

指標探索のための人間と機械の建設的相互作用 Constructive Interaction between Humans and Machines in Search of Measures to Capture Collaborative Learning

中山 隆弘[†], 白水 始[†]
Takahiro Nakayama, Hajime Shirouzu

[†] 東京大学CoREF
CoREF, The University of Tokyo
info@coref.u-tokyo.ac.jp

Abstract

協調学習の分析ツールは教員が検索した表現をハイライトする「ミラーシステム」と特定表現を自動分析する「ガイドシステム」に大別できる。教員のエンパワーにはその両方が必要だろう。ガイドを参考に教員が検索を行い、結果を後人のガイドにフィードバックする体制を作ることができれば、人間と機械の相互作用をより建設的にできるのではないかと。その可能性を検討するため、まずは本OSの人手による分類が機械化できないかを検証した。

Keywords — Constructive Interaction, Learning Models, Quantitative Analysis, Mirroring System

1. はじめに

本稿では、イントロの白水ら[1]に記したとおり、質的評価の名のもとに研究者自身がやっていることの仕様をはっきりさせることと、研究者や教員の質的評価を蓄積し将来的に分析の自動化や機械化をどう図っていくかという二点を検討する。

協調学習の分析ツールを Soller ら[2]に従って、「ミラーシステム」と「ガイドシステム」に分けるとすれば、益川ら[3]の分析は対話の質が一定程度自動的に可視化されるガイドシステム、齊藤ら[4]の分析は教員が対話に見とりたい言葉や表現（指標）を検索するとハイライトするミラーシステムへと展開するものと大別できる。やりたいことは教員のエンパワーメントだと考えると、教員の検索・分析の自由裁量と研究者が質保証した検索ガイドはどちらも必要であり、どちらも単体で十分ということはない。研究者やシステムによる検索・分析の例示を（正解ではなく例示と取って）参考に、教員が本人にとって意味のある自由な検索を行って、その検索結果を後人の例示へとフィードバックするような相互強化体制、あるいは実践を行った教員の分析履歴を取って、後に同じ、あるいは同種の実践を行う教員に対してその履歴を直接、あるいは抽象化

してガイドとして活用できるような、システムを介した螺旋的協力体制を作ることができれば、人間と機械の相互作用はより建設的なものになるのではないかと。

そこで下記では、益川ら[3]と齊藤ら[4]の報告に含まれた分析や、それらの分析の前提となる基礎分析としてどのような工程が必要か、その機械的な手順の書き出しや、実際に既存分析システムを使った分析を行うことで、上記の可能性を検討する。

2. データと分析概要

本稿で、主に評価の対象としたデータ（機械によるテストデータ）は、堀[5]の授業実践（中学校2年生理科「運動の仕組」の授業）2事例である。そのための訓練データ（機械による適応データ）には、東京大学CoREFと自治体学校等との連携による「協調学習の授業づくりプロジェクト」の他の授業実践や益川ら[6]を用いた。

以下、益川ら[3]の「チーム構築調整発話」の自動分類のやり方、及びその実施結果をまず検討する。チーム構築調整発話とは、協調学習中の内容の学びとは関係ない、作業の進め方の相談や指示に関する「マネジメントトーク」のことであり、この多さは、発話が活発でも内容についての児童生徒の学びが進まない要因としてよく知られている。益川らはこの発話を同定し、それ以外の「コンテンツトーク」を分析するという論文展開をしている。

もう一つは、齊藤ら[4]の「問いの共同探究」の自動分析のやり方である。齊藤らの分析では、「問い」自体の同定やその仲間による共同探究を、文脈を読み込みながら行っているが、「共同」探究であるということは、発話前後の話者関係でそれらを機械的に分析し得る可能性を示唆する。それがどう可能かを検討する。

以下、分析方法と結果を合わせて報告する。

3. 分析方法と結果

3.1 「チーム構築調整」の自動抽出

(1) 分析方法

過去の CoREF の実践で収集された知識構成型ジグソー法の対話データ (88 授業, 1271 グループ, 534,951 行, 2,389,909 単語) から機械学習 word2vec[7]を用いて, ジグソー対話における単語の多次元ベクトル値を作るモデルを作成した. このモデルを利用し, 各行のテキストを Mecab で分かち書きし, 各行に含まれている単語のベクトル平均値出し, 文のベクトル値とした. この文のベクトル値を説明変数とし, 教師あり学習の分類器, サポートベクターマシン (SVM: [8]) とランダムフォレスト (RF: [9]) を用いて, チーム構築調整とそれ以外の二項分類を行った.

教師データとして, 益川ら[6]や上記 CoREF 授業について第二著者が「チーム構築調整」かどうかを判断した分類結果を用いた. 教師データ 2037 行の内訳は, チーム構築調整が 498 行, それ以外が 1529 行であった.

比較対象として, 機械学習を用いない, 文にある単語の部分集合で包含関係を作り, その有無で分類する単純パターンマッチ (Subset) も行った.

対象は, 運動の仕組み「2014 授業」と「2016 授業」の対話 3799 行で, 正解分類結果は, 本 OS の益川ら[3]の分類結果を利用した. その内訳は, チーム構築調整が 531 行, それ以外が 3268 行であった (益川ら[3]が記すとおり, 元々チーム構築調整の割合が低い授業であった).

(2) 分析結果

分類器が「チーム構築調整」に分類した結果は, 表 1 の通りである. 左から「チーム構築調整」に分類した行数, 精度 (precision): 正と予測したデータのうち, 実際に正であるものの割合, 再現率 (recall): 実際に正であるもののうち正であると予測されたものの割合, f1-score: 精度と再現率の調和平均である. 分類結果の例は, 後続の頁に掲載した表 1-1 及び 1-2 に示した. なお, チーム構築調整が「1」, その他が「0」である.

表 1 に見るように, トレーニングとテストデータとが, 全く違うコンテンツの授業であっても, 分類器での分類は可能であった.

表 1 「チーム構築調整」発話の分類器結果

	分類行数	精度 precision	再現率 recall	f1- score
Subset	281	0.25	0.13	0.17
SVM	1427	0.23	0.62	0.34
RF	518	0.22	0.21	0.22

しかし, 決して高い精度・再現率ではなかった. 単語の単純パターンマッチである Subset は教師データそのもののパターンに強く縛られてしまい, 再現率が 0.13 であった. 対して, SVM, RF は Subset と比較すれば高い再現率であった. SVM の再現率は 0.62 であるが, チーム構築調整に多くを分類しており, 精度が 0.2 程度であった. RF は精度・再現率共に 0.22 程度であった.

今回は 1 行ずつを分類器に掛けたが, データが 1 ブレスで改行されており非常に短い行もあった. 本来分析者が人手で分類するときは, 短い文は前後の流れを参考に分類しており, 同じ構成の文でも分類結果が変わっていた (「はい」とか「で」とか). このように分析者が複数行を見て分類しているのであれば, 分類器の説明変数に複数行を追加すれば, より精度が上がる可能性がある.

さらに, 自然な対話をそのまま機械学習に利用したが, より精度を上げるには, 口語体に対応した辞書や表記揺れや方言への対応などが必要と考えられる.

ただし, チーム構築調整, すなわちマネジメントトークで使われる単語は, 前後関係があまりなく (表 1-1, 1-2) 突如として出てくることが多く, そうすると前後を見る “word2vec” はベクトル化する方法としては不適切である可能性もあり, 今後の検討を要する.

3.2 「問いと共同探求」の自動抽出

まず齊藤論文[4]の分析方法を囲いに再掲する.

以下が問いの同定方法である.

手順 1) 対話中の「問」の同定
まず, 各班の対話の内容と流れを確認し, 班員から自発した「問」と判別できる発話を同定した. 分析は本稿の筆者を含む 2 名で行い, 疑問, 質問, 問いかけ, 確認等, 日本語でテキスト化したときに「?」が付く発話を「問」と同定した.

次が問いの共同探究についてである。

手順 2)「問」の後に生まれた学習活動のコーディング

次に、「問」の後に生まれた学習活動に注目し、班員から自発した「問」が共同的に探究されたかを判別した。「問」後の学習活動のコーディングは、「大カテゴリ」と「小カテゴリ」の2段階で行った。

「大カテゴリ」は、「問」に対し、班の他のメンバーから何らかの反応があったかに着目し、「問」が「共有」されたかどうかを判別するためのカテゴリである。

「問」の後に、解の提案、疑問の言いなおし、問い返しなど、班員から自発した「問」に対する発話であることが確認できる発話があった場合に、問が「共有」されたと判断した。対して「問」の後に、「問」に対する発話以外の関係ない発話のみしか見られない場合は「非共有」とした。更に、発話の内容ができないなど、「問」に対する発話であるかないかが判別できない場合は「同定できず」とした。

「小カテゴリ」は、「大カテゴリ」で確認された「反応」の内容に着目し、「共有」された問が共同的に探究されたかを判別するためのカテゴリである。「問」の後、他のメンバーから一度応答があり、その後はその「問」の解を探究する発言や動作が確認できない場合を「応答完結」、一度の応答で完結せずに、問の提出後、問の提出者を含む複数のメンバーが、解を探究する発言や動作を行っていることを確認できる場合を「共同探究」とした。

囲いに見るように、問いから次に他者がどういった振る舞いや発話をしたかを質的に解釈して分類している。この分析を機械に再現させるには、まず問いの同定が必要である。問いの同定には、過去の対話データに疑問符をつけ、それを教師データとして、先の機械学習の分類器を使って分類することが考えられる。次に、問いから次の他者の振る舞いへの移行については、表 2 のルールセットを用いれば分類可能だと考える

もしこの機械的分類で取り逃す人手の分類があるとすれば、このルールが単純に過ぎるということになる。逆にそうであれば、人がこれ以上の判断をしているということになる。セッション当日までに機械による分析を行い、結果を報告する。

4. 結論

将来の対話のビッグデータに対する機械的分析に備え、今回、試験的な分析を行った。機械学習分類器の利点としては、当然、教師データが用意できれば機械が何らかの結果を出すことができる。その一方で、word2vec などディープラーニング系の分類器では、結果が予想外なものだったときに、なぜそうなったのかを解釈できず、引いてはデータの振り返りや授業における学びに対する気づきに繋がらない欠点がある。

ルールベース分類器の利点は、ルールを書き出す労力は要るが、ルールができれば結果が自明であることである。結果が意図に反したときも、ルールの不備か想定外のデータが出力されるために、データに対する新たな気づきに繋がりがやすい。対話において、ルール化できない例外ケースが増え続けるかどうかを今後検討することで、質的評価の仕様をはっきりさせたい。

今後の分析においては、少数ケースは実際に読んで目視・意味判断、中程度のケースはルール化して分類、大量ケースは「ルールベースの結果+例外ケース」を教師データにして機械学習分類器で処理するという組み合わせが考えられる。

中程度のケースからビッグデータに展開する際、ルール化した分析結果を教員が簡易に使える分析ツールに反映し、例えば問いの候補のハイライトや、マネジメントトークとコンテンツトーク、あるいはわかる／わからないの色分け、エキスパート資料とのマッチした単語、各時点での新規出現単語、児童生徒の使うホットワードのハイライトなど、ある程度の「ガイド」を設けてツールを使ってもらい、その使用履歴を「実用的なルール」と見なして反映し、分類器を賢くしていくという方法も考えられる。当日報告する。

表 2 共同探求の分類ルール

カテゴリ	定義（優先順位は上位から）
共有 共同探求	他者から、問いと同じ「名詞／動詞」がある
共有 応答完結	他者、問いへの応答「うん／はい／いいえ／そう」がある
非共有 自己探求	自分から、問いと同じ「名詞／動詞」がある
非共有 探求なし	上記以外

参考文献

- [1] 白水始・益川弘如 (2018) “協調学習の評価の刷新：指標を探す”，日本認知科学会第 35 回大会論文集。
- [2] Soller, A., Martinez, A., Jermann, P. & Muehlenbrock, M. (2005). “From mirroring to guiding: A review of state of the art technology for supporting collaborative learning.” *International Journal of Artificial Intelligence in Education (IJAIED)*, 15, 261-290.
- [3] 益川弘如・河崎美保 (2018) “学びのモデルに基づく指標を複数重ねることから見える対話の質”，日本認知科学会第 35 回大会論文集。
- [4] 齊藤萌木・飯窪真也・白水始 (2018) “理解深化につながる対話を見とる指標の提案：対話中の疑問を軸として”，日本認知科学会第 35 回大会論文集。
- [5] 堀公彦 (2018) “知識構成型ジグソー法の授業デザイン 中学校理科「運動の仕組み」”，日本認知科学会第 35 回大会論文集。
- [6] 益川弘如・河崎美保・白水始 (2016) “建設的相互作用経験の蓄積が協調的問題解決能力の育成につながるか”，*認知科学* 23(3), pp237-254.
- [7] <https://code.google.com/p/word2vec/>
- [8] Cortes, C. & Vapnik, V. (1995). “Support Vector Networks.” *Machine Learning*, 20, 273-297.
- [9] Breiman, L. (2001). “Random forests.” *Machine Learning*, 45, 5-23.

表 1-1 チーム構築調整自動抽出例（運動の仕組み 2016 開始直後）

話者 A	話者 B	話者 C	分析者	SVM	RF	Sub
	はい、どうぞ。		1	1	1	1
誰から。			1	1	0	0
		どこから。	1	1	0	0
	え？		1	0	0	0
	ここから。		1	0	0	0
誰から？			1	1	0	0
		そう意味じゃなくて。	1	0	0	0
	え？		1	0	0	0
誰かって。			1	1	0	1
	ここからや。		1	0	0	0
		え？ どこから。	1	1	1	1
だから。			1	0	0	0
	だから、		1	0	0	0
	ここにおった人から。		1	1	0	0
		ああ、そういうこと。	1	0	0	0
ああ。			1	0	0	0
雑だねえ。			1	1	0	0
ヒトの体には、神経細胞が			0	0	0	0
網			0	0	1	0
目 の よ う に 張 り 巡 ら さ れ て い て、			0	0	0	0

表 1-2 チーム構築調整自動抽出例（運動の仕組み 2016 資料説明の切り替え時）

話者 A	話者 B	話者 C	分析者	SVM	RF	Sub
		骨を動かしています.	0	1	0	0
		はい.	1	1	1	1
	はい.		1	1	1	1
	終わり?		1	1	0	0
		うん.	1	0	0	0
	はい.		1	1	1	1
	どうぞ.		1	1	0	1
あ?			1	0	0	0
	どうぞ.		1	1	0	1
何を.			1	1	0	0
	まとめてください.		1	1	1	1
無理です.			1	1	0	0
	全てをどうぞ.		1	1	0	1
	まとめてください.		1	1	1	1
あなたがまとめてください.			1	1	1	1
	だけんってどういうこと?		0	0	0	0
	ていうことやろ.		0	0	0	0
		刺激を受けて感覚器官で、感覚神経、せきずい、脳、せきずい、運動神経、筋肉を	0	0	0	0
		使ってこんなふうに動かす.	0	1	0	0
	ほんで?		0	1	0	0
		いいよ.	1	1	0	0
	じゃあ、言って. はい. どうぞ.		1	1	1	1
	何を書けばいいん?		1	1	0	0
	全部?		1	0	0	0
		こうやって.	0	1	0	1
	え、もうね、何かまとめてよ.		1	1	1	0
		これはまとめられてんな.	0	1	1	0
	でも、それって何か神経とかだけやん.		0	1	0	1
		筋肉を	0	0	0	0
		動かすのは、	0	0	0	0