原著論文

属性知識の高速オンライン学習と 転移による未知クラス推定†

木村 大毅 *1・Kankuekul PICHAI *1・長谷川 修 *2

人のために働く自律知能ロボットの開発には、人の生活環境下で人や実環境、インターネットとの相互作用から知識を追加的に獲得し、知的に発達する知能情報処理機構の実現が必要である。本稿では、その基盤技術の一つとして、オンライン学習が可能で非常に高速な属性知識の学習と転移手法を提案する。これは、実世界のノイズに頑健な学習器であるSelf-Organizing and Incremental Neural Networkを改良し、それを活用して構築する。動物の属性知識を転移し、未学習の動物の種類を推定する実験を用いて本手法を評価した結果、従来のオンライン学習手法と同等の認識率を保ちつつ、学習時間と認識時間を大幅に短縮した。更に、従来手法と比べてノードの数も大きく削減し、情報量の軽量化も実現した。また、提案手法が持つノイズ耐性を用いることで、インターネット上のデータも活用することができた。

キーワード：属性ベース、オンライン追加学習、未知クラス、ロボット

1. はじめに

近年、我が国の少子高齢化に伴う労働力不足や介護士不足を補うため、人と同様の生活環境下で稼働する自律知能ロボットの開発が急務である。その他にも、災害時の救助や、軍事、癒しへの利用のために、様々な研究機関でロボットに関する多くの研究が現在行われている。その一例として、本田技研のASIMOに代表されるような二足歩行に関する研究や、小型のヒューマノイド・ロボットを用いた物体操作の学習に関する研究[1]が挙げられる。ところが、現在開発されているヒューマノイド・ロボットは、限定された環境での動作に留まっており、先に挙げた人と同様の生活環境下では稼働できない。その理由として、人の生活環境(実環境)で知識を獲得する知的発達機構に関する研究が発展途上であることが挙げられる。この知的発達機構の実現のために、人の生活環境下に存在する多種多様な物体を認識する機構の構築が必要であると、我々は考えている。更に、ノイズを多く所有するが、大量の情報を瞬時に取得可能なインターネットからも、新たな知識を追加的に獲得する機構を取り入れ

ることで、学習データの容易な収集が可能となる。

現在、物体認識の分野では画像を用いた研究が主流である。特に、特定の物体や顔などに対して個別の学習器を作成して、それらを認識するものが主流である[2-5]。ところが、人が日頃判別している物体は少なくとも30,000種類あると言われており[6]、すべての物体に対して個別の学習器を作成することは現実的に不可能である。そのため近年、属性ベースの学習手法が注目されている[7]。属性ベースの学習手法とは、物体などがもつ属性(大まかなカテゴリ)を基準にして学習する手法であり、属性の組み合わせにより物体を表現する。そのため、少ない属性の学習で多くの物体を表現することが可能となる。また、学習した属性を用いることで、実世界に多く存在する未学習物体に対しても、属性を推測することができる。その後、更に、求められた属性と、属性と物体の種類を列挙した辞書を用いることで、物体の種類も推定することができる。

ここで、具体的に従来研究について記述する。現在、属性ベースの学習手法は多く存在し[7-10]、その代表例としてLampertらの研究が挙げられる[7]。彼らは、属性ベースの学習に関する2つの学習方式を提案した。更に彼らは、実際の動物の画像とその属性を利用して評価実験データセットを作成し、そのデータセットとSVM[11]を用いて、それぞれの学習方式を評価した。なお用いられた属性の例として、「肉食動物である」や「飛ぶことができる」、「凶暴である」が挙げられる。ところが、彼らの手法はバッチ学習であ

† Guessing an Unknown Class by Online Fast Attributes Learning and Transfer

Daiki KIMURA, Pichai KANKUEKUL and Osamu HASEGAWA

*1 東京工業大学 総合理工学研究科
Interdisciplinary Graduate School of Science and Engineering,
Tokyo Institute of Technology

*2 東京工業大学 像情報工学研究所
Imaging Science and Engineering Laboratory, Tokyo Institute
of Technology

り、追加的に知識を増やすことが出来ない、実環境で稼働するロボット上での運用を想定した時、予め学習物体を全て与えることは困難であり、環境に応じて、その場で追加学習できる機構が必要となる。そのため、オンライン追加学習が、前述したロボットのための知的発達機構に対して必須となる。

そこで、Kawewong らは、オンライン追加学習が可能な属性ベースの学習手法を提案した[10]。そして、従来手法[7]と比較実験を行って評価している。ところが彼らの手法は、属性の増加に伴い学習や認識の処理時間などが増大する問題が存在する。これは、学習器の数が属性の数と比例しているためである。彼らが用いている学習器[12]は、教師なしクラスタリングのために設計されたものであり、教師情報を管理するアルゴリズムが存在しなく、1つの学習器で1つの概念しか学習できない。

そこで我々は、まず、自己増殖型ニューラルネットワーク[12]に対して、教師情報に対する統計情報の算出アルゴリズムを追加し、教師情報の統計的な推定機構を取り入れる。これにより、教師あり学習に対応することができ、1つの学習器で複数の教師情報に対する学習ができるようになる。それを活用することで、前述の知的発達機構の実現のために、オンライン学習が可能で、学習器の数が少ない属性ベースの学習と転移手法を提案する。この手法は、従来手法と比べて学習器の数が少なくなるため、計算回数が少なくなり、高速な属性知識の学習と転移を実現できると考えられる。なお、この時の“転移”とは、学習した属性知識を未学習の物体や動物の識別へ活用することを意味する。そして、[7]や[10]と同様の実験で本手法の評価を行う。考察では、ノイズは多いが大量の学習データを保持しているインターネットからも知識を獲得できることを確認し、本手法が持つ有用性についても考察する。

なお現在、オンライン学習可能なSVMが提案されている[13]。ところが、この学習器を用いた場合にも、学習器の数が属性の数に比例しており、属性の増加に伴い学習や認識の処理時間が増大する。本稿では、ベースラインとしてSVMを用いた手法の結果を参照するために、オンライン手法に比べて安定性があるバッチ学習[7]との比較を行う。

本稿は次のように構成されている。まず、2章では提案する属性ベースの学習手法について述べる。次に、3章では本提案手法の評価について述べ、4章では考察と今後の課題について述べる。最後に、5章では本稿のまとめについて述べる。

2. 提案手法

本稿の提案手法は、2.2節にて提案する自己増殖型ニューラルネットワーク(SOINN)を活用した、オンライン学習可能で非常に軽量の属性情報の学習と転移手法である。ここでは、まず、提案する属性の学習および転移手法について述べ、その後提案するSOINNの具体的な改良点を記述する。

なお、この時、我々が学習器としてSOINNを選んだ理由は、実世界のデータに対して非常に有用な次の3つの特性を持つためである。

- 学習に対する事前情報が不要である特性
- オンライン追加学習が可能である特性
- 実世界のノイズに対応する強いノイズ耐性

また本稿では、以下の単語を次のように定義し用いている。

クラス 物体や動物の種類(例：リンゴ)

属性 物体や動物の大まかなカテゴリ(例：球体)

ラベル 学習器が持つ学習カテゴリに関する情報

2.1 提案手法のアルゴリズム

まず、概要図を図1に示す。本手法は、原理的にはロボットなどで取得できる様々な情報(画像や音声、圧力情報など)を用いた学習に応用可能である。ところが、理解の補助のため、図1および以降の説明は、後述する実験に合わせて画像を用いた例で記述する。本手法には“学習フェーズ”と“推定フェーズ”があり、前者では入力されたデータを用いて属性ベースの学習を、後者では学習属性を転移し、クラスの推定を行う。図1では、学習フェーズのステップを実線と二重線で示し、認識フェーズのステップを実線と点線で表している。

2.1.1 学習フェーズ

始めに、学習画像から予め設定された特徴を抽出し、その特徴に学習する画像のクラスが持つ属性値を付加する。この時、“クラスと属性の関連度の表”(図1の“Classes & Attributes matrix”)を利用し属性値を参照する。これは、全てのクラス(動物や物体など)が持っている属性を列挙したものであり、人が事前に定義して本システムに与える。この時、関連度は、“関連がある”と“関連がない”の二値や、実数値(確率のように0から1の間)で定義される。次に、抽出された特徴ベクトルをパターン情報、属性値をクラスの教

図1 提案手法の概要図

師信号として、特徴毎に生成された提案するSOINNへ入力し、学習させる。また、事前に定義するパラメータ ρ 番目より後のクラスを学習する際には、SOINNへの入力を行う前に後述する認識処理を実施して、各々の属性の関連度を算出する。これは、認識時に必要となる属性判定の閾値決定に用いており、利用方法と計算式は後述する。

2.1.2 推定フェーズ

まず、認識画像から学習フェーズと同じ特徴を抽出する。抽出された特徴と、その特徴に対応するSOINN

を用いて、それぞれの属性の関連度を求める。次に、“クラスと属性の関連度の表”を用いて、認識画像のクラス g を推定する。この一連のクラス g に関する推定は、式(1)に従う。 $\mathcal{Z}$ はクラスの集合、 $\mathcal{M}$ は属性の集合、 $\mathcal{Q}$ は特徴の集合、 V_q は入力画像の特徴 q のパターン、 a_z^m はクラス z と属性 m の関連度を表す。関連度は、 $0 \leq a_z^m \leq 1$ で、1に近いほど関連度が高いことを表す。 R_q^{m+} および R_q^{m-} は、学習クラス中の属性 m と関連性が高いクラスの属性値の平均および低いクラスの属性値の平均であり、式(2)および式(3)で決定される。 $\mathcal{Z}_l$ は学習クラスの集合、 $\mathcal{Z}_\rho$ は

$$g = \operatorname{argmax}_{z \in \mathcal{Z}} \left[\sum_{m \in \mathcal{M}} \left(\sum_{q \in \mathcal{Q}} \left[P^m(V_q) - \frac{1}{2} (R_q^{m+} + R_q^{m-}) \right] \cdot (2a_z^m - 1) \right) \right] \quad (1)$$

$$R_q^{m+} = \frac{1}{\sum_{q \in \mathcal{Q}} \sum_{z \in \mathcal{Z}_l \setminus \mathcal{Z}_\rho} \sum_{m \in \mathcal{M}} a_z^m} \sum_{q \in \mathcal{Q}} \sum_{z \in \mathcal{Z}_l \setminus \mathcal{Z}_\rho} \sum_{m \in \mathcal{M}} [P^m(V_q) \cdot a_z^m] \quad (2)$$

$$R_q^{m-} = \frac{1}{\sum_{q \in \mathcal{Q}} \sum_{z \in \mathcal{Z}_l \setminus \mathcal{Z}_\rho} \sum_{m \in \mathcal{M}} (1 - a_z^m)} \sum_{q \in \mathcal{Q}} \sum_{z \in \mathcal{Z}_l \setminus \mathcal{Z}_\rho} \sum_{m \in \mathcal{M}} [P^m(V_q) \cdot (1 - a_z^m)] \quad (3)$$

$$P^m(V_q) = \frac{1}{k} \sum_{i \in \mathcal{N}_k(V_q)} \left[0.5 + \frac{L_{q,i}^m - 0.5}{\max \left(1, \frac{\|V_q - W_{q,i}\|}{T_{q,i}} \right)} \right] \quad (4)$$

学習クラスの1番目から ρ 番目までのクラスの集合を表す。式(1)や式(2)、式(3)の中で用いられている $P^m(V_q)$ の定義を式(4)に示す。これは、特徴パターンの k 個の近傍重みベクトルを求め(k -近傍法: k -nearest neighbor [14]), それらのノードが持つラベルの平均である。平均を求める際には、ノードが持つ閾値で正規化された距離を用いて重み付き平均を用いている。なお、 $\mathcal{N}_k(V_q)$ は V_q の k 個の近傍ノードの集合、 $L_{q,i}^m(W_{q,i}, T_{q,i})$ は特徴 q のSOINN内にあるノード i の属性 m のラベル重み(重みベクトル, 類似度閾値)を表す。また、 $\max(a, b)$ は a と b を比較した時の大きな値を返す。

2.2 提案するSOINN

SOINN [12]に対して統計を取り入れ、一つのネットワークで複数概念を学習できるように改良する。まず、従来のSOINNについて記述する。

2.2.1 SOINN

SOINNとは、SOM[15]とGNG[16]に着想を得て構築された学習手法である。具体的には、パターン情報の入力に合わせてニューロン(ノード)が増殖や消滅を行い、仮想的なノード同士の繋がり(エッジ)を用いて構造化されるネットワークモデルである。現在SOINNには複数の構造(アルゴリズム)が存在し、多

くの論文で紹介されている[12, 17-21]。その中でもAdjusted SOINN[12]は、他のSOINNと比べて、パラメータ数が少なく、計算時間が短いなどの優位点があり、SOINNを用いた応用研究で使用されている[22-25]。そのため本稿では、Adjusted SOINNを改良する。参考のために、Adjusted SOINNの学習アルゴリズムのフローチャートを図2に示す。

2.2.2 改良の概要

本稿で提案するSOINNは、従来のSOINNに対して、大きく2つの改良を行なっている。

1つは、ノードの挿入や削除、及びノードの重みベクトルを調節する処理の改良である。これは、従来のSOINNのネットワークがデータの入力順序に依存する問題(入力に対する不安定性)を軽減する目的で行なっている。従来のSOINNでは、入力されたデータ数がある定数の倍数になった時に、ノイズの削除処理を行っていた。ところが、削除処理のサイクルの中で、早い段階で入力されたデータと、遅い段階で入力されたデータで、ノイズの削除処理を行うまでのステップ数が異なっていた。我々は、これが不安定である一つの原因と考え、それぞれのノードに対して年齢を導入した。これにより、ノイズの削除処理までのステップ数が全ノードで同じになる。また、ノイズに対する判定条件の明確化や、重みベクトルの更新処理のヒューリスティックな部分、例えば第2勝者ノードの更新に対する1/100の乗算の廃止などを行なっている。

もう1つは、ノードにクラスの統計情報を追加した改良である。SOINNは、ネットワークモデルを用いたパターン情報の記録という観点では様々な優位点を保持している。しかし、クラスに関する教師ありデータの学習に関しては、未だ議論がされていない。そのため従来は、クラスごとにネットワーク(SOINN)を作成し、認識データと、それぞれのネットワーク内のノードとの距離を測定することで、クラスの推定を行っていた。具体的には、認識データと各ネットワークに対して固定数の近傍ノードを探し、それらノードと認識データとの距離の総和が最も短いネットワークのクラスを、推定されたクラスと設定していた。ところが、この手法では、クラスと同数のSOINNが必要となり、学習後のデータの量や学習時間がクラス数と比例する問題点があった。更には、他のクラスでの学習結果を活用できない問題点が存在していた。そこで我々は、クラスに関する教師信号を「ラベル」としてノードに付加する機構と、ネットワークの変化に合わせてラベル情報の変更を行う機構を追加する。これにより、1つのネットワークで全てのクラスに関する情

図2 [12]のフローチャート (学習時)

報を学習できるため、ネットワークの削減ができ、更には学習時に類似性の高い他のクラスの学習結果を活用することができる。

2.2.3 アルゴリズム

提案するSOINNの学習アルゴリズムのフローチャートを図3に示す。認識アルゴリズムは、使用目的などにより異なるため、ここでは言及しない。

学習は、大きく2つの処理に分けることができる。ノードやエッジなどのネットワーク構築に関する更新処理と、ノイズの除去を行うノードの削除処理である。これらは、パターン情報とクラスに関する教師信号の組が入力される度に逐次実行される。例えば、パターン情報として画像の特徴ベクトル、教師信号として学習画像に対応する物体の種類や属性などが当てまる。

1. ネットワークの更新

学習データ(パターン情報と、クラスの教師信号)が入力されると、全ノードの年齢をインクリメント(+1)する。ノードの年齢とは、ノードが存在

していた期間を表す。次に、入力パターンと全ノードの重みベクトルとの距離を算出する。それを基に、入力パターンに最も近い重みベクトルを持つノード(第1勝者ノード: $winner = "w"$)と、2番目に近い重みベクトルを持つノード(第2勝者ノード: $2^{nd} winner = "sw"$)を決定する。式(5)を用いて、これら2つのノードに対する類似度閾値 T_w, T_{sw} を算出する。 $\mathcal{N}_i$ はノード i の全隣接ノード、 W_i はノード i の重みベクトル、 $\mathcal{A}$ はSOINNがもつ全てのノードである。なお「隣接ノード」とは、エッジで接続されているノードを表す。

$$T_i = \begin{cases} \max_{k \in \mathcal{N}_i} \|W_i - W_k\| & (\mathcal{N}_i \neq \phi) \\ \min_{k \in \mathcal{A} \setminus \{i\}} \|W_i - W_k\| & (\mathcal{N}_i = \phi) \end{cases} \quad (5)$$

これらの類似度閾値から、入力情報が属するクラスタを決定する。入力パターンと第1勝者ノードの重みベクトルとの距離が、第1勝者ノードの類似度閾値よりも大きい、もしくは、第2勝者との距離が、第2勝者の類似度閾値よりも大きい場合には、入力情報は新しいクラスタであると判定する。その場合には、入力パターンを重みベクトル、クラスの教師信号をラベル情報とする新しいノードを挿入する。ラベル情報とは、各々のクラスとの関連度合いを保持している。例えば、“クラス α とは75%の関連があり、クラス β とは関連がない(0%)”などが記録されている。

一方で上記以外の場合には、入力情報と、第1勝者ノード、第2勝者ノードは同一クラスタに属すると判定し、第1勝者と第2勝者の間に年齢が零のエッジを作成する。既にエッジが存在する場合には、そのエッジの年齢を零とする。エッジは、2つのノードが同一クラスタであることを表し、エッジの年齢は、それらのノードが他のクラスタと判定された回数を表す。次に、第1勝者の隣接ノードの中で、第2勝者以外と接続しているエッジの年齢をインクリメント(+1)する。この時、年齢が、事前に定義するパラメータ age_{max} を超える場合、そのエッジを削除する。次に、第1勝者ノードの重みベクトル W_w を式(6)に従い更新する。 M_w は第1勝者ノードが過去に第1勝者になった回数、 V は入力パターンを表す。最後に、入力されたクラスの教師信号を基に、周辺ノードのラベル情報 L_i を更新する。これは、第1勝者ノードとその隣接ノード、及び、第2勝者ノードとその隣接ノードに対して行われ、前者は式(7)

図3 提案するSOINNのフローチャート(学習時)

に、後者は式(8)に従う。 $\mathcal{C}$ はクラスの集合、 $\mathcal{N}_w(\mathcal{N}_{sw})$ は第1勝者(第2勝者)ノードの隣接ノードの集合、 L_i^m はノード i のクラス m のラベル重み、 S^m はクラス m の入力された教師信号、 U_i はノード i のラベルが過去に更新された回数である。

$$\Delta W_w = \frac{1}{M_w + 1} (V - W_w) \quad (6)$$

$$\Delta L_i^m = \frac{1}{U_i + 1} (S^m - L_i^m) \quad (\forall i \in \{winner\} \cup \mathcal{N}_w, \forall m \in \mathcal{C}) \quad (7)$$

$$\Delta L_i^m = \frac{1}{2(U_i + 1)} (S^m - L_i^m) \quad (\forall i \in \{2^{nd} \text{ winner}\} \cup \mathcal{N}_{sw}, \forall m \in \mathcal{C}) \quad (8)$$

2. ノイズの除去

ネットワークの更新後、ノイズの可能性が高いノードの削除を行う。本稿での“ノイズ”とは、出現頻度が低いデータのことを指す。先の更新法より、ある領域に対して多くの入力が行われると、ノードやエッジが次々と作成され、隣接ノードの数が増える。そのため、隣接ノードの数が増えないことは、その領域に入力が少ないことを示している。この特性を用いて「ノードの年齢がパラメータ λ の倍数」と「隣接ノードの数がパラメータ η より小さい」の2つの条件式が共に真であるノードをノイズとみなし削除する。 λ は判定を行う間隔を、 η はノイズの定義を調節する。

ノードの削除時には、そのノードが持つラベル情報を周辺ノードに伝播させる。これは、上記条件では入力回数が少ない重要な分布も削除してしまう可能性があり、削除による影響を軽減する目的で行う。伝播は、削除するノードの重みベクトルと最も近い重みベクトルを持つノード fn のラベル L_{fn} 、及び2番目に近いノード sn のラベル L_{sn} に対して行う。前者は式(9)、後者は式(10)の更新式に従う。 $\mathcal{K}_d$ は削除するノード d が持つラベルの集合、 L_{fn}^m (L_{sn}^m , L_d^m) はノード fn (sn , d) のクラス m のラベル重み、 T_{fn} (T_{sn}) はノード fn (sn) の類似度閾値、 U_{fn} (U_{sn}) は L_{fn} (L_{sn}) が過去に更新された回数を表す。また、 D_{fn} , D_{sn} は、式(11), 式(12)で定め、 W_{fn} (W_{sn} , W_d) はノード fn (sn , d) の重みベクトルを表す。

$$\Delta L_{fn}^m = \begin{cases} \frac{L_d^m - L_{fn}^m}{U_{fn} + 1} & (D_{fn} \leq T_{fn}) \\ \frac{T_{fn} (L_d^m - L_{fn}^m)}{D_{fn} (U_{fn} + 1)} & (D_{fn} > T_{fn}) \end{cases} \quad (\forall m \in \mathcal{K}_d) \quad (9)$$

$$\Delta L_{sn}^m = \begin{cases} \frac{L_d^m - L_{sn}^m}{2(U_{sn} + 1)} & (D_{sn} \leq T_{sn}) \\ \frac{T_{sn} (L_d^m - L_{sn}^m)}{2D_{sn} (U_{sn} + 1)} & (D_{sn} > T_{sn}) \end{cases} \quad (\forall m \in \mathcal{K}_d) \quad (10)$$

$$D_{fn} = \|W_{fn} - W_d\| \quad (11)$$

$$D_{sn} = \|W_{sn} - W_d\| \quad (12)$$

3. 実験と評価

3.1 実験概要

本実験の概要図を図4に示す。本実験で用いるデータセットは、Lampert らが作成した“Animals with attributes - A dataset for Attribute Based Classification”¹である。このデータセットには、50種類の動物の画像(図5)と、各々の動物と85種類の属性との関連度を記した一覧表(図6)が含まれている。なお、本実験で用いられる関連度は、“関連がある”と“関連がない”の二値のみである。50種類の動物の内、40種類を学習用、10種類を認識用として定められており、それ

図4 実験の概要図

既習動物の学習結果を転用して、未学習動物の属性の割合を算出し、未学習動物を推定する

1 <http://attributes.kyb.tuebingen.mpg.de/>

図7 AT-SOINN [10] の概要図

フェーズと認識フェーズについて述べる。

学習フェーズでは、画像から特徴を抽出して、動物の属性に従ってSOINNを用いて学習する。具体的には、全ての属性に対して、関連性がある属性については、その属性の *pos*-SOINN へ、関連性がない属性については、*neg*-SOINN へ特徴を入力する。そのため AT-SOINN では、属性の数×特徴の数×2 個のネットワークが必要である。

認識フェーズでは、抽出された特徴パターンと、全ての属性の *pos*-SOINN と *neg*-SOINN に対して、重みベクトルとの距離を求め、小さな値を k 個選択して、その総和を算出する。その後、動物と属性の関連度を用いて距離を足しあわせ、動物を推定する。一連の判定式を式(13)に示す。 g は推定された動物、 $\mathcal{Z}$ は動物の集合、 $\mathcal{M}$ は属性の集合、 $\mathcal{Q}$ は特徴の種類の集合を表す。 a_z^m は動物 z と属性 m の関連性を表し、「関連がある」場合には $a_z^m = \text{positive}$ 、「関連がない」場合には $a_z^m = \text{negative}$ とする。 $[P]$ は、条件 P が真であるときは“1”、偽であるときは“0”とする。 $d_{q,m}^p$, $d_{q,m}^n$ は、それぞれ式(14), 式(15)より求める。 V_q は入力画像の特徴 q のパターン、 $W_{q,m,t}^p(V_q)$ ($W_{q,m,t}^n(V_q)$) は特徴が q で属性が m の *pos*-SOINN (*neg*-SOINN) の中で V_q と t 番目に近い重みベクトルを表す。

$$g = \operatorname{argmax}_{z \in \mathcal{Z}} \left(\sum_{q \in \mathcal{Q}} \left(\sum_{m \in \mathcal{M}} \left(\frac{d_{q,m}^p [a_z^m = \text{positive?}] + d_{q,m}^n [a_z^m = \text{negative?}]}{d_{q,m}^p + d_{q,m}^n} \right) \right) \right) \quad (13)$$

$$d_{q,m}^p = \frac{1}{k} \sum_{t=1}^k \|V_q - W_{q,m,t}^p(V_q)\| \quad (14)$$

$$d_{q,m}^n = \frac{1}{k} \sum_{t=1}^k \|V_q - W_{q,m,t}^n(V_q)\| \quad (15)$$

図8 学習画像数の割合と認識率の関係

表1 未学習動物（10 種類）の認識率

Method		Accuracy (avg)
Online	Proposed	25.72%
	AT-SOINN [10]	26.96%
Batch	DAP(SVM) [7]	40.51%
	IAP(SVM) [7]	27.83%

3.3 評価方法と結果

提案手法を、従来手法のDAPおよび IAP[7] と、AT-SOINN[10] との比較で評価する。提案手法のパラメータは、 $\lambda = 600$, $age_{max} = 100$, $\eta = 1$, $k = 13$, $\rho = 25$ である。これらは、動物の種類を少なくした予備実験の結果に基づき決定した。なお実験した候補は、 $\lambda = 200, 400, 600, 800, 1000$, $age_{max} = 50, 100, 150$, $\eta = 1, 2$, $k = 9, 11, 13, 15$, $\rho = 10, 25$ である。また、提案手法はオンライン学習であるため、学習順序により結果が異なる。そのため、以降の結果は、順序をランダムに変更し、10回実行した平均値を記載する。

各々の認識率を表1に記し、学習及び認識にかかる処理時間を図9, 図10に示す。なお、ここでの処理時間は、特徴抽出時間など全ての手法において同様に必要となる処理の時間を含めていない。次に、提案手法とAT-SOINNが保持するネットワークの数と、それ

図9 24,295 枚の画像の学習にかかる時間（特徴抽出時間等を除く）
提案手法では1秒間に52枚の画像を学習可能

図10 6,180 枚の画像の認識にかかる時間（特徴抽出時間等を除く）
提案手法では1枚の画像を17ミリ秒で認識可能

らが保有する全ノード数を表2に記す。また、学習画像の枚数を変化させた時の未知動物に対する認識率の推移を図8に示す²。横軸は全ての画像を学習した場合を100%とした学習画像の割合であり、本図は学習量が少ない時(50%以下)の認識率の変化を示している。点は実際の認識率を、実線はこれらの結果から対数近似を行った結果を示す。なお、実験は同一環境(CPU: 2.93GHz, メモリ: 8GB)で実施した。

4. 考察

4.1 提案手法の評価

3.3節より、提案手法の認識率は、バッチで学習したDAPには劣るが³、従来のSOINNを用いたAT-SOINNと同等であった。そのような認識性能を保ちながら、学習時間をDAPに比べ約13,034倍速く、AT-SOINNに比べ約45倍速くすることができた。更に、認識時間をDAPに比べ約1,765倍速く、AT-SOINNに比べ約146倍速くすることができた。また、ノード数をAT-SOINNの0.56%に抑えた。つまり、提案手法は、オンライン学習であるAT-SOINNと同等の認識率を保ちつつ、非常に高速で軽量の学習および認識を

表2 作成されたネットワーク(SOINN)の数と、学習後のSOINNが保有するノードの数

Method	# of SOINNs	# of Nodes
Proposed	6	1,952
AT-SOINN [10]	1020	347,082

実現している。

従来手法[10]と比べて、学習時間と認識時間を削減できた理由は、学習器が持つネットワーク数と大きく関係している。ネットワークの数が少ない場合、学習時と認識時に距離計算を行う回数が少なくなる。従来手法[10]は属性ごとにネットワークを作成し学習していたが、提案手法はネットワークに全ての属性情報を保管するラベルを導入して、ネットワーク数が属性の数に比例しない構造を持っている。そのため、ネットワーク数が表2のように減り、高速化を実現できた。

更に、DAP[7]とAT-SOINN[10]では属性ごとに学習器を作成して学習しているが、提案手法では1つの学習器(ネットワーク)で全ての属性を学習している。そのため、似通った属性⁴が存在する時、それぞれの

² [7]は、バッチ学習であり追加学習性が無いため計測していない

³ バッチ学習では実環境で稼働するロボットへの実装が不可能

⁴ 正確には特徴パターンの近い属性群

属性を学習する際に、他の属性の学習結果を活用しながらの学習が可能である。つまり、ある属性を学習するときに、似通った属性が学習済みであれば、早く属性の学習を完了することができる⁵。このように提案手法は、従来手法に比べて、学習効率を高めた学習が可能である。実際に図8を確認すると、提案手法は早い段階で認識率が向上している⁶。この特性は、少ない学習画像による学習を実現し、実世界のような膨大な数の属性や物体の学習が必要となる環境において、非常に有用である。

前述したような時間と情報量の削減や、学習効率の向上は、ヒューマノイド・ロボットへの実装を考慮した実世界知能に相応しい知的な情報処理を実現する基盤技術の必要不可欠な特徴である。そのため我々は、本提案手法がこれらの知的発達機構の基盤技術の一つになりうると考えている。

その一方で、認識率に関しては、バッチ学習手法であるSVMを用いた手法[7]と比べて低下している。この原因として考えられるのは、SOINNが持つオンライン学習性による安定性の低下が関係していると考えている。また、オンライン追加学習で動作している本提案手法は、認識精度にばらつきが存在している。これら点について、今後改良及び検討すべき課題であると我々は考えている。

4.2 インターネットの活用

先の実験の学習データは、全て人が選別し与えていた。ところが、学習すべき物体数が膨大である実世界での運用を考慮した時、その選別は非常に困難となる。一方で、近年普及したインターネットを用いることで、多くの画像を瞬時に入手できる。ところが、不特定多数の人々によって提供されたものであるため、ノイズと思われる画像も多い。そこで、本提案手法の学習器SOINNが持つノイズ耐性を用いてインターネット上のノイズを削除することで、人が学習データを準備せずに学習できることを検証する。また、インターネット上のデータは数が膨大であり、本手法の高速性と軽量性も必要となる。

実際に、インターネットの画像を用いて学習するシステムを構築した。まず事前に、学習物体と認識物体のそれぞれの属性は与えられている。学習フェーズは、検索エンジンで学習物体の名前を検索しダウンロードしてオンライン学習する。認識フェーズは、カメラの前に置かれているものの属性を推定し、辞書を用いて物体を推定する。システムの概要は、図1と同

様である。物体の認識結果と、学習や認識に所要する時間で評価を行う。

具体的な実験の内容と、その結果を以下に記す。学習物体を25種類、認識物体を8種類、属性を18種類定義した。次に、学習物体に対して実際に検索エンジン⁷を用いて100枚のサムネイル画像(低画質画像)を取得した。その画像から、画像の全体に対するpHOG特徴量[28]と、物体の輪郭のみのpHOG、色ヒストグラムを抽出して、それらを本提案手法を用いてオンライン学習させた。本実験で使用した特徴は、認識率よりも速度を重視して選択している。そのため、認識精度が不十分であれば特徴を再度選択することにより、より良好な結果が得られると考えられる。この時に所要した学習時間(ダウンロードと特徴抽出を含まない)は、4秒ほどであった。次に、認識物体をカメラの前におき、そこから上記特徴を抽出した後、本提案手法の認識処理を行った。例として、認識物体である“腕時計”を見せた時のシステムの結果を図11に示す。図中の左図が入力画像であり、右上が推定された属性の中で関連度が高い上位3つの属性、右下が推定された属性から導かれた物体の尤もらしい上位3つである。この時、特徴抽出も含めた、全ての処理の所要時間は0.3秒未満であった。認識結果としては、8種類の全ての認識物体に対して同様の認識処理を行なってみたところ、9割以上の確率で正確に推定することができた。

本システムは、“画像センシングシンポジウム(SSH2012)”[32]と“画像の認識・理解シンポジウム(MIRU2012)”[33]や、複数回の講演や展示にて実演デモを実施した。そこで来場者の身に着けている物を瞬時に認識できることを披露し、その安定性を確認した。認識率に加えて、それぞれ時期が異なっていたため、検索結果の変動に対する頑健性も確認した。この

図11 インターネット上の画像から学習し、未学習の腕時計を与えた時の結果
左：認識画像、右：推定結果（上が属性、下が物体）

5 詳細について、4.3節で考察している

6 学習画像を全て学習すると、[10]の認識率の方が若干高い

7 Google Images (<http://www.google.com/imghp>)

システムの実現により、人が物体と属性に関する関係性（キーワードと、その物体に関する概念）を与えるだけで、自動的にシステムが知識を獲得して、それを用いてカメラの前のものを認識する基盤技術が確立できたと言える。また、物体と属性に関する関係性を自動で獲得する方法も現在検討されている[9]。このような手法と組み合わせることで、完全に“キーワード”のみでシステムが知識を獲得できる技術が実現できると考えている。

4.3 属性の追加と削除

提案手法は、属性の追加が容易に可能である。従来手法のAT-SOINNやDAPなどで属性を追加させる場合には、一から新しく学習器を作成する必要がある。一方で、提案手法ではSOINNが保有しているラベルの追加で対応する。また、類似した学習データに関する既存のネットワークが存在すれば、ラベルの拡散により瞬時に対応できる。これは、新たに学習する属性とパターンの似通った属性が、既に学習済みである時に効果的である。ある知識を受けると、ネットワーク上で繋がっているノードに対してもラベル情報を拡散する。つまり、新たに学習する属性の知識を与えられた時、類似の属性により構築されたネットワークを用いてラベルを拡散することで、少ない学習回数で早く学習が完了する。

図12に具体的な例を示している。上図のような属性を学習すると仮定し、現在左下図のようなネットワークをSOINNが保持しているとする。ここで、新たな属性Bについて学習するために、右下図のような入力があったとする。すると、まずノード1に対して属性Bがポジティブである情報が学習される。その後、ラベル拡散により、ノード2から7に対しても同じよう

にポジティブである情報が伝播される。つまり、ノード2から7の近傍には入力されていないにも関わらず、学習が可能となる。属性Bの右側の小さな2つの分布については、事前にネットワークが形成されていないため、一から学習する必要が存在する。このように、ラベル拡散を行うことにより属性の追加に対して早期に対応可能である。なお、その一方で、この効果の副作用として、パターンの似ている場所に異なった構造（ネットワーク）を持つ知識を学習しにくいという点が存在する。この点が、学習器を重ねた時の難しさであると考えられる。

また、重複した属性の削除も容易に可能である。2つの属性の重複度を調べるとき、従来手法では2つの学習器(AT-SOINNでは4つのネットワーク：それぞれのpos-SOINNとneg-SOINN)を参照する必要がある。ところが、提案手法では1つのSOINNが保有しているラベルの類似度を調べることで対応ができ、少ない計算コストで実行可能である。属性を減らすことで、過学習の抑制や認識時間の短縮、情報量の削減などが期待できる。

4.4 曖昧性のある属性

提案手法には、従来手法の[10]と比べて属性の定義の仕方に柔軟性がある。具体的には、動物と属性の関連度を連続値で定義可能である。従来手法では、属性に対して“関連あり”と“関連なし”のそれぞれについてネットワークを作成するため、連続値の属性を扱うことが出来ない。一方で提案手法では、連続値で定義された属性をそのまま扱うことが可能である。また、[7]では回帰手法を用いることで対応可能であると述べられている。しかし、回帰処理を用いることで計算時間の増加が懸念される。一方で、提案手法では、提案したネットワーク構成で対応可能である。なお本稿の比較実験では、比較のために、二値を利用した。連続値による属性の定義は、詳しい物体の定義に繋がり、複雑な実世界の物体に対して特に重要であると考えられる。

4.5 今後の課題

今後の課題として、SOINNや本稿で提案するSOINNの理論的な解析が挙げられる。本稿では、SOINNが理論的にどのようなネットワークを作成しているのか、作成されたネットワークをどのような理論で活用しているのかまでは言及していない。我々が最も効果的であると考えたネットワーク構築方法を用いている。そのため、今後SOINNの理論的な解析が必要であると考えており、それにより本手法の適用可能範囲

図12 ラベル拡散による効果の例

上図：属性に関する真の確率密度（濃淡で強弱）

下図：属性Aを十分学習した後に属性Bを新たに学習した時のSOINN

表3 それぞれの手法の特徴についての比較

Method	Proposed	[10]	SVM [7]
Online learning	○	○	×
Speed of learning and test	○	△	×
Easy to add and delete attributes	○	×	×
To define a real valued attribute	○	×	△

や更なる改善点などの発見に繋がり、認識率の向上も可能であると考えている。

最後に、提案手法と従来手法が持つ特徴について表3にまとめる。

5. おわりに

本稿では、オンラインで追加学習可能なSOINN [12] に教師情報に関する統計情報を算出する拡張を行い、新しいSOINNを提案した。次に、提案したSOINNを用いた属性ベースの軽量な学習と転移手法を提案した。この手法を用いることで、従来手法に比べて、非常に高速な学習と認識を実現した。更には、SOINNが持つノイズ耐性により、インターネット上のデータも扱うことができ、人が学習データを準備せずに学習する知的発達システムを構築した。また、柔軟な属性の追加や削除、曖昧な属性の定義についても議論した。我々は、本手法が人の生活環境下で人間や実環境、インターネットとの相互作用から追加的に知識を獲得する知的発達機構の実現に必要な基盤技術の一つであると考えている。

謝辞

本研究は独立行政法人科学技術振興機構のCRESTから支援を頂きました。記して感謝致します。

参考文献

- [1] M. Ito, K. Noda, Y. Hoshino, and J. Tani: "Dynamic and interactive generation of object handling behaviors by a small humanoid robot using a dynamic neural network model", *Neural Networks*, Vol.19(3), pp.323-337, Apr. 2006.
- [2] R. Fergus, P. Perona, and A. Zisserman: "Object class recognition by unsupervised scale-invariant learning", *IEEE Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.264-271, Jun. 2003.
- [3] H. Schneiderman, and T. Kanade: "A statistical approach

- to 3D object detection applied to faces and cars", *IEEE Conference on Computer Vision and Pattern Recognition*, No.CMU-RITR-00-06, May. 2000.
- [4] P. Viola, and M. Jones: "Rapid object detection using a boosted cascade of simple features", *IEEE Conference on Computer Vision and Pattern Recognition*, Vol.1, pp.511-518, 2001.
- [5] M. Weber, M. Welling, and P. Perona: "Unsupervised learning of models for recognition", *European Conference on Computer Vision*, pp.1832, 2000.
- [6] I. Biederman: "Recognition by components—a theory of human image understanding", *Psychological Review*, Vol.94, pp.115-147, 1987.
- [7] C. H. Lampert, H. Nickisch, and S. Harmeling: "Learning to detect unseen object classes by between-class attribute transfer", *IEEE Conference on Computer Vision and Pattern Recognition*, pp.951-958, Jun. 2009.
- [8] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar: "Attribute and simile classifiers for face verification", *IEEE International Conference on Computer Vision*, Oct. 2009.
- [9] M. Rohrbach, M. Stark, G. Szarvas, I. Gurevych, and B. Schiele: "What helps where — and why? semantic relatedness for knowledge transfer", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 910-917, Jun. 2010.
- [10] A. Kawewong, and O. Hasegawa: "Fast and incremental attribute transferring and classifying system for detecting unseen object classes", *International Conference on Artificial Neural Networks*, Vol.6354, pp.563-568, 2010.
- [11] Corinna Cortes, Vladimir Vapnik: "Support-vector networks," *Machine Learning, Volume 20, Issue 3*, 1995.
- [12] F. Shen, and O. Hasegawa: "A fast nearest neighbor classifier based on self-organizing incremental neural network", *Neural Networks*, Vol.21, No.10, pp.1537-1547, Dec. 2008.
- [13] L. Bottou, and O. Bousquet: "The Tradeoffs of Large Scale Learning", *Optimization for Machine Learning*, pp.351-368, 2011.
- [14] A. K. Jain, R. P. W. Duin, and J. Mao: "Statistical pattern recognition: a review", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.22, No.1, pp.4-37, Jan. 2000.
- [15] T. Kohonen: "Self-organized formation of topologically correct feature maps", *Biological Cybernetics*, Vol.43, No.1, pp.59-69, Jan. 1982.
- [16] B. Fritzke: "A growing neural gas network learns topologies", *Advances in Neural Information Processing Systems*, MIT Press, 1995.
- [17] F. Shen, and O. Hasegawa: "An incremental network for on-line unsupervised classification and topology learning", *Neural Networks*, Vol.19, No.1, pp.90-106, Jan. 2006.
- [18] S. Okada, and O. Hasegawa: "Incremental learning, recognition, and generation of time-series patterns based on self-organizing segmentation", *Journal of Advanced Computational Intelligence and Intelligent*

- Informatics*, Vol.10, No.3, pp.395–408, 2006.
- [19] F. Shen, T. Ogura, and O. Hasegawa: “An enhanced self-organizing incremental neural network for online unsupervised learning”, *Neural Networks*, Vol.20, No.8, pp.893–903, Oct. 2007.
 - [20] F. Shen, H. Yu, K. Sakurai, and O. Hasegawa: “An incremental online semi-supervised active learning algorithm based on self-organizing incremental neural network”, *Neural Computing and Applications*, Vol.20 No.7, pp.1061–1074, Oct. 2010.
 - [21] F. Shen, and O. Hasegawa: “Self-organizing Incremental Neural Network and its Applications”, *International Joint Conference on Neural Networks*, Jun. 2009.
 - [22] A. Sudo, A. Sato, and O. Hasegawa: “Associative memory for online learning in noisy environments using self-organizing incremental neural network”, *IEEE Trans. Neural Networks*, Vol.20, No.6, pp.964–972, Jun. 2009.
 - [23] F. Shen, A. Sudo, and O. Hasegawa: “An online incremental learning pattern-based reasoning system”, *Neural Networks*, Vol.23, No.1, pp.135–143, Jan. 2009.
 - [24] X. He, R. Kojima, and O. Hasegawa: “Developmental word grounding through a growing neural network with a humanoid robot”, *IEEE Trans. Systems, Man, and Cybernetics -Part B*, Vol.37, No.2, pp.451–462, Apr. 2007.
 - [25] X. He, T. Ogura, A. Satou, and O. Hasegawa: “Developmental word acquisition and grammar learning by humanoid robots through a self-organizing incremental neural network”, *IEEE Trans. Systems, Man, and Cybernetics -Part B*, Vol.37, No.5, pp.1357–1372, Oct. 2007.
 - [26] D. G. Lowe: “Object recognition from local scale-invariant features”, *International Conference on Computer Vision*, Vol.2, pp.1150–1157, Sep. 1999.
 - [27] H. Bay, T. Tuytelaars, and L. V. Gool: “Surf: Speeded up robust features”, *European Conference on Computer Vision*, Vol.3951, pp.404–417, 2006.
 - [28] A. Bosch, A. Zisserman, and X. Munoz: “Representing shape with a spatial pyramid kernel”, *The 6th ACM International Conference on Image and Video Retrieval*, pp.401–408, 2007.
 - [29] K. E. A. Sande, T. Gevers, and C. G. M. Snoek: “Evaluating color descriptors for object and scene recognition”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.32, No.9, pp.1582–159, 2008.
 - [30] E. Shechtman, and M. Irani: “Matching local self-similarities across images and videos”, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1–8, Jun. 2007.
 - [31] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray.: “Visual categorization with bags of keypoints”, *In Workshop on statistical learning in European Conference on Computer Vision*, pp.1–24, 2004.
 - [32] 木村大毅, Kankuekul Pichai, 長谷川修: “インターネットを活用した超高速オンライン転移学習”, 画像センシングシンポジウム (SSII), Jun. 2012.
 - [33] 木村大毅, Kankuekul Pichai, 長谷川修: “ネット画像を用いた超高速オンライン転移学習による未知物体推定”, 画像の認識・理解シンポジウム (MIRU), Aug. 2012.

(2013年11月26日 受付)

(2014年 7 月30日 採録)

[問い合わせ先]

〒226-8503 横浜市緑区長津田町4259 J3-13

東京工業大学 大学院総合理工学研究科

知能システム科学専攻 長谷川修研究室

木村 大毅

E-mail : kimura.d.aa@m.titech.ac.jp

Homepage : <http://haselab.info>

著者紹介

きむら だいき
木村 大毅 [非会員]

2011年，東京工業大学工学部情報工学科卒業，2012年，同大学院総合理工学研究科知能システム科学専攻博士前期課程修了，修士(工学)。現在，同大同研究科同専攻博士後期課程在籍中，及び，日本学術振興会特別研究員，画像認識，ヒューマノイドロボットの知能情報処理などの研究に従事，日本ロボット学会，計測自動制御学会各学生会員。

Kankuekul PICHAI [非会員]

2010年，職業能力開発総合大学校東京校卒業，2012年，東京工業大学大学院総合理工学研究科知能システム科学専攻博士前期課程修了，修士(工学)。同年，株式会社クロスコンパスに就職，現在，同社勤務中。

はせがわ おさむ
長谷川 修 [正会員]

1993年，東京大学大学院博士課程修了，博士(工学)。同年，電子技術総合研究所入所，1999年米国カーネギーメロン大学客員研究員，2001年，産業技術総合研究所主任研究員，2002年，東京工業大学大学院理工学研究科附属情報工学研究施設助教授，2010年，同准教授，現在に至る，SOINN株式会社代表取締役(兼職)，パターンベース人工知能，知能ロボットなどの研究に従事，電子情報通信学会，人工知能学会，情報処理学会，日本ロボット学会，IEEE CS 等各会員。

Guessing an Unknown Class by Online Fast Attributes Learning and Transfer

by

Daiki KIMURA, Pichai KANKUEKUL and Osamu HASEGAWA

Abstract :

Towards building an intelligent robot, we have to create an autonomous mental development system that incrementally and speedily learns from humans, their environments, and the Internet. In this paper, we propose an ultra-fast and online incremental attributes learning and transfer method using new SOINN; stand for Self-Organizing and Incremental Neural Network. We conducted a comparison experiment with previous methods. Based on these results, our proposed method can keep an equivalent recognition rate of an online method, and shorten the learning time and test time. And another advantage is to be able to use the Internet data, and be dened real valued attributes.

Keywords : Attribute-based, Online incremental learning, Unknown class, Robot

Contact Address : **Daiki KIMURA**

*O. Hasegawa laboratory, Interdisciplinary Graduate School of Science and Engineering,
Tokyo Institute of Technology
J3-13, 4259, Nagatsuta-cho, Midori-ku, Yokohama, 226-8503, Japan
E-mail : kimura.d.aa@m.titech.ac.jp
Homepage : <http://haselab.info>*