アクティブラーニング型授業の分析のための深層学習
Deep Learning for Analysis of Active Learning Classes

阪口 真也人[*1], 豊浦 正広[*1], 赤穂 大樹[*1], 茅 暁陽[*1], 西口 敏司[*2], 塙 雅典[*1], 村上 正行[*3]
Mayato SAKAGUCHI[*1], Masahiro TOYOURA[*1], Daiki AKOH[*1], Xiaoyang MAO[*1],
Satoshi NISHIGUCHI[*2], Masanori HANAWA[*1], Masayuki MURAKAMI[*3]
[*1] 山梨大学 [*2] 大阪工業大学 [*3] 京都外国語大学
[*1] University of Yamanashi [*2] Osaka Institute of Technology [*3]Kyoto University of Foreign Studies
Email: g16tk007@yamanashi.ac.jp

あらまし：我々はこれまでに，機械学習による授業状況推定を行い，授業全体の状況を俯瞰できる画像を合成して可視化することによって，アクティブラーニング型授業の振り返りの支援を行ってきた．本研究ではさらに，メディア認識の課題に対して大きな成果を挙げてきた**深層学習を用いて状況推定精度の向上**を実現した．深層学習に合わせて授業映像データを整形し，授業状況解析に適したネットワーク構造を設計することで，従来よりも高い推定精度を得た．
キーワード：アクティブラーニング，可視化，授業改善，映像解析，深層学習，ニューラルネットワーク

1. 背景

アクティブラーニング型授業は講師が一方的に知識伝達を行わず，学生が能動的に学習することにより，認知的，倫理的，社会的能力，教養，知識，経験を含めた汎用的能力の育成を図ることを目的としている．課題研究や PBL(Problem Based Learning, Project Based Learning)，ディスカッション，プレゼンテーションなどが積極的に取り入れられる．アクティブラーニング型授業の実施に関しては，経験的または理論的な知識が共有されてきてはいるものの，多くの講師にとってはその授業スタイルが確立してはいない．また，議題の提案や学生の誘導といった従来形式の授業では必要なかった指導法が講師に要求されるが，講師自身が授業の成否を客観的に認識することは難しい．授業の振り返り方法としては，撮影された授業映像を見ることが有効であるとされているが，長時間の動画視聴の負担を強いることになり，授業へのフィードバックまでのサイクルを短期間で回すのは難しい．これらのことからアクティブラーニング型授業の客観的で即時的な分析の必要性は高い．

他方で，映像認識の技術としては，人工知能を用いた人間の行動解析や状況推定が，近年になって注目を集めている．中でも生物の脳を模した神経ネットワークを最適化することによって学習を行う**深層学習**では，認識に必要な特徴を自動で抽出でき，多くの分野に渡って，従来の認識精度を大きく越える結果が得られてきている．特に生物が認識を得意とする言語・音声・画像などのメディアに関わる問題対しては，深層学習との親和性が高いことがわかっている．

本研究では**深層学習をアクティブラーニング型授業の解析に適用する**ことで，授業の高精度な分析結果を得ることを目指す．

2. 既存手法と問題点

我々はこれまでに機械学習の一種である Bag-of-Words を用いて，アクティブラーニング型授業の状況推定を行ってきた[1]．授業映像から得られる音声や動作の特徴から，移動，各自作業，解説，発表，グループワークの 5 カテゴリの推定を可能にした．さらに，授業映像ごとの授業状況の推移や各カテゴリの所要時間を可視化し，アクティブラーニング型授業の分析の手掛かりとすることを提案した．既存手法の推定精度は約 73%を示しているが，分析に十分な推定精度であるとは言えない．講師が自身の授業の客観的な振り返りをするためには，さらに推定精度を高める必要があった．

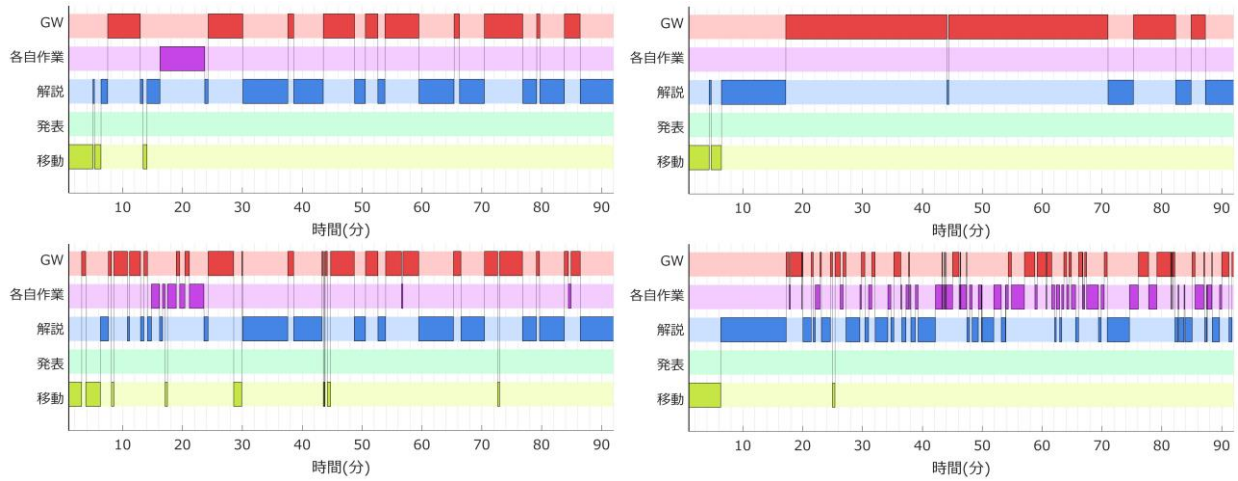
3. 深層学習による授業状況認識

本研究では，授業状況の推定精度を高めるために，深層学習を実現する畳み込みニューラルネットワーク(CNN) ，時系列データの学習に利用される再帰型ニューラルネットワーク(RNN) を利用する[2]．CNN では画像中で特定の周波数や方向を持つ領域に強く反応するように畳み込み層が設定される．RNN では，系列データをよく表現するために LSTM (Long short-term memory) 層が設定される．両者を含むネットワークは C-RNN と呼ばれる．

以降，どのようにしてこれらのネットワークモデルが利用可能な形式に授業映像データを**整形**し，さらに，どのような**構造**を取れば高精度な認識が実現されるかについて，説明する．

3.1 授業映像データセットの整形

アクティブラーニング型授業の状況を認識するためには，ある 1 時刻だけでは状況は把握することはできず，その前後の時刻の状況を合わせて考える必要がある．そこで，入力信号として任意時刻 n の画像・音声の特徴に加え，前後時刻の特徴と合わせて 1 入力とするデータセットを構築する．既存手法と同様に 1 秒ごとのフレーム間差分，環境音，マイク音の 3 種類の特徴を，前後 2n+1 秒分まとめて 1 入力とする．教師信号となる出力は，認識対象時刻で



(a)推定結果 I （講師 A，推定率 85.5%）　(b) 推定結果 II （講師 B，推定率 50.3%）

図 1　CNN を用いた推定結果例．上段は真値を表す．

ある中央の n 秒目のものとし，人手によって正解を与える．アクティブラーニング型授業映像を視聴することで各時刻に何が行われているかを決定した．正解は移動，発表，解説，各自作業，グループワークの 5 種類のいずれかとする．

3.2 ネットワークの構築

深層学習のフレームワークである Caffe に RNN を合わせて実装した．CNN で用いる畳み込みフィルタでは，画像・音声の情報が互いに混じることがないように，時間方向にのみ畳み込みを行うフィルタを設定した．

ネットワークに含む層の内容や数を用意した学習用・評価用データで高い精度を出すようにチューニングを繰り返した．最終的に得られたネットワーク構成の CNN, RNN, C-RNN のネットワークを表 1 に示す．

表 1 最終的に得たネットワーク構成

CNN		RNN		C-RNN	
畳み込み	16×1×15			畳み込み	16×1×15
プーリング	1×6			プーリング	1×6
畳み込み	32×1×5			畳み込み	32×1×15
		LSTM	15	LSTM	15
全結合	125	全結合	125	全結合	125
全結合	25	全結合	25	全結合	25
全結合	5	全結合	5	全結合	5
Softmax		Softmax		Softmax	

4.　評価

各ネットワークにおける推定精度を Leave-one-out 交差検証(LOOCV)で評価し，比較を行った．標本には講師二人のアクティブラーニング型授業映像を使用した．各講師とも認識の対象とした授業数は 10 であり，授業時間は約 90 分である．表 2 に各ネットワークによる推定結果を示す．

CNN を用いたネットワークが最も高い推定率を示した．また RNN と C-RNN においても，既存手法を上回る推定精度が得られた．CNN における推定結果の一部を図 1 に示す．縦軸に分類された 5 種類のカテゴリ，横軸に授業の経過時間を示すタイムラインとして可視化したグラフである．図の上段は正解(真値)を示し，図 1(a)の授業は推定率 85.5%，図 1(b)は 50.3%である．最小推定率となった図 1(b)の授業では，グループワークと各自作業のカテゴリで誤分類が多く発生している．これは講師 B が各自作業中にコメントを入れる場面が多く，グループワークとの区別がつきにくかったためであると考える．

表 2 推定結果

手法		講師A	講師B	2講師平均
既存手法	Bag-of-Words	77.60%	67.80%	72.70%
提案手法	C-RNN	80.30%	71.70%	76.00%
	CNN	80.90%	72.60%	76.80%
	RNN	79.40%	64.80%	72.10%

5.　まとめ

アクティブラーニング型授業に深層学習による分析のために，データの整形方法とネットワークの具体的な構築方法を示した．CNN を用いる手法において既存手法を上回る推定率を得ることができた．入力信号により多くの情報を含めることで，さらに推定精度を向上させられると考えている．

謝辞

本研究の一部は，JSPS 科研費 JP16K12784, JP26282062, JP15K00499 の助成を受けたものである．

参考文献

(1) M. Toyoura, M. Sakaguchi, X. Mao, M. Hanawa, Masayuki Murakami, “Visualizing the Lesson Process in Active Learning Classes,” IEEE FIE, 2016.
(2) J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, T. Darrell, “Long-term recurrent convolutional networks for visual recognition and description, ” Proc. CVPR, pp. 2625 -2634, 2015.