーバーラップ率 (bor) で計算する. これらは, 「他の選択肢に比べて表層的に大きな差がある選択肢は正解である可能性が低い」という指摘 [津田 13] に基づく. 背後にある考え方は, 選択肢の事前選抜と同じである.

特徴量 f_{10} と f_{11} は, 注目選択肢とその他の選択肢に含まれる名詞の違いを表す. 特徴量 f_{13} では, 名詞以外に動詞も考慮し, かつ, それぞれがいくつの選択肢に現れるかも考慮する. これらの特徴量は, 「他の選択肢に比べて話題の欠けている選択肢は正解である可能性が低い」という指摘 [津田 13] に基づく. なお, これらの特徴量の計算では, 形式的な名詞・動詞は考慮しない*3.

*2 各文の文番号と傍線部の文番号の差の絶対値をとってから平均を求めるべきであったが, 実装は本文のとおりである.

*3 用いる名詞は, IPADic の品詞細分類が「一般」, 「サ変接続」, 「形容動詞語幹」, 「固有名詞」のいずれかであるもの. 動詞は, 細分類が「自立」のもの.

表 2 17 個の特徴量

f_1	照合領域と注目選択肢との文字オーバーラップ率 ($dor(A)$)
f_2	照合領域と注目選択肢との漢字・カタカナオーバーラップ率 ($dor(K)$)
f_3	照合領域と注目選択肢との文字 bigram オーバーラップ率 ($dor(A^2)$)
f_4	照合領域と注目選択肢との形態素原形オーバーラップ率 ($dor(L)$)
f_5	照合領域と注目選択肢との形態素表層形オーバーラップ率 ($dor(W)$)
f_6	本文中で注目選択肢と最も類似する 10 文と、注目選択肢との文字オーバーラップ率 ($dor(A)$)
f_7	本文中で注目選択肢と最も類似する 10 文が、平均的に傍線部からどれほど離れているか
f_8	注目選択肢と他の選択肢との文字オーバーラップ率 ($bor(A)$) の平均値
f_9	注目選択肢と他の選択肢との文字オーバーラップ率 ($bor(A)$) の最大値
f_{10}	注目選択肢には出現しないが、他の全ての選択肢に出現する名詞の数
f_{11}	注目選択肢には出現するが、他の全ての選択肢に出現しない名詞の数
f_{12}	注目選択肢に出現する「極端な表現」の数
f_{13}	注目選択肢に出現する名詞および動詞の、普遍さ (5 つの選択肢のうちのいくつに現れるか) の総和
f_{14}	照合領域と注目選択肢との文字 CN スコア ($score_{CN}(A)$)
f_{15}	照合領域と注目選択肢との漢字・カタカナ CN スコア ($score_{CN}(K)$)
f_{16}	照合領域と注目選択肢との形態素原形 CN スコア ($score_{CN}(L)$)
f_{17}	照合領域と注目選択肢との CN スコア ($score_{CN}(W)$)

特徴量 f_{12} は、注目選択肢に含まれる以下の表現の数である。

絶対、必ず、かならず、つねに、常に、だけ、のみ、すべて、全て

この特徴量は、「『必ず』などの極端な表現を用いて強い主張をしている選択肢は、正解である可能性が低い」という指摘 [津田 13] に基づく。

§4 グループ 4: CN スコア

このグループの特徴量は、節境界法の CN スコア (式 (8)) である。 f_{13} から f_{17} は、どのような単位で CN スコア (の計算に必要なオーバーラップ率) を計算するかのみが異なる。

図 3 BCBM の概要

4.3 BCBM

5 つの選択肢のそれぞれを特徴ベクトルに変換した後に行うことは、それらの中から 1 つを選ぶことである。機械学習を用いた 1 つ目の解法である **Binary Classifier-Based Method (BCBM)** は、「2 つの選択肢のうち、より本文に合致する方を出力する」2 値分類器を学習し、それを用いた選択肢の総当たり比較により、これを実現する方法である (図 3 参照)。選択肢は 5 つあるので、総当たり比較は 10 ペアとなる。それぞれの比較で「より本文に合致する」と判定された選択肢に 1 ポイントを与え、最も獲得ポイントが多かった選択肢を解答として出力する。

BCBM の中核要素である 2 値分類器は、サポートベクターマシン (SVM) を用いて構成する。SVM の入力には、2 つの選択肢 c_i と c_j の特徴ベクトルの差 $x = f^{(i)} - f^{(j)}$ を用いる。SVM の学習では、読解問題 1 問 (5 つの選択肢) に対して、正解選択肢と間違い選択肢のペア 8 個 (4 個 $\times$ 2 方向) を作成し、学習データとして使用する。

4.4 Ranking 法

5 つの選択肢から 1 つを選ぶもう一つの方法は、ランキング学習を用いる方法である。機械学習を用いた 2 つ

目の解法である **Ranking 法**^{*4}は、Ranking SVM を用いて、本文に合致する選択肢の順位が高くなるようにランキング学習を行い、順位が 1 位になった選択肢を出力する方法である (図 4 参照)。問題解答時は、学習したスコア関数を用いて 5 つの選択肢をスコアの大きい順に並べ替え、1 位の選択肢を解答として出力する。学習時は、各設問の 5 つの選択肢のうち、本文に合致する選択肢を 1 位、合致しない選択肢を 2 位として学習データを構成する^{*5}。

5. 実験

5.1 実験条件

実験には、センター試験の問題^{*6}、および各種予備校のセンター試験対策問題集や模擬試験の問題を典拠とする大問 65 問、設問数 276 問を使用した (表 3)。実験に使

*4 査読者の示唆に基づき、Machine Learned Ranking 法 (MLR) から改名した。

*5 「本文に合致しない選択肢を選ぶ」問題では、不正解 (4 選択肢) を 1 位、正解 (1 選択肢) を 2 位にすることが必要である。

*6 奇数年度の本試験と追試験を使用 (2009 年度追試験と 2011 年度追試験を除く)。2005 年度以前は『国語 I・II』を使用。

図4 Ranking法の概要

表3 使用した読解問題

出典 (20xx 年)	大問	設問	選択肢
センター試験過去問 (01-15)	14	56	280
代ゼミ問題集+模試 (05, 14, 15)	19	76	380
駿台問題集 (06, 14)	14	56	280
河合塾問題集 (14, 15)	10	40	200
旺文社問題集 (14)	4	16	80
ベネッセ模試 (14, 15)	4	32	160
計	65	276	1380

用した読解問題ソルバーの設定と実装の詳細、および評価法を表4に示す。

5.2 実験1: 照合領域と性能

すべてのソルバーは、照合領域パラメータ α を持つ。表5に8通りの α における正解数を示す。この α の値は、これまでの実験結果[佐藤 14a, 加納 15a, 加納 15b]に基づいて定めた。表のボールド体は、それぞれのソルバーの最大の正解数である。なお、本文照合法と節境界法では、パラメータ β と γ を定める必要があるが、 β は4通り ($\mathcal{A}, \mathcal{A}^2, \mathcal{L}, \mathcal{W}$)、 γ は2通り (on と off) の合計8通りに対して正解数を求め、最大正解数が最も多かったものを採用した。

5.3 実験2: BCBMにおける特徴選択

機械学習では、特徴選択(不要な特徴量を除去)によって性能が向上する場合があることが、よく知られている。これを調べる実験を、BCBMに対して実施した*7。

実験手順は、次の通りである。

- (1) 全特徴集合(17要素)から、空集合を除いた全ての部分集合 ($2^{17} - 1 = 131,071$ 通り)を作成する。
- (2) そのそれぞれに対し、2値分類器の性能を 10-fold cross-validation で測定する。
- (3) 一定以上の性能を記録した特徴集合に対して、大問単位の leave-one-out cross-validation でソルバーの性能を測定する。

*7 Ranking法に対しては、現実的な実行時間で実施できる見込みがなかったため、実施しなかった。

表4 ソルバーの設定、実装、評価法

タイプブレーク	同点の場合は、選択肢番号の小さいものを選ぶ
順位反転	「本文と合致しないものを選ぶ」問題では、スコアが大きいものではなく、小さいものを選ぶ
形態素解析	MeCab/IPADic† を使用
SVM	WEKA の classifiers/functions/SMO パッケージを使用 [Hall 09] カーネル: 線形カーネル パラメータ: すべてデフォルト値を使用 前処理: 特徴ベクトルの各要素の値を、0 から 1 の範囲の実数に正規化
Ranking SVM	SVM ^{RANK} を使用 ‡ [Joachims 06] カーネル: 線形カーネル 正則化パラメータ: $C = 39.8$ (実験的に決定) 前処理: 特徴ベクトルの各要素の値を、0 から 1 の範囲の実数に正規化
評価方法	機械学習を用いない方法は、すべての設問をそのまま解く。 機械学習に基づく方法では、大問単位の leave-one-out cross-validation、すなわち、同一の本文を用いる設問を除いて学習したソルバーで、設問を解く。

†<http://taku910.github.io/mecab/>

‡https://www.cs.cornell.edu/people/tj/svm_light/svm_rank.html

表5 実験1: 照合領域と性能

解法	本文 照合法	節境界法		機械学習	
		CN	LN	BCBM	Ranking
β	$\mathcal{A}$	$\mathcal{W}$	$\mathcal{W}$	—	—
γ	on	on	on	—	—
$\alpha=P-0-0$	87	82	74	97	108
$\alpha=P-1-0$	82	80	79	101	112
$\alpha=P-b-0$	92	78	69	98	106
$\alpha=P-a-0$	99	73	69	105	109
$\alpha=P-1-1$	84	89	72	104	107
$\alpha=P-b-1$	87	86	69	97	102
$\alpha=P-b-c$	97	89	71	102	112
$\alpha=P-a-c$	104	79	67	108	110

照合領域パラメータ α には、実験1で最も性能が良かった P-a-c を採用した。

実験結果を表6に示す。この表には、2値分類器の精度が70.5%を超えた32通りの場合における、特徴集合の構成(使用された特徴量を1で表示)、2値分類器の精度(正解率)、ソルバー性能(正解数)を示している。この実験で得られた最大の正解数は117問であり、それは ID=23 の特徴集合(12の特徴量)において得られた。

5.4 実験3: 出典別の正解数

これまでの研究において、ソルバーの性能は、センター試験を対象とした場合と、予備校の模試を対象とした場合では、かなり異なることが明らかになっている。そこ

表6 実験2: BCBMの属性選択

ID	12345	67890	12345	67	精度	正解		
1	1	11	11	1	111	11	70.74	111
2	1		11	1	111	11	70.70	111
3	1	1	11	1	111	11	70.70	112
4	1	11	11		111		70.70	112
5	1	1	11	11	111	11	70.70	113
6	1		1	11	111	1	70.65	113
7	1		11	11	111	1	70.65	113
8	1		1	1	111	1	70.61	115
9	1		1	11	111	11	70.61	114
10	1	1	11		111		70.61	112
11	1	1	11111	111	11		70.61	116
12	1	11	1111	111	1		70.61	114
13	1	1	1	11	111	11	70.61	112
14	11		1	11	111	1	70.61	113
15	1111	1	11	11111	11		70.61	106
16	1		11	11	111	11	70.56	113
17	1	11	11	11	111	11	70.56	114
18	1	1	11	1	111	11	70.56	111
19	11	11	11	11	111	11	70.56	111
20	11111		1	1111	1		70.56	107
21	1		1	1	111	11	70.52	115
22	1	1	1	11	111	11	70.52	113
23	1	1	11111	111	11		70.52	117
24	1	1	11	11	111	11	70.52	113
25	1	1	1111	111	1		70.52	114
26	1	1	11	1	1	11	70.52	110
27	1	11	11111	111	11		70.52	115
28	1	1	1	11	11	111	70.52	112
29	1	111	1111	1111	1		70.52	110
30	11	1	11	11	111	11	70.52	114
31	1111	1	111	1111	11		70.52	114
32	11111	1	11	1111	1		70.52	108
base	11111	11111	11111	11			70.02	108

表7 実験3: 出典別の正解数

	本文 照合法	節境界法		機械学習	
		CN	LN	BCBM	Ranking
α	P-a-c	P-b-c	P-1-0	P-a-c	P-b-c
β	$\mathcal{A}$	$\mathcal{W}$	$\mathcal{W}$	—	—
γ	on	on	on	—	—
すべて (276)	104	89	79	117	112
センター (56)	25	32	25	25	28
代ゼミ (76)	31	24	22	31	27
駿台 (56)	16	17	13	26	24
河合塾 (40)	10	7	8	11	11
旺文社 (16)	8	1	3	8	8
ベネッセ (32)	14	8	8	16	14

で、それぞれの解法における、最高性能のソルバーに対し、問題の出典別の正解数を調べた。その結果を表7に示す。

5.5 検 討

現時点における最終的な結果は表7である。全設問に対して最も成績が良かったのはBCBM(特徴選択あり)で、

表8 McNemar 検定の行列

		BCBM		計
		正解	不正解	
本文 照合法	正解	73	31	104
	不正解	44	128	172
計		117	159	276

276 問中 117 設問 (42%) に正解した。センター試験 (過去問) に対して最も成績が良かったのは節境界法 (CN) であり、56 問中 32 問 (57%) に正解した。センター試験以外の出典に対しては、すべて BCBM の成績が最も良かった。出典別で見ると、他と比較して、河合塾の問題に対する成績が悪いように見える (40 問中 11 問正解)。

最も成績が良かった BCBM と一番最初の解法である本文照合法に McNemar 検定を適用したところ (表 8)、統計的に有意な差は見られなかった。

照合パラメータ α が同一の場合、本文照合法の判定基準は、BCBM で f_1 のみで判定している場合に相当する*8。このことは、 f_1 が非常に強力な特徴量であることを示唆する。事実、BCBM の特徴選択 (表 6) において、 f_1 は必ず選ばれている。

この他に特徴選択でよく選ばれている特徴量は、 f_{11} , f_{12} , f_{13} , f_6 , f_{16} , f_9 である。これらのうち、 f_{16} (形態素原形 CN スコア) 以外は、照合領域とは無関係、つまり f_1 とは独立の特徴量である。性能に統計的に有意な差をもたらさなかったにせよ、これらの特徴量は一定の効果を持つと考えられる。

照合領域に目を向ける (表 5) と、P-b-c (当該設問の一つ前の設問の傍線部を含む段落から、一つ後の設問の傍線部を含む段落まで) と P-a-c (本文の先頭から一つ後の設問の傍線部を含む段落まで) が、比較的成绩が良い。これは、傍線部の後方を参照しないと正しい解答を導けない場合があることに対応していると考えられる。我々の経験では、一般に、長いテキストの照合の場合は文字オーバーラップ率が効果的で、短いテキストの照合の場合は、文字 bigram や形態素のオーバーラップ率の方が効果的である。節境界法の最良の照合単位パラメータが $\beta = \mathcal{W}$ であるのは、節という短い単位で照合を行うことによると考えている。

これらの結果を受けて、どのような結論を導くべきかは、それほど明白ではない。まず、試験問題は、それぞれの問題で本文が異なり、(人間にとっての) 難易度も異なる。特に、予備校の模試は、実施時期 (春、秋、本番直前の冬) で難易度が異なるようである。

センター試験は、受験者の平均点がおおよそ配点の 50% となるように作られていると言われている。これを敷衍すると、評論読解問題では 4 問中 2 問正解が、おおよそその平均となる。センター試験過去問に限定すれば、我々のソルバーは、すでにそれと同等と考えて良いレベルに

*8 事前選択が存在するので、完全には一致しない。

表 9 公開性能評価の結果

ソルバー	解法	α	β	γ	2014 年				2015 年							
					第 1 問				第 1 問				第 5 問			
					問 2	問 3	問 4	問 5	問 2	問 3	問 4	問 5	問 2	問 3	問 4	問 5
2014-1	本文照合法	P-b-0	$\mathcal{A}$	off	3	2	3	4	2	2	4	1	2	4	1	4
2014-2	本文照合法 + α	$\dagger$	$\mathcal{A}$	off	4	2	1	1	—	—	—	—	—	—	—	—
2014-3	節境界法 (CN)	P-b-0	$\mathcal{W}$	off	5	1	1	5	3	4	1	1	4	2	1	4
2014-4	節境界法 (CN)	P-1-1	$\mathcal{W}$	off	3	1	1	5	3	4	5	2	3	2	3	4
2014-5	節境界法 (LN)	P-1-1	$\mathcal{W}$	off	3	2	3	5	1	4	5	3	2	2	2	4
2015-1	BCBM	P-a-c							3	3	4	2	4	1	1	1
2015-2	BCBM	P-b-0							3	3	4	2	4	1	1	1
2015-3	Ranking 法	P-a-c							3	2	4	2	2	1	1	1
2015-4	Ranking 法	P-b-1							3	2	4	2	4	1	1	5
2015-5	Ranking 法 (RBF)	P-b-c							3	2	1	2	2	1	1	4

 $\dagger$ 設問ごとに自動変更

到達している。この事実をどう解釈すべきであろうか。

我々のソルバーに対する強い拒否反応に、「本文をまったく理解していないのではないか」という意見がある。確かに我々のソルバーは、表層的な手がかりのみを使用し、多くの人間が「理解」という言葉から期待するような形では本文を分析してはいない。しかし、そのソルバーの成績が平均的な高校生の成績と同等であれば、逆に「はたして平均的な高校生は本文を理解しているといえるのか」という疑問が生じる。

文章の理解の度合いを測る標準的な方法は、その文章の内容に対する質問をいくつか用意し、それらに正しく答えられるかを調べる方法である。そして、この方法で読解力が計測できるという確信こそが、『国語』の読解問題の存在基盤である。この方法による読解力の計測を受け入れるのであれば、その成績と読解力には強い相関があることを受け入れるしかない。

石原の分析によれば、センター『国語』の評論問題を解くために求められる能力は、次の 3 つである ([石原 14], p.17)。

- (1) 書いてあることを過不足なくまとめること＝情報整理能力
- (2) 書いてあることを別の言葉に言い換えること＝翻訳能力
- (3) 二項対立にそって論述されている道筋を整理すること＝二項対立参照能力

これらの能力のうち、現在のソルバーに組み込まれているものは「翻訳能力」であり、より正確には「選択肢に書かれてあることが、本文に書いてあることを別の言葉に言い換えたものであるかどうかを判定すること (= 言い換え認識能力)」である。

試験問題では、「選択肢が問題文をそのまま書いていたら『正解』丸見えで問題にならないから、出題者は常識的に許される範囲で問題文の表現を別の言葉で言い換える ([石原 14], p.14)」^{*}。しかしながら、他の言語に翻訳するわけでないので、非常に巧妙に言い換ええない限り、表

層的な類似を消し去ることができない。表層的な類似を主な手がかりとする現在のソルバーがまがりなりにも問題が解けるのは、このような理由によるものと推測する。この仮説が正しいと仮定すれば、読み手を意図的に惑わす必要がない場合の内容照合問題 (たとえば、歴史の正誤問題を解くために行う教科書との照合) に対しては、表層的方法はより有効に機能すると考えられる。

6. 公開性能評価

6.1 東ロボ・フォーマルラン概要

東ロボ・プロジェクトの公開性能評価は、作成したソルバーを用いて新規の模試^{*9}を解く形式 (評価型ワークショップのフォーマルランと同じ方式) で実施されている。我々は、2013 年から 2015 年の 3 回のフォーマルランにおいて、『国語』現代文の解答を提出した。フォーマルランでは、複数のソルバーの解答結果を提出することが認められている。『国語』評論読解問題に対しては、2013 年度と 2014 年度はそれぞれ 5 種のソルバーの結果を、2015 年度は 9 種 (うち 4 種類は前年との比較のため) のソルバーの結果を提出した。

2014 年と 2015 年のフォーマルランの結果を表 9 に示す。なお、2015 年で使用した進研模試では、評論読解問題が中心となる大問 5 が存在し、大問 4 (漢文) とどちらかを選択する形式となっている。この表の数字は、ソルバーが解答した選択肢番号を表し、四角で囲まれているものは、正解したことを表す。

フォーマルランに使用したソルバーは、それぞれの年の秋の時点で有望と考えたものであり、先の実験に示した最高性能を達成したソルバーとは設定が異なる。

- (1) 2014-2 は、本文照合法にいくつかの機能を追加したものである ([佐藤 14a] の E3 ソルバー)。
- (2) 2015- n は、表 2 に示した 17 個の特徴量のうち、 $f_6, f_7, f_9, f_{12}, f_{13}$ を除いた 12 個の特徴量を使用して

^{*9} 2013 年度と 2014 年度は代々木ゼミナールの模試、2015 年度はベネッセの模試 (進研模試) が使用された。

いる。

- (3) 2015-5 は、Ranking SVM のカーネルに RBF カーネルを採用した Ranking 法 (SVM の正則化パラメータは $C = 6.31$, RBF カーネルのパラメータは $\gamma = 0.001$) である。

東ロボ・プロジェクトの公式発表では、これらのソルバーのうち、最も成績が良かったものを公表している。2014 年は 2014-5 が、2015 年は 2015-1 が選ばれ、それぞれ 4 問中 2 問、8 問中 5 問の正解であった。

表面上の成績は毎年向上しており、かつ、前年のソルバーで解いた場合より、良い成績を取めている。ただし、対象問題数が少ないため、確定的なことは何も導くことはできない。

7. 関連研究

7.1 東ロボ

センター試験形式の『国語』評論読解問題に最も近い問題は、同じくセンター試験形式の『国語』小説読解問題である。本文照合法を小説読解問題に対して適用したの成績はチャンスレベルである [佐藤 14b]。2013 年の公開性能評価では、本文照合法と心情語辞書を組み合わせたソルバーを用い、4 問中 2 問正解 [佐藤 14a]。2014 年以降は、心情を問う設問に対しては新しい解法 (未発表) を、その他の設問は本文照合法を適用するソルバーを用い、2014 年は 4 問中 1 問正解、2015 年は 4 問中 2 問正解であった。小説読解問題は、心情を問う設問以外は評論読解問題とかなり似ており、評論読解問題の解法がそのまま適用できると予想されるが、研究は遅れている。

センター試験の『英語』にも長文読解問題が存在する。菊井ら [菊井 15] はランキング学習を用いた方法を適用し、センター試験過去問 74 問に対して 42% の正解率が得られたことを報告している。なお、我々の Ranking 法は、この研究報告に基づき、『国語』評論読解問題にランキング学習を適用したものである。

センター試験の『社会』の世界史や日本史は、現時点で成績が良い教科である。これらの歴史の問題の中心は、歴史的事実を問う選択式の正誤問題であり、そのほとんどは、教科書の記述との対応付けで解ける問題である。研究の初期は、正誤問題を含意認識問題に落とし込んで解くことが多かったが、2015 年のプロジェクトの成果報告会 (2015 年 11 月 14 日) では、選択肢の一部を空欄にして質問応答に落とし込んで解く方法 [Kanayama 12] が有効であると報告された。本文照合法は、選択肢の選択を含意認識問題に落とし込んで解いているので、歴史問題の初期の解法とは類似性が高い。

『数学』は、東ロボ・プロジェクトのフラグシップ科目であり、最も労力と人員が投入されている科目である。公式発表では高い正答率 (偏差値) が報告されているが、それは、設問文の解析に人間がヒントを与えている場合

の正答率であり、このヒントを削除し、設問文の解析を完全に自動的に行った場合は、正答率はかなり低下する。『物理』では、設問文の解析の自動化はまったく実現されておらず、問われている内容は人間が内部表現に翻訳してから、それを物理シミュレーターに入力する方法が採られている。

以上が東ロボ・プロジェクトの他の科目の現状である。正統的な文解析のアプローチで設問文を理解しようとしている『数学』では、背後に明確な数学モデルが存在するにもかかわらず、数文からなる文章の理解も完全には実現できていない。3500 字からなる長い文章で、かつ、毎回テーマや内容が変わる評論の本文を、同様のアプローチで解析することは、当面、不可能であろう。

7.2 Machine Comprehension

センター試験の英語の読解問題は、CLEF QA Entrance Exams Task でも取り上げられ、すでに 3 回 (2013 年、2014 年、2015 年) のフォーマルランが実施されている [Peñas 13, Peñas 14, Rodrigo 15]。それぞれの年のベストの成績は、それぞれ、46 設問中 16 問正解 (35%)、56 設問中 25 問正解 (45%)、89 設問中 52 問正解 (58%) であり、3 年目によりやく 50% を超えた。これらの読解問題は 4 択の選択式であり、ランダム選択によるベースラインは、5 択の国語読解問題よりも高い。

テキストとそれに対する選択式の問題によって (機械の) 読解力を測定する取り組みは、評価が明確かつ容易という利点があるため、今後、Machine Comprehension の研究の一つの方向性になる可能性が高い。Richardson ら [Richardson 13] は、7 歳児が解けるような英語読解問題 (4 択式) のテストセットの作成を報告している。

8. まとめ

本論文では、センター試験『国語』評論読解問題に対する、我々のこの 2 年間 (2014–2015 年) の取り組みについて述べた。最後に、その総括として、結局何がわかったのかをまとめてみよう。

- (1) 表層的な手がかりのみを用いるソルバーの正解率は、平均的な高校生とほぼ同等レベルに達していると考えてよい。
- (2) 現在のソルバーに組み込まれている能力は、表層的な手がかりに基づく「言い換え認識能力」である。表層的な手がかりのみで言い換え認識が可能なのは、非常に巧妙に言い換えない限り、表層的な類似を消し去ることができないからであろう。
- (3) 文字オーバーラップ率という特徴量は、非常に強力な特徴量である。

現時点でベストと考えられる BCBM は、今回の実験では、本文照合法より有意に優れているという結論は導けなかった。しかしながら、工夫の余地は色々残っており、

表層の手がかりを用いる方法でも、性能向上の可能性は残っていると思われる。受験参考書(あるいは、代々木ゼミナールの講評)を信じるならば、消去法をマスターすることが、得点向上の大きな鍵となる可能性が高い。消去法の根幹は、矛盾検出である。このあたりに的を絞った特徴量を考えていくのが、一つの方向性である。

もう一つの方向性は、本文や設問に応じた照合領域の自動設定である。これまでの解法の最大の弱点は、外からのパラメータにより照合領域が固定化されている点にある。板野 [板野 10] は、正解を導くための根拠が傍線部の前後 2 行の範囲にある設問は全体の 63.1% と分析しており、傍線部付近に広めに網を張る現在の戦略は、全体としては正しい選択であろう。しかし、個別の設問においては、必ずしも最良とは限らない。それぞれの設問に対して、解答の根拠となる部分を正確に捉えることは、今後の最重要課題の一つであり、記述式の読解問題攻略への出発点でもある。

本論文では、センター『国語』評論読解問題を対象に、そのソルバーを作ることを通して読解力について検討したが、小説読解問題や他教科の問題で必要な読解力も加味して考えるならば、「読解力」を単一の能力と考えるのではなく、多くの能力を統合した総合力と捉えるほうが適切であろう。たとえば、小説の読解では、「書かれていない『心情』を多くの人がそう思うようだろうに想像すること＝小市民的感情力 ([石原 14], p.13)」が必要であろうし、数学や物理の設問文の読解では、日本語表現を計算可能な式に翻訳する能力が必要になる。我々が今後すべきことは、「読解力」の分解を進め、読解力を構成するそれぞれの能力に対する理解を深めるとともに、総合力としての「読解力」の把握につとめることであろう。

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。本研究では、次の研究助成等の支援を受けた。科学研究費基盤研究 (B)24300052, 人工知能知能学会実践 AI チャレンジ研究助成, 科学研究費基盤研究 (B)15H02748。

◇ 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)
- [船口 97] 船口 明: きめる! センター国語現代文, 学研教育出版 (1997)
- [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(4)/2013-SLP-96(4), 情報処理学会 (2013)
- [石原 14] 石原 千秋: 打倒! センター試験の現代文, ちくまプリマー新書 217, 筑摩書房 (2014)
- [板野 10] 板野 博行: ゴロゴ板野のセンター現代文解法パターン

集, 星雲社 (2010)

- [Joachims 06] Joachims, T.: Training Linear SVMs in Linear Time, in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '06, pp. 217–226 (2006)
- [Kanayama 12] Kanayama, H., Miyao, Y., and Prager, J.: Answering Yes/No Question via Question inversion, in *Proceedings of International Conference on Computational Linguistics*, COLING '12, pp. 1377–1392 (2012)
- [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 15a] 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-NL-220(8), 情報処理学会 (2015)
- [加納 15b] 加納 隼人, 佐藤 理史, 松崎 拓也: センター試験『国語』評論読解問題ソルバーの改良の検討, 2015 年度人工知能学会全国大会論文集, 1K2-3 (2015)
- [菊井 15] 菊井 玄一郎, 三村 奈央, 藤川 哲志, 但馬 康宏: 表層類似性と含意認識結果の組み合わせによる英語長文問題解答の試み, 2015 年度人工知能学会全国大会論文集, 1K2-4 (2015)
- [益岡 92] 益岡 隆志, 田窪 行則: 基礎日本語文法—改訂版—, くろしお出版 (1992)
- [Peñas 13] Peñas, A., Miyao, Y., Hovy, E., Forner, P., and Kando, N.: Overview of QA4MRE 2013 Entrance Exams Task, in *Working Notes of CLEF 2013 (Conference and Labs of the Evaluation Forum)* (2013)
- [Peñas 14] Peñas, A., Miyao, Y., Rodrigo, A., Hovy, E., and Kando, N.: Overview of CLEF QA Entrance Exams Task 2014, in *Working Notes of CLEF 2014 (Conference and Labs of the Evaluation Forum)*, pp. 1194–1200 (2014)
- [Richardson 13] Richardson, M., Burges, C. J., and Renshaw, E.: MCTest: A Challenge Dataset for the Open-Domain Machine Comprehension of Text, in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, EMNLP '13, pp. 193–203 (2013)
- [Rodrigo 15] Rodrigo, A., Peñas, A., Miyao, Y., Hovy, E., and Kando, N.: Overview of CLEF QA Entrance Exams Task 2015, in *Working Notes of CLEF 2015 (Conference and Labs of the Evaluation Forum)* (2015)
- [佐藤 14a] 佐藤 理史, 加納 隼人, 西村 翔平: 代ゼミ模試に挑戦 2013—『国語』現代文, 情報処理学会研究報告, 2014-NL-215(10), 情報処理学会 (2014)
- [佐藤 14b] 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 (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)

[担当委員: 岡崎 直観]

2016 年 6 月 13 日 受理

—— 著 者 紹 介 ——

加納 隼人

2014 年名古屋大学工学部電気電子・情報工学科卒業。2016 年名古屋大学大学院工学研究科電子情報システム専攻博士前期課程修了。在学中は、主に自然言語処理の研究に従事。

佐藤 理史(正会員)

1988 年京都大学大学院工学研究科博士後期課程電気工学第二専攻研究指導認定退学，京都大学工学部助手，北陸先端科学技術大学院大学助教授，京都大学大学院情報学研究科助教授を経て，2005 年より名古屋大学大学院工学研究科教授，工学博士，自然言語処理，人工知能の研究に従事，

松崎 拓也(正会員)

名古屋大学大学院工学研究科准教授，博士（情報理工学），1977 年生まれ，2007 年東京大学大学院情報理工学系研究科コンピュータ科学専攻博士課程修了，2007 年から 2012 年，東京大学大学院情報理工学系研究科助教，国立情報学研究所特任准教授を経て 2014 年より現職，統語解析・意味解析を中心とした自然言語処理に関する研究に従事，