階層型のラベル付けマイクロタスクにおける能動学習戦略の比較

鵜尾　厚佑† 　小林　正樹†† 　松原　正樹††† 　馬場　雪乃†††† 　森嶋　厚行†††

† 筑波大学 情報学群情報メディア創成学類　〒 305-8550 茨城県つくば市春日 1-2
†† 筑波大学 図書館情報メディア研究科　〒 305-8550 茨城県つくば市春日 1-2
††† 筑波大学 図書館情報メディア系　〒 305-8550 茨城県つくば市春日 1-2
†††† 筑波大学 システム情報系　〒 305-8573 茨城県つくば市天王台 1-1-1
E-mail: †kousuke.uo.2019b@mlab.info, ††makky@klis.tsukuba.ac.jp, †††{masaki,mori}@slis.tsukuba.ac.jp,
††††baba@cs.tsukuba.ac.jp

あらまし　本論文では，階層型のラベル付けマイクロタスクを対象とした能動学習の戦略を複数提示し，これらを比較する．階層型のラベル付けマイクロタスクの典型的な例は，航空写真のようなサイズの大きな画像の，メッシュ状に区切られた部分それぞれに対してラベルを付けるためのタスクの集合である．それらのタスクは，大きな画像から始まり，タスクを行う人は, それにラベルを与えるか，小さな画像に分割していくかを選ぶことができる．本論文では，これを対象としたタスクでいくつかの能動学習の戦略を比較した実験の結果を示す．実験結果は，戦略の違いによって分類器の早い段階の性能に影響を与えることを示唆するものであった．

キーワード　能動学習，ユーザーインターフェース，階層型のラベル付けマイクロタスク

1　はじめに

マイクロタスクとは，短い時間で簡単に行える作業のことであり，多くのクラウドソーシングアプリケーションの基本的な構成要素の一つである．その中でも，機械学習の訓練データを作成するために，ある画像や文章などに対して人が正解ラベルを付けるものを，ラベル付けマイクロタスクと呼ぶ．

本論文では，階層型のラベル付けマイクロタスクを対象とした能動学習の戦略を比較する．

階層型のラベル付けマイクロタスクとは，すべてのデータ要素の集合から始まり，徐々に小さな部分のデータ要素へと対象を移しながらラベル付けをしていくという，マイクロタスクの集合である．例えば，そのデータが大きな航空写真で，その写真から均等な大きさに分割された複数枚の小さな画像を最終的に訓練データとして手に入れたい場合，その写真を段々と小さな部分へと分割していきながらラベル付けをすることになる．(図 1).

階層型のラベル付けマイクロタスクは，ラベルの分布や出現確率が偏っているような状況で，効率的にラベルを手に入れることを狙いとしている．例えば，大きな航空写真から，均等な大きさに分割された複数枚の小さな画像を最終的に訓練データとして手に入れたい場合をもう一度想定してほしい．その航空写真は，地方で撮影されたものであり，付けるラベルは「建物を含んでいる」と「建物がない」の 2 つであるとする．この写真のラベルの分布はとても偏っている．なぜなら都市部とは異なり，建物のない土地の割合が多いため，画像のほとんどの部分は建物を含んでいないからである．さらにラベルの出現確率も偏っている．なぜなら，建物はしばしばほかの建物の近くに出現し，建物がない範囲が広く存在するからである．このような状況では，階層型のラベル付けマイクロタスクが明らかに効果的である．なぜなら，図 1 で示しているように，多くの小さな画像に対して一度にラベルを与えることができるからである．

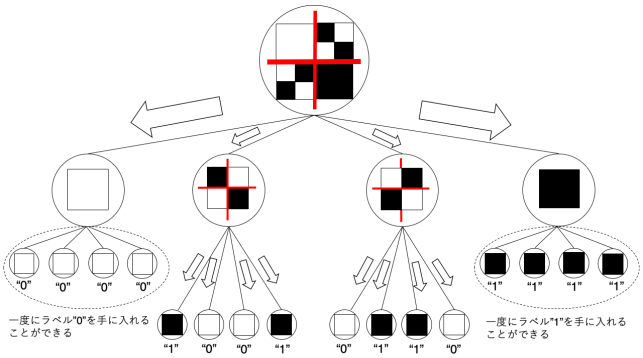


図 1　階層型のラベル付けマイクロタスクの例

しかし，階層型のラベル付けマイクロタスクに能動学習を適用する場合，一般的な能動学習の設定とは異なっているという問題がある．

能動学習とは，機械学習の研究分野の一つである．そこでは，機械学習モデルが，ラベル無しデータ集合から次にラベル付けをするべき適切なデータを選択し，人間にラベル付けマイクロタスクとして提示する．そのようにして，人間がラベル付けして得たトレーニングデータのみを機械学習モデルが学習することによって，少ないトレーニングデータで高い正解率を算出することを目的としている．

一般的な能動学習では，ラベルなしデータ集合に，MNIST の手書き数字の画像データセット [1] のような，ラベルが一意に決定するようなデータ要素をおき，それに対して適切なデータを選択するアルゴリズムを適用する．しかし，階層型のラベル付けマイクロタスクを用いた能動学習の場合，ラベルなしデー

タ集合には，複数のデータ要素が統合されたデータが存在し，タスクのインターフェースが異なるため，一般的な能動学習のアルゴリズムが適用できない．

そこで本論文では，階層型のラベル付けマイクロタスクを用いた能動学習の複数の戦略を比較した実験の結果を示す．

本論文の貢献は，以下の通りである．(1) 階層型ラベル付けマイクロタスクを対象とした，能動学習のいくつかの戦略を比較した．(2) 実験の結果，戦略の違いによって分類器の早い段階の性能に影響を与えることを示唆した．

本論文の構成は次のとおりである．第 2 節では，関連研究ついて説明する．第 3 節では，階層型のラベル付けマイクロタスクを用いた能動学習の定義と，いくつかの戦略を説明する．第 4 節では，実験の結果を示す． 第 5 節では，結論を述べる．

2 関 連 研 究

能動学習で行われるラベル付けマイクロタスクを，クラウドソーシングを用いて行った研究がいくつか存在する．Florian, Christian，Hinrich (2011) [2] は，固有表現抽出や，感情分析のクラウドソーシングタスクを用いて能動学習を行い，その有効性を示した．Liyue，Gita，Rahul (2011) [3] は，クラウドソーシングによって得られた不正確なデータを用いたロバストな能動学習の手法を提案した．Vamshi (2011) [4] は，機械翻訳システムを作るための，クラウドソーシングタスクを用いた能動学習の手法を提案した．

これらの研究で使用されているデータセットは，ラベルなしデータ集合にラベルが一意に決定するようなデータをおいているため，本論文の設定とは異なっている．また，クラウドソーシングによって手に入るデータ品質の問題も扱っている．データ品質に関しては本論文の扱うところではないが，これらの能動学習をクラウドソーシングで行う設定で，データ品質を高める研究の背景にあるアイデアは，階層型のラベル付けマイクロタスクに対して適用されうるものであり，今後の課題となるだろう．

また，ディープラーニングを用いた能動学習の手法もいくつか提案されている．William，Tim，Andreas(2018) [5] らは，MNIST や CIFAR-10 といったデータセットに対する CNN のアンサンブル学習を用いた能動学習を提案した．Lin，Yizhe，Jianxu ら (2017) [6] は，生物医学画像に対して FCN を用いたアクティブラーニングを提案した．このような，FCN を用いたセマンティックセグメンテーションを我々の研究に対して適用することは，興味深い課題となるだろう．

3 定義と戦略

この章では，階層型のラベル付けマイクロタスクの定義と，そのタスクを用いた能動学習のいくつかの戦略を与える．

3.1 階層型のラベル付けマイクロタスク

［定義 1］（階層型のラベル付けマイクロタスク） D をデータ要素の集合とし，L をラベルの集合とする． ラベルは，D のデータ要素それぞれに与えられるものとする．このとき，階層型のラベル付けマイクロタスクとは，木 $t = (V, E)$ であり，その葉の集合 V は，D と一致する．

もし t が階層型のラベル付けマイクロタスクならば，すべての t の部分木もまた，階層型のラベル付けマイクロタスクであることに注意する．

図 2 は，階層型ラベル付けマイクロタスクのラベル付けの過程を示している．ラベル付けの過程で，t のタスク表現である Task(t) 生成される．これは，t のすべての葉を集約させた情報をワーカーに示す．例えば，葉がそれぞれ小さな画像だった場合，Task(t) は，それらを統合した画像となる．この時，Task(t) は，ワーカーに以下の選択肢のうち一つを選ばせる．

- 「全て $l \in L$ である」(t の葉全てにラベル l が与えられる場合)
- 「どれでもない」

もし，ワーカーが「全て $l \in L$ である」を選んだ場合，t の全ての葉にラベル l をつけて，ラベル付けの過程を終える．もし，ワーカーが「どれでもない」を選んだ場合，t の根の子が，根である部分木に t を分割し，それらの部分木を *available tasks* と呼ばれる集合に追加し，ラベル付けの過程を終える．はじめ，available tasks は t のみである．ラベル付けの過程が進んでいくにつれて，available tasks は t の部分木を含んでいく．それらもまた，階層型のラベル付けマイクロタスクである．ラベル付けは，available tasks に階層型のラベル付けマイクロタスクがなくなるまで，繰り返される．

3.2 能動学習戦略

能動学習では，次にラベルをつけるデータを選ばなければならない． 通常の能動学習の設定での典型的なアプローチは，機械学習モデルを最適に改善するもののひとつを選択することである．しかし，階層型のラベル付けマイクロタスクでは，ラベルなしデータ集合 (available tasks) に，データ要素が複数存在することがあるため，そのアプローチを直接適用することができない．

ここでは，available tasks からタスクを選ぶためのいくつかの戦略を列挙する．それらは，Uncertainty Sampling [7] に基づいたものである．Uncertainty Sampling とは，最も単純で一般的に使用されるデータを選択するアルゴリズムである．そ

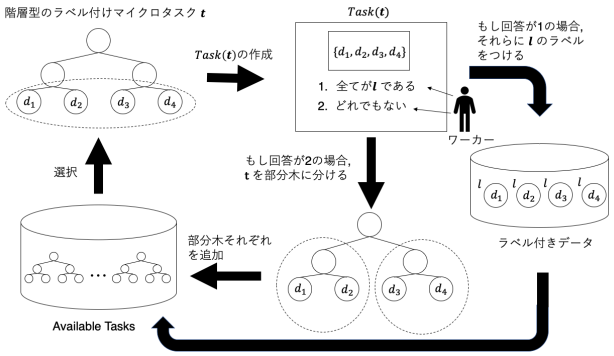


図 2 階層型のラベル付けマイクロタスクのラベル付けの過程

こでは，機械学習モデルが算出したラベルが最も不確実なデータを選択する．その不確実性値は $1 - \max(\{\text{prob}_l(d)|l \in L\})$ によって計算される．ただし $\text{prob}_l(d)$ は d に，ラベル l がつけられるときの確率を表す．

それぞれの戦略では，available tasks のそれぞれに対して，score(t) が計算される． それは ，新しいラベルが手に入った後，それまでで得られたラベルで機械学習モデルが学習されるたびに行われる．その後，score(t) が最大であるものの一つを選択し，次のタスクとする．

戦略は以下の通りである．

(1) Item-Average-Max この戦略では，以下の公式を用いて不確実性値の平均を計算し，それを score(t) とする．

$$\text{score}(t) = \sum_{d \in D} (1 - \max(\{\text{prob}_l(d)|l \in L\}))/|D|.$$

ここでの根底にある仮定は，可能な限り高い不確実性値を持つデータ要素をできるだけ多く取得することが望ましいということである．

(2) Item-Top-k-Average-Max これは Item-Average-Max の一般化であり，D のすべてのデータ項目の不確実性値の平均を計算しない． 代わりに，上位 k 個のデータ要素の不確実性値のみを考慮する．

$$\text{score}(t) = \sum_{d \in D_k} (1 - \max(\{\text{prob}_l(d)|l \in L\}))/|D_k|,$$

$$\text{where } D_k = \{d|\text{argTop}k_{d \in D}\ (1 - \max\{\text{prob}_l(d)|l \in L\}\}.$$

ここでの根底にある仮定は，不確実性値が高いが最大ではない多くのデータ要素を平均するよりも，不確実性の値がより高いデータ要素を平均する方が重要だということである．

(3) Label-Max この戦略では，次の式で計算されるラベルの累積不確実性値が最大になるよう score(t) を計算する．

$$\text{score}(t) = 1 - \max(\{\sum_{d \in D} \text{prob}_l(d)/|D|\ |l \in L\}).$$

ここでの根底にある仮定は，異なるラベルを持つデータ要素を取得するほうが良いということである．

(4) Random この戦略では，score(t) はランダムな数値となる．つまり，available tasks からランダムにタスクを選ぶということである．能動学習では通常，ラベル無しデータ集合の中からランダムにデータを選ぶよりも，何らかの戦略を使ってデータを選んだほうが少ないデータで高い正解率が得られるため，この戦略よりも上記の戦略のほうが性能がいいことが期待される．

4 実　　験

我々は，5 つの戦略 (Item-Average-Max，Item-Top-1-Average-Max，Item-Top-3-Average-Max，Label-Max，Random) を比較する実験を行った．この実験では，最初に既知のラベルを使用してデータセットを作成し，これら戦略を使用してシミュレーションを実行した．

4.1 データとタスク

我々は，国土地理院によって撮影された，西日本豪雨の航空写真を利用した [8] (図 3)．この航空写真には，雨により土地が浸水している場所が存在している．

はじめに，その航空写真を 2x2 の 4 枚の画像 (P_1, P_2, P_3, P_4) に分割した． その 4 枚の画像それぞれをさらに， 256 枚の画像 (16x16) に分割した．

次に，P_i の中の 256 枚の画像全てに人手によって正解ラベルをつけた．つけたラベルは,「浸水している」,「浸水していない」,「雲で隠れている」の 3 つである．(1)2 人が P_i の中の 256 枚の画像全てにラベルを付けた． (2)2 人のラベルが異なった部分については，そのラベルについて議論し，合意したものをつけた．議論の対象となった画像は，18 個であった．

最後に，それぞれの P_i に対して，階層型のラベル付けマイクロタスク $t_i(i \in [1, 2, 3, 4])$ を，図 4 のように, 完全四分木で, 256 個の画像の集合が葉 D となるように作成した．

まとめると，我々は 4 つのデータセット (P_1, P_2, P_3, P_4) を作成した．各データセットは，256 枚の, ラベルがついた画像となっている．また，4 つのデータセットから構築した，4 つの階層型のラベル付けマイクロタスク $t_i(i \in [1, 2, 3, 4])$ を作成した．

4.2 モ デ ル

我々は，機械学習アルゴリズムとして CNN と，ロジスティック回帰を使用して機械学習モデルを作成した．

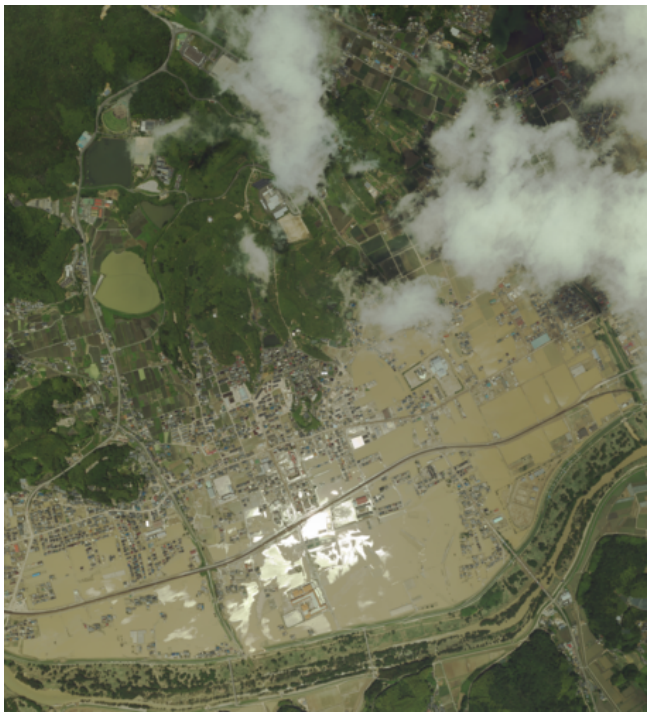


図 3　西日本豪雨の航空写真

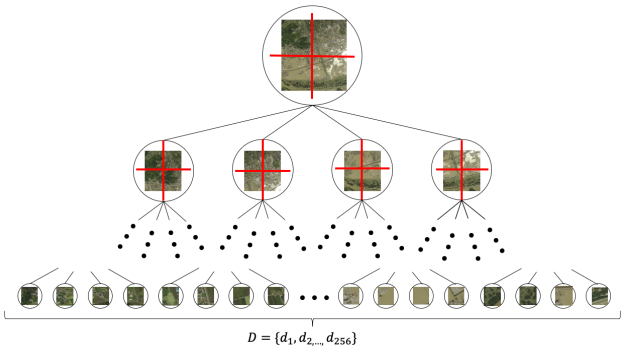


図 4　実験で使用した階層型のラベル付けマイクロタスク t_i

CNN モデルの構築には Keras [9] を使用した．CNN モデルは入力層，畳み込み層 1，プーリング層 1，ドロップアウト層，畳み込み層 2，プーリング層 2，全結合層，および出力層の順で構成される．各畳み込み層は，3x3 のカーネルと，ReLU 関数をアクティベーション関数として使用する． 1 番目と 2 番目の畳み込み層では，それぞれ 32 個と 64 個の特徴マップを出力する． 各プーリング層では，画像のサイズが半分に縮小されるようにマックスプーリングが実行される． ドロップアウト層では，プーリングレイヤー 1 からの出力の 25%をドロップアウトした． 出力層のアクティベーション関数として softmax を使用した．

ロジスティック回帰モデルの構築には，scikit-learn [10] を使用した．

4.3　実験の手順

ここでは，実際に行った実験の手順を説明する．

（1） 4 つのデータセットのうち，3 つのデータセット (例えば P_1，P_2，P_3) から作られた階層型のラベル付けマイクロタスク (例えば t_1，t_2，t_3) を available tasks に入れる．この 3 つのデータセットの画像は，訓練データとなる．

（2） 初期の訓練データとして，選ばれた 3 つのデータセットの中から各ラベル (「浸水している」,「浸水していない」,「雲で隠れている」) ごとに 1 つの画像を選び，それらを機械学習モデルに学習させる．

（3） available tasks の中のすべての階層型のラベル付けマイクロタスクに対して，機械学習モデルの予測結果を用いて $score(t)$ を計算する．その戦略は 3 章で説明した戦略のうち一つを用いる．

（4） その $score(t)$ が最大の階層型のラベル付けマイクロタスク t を選び，$Task(t)$ を発行する．その $Task(t)$ は，3.1 節で説明したとおり，t の葉を統合させた画像である．

（5） 3.1 節で説明したとおり，$Task(t)$ の回答結果によって，以下の (a),(b) いずれかに分岐する．本実験では．タスクの回答は正解ラベルによってシミュレーションされる．

（a） 「全て $l \in L$ である」が選ばれた場合，つまり $Task(t)$ の画像がすべて同じラベル l ならば，それらすべての画像にラベル l をつけたものを訓練データとして手に入れ，それを用いて機械学習モデルを再度学習させる．

（b） 「どれでもない」が選ばれた場合，t の根の子が，根である部分木に t を分割し，それらの部分木を available tasks に追加する．本実験では，t が完全四分木であるので，画像が 4 分割される．

（6） その時点の機械学習モデルで，4 つのデータセットのうちの残りの一つ (例えば P_4) の 256 枚の画像をテストデータとして予測し，その正解率を算出する．

（7） available tasks にタスクがなくなるまで，3〜6 を繰り返す．

以上のワークフローをまとめたものが，図 5 である．この図では簡単のために，階層型のラベル付けマイクロタスクを木の表現ではなくその葉の画像を統合したものとして表示してある．

実験は，P_1，P_2，P_3，P_4 のうちの 3 つを訓練データ，1 つをテストデータとした全ての組み合わせ 4 回を，戦略 5 つとモデル 2 つを変えながらおこなったので，計 40 回実験を行ったことになる．

4.4　結　　果

図 6，図 7 はそれぞれ，モデルとして CNN を用いてシミュレーションを行った場合と，ロジスティック回帰によってシミュレーションを行った場合の実験結果である．各図では，実行されるタスクの数が増えるにつれて正解率がどのように変化していったかを示す．各図の [1]〜[4] では，トレーニングデータとテストデータの 4 つ組み合わせ全ての結果を表示している．各線は，それぞれの戦略の結果を，10 区間移動平均をとって表示してある．

これらの結果は，戦略の違いによって特に初期段階 (最初の 100 タスクほど) の機械学習モデルの性能に影響がうまれ，最終的に結果は収束していくことを示唆している．例えば，図 6[2] の，50 タスクを行った時点での結果を見ると，Rondom や，Label-Max では，45%ほどの正解率であるのに対し，それ以外の戦略では 80%ほどの正解率であり，差が見られるのがわかる．また，250 タスクを行った時点での結果を見ると，結果は両者とも 90%ほどに収束している．上記の例に限らず，戦略の違いによって，初期段階に正解率の差が見られ，最終的に結果が収束する傾向は，程度が違えど全ての図に見られる．このように，戦略の違いによって，初期段階の機械学習モデルの性能の違いが生まれると把握できたことで，それが重要なアプリケー

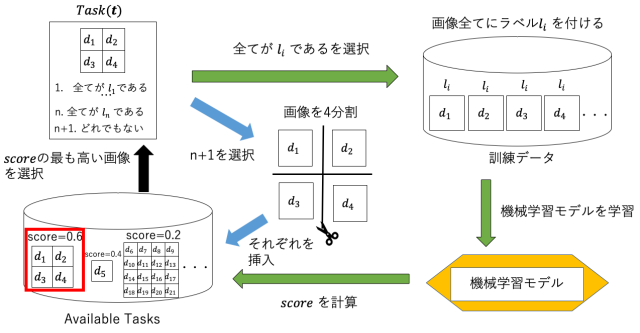


図 5　実験のワークフロー

ション（例えば，航空写真から自然災害の状況を把握するアプリケーションなど）に本研究を適用する場合に，どの戦略を選べばよいかが重要な問題になることが分かった．しかし，本実験ではどの戦略がもっとも良い性能を発揮するかはどうかまでは分からなかった．各図を見ればわかるように，トレーニングデータやテストデータの違い，また，モデルの違いによって，早い段階で良い正解率を発揮する戦略は異なっている．また，ランダムにタスクを選んだ時よりもその他の戦略 (Item-Average-Max，Item-Top-1-Average-Max，Item-Top-3-Average-Max，Label-Max) を用いたほうが，常によい正解率が出せることが期待されたが，必ずしもそうではないことが分かった．我々は，最も良い戦略というのは訓練データのラベルの分布や，出現確率，また，機械学習モデルの性能が影響されると想定している．今後の課題としては，どのようなデータに対してどの戦略がもっとも良い性能を発揮するかを確かめるために，ラベルの分布や出現確率が違うような様々なデータに対して戦略を適用し，実験を行うことである．また，より良い性能を発揮できるような新たな戦略を考え，それを実際のデータに適用することである．

5 結　　論

本論文は，階層型のラベル付けマイクロタスクにおける能動学習戦略を比較した実験の結果を報告した． 実験の結果，戦略の違いによって，初期段階の結果に大きな影響を与えることを示唆した．今後の課題は，より詳細な分析を実施して，適切な戦略を選択できる方法を見つけることである．

謝　　辞

本研究の一部は JST CREST の支援 (Grant Number JPMJCR16E3) による．ここに謝意を示す．

文　　献

[1] Mnist handwritten digit database, yann lecun, corinna cortes and chris burges. http://yann.lecun.com/exdb/mnist/.

[2] Florian Laws, Christian Scheible, and Hinrich Schütze. Active learning with amazon mechanical turk. In *Proceedings of the conference on empirical methods in natural language processing*, pp. 1546–1556. Association for Computational Linguistics, 2011.

[3] Liyue Zhao, Gita Sukthankar, and Rahul Sukthankar. Robust active learning using crowdsourced annotations for activity recognition. In *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*, 2011.

[4] Vamshi Ambati. Active learning and crowdsourcing for machine translation in low resource scenarios. *2011*, 2011.

[5] William H. Beluch, Tim Genewein, Andreas Nürnberger, and Jan M. Köhler. The power of ensembles for active learning in image classification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

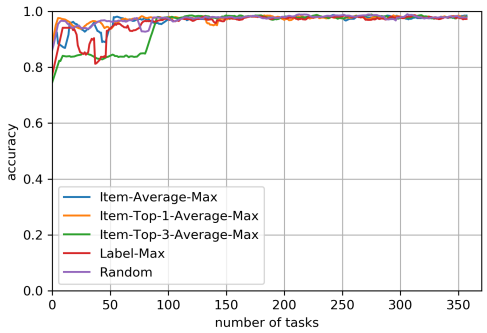
[6] Lin Yang, Yizhe Zhang, Jianxu Chen, Siyuan Zhang, and Danny Z Chen. Suggestive annotation: A deep active learning framework for biomedical image segmentation. In *International conference on medical image computing and computer-assisted intervention*, pp. 399–407. Springer, 2017.

[7] Burr Settles. Active learning literature survey. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 2009.

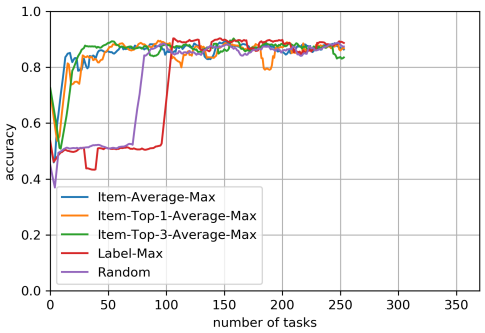
[8] Information about heavy rain in july 2018 | geo spatial information authority of japan. https://www.gsi.go.jp/BOUSAI/H30.taihuu7gou.html.

[9] Home - keras documentation. https://keras.io/.

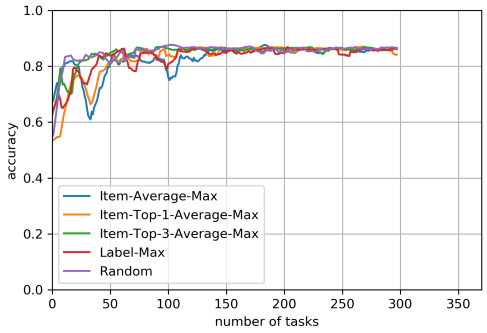
[10] scikit-learn: machine learning in python — scikit-learn 0.22.1 documentation. https://scikit-learn.org/stable/.



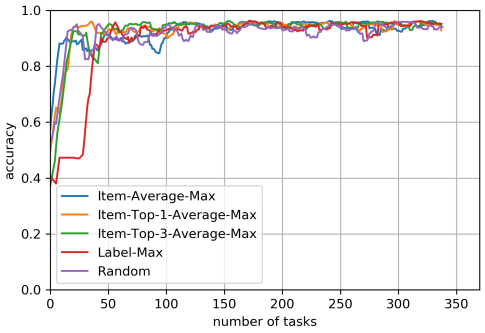
[1]P_1 をテストデータ，P_2, P_3, P_4 を訓練データとして用いた場合



[2]P_2 をテストデータ，P_1, P_3, P_4 を訓練データとして用いた場合

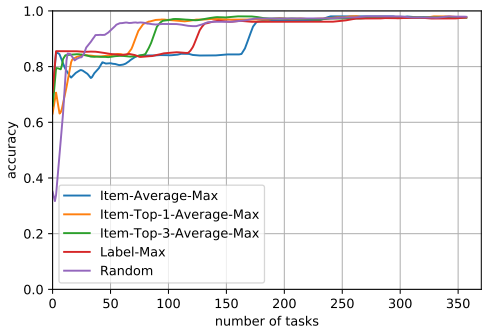


[3]P_3 をテストデータ，P_1, P_2, P_4 を訓練データとして用いた場合

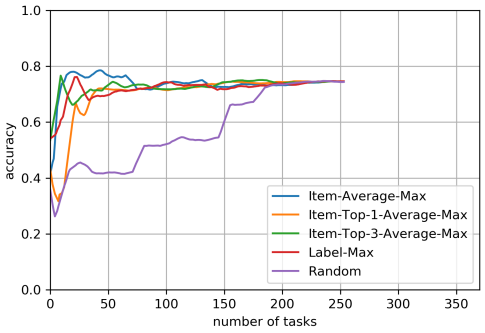


[4]P_4 をテストデータ，P_1, P_2, P_3 を訓練データとして用いた場合

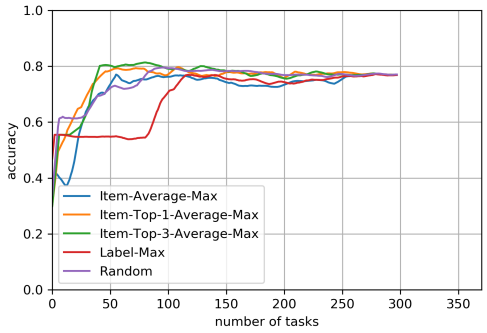
図 6　CNN を使用して，5 つの戦略を比較した実験結果



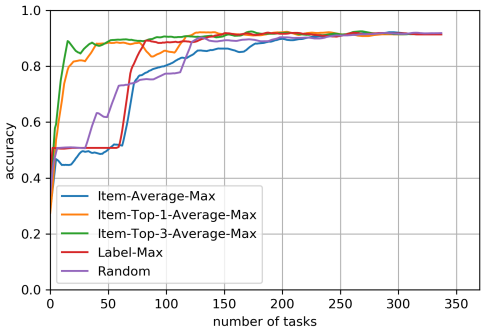
[1]P_1 をテストデータ，P_2, P_3, P_4 を訓練データとして用いた場合



[2]P_2 をテストデータ，P_1, P_3, P_4 を訓練データとして用いた場合



[3]P_3 をテストデータ，P_1, P_2, P_4 を訓練データとして用いた場合



[4]P_4 をテストデータ，P_1, P_2, P_3 を訓練データとして用いた場合

図 7　ロジスティック回帰を使用して，5 つの戦略を比較した実験結果