

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 7 日現在

機関番号：17104

研究種目：挑戦的萌芽研究

研究期間：2013～2015

課題番号：25540164

研究課題名(和文)プロジェクト型学習における自動相互評価方式

研究課題名(英文)Automatic mutual evaluation method for project based learning

研究代表者

井上 創造 (Inoue, Sozo)

九州工業大学・大学院工学研究院・准教授

研究者番号：90346825

交付決定額(研究期間全体)：(直接経費) 2,900,000円

研究成果の概要(和文)：本研究では、プロジェクト型学習において、学習者及び指導者の利用を負担を下げながら学習者の評価を行う手法を提案し、定式化・システム開発を行い、実際に大学における講義内で実験を行うことで評価した。今後の課題としては、定式化において既存の項目反応理論のようなテスト理論との整合性を検証したり、更なるユーザインタフェースにより問題を改善していくシステムが望まれる。

研究成果の概要(英文)：In this work, we proposed a method for assessing learners without increasing the workload of learners and teachers, mathematically formalized, developed a web-based system, and evaluated in real classes in the university. As future work, validating the compatibility with the traditional item response theory in test theory, and the method for improving quizzes with additional user interfaces, are required.

研究分野：情報工学

キーワード：プロジェクト型学習支援システム Webシステム eラーニング クラウドソーシング MOOCs

1. 研究開始当初の背景

近年、目標に向かってプロジェクトを進めながら自発的な学習を狙うプロジェクト型学習が注目されているが、その公平かつ客観的な評価の仕方には決定的な解はまだ無い。市民権を得ている方法として、学習者どうしで評価しあう相互評価があるが、これには以下のような問題がある。

- 一人一人の評価を集計する指導者の負担が大きい。原理的には、全員が全員を評価すれば、学習者数を n とすると n^2 個の相互評価を処理しなければならなくなる。
- その結果として、相互評価を頻繁に行うことができない。過程を評価するために、相互評価を高頻度に行うことが望ましいが、現実には1, 2回程度の相互評価ですますことになる。
- 多様な指標で評価しにくい。例えば「プレゼンテーション能力」、「プロジェクトの工夫」、「学習内容」のように、公平性と客観性を確保するために多様な評価指標を設けることが望ましいが、これも学習者と指導者の負担が増すことにつながり、困難を伴う。

2. 研究の目的

上記問題に対して、本研究では、学習者および指導者の利用の負担を極力下げ、かつ自動的に個人およびチームの評価を行う手法を研究する。それは、利用者には他人の活動に対して評価を簡単かつ頻繁に入力できるようにしながら、収集した評価を意味を考慮して機械学習することにより個人とチームを自動評価する方式である。他の手法と同等以上の、精度と評価尺度の多様性を、利用者の負担を増やすこと無く実現することを目標とする。

3. 研究の方法

作問・自動採点評価 Web システム 3 章ではアルゴリズムを紹介し、シミュレーションを行った。それを実運用するために、実際にユーザが作問し、問題を解答するための Web システムが必要である。実際に作問し、問題を解答してもらうための作問・自動採点評価 Web システム、“Atquiz”を作成した。

問題作成時の画面を図1、図2に、回答時の画面を図3に示す。

図 1: Atquiz の作問画面

図 2: Atquiz の作問画面

採点計算方法 この節ではユーザが生成するの各概念を定式化し、自動採点システムをモデル化する。最後に評価計算をするため反復アルゴリズムを記述する。

以下では、 U をユーザ集合、 Q を問題集合、 q を問題、 u を回答者とする。また、問



図 3: 解答終了時の画面

問題 q に回答者 u が回答して自動的に採点されたスコアを $score(q, u)$ と表す．スコアは問題や試験ごとに範囲がそれぞれあるが 0 から 1 の範囲にした．

- 問題の難易度

難易度の高い問題は多くの回答者のスコアが低く，簡単な問題は点数が高いと考えられるため，問題の難しさはその問題のスコアの平均に反比例するものとする．

$$\text{diff}(q) := -\frac{\sum_{u \in U} score(q, u)}{|U|}$$

ただし，後に乗算される際に難易度が最低の問題は評点を 0 にするため，この値から最小値 $\min_{q \in Q} \text{diff}(q)$ を引いても良い．

- 問題の妥当性

妥当性はその問題の作問者の評点の平均とする．問題と試験の目標の依存関係のモデル化ができないため妥当性の定式化が少し難しい．従来の授業から推察すると，教師は良い問題を，アシスタントはやや良い問題を，生徒はあまり本格的でない問題を作成するはずである．つまり、評点の高い作問者ほど妥当性の高い問題を作成するものと

仮定する．問題 q の妥当性は q の作問者 ($author(q)$ とする) によって解かれた問題の評点の平均と定義できる．問題 q の妥当性は q の作成者によって解かれた問題の評価の平均であると定義する．

$$\text{valid}(q) := \frac{\sum_{q' \in Q} \text{eval}(q', author(q))}{|Q|}$$

ただし，作問者 $author(q)$ が解いた問題が少ない場合は，この値がたまたま良い値となることも考えられるため， t 検定などにより母集団の区間推定を行い，その信頼区間の最小値を用いても良い．

また，後に乗算される際に妥当性が低すぎる問題は，むしろ不正解の方が評点が高いことが望ましい場合には，妥当性が低すぎる問題について負の値とするために，例えば平均値より 2σ 小さい値が 0 となるように値を調整しても良い．

- 問題へのユーザの解答に対する評点

問題の難しさや妥当性を考慮した上で回答者に与えられる評価点で，問題 q に対するユーザ u の回答に対する評点は

$$\text{eval}(q, u) := score(q, u) \cdot \text{diff}(q) \cdot \text{valid}(q)$$

と定義する．

ただし，この値は例えば平均値 50，標準偏差 20 となるように正規化されても良い．

アルゴリズム 難易度，妥当性，評点は前節で定義した．しかしながら，それらの定義は再帰的である．したがって私たちは収束するまで計算を繰り返すヒューリスティックな手法を用いる．

1. 各 $q \in Q$ について，難易度 $\text{diff}(q)$ を計算する．

2. 各ペア $(q, u) \in Q \times U$ について，評点の初期値を

$$\text{eval}(q, u) \leftarrow \text{score}(q, u)$$

とする．

3. 各 $q \in Q$ について，妥当性 $\text{valid}(q)$ を計算する．ただし，教師が作った問題のように明らかに妥当な問題については，妥当性を最高値 $\max_{q \in Q} \text{valid}(q)$ とする．

4. 各ペア $(q, u) \in Q \times U$ について，評点を

$$\text{eval}(q, u) \leftarrow \text{score}(q, u) \cdot \text{valid}(q)$$

により更新する．

5. 収束するまで (3)-(4) を繰り返す．

6. 各ペア $(q, u) \in Q \times U$ について，評点を

$$\text{eval}(q, u) \leftarrow \text{eval}(q, u) \cdot \text{diff}(q)$$

により更新する．

ここで評価は初期値としてスコアを与えており，ステップ (3)-(4) において収束するまで $\text{diff}(q)$ 以外の因子を繰り返し計算している．難易度 $\text{diff}(q)$ は繰り返し乗算されるのを防ぐために，最後のステップ (6) においてのみ乗算される．

上記のアルゴリズムについて，各問題の妥当性にばらつきがある場合と難易度にばらつきがある時で計算機シミュレーションを行い，アルゴリズムが収束することと，難易度と妥当性を適切に導くことを確認した．

4. 研究成果

今回の評点方法において，難しさ，妥当性がそれぞれ評点に影響しているかどうかを評価するため，大学内での講義での評価実験を行うことで評価を行った．グラフ (図 4) は解かれた問題の評点と妥当性を示している．

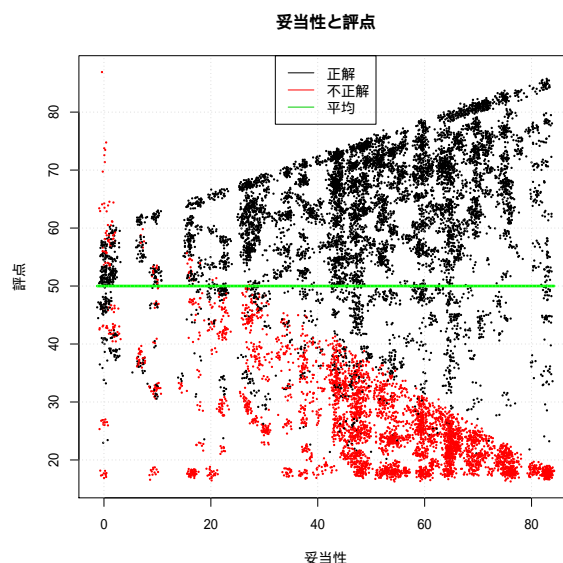


図 4: 評点と妥当性のグラフ

図 4 の x 軸は解かれた問題の評点，y 軸は解かれた問題の妥当性である．正解した問題は妥当性が上がるに従い，評点が上がっていることがわかる．また正解していない問題は妥当性が高くなれば評点も下がっていることが見受けられる．

まとめ このように，本研究では定式化・システム開発・実験により，学習者及び指導者の利用を負担を下げながら学習者の評価を行う手法を提案・評価した．今後の課題としては，定式化において既存の項目反応理論のようなテスト理論との整合性を検証したり，更なるユーザインタフェースにより問題を改善していくシステムが望まれる．

5. 主な発表論文等

【雑誌論文】(9 件)

1. (査読有) 谷口 敦, 井上 創造, "Experiment for Automatic Assessment of User-generated Tests", IEEE International Conference on MOOCs, Innovation and Technology in Education (MITE), pp. 152-155, 2015/10/01,

Amiritsar, India.

2. (査読有) 谷口 敦, 井上 創造, "A Method for Automatic Assessment of User-generated Tests and Its Evaluation", ACM Int'l Conf. Pervasive and Ubiquitous Computing (Ubicomp) Poster, pp. 225-228, 2015/09/09, Osaka.
3. (査読有) 谷口 敦, 井上 創造, "ユーザに作問を任せるテストの自動評点手法とその評価", マルチメディア, 分散, 協調とモバイル (DICOMO2015) シンポジウム, pp. 602-608, 2015/07/08, Iwate.
4. (査読有) 戸田 隆道, 井上 創造, 田中 翔太, 上田 修功, "Training Human Activity Recognition for Labels with Inaccurate Time Stamps", Ubicomp Workshop for Human Activity Sensing Corpus and its Application (HASCA), pp. 863-872, 2014/09/13, Seattle, USA.
5. (査読有) Ghada Farouk Naiem, 井上 創造, "A Method for Assessing User-generated Tests for Online Courses Exploiting Crowdsourcing Concept", International Workshop on Web Intelligence and Smart Sensing (IWWISS), pp. 1-6, 2014/09/01, Saint Etienne, France.
6. (査読有) 戸田 隆道, 田中 翔太, 林田 興祐, 井上 創造, 上田 修功, "ラベルの時刻ずれに対応した携帯センサ行動認識手法", マルチメディア, 分散, 協調とモバイル (DICOMO2014) シンポジウム, pp. 394-400, 2014/07/09, Niigata.
7. (査読有) 右田 尚人, 服部 祐一, 田中 翔太, 井上 創造, "動画像と加速度データを用いた行動類似度評価システム MimicMotion の開発と評価 (Development and Evaluation of Behavioral

Similarity Evaluation System MimicMotion Using Acceleration Data and Video)", マルチメディア, 分散, 協調とモバイル (DICOMO2013) シンポジウム, pp. 1207-1216, 2013/07/10, Tokachi, Hokkaido.

【学会発表】(9 件)

1. 松木 萌, 谷口 敦, 井上 創造, "文書のベクトル表現を用いて学習者が作成した問題を分類する試み", SOFT九州支部学術講演会, pp. 25-26, 2015/12/12, Kumamoto.
2. 松木 萌, 谷口 敦, 井上 創造, "問題作成システムにおける問題自動分類に向けて (Towards Automatic Classification of Quizzes for User-generated Test Systems)", SOFT九州支部夏季ワークショップ, poster, 2015/08/27, Kumamoto.
3. 井上 創造, "(Invited) Pervasive Sensing for Nursing and Smart Energy Applications", International Conference on Informatics, Electronics & Vision (ICIEV), 2015/06/16, Kitakyushu, Japan.
4. 井上 創造, "センサビッグデータを用いた看護行動/節電行動理解", パーティクルフィルタ研究会 発足十周年記念イベント, 2015/04/24, Kitakyushu.
5. 谷口 敦, Ghada Farouk Naiem, 井上 創造, "ユーザに作問を任せるテストの自動評点に向けて", 日本知能情報ファジィ学会九州支部学術講演会予稿集, pp. 81-82, 2015/03/09, 北九州.
6. 潘 新程, 峯崎 智裕, 磯田 達也, 田中 翔太, 内野 百里, 井上 創造, "生活行動と消費電力の関係を調べるためのタブレット端末センシングおよび行動入力シ

ステム”, 日本知能情報ファジィ学会九州支部学術講演会予稿集, pp. 99-100, 2015/03/09, 北九州.

7. 戸田 隆道, 井上 創造, 上田 修功, ”ラベルの時刻ずれに対応した携帯センサ行動認識におけるクラスタリングによる特徴量境界の明確化 (Feature Clustering for Mobile Activity Recognition with Time-lagged Labels)”, 情報処理学会 第160回ヒューマンコンピュータインタラクション・第44回ユビキタスコンピューティング合同研究発表会, pp. 1-7, 2014/10/14, 沖縄.
8. 井上 創造, ”(Invited) Mobile Activity Recognition and Large-scale Healthcare Sensing”, International Symposium on Advanced Intelligent Systems, 2013/11/13, Daejeon, Korea.
9. 井上 創造, 中島 直樹, ”サイバーフィジカルヘルスケア”, 電子情報通信学会ソサイエティ大会「CPSを支える / CPSが変えるインターネットアーキテクチャ」, 2013/09/19, Fukuoka.

【図書】(0 件)

【産業財産権】(0 件)

【その他】

ホームページ: <http://atquiz.sozolab.jp/>

6. 研究組織

(1) 研究代表者

井上 創造 (INOUE, Sozo)
九州工業大学・大学院工学研究院・准教授
研究者番号 : 90346825

(2) 研究分担者

中尾 基 (NAKAO, Motoi)
九州工業大学・大学院工学研究院・教授
研究者番号 : 70336816