Active Learning における不確実サンプル選択による アノテーション効率化*

川野恭史** 野田祥希*** 望月凜平*** 青木義満†

Improving Annotation Efficiency through Uncertain Sample Selection in Active Learning

Yasufumi KAWANO, Yoshiki NOTA, Rinpei MOCHIZUKI and Yoshimitsu AOKI

Active learning refers to label-efficient algorithms that use the most representative samples for labeling when creating training data. In this paper, we propose a model that derives the most informative unlabeled samples from the output of a task model. The tasks are a classification problem, multi-label classification and a semantic segmentation problem. The model consists of an uncertainty indicator generator and a task model. After training the task model with labeled samples, the model predicts unlabeled samples, and based on the prediction results, the uncertainty indicator generator outputs an uncertainty indicator for each unlabeled sample. Samples with a higher uncertainty indicator are considered to be more informative and selected. As a result of experiments using multiple datasets, our model achieved better accuracy than conventional active learning methods and reduced execution time by a factor of 10.

Key words: deep learning, annotation, active learning, uncertainty sample selection

1. 序 論

1.1 研究背景

近年、深層学習モデルは多くの分野で大きな成功を収めているが、その学習は大規模なラベル付きデータに依存している。ラベル付きのサンプルを得るためのアノテーション作業は困難であり、時間とコストがかかるため、深層学習の主な問題となっている。

アノテーション作業のコストを減らすために教師なし学習、半教師あり学習、Active Learning などの学習手法が注目されている。教師なし学習や、半教師あり学習はラベル付けされていないサンプルを十分に利用してモデルを学習させることを目的としているのに対し、Active Learning は教師データを作成する際、大量のデータの全てに対してアノテーションをするのではなく、モデルの学習に効果的なサンプルを優先してアノテーションする事で、アノテーションの総量を少なく抑えつつ教師データの作成、モデルの学習を行うことを目的としている。

1.2 関連研究

Active Learning には主に 3 つのシナリオがある。モデルの学習に有効なデータを生成するアプローチ(membership query synthesis)、データに対してラベル付するか破棄するかをふるい分けを行うアプローチ(stream-based selective sampling)、ラベル無しデータプールの中からモデルを学習させる上で最も効果が高いサンプルを選択するアプローチ(pool-based sampling)である。3 つの中でも図 1 に示す pool-based sampling は最も一般的な手法である。具体的には、大

Fig.1 Traditional Pool-based Active Learning Cycle

量のラベル付けされていないデータから代表的なサンプルを選び、oracle、いわゆるアノテータがアノテーションを行い、ラベル付きデータとする。この中でも複数のアプローチが存在する。近年主流になりつつある方法は敵対的学習を用いた方法^{2,3)}である。これらの手法では VAE(Variational Auto Encoder)で画像を低次元表現に落とし込み、得られた低次元表現がラベル付きサンプル由来なのか、ラベル無しサンプル由来なのかを Discriminator が判別する。VAE と Discriminator を敵対的に学習させることでラベル無しデータプールからモデルがまだ得ていない特徴を持つラベル無しサンプルを選択することができる。これらの手法は画像特徴からサンプリングを行なっているため、分類やセグメンテーションといったタスクによらない Active Learning を行うことができる利点がある一方でタスクの学習以上に時間がかかるという欠点もある。本研究では不確実性サンプリングと呼ばれる、モデルで推論を行う際に推論結果の不確実性が最も高いデータからラベル付けをしていくアプローチをとっている。モデルにとって不確実性が高いということは新たな情報が含まれており学習に優位に働く可能性。

* 原稿受付 令和 3 年 5 月 11 日

掲載決定 令和 3 年 10 月 29 日

** 慶應義塾大学大学院理工学研究科(神奈川県横浜市港北区日吉 3 丁目 1 4-1)

*** 株式会社明電舎(東京都品川区大崎 2 丁目 1-1)

† 正 会 員 慶應義塾大学大学院理工学研究科

Fig.2 Model of proposed method

が高いという考えに、不確実性サンプリングは基づいている。

1.3 研究目的

本研究の目的は、ラベル無しプールから最もモデルの学習に効果的なサンプルを優先して選択し、使用するラベル付きデータの量を減らすことでアノテーションコストを低下させることである。それにあたり、タスクに依らず従来手法よりも精度と時間の面でアノテーションコスト削減が可能なモデルを提案する。

本研究における貢献は以下の3点である。(i)複数のタスクに対して本手法を適応させることができる、(ii)従来法と比較して同じアノテーション量でより高い精度での推論が可能、(iii)従来法と比較して最大で10倍の高速化。

1.4 本論文の構成

本論文において2章では提案手法について、3章では実験の概要について、4章ではその実験結果と考察について、5章ではまとめと展望について述べる。

2. 提案手法

2.1 提案手法の概要

本研究では図2に示すように、プールベースにおける不確実性サンプリングを用いたActive Learningを行なっている。また、タスクは分類問題とセマンティックセグメンテーションを扱う。特に分類問題では、単一ラベル分類と、マルチラベル分類の二つを扱う。

まず、少数のラベル付きプールのデータを用いてタスクモデルを学習させる。その後、学習済みのタスクモデルを用いた不確実指標生成器にラベル無しプールのサンプルを入力し、サンプルごとの不確実指標を出力する。この不確実指標の上位K個を選択しアノテーションを行う。アノテーションされてラベル付きサンプルになったデータはラベル付きプールに加わる。

以上の工程を繰り返すことで、最小限のアノテーションコストで高い性能を得ることができる。

2.2 不確実性指標生成器

本手法はラベル付データのアノテーション情報を最大限利用するために②をもとに不確実性指標生成器を提案する。ラベル無しプールのサンプルごとの不確実性を[0,1]の連続値を用いて異なる重要度で表すように設計されている。不確実性指標生成器は、対象モデル(画像分類器、セマンティックセグメンテーションモデル)の予測ベクトルに基づいて不確実性指標を計算する。具体的には、画像分類の場合、予測は各カテゴリの確率ベクトルである。セグメンテーション

ョンでは、各画素は確率ベクトルを持ち、予測ベクトルは各確率ベクトルの平均値である。クラスの数 C とし、サンプルを $y_i \in R^C$ でラベリングしたと仮定すると不確実性指標の計算は式(1)のように定式化される。

$$\text{Indicator}(x_U) = 1 - \frac{\text{MINVar}(V)}{\text{Var}(V)} \times \max(V) \quad (1)$$

ここで x_U はラベル無しサンプルであり、 $V = p(x_U|D_L)$ は現在のラベル付けされたプール D_L で訓練されたタスクモデルに基づく x_U の確率ベクトルである。

$\text{MINVar}(V)$ は式(2)のように定式化できる。

$$\begin{aligned} \text{MINVar}(V) &= \text{Var}(V') \\ &= \frac{1}{C} \left(\left(\frac{1}{C} - \max(V) \right)^2 + (C-1) \left(\frac{1}{C} - \frac{1 - \max(V)}{C-1} \right)^2 \right) \end{aligned} \quad (2)$$

$\text{MINVar}(V)$ は最大の要素が V と同じで、その他の $C-1$ 個の要素が $1 - \max(V)/(C-1)$ であるベクトル V' の分散である。 $\text{MINVar}(V)$ は最大値が V と同じで他の要素が同じ値を持つベクトル V' の分散であるため、 $\text{MINVar}(V)/\text{Var}(V)$ は確率分布の集中度を示している。以上より、確実性指標は[0,1]の境界を持ち、最大確率の値と負の相関を持ち、確率分布の集中度と負の相関を持つという三つの特徴があることがわかる。つまり、最大確率が小さいサンプルは不確実性が高いとみなすLeast Confidentと、最大確率と二番目に高い確率との差分が小さいサンプルの不確実性が高いとみなすMargin Samplingの二つの視点からサンプルを不確実性で測ることができる。不確実性指標が1に近い値であるほどそのサンプルは不確実なものであるとして、選択されやすくなる。

しかしマルチラベル分類では答えとなるラベルが複数あるため、式(1)を使用することができない。そこで、マルチラベル分類に対応する不確実性指標の計算を式(3)のように定式化する。

$$\text{Indicator}(x_U) = 1 - \frac{\text{MINVar}(V)}{\text{Var}(V)} \times \prod_{i=1}^m \max_i(V) \quad (3)$$

ここで、 $\max_i(V)$ を i 番目に大きい V の値とすると、 m は $\max_m(V) \geq 0.5 > \max_{m+1}(V)$ で表される数値である。 $\text{MINVar}(V)$ は式(4)のように定式化できる。

$$\begin{aligned} \text{MINVar}(V) &= \text{Var}(V'') \\ &= \frac{1}{C} \left(\sum_{i=1}^m \left(\frac{\text{sum}(V)}{C} - \max_i(V) \right)^2 + (C-m) \left(\frac{C \sum_{i=1}^m \max_i(V) - m \text{sum}(V)}{C(C-m)} \right)^2 \right) \end{aligned} \quad (4)$$

$\text{MINVar}(V)$ は0.5以上の上位 m 個の要素が V と同じで、その他の $C-m$ 個の要素が $(\text{sum}(V) - \sum_{i=1}^m \max_i(V))/(C-m)$ であるベクトル V'' の分散である。以上のようにすることで、マルチラベル分類においてもLeast ConfidentとMargin Samplingの二つの視点からサンプルを不確実性で測ることができる。

3. 実験概要

本研究では単一ラベル分類とマルチラベル分類、セマンティックセグメンテーションの三つのタスクにおいて提案手法をActive Learningの先行研究であるVAAL³⁾、SRAAL²⁾と、entropy¹⁾、ランダム⁴⁾の4種類の手法によってサンプリングした際の結果と比較して評価する。評価方法は従来のActive Learning^{2,3)}と同様の方法を取る。単一ラベル分類とセ

Fig.3 Result of CIFAR-10

マンティックセグメンテーションではデータセットの全体10%のサンプルをランダムにサンプリングしてラベル付きデータ, 残りの90%をラベル無しデータとして実験を開始する。ラベル付きデータが全体の40%になるまで5%ずつサンプリングを行い, 各データセットにおける性能を先行研究, ランダムサンプリングの結果と比較する。Active Learning は少ないアノテーションでも高い精度を実現することを目的としているため, 本実験では先行研究^{2,3)}と同様にラベル付きデータを全体の40%までにしている。マルチラベル分類ではラベル付きデータを1%から始めて, 4%まで0.5%ずつサンプリングを行う。実験は各5回ずつ行い, その平均を結果とする。評価指標は単一ラベル分類では正解率(Accuracy), セマンティックセグメンテーションでは mean Intersection Over Union (mIOU), マルチラベル分類では画像ごとの各ラベルの正解率の平均を採用した。

3.1 データセット

本研究では単一ラベル分類とマルチラベル分類, セマンティックセグメンテーションの三つのタスクにおいて実験を行う。単一ラベル分類では CIFAR-10⁴⁾, CIFAR-100⁴⁾を用い, セマンティックセグメンテーションでは Cityscapes⁵⁾, マルチラベル分類では CelebA⁶⁾を用いる。

CIFAR-10, CIFAR-100 は, いずれも 32×32×3 の画像が 6 万枚あり, 5 万枚が学習データ, 1 万枚がテストデータとなっており, CIFAR-10 は 10 クラス, CIFAR-100 は 100 クラスのカテゴリからなる。

Cityscapes はモデルに学習させる際に, 画像サイズを 688×688×3 に変換しており, 2975 枚を学習データ, 500 枚をテストデータとして使用している。また, Cityscapes は曖昧なクラスを含めて 34 クラス分アノテーションされているため, 曖昧なクラスを除いて 19 クラスに変換している。

CelebA はモデルに学習させる際に, 画像サイズを 64×64×3 に変換している。クラス数は 40 あり, この中の複数のラベルが答えとなる。また, データ数は 162770 枚であり, 他データセットに比べて非常に多いため本実験では 1%から 4%まで 0.5%ずつのサンプリングにしている。

3.2 学 習

各タスクにおいてデータセットの学習データを用いて 100epoch 程学習を行なった。学習に用いたタスクモデルは単一ラベル分類とマルチラベル分類では VGG16⁷⁾, セマンティックセグメンテーションでは DRN⁸⁾である。

最適化アルゴリズムは単一ラベル分類とマルチラベル分類では SGD, セマンティックセグメンテーションでは Adam

Fig.4 Result of CIFAR-100

で, 学習率はそれぞれ 0.01 と 0.0003 である。損失関数には単一ラベル分類とマルチラベル分類, セマンティックセグメンテーションのそれぞれに交差エントロピー誤差, バイナリ交差エントロピー誤差, NLLLoss を用いている。

4. 実験結果と考察

4.1 CIFAR-10 における実験結果と考察

CIFAR-10 での実験結果は図 3 のようになった。本研究で用いた VGG16 で CIFAR-10 の全データを学習させると 84.50%になる。提案手法では同一のタスクモデルを使用して 40%のデータで 81.83%を達成している。ランダムに 40%をサンプリングした場合, 精度は 76.69%になる。これは提案手法で 25%のサンプリングをした際の精度と近い値である。また, 従来手法の VAAL では 40%のデータで精度は 78.96%になり, 提案手法の 30%に近い値である。

VAAL の数値は論文内の結果よりやや低い値になっているが, これは全データを使用した際の数値が³⁾では 90.16%であるのに対して本実験では 84.50%であったこと, つまりモデルの性能に起因すると考えられる。

SRAAL の精度は VAAL を上回るものの, より簡略化された構造である提案手法より低いという結果になった。SRAAL は構造的に本手法における不確実性指標生成器の出力を正解ラベルとしてモデルの学習を行っており, 結果的に不確実性指標生成器を単体で使用した方がより正確に不確実性を表現することができるからだと考えられる。

4.2 CIFAR-100 における実験結果と考察

CIFAR-100 での実験結果は図 4 のようになった。本研究で用いた VGG16 で CIFAR-100 の全データを学習させると 51.59%になる。提案手法では同一のタスクモデルを使用して 40%のデータで 39.66%を達成している。ランダムに 40%をサンプリングした場合, 精度は 36.52%になる。これは提案手法で 35%のサンプリングをした際の精度とほぼ等しい。また, VAAL では 40%のデータで精度は 37.43%になり, これも同様に提案手法の 35%に近い値である。

以上の結果から CIFAR-10 と同様のデータ数でクラス数が増えた, より複雑なデータセットに対しても VAAL, SRAAL は entropy の性能を上回ったものの, 提案手法の方がより良い性能であることが示された。また, 精度が収束する以前においても提案手法の効果があることも示している。

Fig.5 Result of Cityscapes

4.3 Cityscapes における実験結果と考察

Cityscapes での実験結果は図 5 のようになった。本研究で用いた DRN で Cityscapes の全データを学習させると 58.14% になる。提案手法では同一のタスクモデルを使用して 40% のデータで 56.72% を達成している。ランダムに 40% をサンプリングした場合、精度は 54.98% になる。これは提案手法で 20% のサンプリングをした際の精度とほぼ等しい。また、従来手法である VAAL では 40% のデータで精度は 56.10% になり、これは提案手法の 30% に近い値である。以上からセマンティックセグメンテーションにおいても提案手法は従来手法よりも優れた性能であることがわかる。

また、ランダムサンプリングと提案手法でそれぞれ 40% のサンプリングを 5 回ずつ行なった結果の各クラスの IOU の平均を比較したものを表 1 に示した。提案手法に比べて、ランダムサンプリングはサンプリングの度に結果が変わりやすいため表 2 にランダムサンプリングにおけるクラスごとの IOU の平均値、最小値、最大値を示した。

表 1 に示した結果から、提案手法は Road と Fence の 2 クラス以外の 17 クラスおよび平均においてランダムサンプリングよりも高い IOU が得られた。また、表 2 の各クラスの最大値の平均が 53.07 に対して提案手法の平均が 52.90 と近い値になっていることから、提案手法は理想的なサンプリングが出来ているといえる。

さらに、定性的に評価するためにラベル画像、ランダムサンプリングの結果、提案手法の結果の画像を図 6 に示した。提案手法はランダムサンプリングと比較して、大きな部分では画像左下の歩道と木の識別ができており、細かい部分では画像左上の標識、画像右中央の歩行者、画像中央のポールを識別できている。

以上定量的、定性的な比較により、提案手法の優位性が示せた。

表 1 に関して、AVERAGE の値が図 5 の 40% サンプリング時の値と異なるのは、図 5 の数値はサンプル 1 枚の mIOU を枚数分平均した数値であるのに対して、表 1 の各クラスの値はクラスごとに全サンプルの mIOU を平均したものを示したものであり、AVERAGE の値は出現回数の異なる各クラスを平均した数値であるためである。

Table 1 IOUs per class

Class	Random40%	Proposed40%	Occurrence
Road	93.75	92.04	96
Sidewalk	65.69	67.90	92
Building	79.06	83.82	98
Wall	22.83	27.68	39
Fence	27.52	27.26	37
Poll	27.77	32.14	99
Traffic light	32.33	46.37	60
Traffic sign	45.89	52.87	96
Vegetation	84.46	88.15	96
Terrain	30.00	32.89	47
Sky	85.26	89.28	91
Person	40.18	55.92	82
Rider	21.25	28.84	48
Car	80.70	83.53	97
Truck	19.67	19.96	21
Bus	29.21	46.36	17
Train	42.01	46.61	2
Motorcycle	20.90	30.76	11
Bicycle	42.82	52.77	65
AVERAGE	46.91	52.90	

Table 2 IOUs at 40% random sampling

Class	average	min	max
Road	93.75	91.36	94.66
Sidewalk	65.69	64.56	66.60
Building	79.06	77.53	83.58
Wall	22.83	19.91	28.11
Fence	27.52	26.32	29.16
Poll	27.77	25.86	30.83
Traffic light	32.33	28.64	42.76
Traffic sign	45.89	43.39	50.16
Vegetation	84.46	83.10	88.00
Terrain	30.00	27.87	35.84
Sky	85.26	82.92	90.44
Person	40.18	35.29	54.52
Rider	21.25	14.25	25.59
Car	80.70	79.06	84.42
Truck	19.67	11.13	27.78
Bus	29.21	25.29	34.78
Train	42.01	31.97	62.79
Motorcycle	20.90	16.39	25.23
Bicycle	42.82	38.56	53.07
AVERAGE	46.91	43.34	53.07

Fig.6 Qualitative Evaluation in Cityscapes

Fig.7 Result of CelebA

4.4 CelebA における実験結果と考察

CelebA での実験結果は図 7 のようになった。本研究で用いた VGG16 で Cityscapes の全データを学習させると 88.40%になる。提案手法では同一のタスクモデルを使用して 40%のデータで 87.71%を達成している。ランダムに 40%をサンプリングした場合、精度は 87.36%になる。また、従来手法の VAAL では 40%のデータで精度は 87.42%になる。

2.5%以降、VAAL とランダムサンプリングはあまり差が見られない結果となっている。対して、提案手法は 2%のサンプリング時以外は他手法と比較して高い精度で推論できていることがわかる。以上からマルチラベル分類においても提案手法は従来手法よりも優れた性能を示していることがわかる。

また、提案手法、VAAL、ランダムサンプリングの全てにおいて他データセットには見られないばらつきがあるように見える。これは 1%から 4%までのサンプリングにおいて、精度は約 2%の上昇と、他データセットよりも少ない伸び幅であるため、0.1%のばらつきであっても大きくばらついているかのように見えてしまうからだと考えられる。CelebA では 1%のサンプリング時点で精度は収束に近づいており、他データセットと比較して、提案手法と他手法の精度の差が小さくなっているが、上がり幅が小さい中でも平均的に他手法よりも高い精度が得られる点で提案手法の優位性が示せていると言える。

Table 3 Execution time comparison

	VAAL	Proposed
CIFAR-10	1.9×10^5	2.0×10^4
CIFAR-100	2.8×10^5	2.2×10^4
Cityscapes	3.0×10^5	1.3×10^5
CelebA	4.0×10^5	8.3×10^4

4.5 モデルの実行時間の比較

提案手法と VAAL を実行した際の時間を比較した結果、表 3 のようになった。CIFAR-10、CIFAR-100 の両方において実行時間を約 10 分の 1 に、Cityscapes では約 2 分の 1 に、CelebA では約 5 分の 1 に短縮することに成功した。

5. まとめと展望

本論文ではラベル無しプールから最も情報量の多いサンプルを選択するために、タスクモデルの出力結果から最も情報量の多いラベル付けされていないサンプルを導出するモデル、不確実性指標生成器を提案し、使用するラベル付きデータの量を減らすことでアノテーションコストを低下させることができることを実験により示した。複数のタスクにおいて、提案手法は従来手法と比較して同じアノテーション量でより高い精度での推論が可能である。また、モデルの実行時間も従来手法である VAAL と比較して CIFAR-10、CIFAR-100 では約 10 分の 1 に、Cityscapes では約 2 分の 1 に、CelebA では約 5 分の 1 に短縮することに成功した。

本研究では不確実性指標が最も高い数値、つまりラベル無しプール内で最も不確実なサンプルをアノテーションするというを行なった。実験の結果、この不確実性指標はサンプルの不確実性を表現できていることがわかった。本研究では画像に対して本手法の優位性を示したが、動画等のより高度なタスクについても応用していくことができるのではないかと考えている。

参 考 文 献

- 1) Burr Settles: Active Learning Literature Survey. (2010).
- 2) Beichen Zhang, Liang Li, Shijie Yang, Shuhui Wang, Zheng-Jun Zha and Qingming Huang: State-Relabeling Adversarial Active Learning, Proceedings of the IEEE conference on computer vision and pattern recognition, (2020) 8756.
- 3) Samarth Sinha, Sayna Ebrahimi, and Trevor Darrell: Variational adversarial active learning, arXiv preprint arXiv:1904.00370, (2019).
- 4) Alex Krizhevsky and Geoffrey Hinton: Learning multiple layers of features from tiny images, Technical report Citeseer, (2009).
- 5) Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele: The cityscapes dataset for semantic urban scene understanding, Proceedings of the IEEE conference on computer vision and pattern recognition, (2016) 3213.
- 6) Z. Liu, P. Luo, X. Wang, and X. Tang: Deep learning face attributes in the wild, Proceedings of the IEEE International Conference on Computer Vision, (2015) 3730.
- 7) Karen Simonyan, Andrew Zisserman: Very Deep Convolutional Networks for Large-Scale Image Recognition, arXiv preprint arXiv:1409.1556, (2014).
- 8) Fisher Yu, Vladlen Koltun and Thomas Funkhouser: Dilated Residual Networks. Proceedings of the IEEE conference on computer vision and pattern recognition, (2017) 472.

Oracle は Oracle Corporation の登録商標