

科学研究費助成事業 研究成果報告書

平成 26 年 5 月 22 日現在

機関番号：82616

研究種目：挑戦的萌芽研究

研究期間：2011～2013

課題番号：23650558

研究課題名(和文) 短答式記述テストにおける実用的な自動採点システムの開発

研究課題名(英文) Automated scoring system for short-answer responses

研究代表者

石岡 恒憲 (Ishioka, Tsunenori)

独立行政法人大学入試センター・研究開発部・教授

研究者番号：80311166

交付決定額(研究期間全体)：(直接経費) 2,700,000円、(間接経費) 810,000円

研究成果の概要(和文)：研究開始当初は、入力解答を模範解答と比較し、正確性を測るために再現率と精度という2つの指標を用いることを考えていた。しかしながら、その後、含意関係認識という文法に基づく正統的な自然言語処理技術の利用が有望であることがわかり、国立情報学研究所のメンバーから知見を得ながら研究を進めた。成果は以下の通り：

1. 小論文の採点・評価について、我々の実証実験をもとに、日本テスト学会第10回大会で発表した。2. エッセイ/作文テストにおけるコンピュータ利用と自動採点について、公開シンポジウムで講演し、広く活動を紹介した。3. 自動採点に向けた統計処理技術の方法について研究を行い、有名雑誌等に多くの論文を掲載した。

研究成果の概要(英文)：At the beginning of our research, we were planning to use two indices of recall and precision in order to measure a similarity between an input answer and the model answer. After that, it turned out that use of the textual entailment recognition based on the legitimate natural-language-processing technology is promising. Our research was advanced, acquiring knowledge from the member of National Institute of Informatics. Main results are follows:

(1) Evaluating or grading essays based on our actual proof experiment was announced at the 10th Japanese Test society convention. (2) About the computer use and automatic grading in an essay or composition test, a lecture was given at open symposium to introduce our activity widely. (3) The method of the statistical technology towards automatic grading was developed, and many papers were published in famous journals.

研究分野：総合領域

科研費の分科・細目：科学教育・教育工学

キーワード：e-ラーニング テキスト評価 自然言語処理 情報システム 統計数学 人工知能

1. 研究開始当初の背景

アメリカでは教育関係者のみならず一般においても主観が入るエッセイ試験においては、コンピュータによる評価採点の考えが受け入れられるようになってきた。実際、アメリカのビジネススクール入学のための共通テストである GMAT における作文(エッセイ)テストでは、1998 年より e-rater が、2006 年よりは IntelliMetric が採点を行なっている。他にも商用のシステムとして、PEG (Project Essay Grade) や IEA (Intelligent Essay Assessor) があり利用に供されている。我が国においても、本研究代表者のグループが日本語で初めての小論文自動採点システム Jess (Ishioka, 2006) を開発し、実用化に向けた研究を進めている。

ところが、質問の回答を 1 文あるいは 2 文で返す、いわゆる短答式記述(short-answer)テスト(以下、短答式テストと略記する)をコンピュータで行なうことについては、その重要性は認められているものの技術的にさまざまな課題が未解決のままである。

2. 研究の目的

我が国で初めての短答式記述(short-answer)テストをコンピュータで自動採点を行なうシステムを構築することを最終目標とし、そのための要素技術を確立する。短答式テストは多肢選択テストに比べ、以下の点が優れていると考えられており、採点のための技術的な課題が克服されてさえいれば、短答式テストを使う潜在的な需要は計りしれない。

- (1) 短答式テストの方がより真正(authentic)で信頼できると広く考えられている。実際、現実世界における質問応答は、多肢選択ではなく、短答式テストに近い。
- (2) 経済性。高品質な多肢選択問題を作成することは、通常、コストと手間がかかる。
- (3) 多肢選択テストはテスト戦略を立てやすく、学生の問題についての理解を正しく評価することが難しい。また当て推量(まぐれ当たり)による効果も無視できない。これが実現すれば、マークシートによらない試験問題の出題の可能性が大きく広がる。

3. 研究の方法

(1) アルゴリズムの検討

我々が取るアプローチは、解答した答えをアンサーキーと比較するものである。アンサーキーは 1 つの質問に対して、1 つ以上の正解からなるもので、予め試験問題作成者が用意しておく。正確性を測るために 2 つの指標を考える。最初の指標はアンサーキーと解答文との単語の重複に基づいたもの、すなわち再現率である。もう一つの指標は精度である。これは、いわば解答におけるゴミの少なさを示す指標で、解答文とアンサーキーとの単語の重複数を解答文の単語数で割る。一般に精度と再現率の間には、(一方を高くすれば他

方が低くなるという)トレードオフの関係があるから、ある一定の再現率のもとで、ある閾値を超えた精度のものを正解と判定することになる。そこで我々は、正解を判定するための再現率 - 精度曲線を求め、これを用いて正解 / 不正解を判定することを考えている。今年度はそのためのこの閾値の洗い出しについての検討を行う。

(2) マシン環境の整備

LAN上でWebブラウザによってシステムが実行できるように、Webサーバーを設置する。可能ならば、今年度の実運用を行い、その問題点を洗い出すとともに、不備について改良をおこなう。

(3) Rによる言語処理ライブラリの一部作成

Rでは自然言語を取り扱うためのパッケージLanguageRの他、日本語処理においても形態素解析RMeCab (Rで動作するMeCab)や構文解析RCaboCha (Rで動作するCaboCha)を利用することができる。またNグラムなどの計量言語学で用いられる各種統計量を計算するためのパッケージ群も同時に提供され、現在も精力的にバージョンアップが重ねられている。本システムでもこれらライブラリの積極的な利用を前提とするが、不足するものについては自作する。

4. 研究成果

研究開始当時は、解答した答えをアンサーキーと比較し、正確性を測るために再現率と精度という 2 つの指標を設定することを考えていた。しかしながら、その後、含意関係認識という文法に基づく正統的な自然言語処理技術の利用が有望であることがわかり、国立情報学研究所のメンバーから知見を得ながら研究を進めた。成果は以下の通り：

- (1) 小論文の採点・評価について、2012 年度末、モニター調査で行った実証実験をもとに、その成果を日本テスト学会 第 10 回大会で発表した。石岡・荒井・大久保論文は従来型の 800 字の小論文についてであり、荒井・大久保・石岡・宮埜論文はそれより短い 400 字の論述テスト 2 題についての評価を行った。報告書「日本語小論文試験についての実証的研究」を作成し、文科省、他関連区に配布した。2014 年度にも評価視点を被験者に示した上での小論文実験を実施した。

- (2) エッセイおよび作文テストにおけるコンピュータ利用と自動採点について、日本テスト学会 第 10 回大会 公開シンポジウムで講演を行い、広く活動を紹介した。また、国立情報学研究所が中心となり、いま人工知能の分野で耳目を集めている「ロボットは東大に入れるか」プロジェクトにおいても、研究代表者が小論文の自動採点について講演を行った。

- (3) 自動採点に向けた統計処理技術の方法について研究を行い、共同研究者(長塚)が有名雑誌の一つである Computational Statistics and Data Analysis 誌らに多くの

論文を掲載した。

(4)短答式記述試験の自動採点については、そもそも課題が困難であることに加え、当該技術分野の世界的技術の進展が停滞したこともあり、当初の目標を十分に達成したとはいえないが、現在、短答式試験におけるコンピュータ・ベースド・テストのシステムを、同僚である大久保博士を中心にホワイトリスト形式で試作中である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 11件)

Hideki Nagatsuka, Toshinari Kamakura and N. Balakrishnan, A consistent method of estimation for the three-parameter Weibull distribution, Computational Statistics and Data Analysis, Elsevier, 査読有, Vol.58, 2013, 210-226.

荒井清佳, 大久保智哉, 石岡恒憲, 宮埜壽夫, 大学入学者選抜における小論文試験と教科・科目試験との関連について, 査読有, Vol.9, 2013, 27-36.

Tsunenori Ishioka, Imputation of Missing Values for Unsupervised Data Using the Proximity in Random Forests, eLmL 2013 : The Fifth International Conference on Mobile, Hybrid, and On-line Learning, ISBN: 978-1-61208-253-0, 査読有, 2013, 30-36.

Tsunenori Ishioka, Imputation of Missing Values for Semi-supervised Data Using the Proximity in Random Forests, iiWAS 2012, ACM International Conference Proceedings Series (ISBN: 978-1-4503-1306-3), 査読有, 2012, 319-325.

Hideki Nagatsuka and N. Balakrishnan, Parameter and quantile estimation for the three-parameter lognormal distribution based on statistics invariant to unknown location, Journal of Statistical Computation and Simulation, 査読有, Vol. 43, 2012, 1629-1647.

DOI: 10.1080/00949655.2012.667410

Hideki Nagatsuka and N. Balakrishnan, A consistent parameter estimation in the three-parameter lognormal distribution, Journal of Statistical Planning and Inference, Elsevier, 査読有, Vol. 142, Issue 7, 2012, 2071-2086.

DOI: 10.1016/j.jspi.2012.01.019

Hideki Nagatsuka and N. Balakrishnan, Parameter and quantile estimation for the three-parameter gamma distribution based on statistics invariant to unknown location, Journal of Statistical Planning and Inference, Elsevier, 査読有, Vol. 142, Issue 7, 2012, 2087-2102.

DOI: 10.1016/j.jspi.2012.01.018

Hideki Nagatsuka, Megumi Uchino and Hisashi Yamamoto, Parameter estimation of multivariate distributions under order restrictions of the parameters: An extension of isotonic regression, Quality Technology and Quantitative Management, 査読有, Vol.9, 2012, 283-293.

Hideki Nagatsuka and N. Balakrishnan, A consistent method of estimation for the parameters of the three-parameter inverse Gaussian distribution, Journal of Statistical Computation and Simulation, 査読有, 83, 1915-1931,

DOI:10.1080/00949655.2012.674130

Tsunenori Ishioka, Ascertain and graphically representing the logical structure of Japanese essays, International Journal of Continuing Engineering Education and Life-Long Learning (IJCEELL) 査読有, Volume 21, Issue 4, 2011, 276-288.

石岡恒憲, Random Forest を用いた欠測データの補完に基づく大学入試センター試験科目間得点差, 応用統計学, 査読有, 40 (3), 2011, 193-209.

[学会発表](計 7件)

石岡恒憲, Random Forests を用いた欠測値補完の半教師データおよび教師なしデータへの拡張, 言語処理学会第 20 回年次大会 (NLP2014), 北海道大学, A7-1. 査読無, 2014, 1047 - 1050 .

Tsunenori Ishioka, Imputation of Missing Values for Unsupervised or Semi-supervised Data Using the Proximity in Random Forests, Incomplete Data Analysis and Causal Inference, International Symposium at Osaka University, Sep. 22-23, 査読無, 2013 .

石岡恒憲, 小論文の自動採点, 国立情報学研究所「ロボットは東大に入れるか」プロジェクト講演, 招待講演, 2013

石岡恒憲, 荒井清佳, 大久保智哉, 小論文試験得点と他教科科目得点および受験者属性との関連, 日本テスト学会 第 10 回大会, 査読無, 2012, 72 - 75 .

石岡恒憲, 教師なしデータおよび半教師データにおけるランダムフォレストによる欠測値補完, 2012 年度 統計関連学会連合大会, 北海道大学, 査読無, 2012, 196 .

石岡恒憲, エッセイおよび作文テストにおけるコンピュータ利用と自動採点, 日本テスト学会 第 10 回大会 公開シンポジウム「パフォーマンス評価の未来」, 招待講演, 2012 .

石岡恒憲, Random Forest によるコンテキスト情報の補完について, 電子情報通信学会, 信学技報, 査読無, 2011.

[図書](計 3件)

石岡恒憲ほか(繁樹算男編),新しい時代の大学入試,金子書房,2014(刊行予定),p240.

荒井清佳,大久保智哉,石岡恒憲,日本語小論文試験についての実証的研究 小論文試験は何を測っているのか,大学入試センター研究開発部,2014,p49.

石岡恒憲ほか(植野真臣・永岡慶三・山内祐平共編),自然言語処理と学習評価,教育工学における学習評価,教育工学選書,2012,p.224.

〔産業財産権〕

出願状況(計 0件)

名称:
発明者:
権利者:
種類:
番号:
出願年月日:
国内外の別:

取得状況(計 0件)

名称:
発明者:
権利者:
種類:
番号:
取得年月日:
国内外の別:

〔その他〕

ホームページ等

Ishioka's Home Page:

<http://www.rd.dnc.ac.jp/~tunenori/>

6. 研究組織

(1)研究代表者

石岡 恒憲 (ISHIOKA TSUNENORI)

大学入試センター・研究開発部・教授

研究者番号: 8 0 3 1 1 1 6 6

(2)研究分担者

長塚 豪己 (NAGATSUKA HIDEKI)

中央大学・理工学部経営システム工学科・

准教授

研究者番号: 3 0 3 8 4 7 3 8

荒井 清佳 (ARAI SAYAKA)

大学入試センター・研究開発部・助教

研究者番号: 0 0 5 6 1 0 3 6