Learning from Cluster Examples

(クラスタ例からの学習)

博士論文 日本語概略説明

神嶌 敏弘

Abstract

クラスタ例からの学習（Learning from Cluster Examples; LCE）とは，一般的な学習タスクであるクラスタリングと例からの学習とを組合わせた新たな学習タスクである．これは，未知の分類対象集合に対する真の分割を求める規則を事例集合から獲得するタスクである．ただし，事例集合の各要素は分類対象集合とそれに対する真の分割の具体例との対で，真の分割とは可能な分割の中で最も利用者の意図に適した分割である．また，未知の分類対象集合とは学習時の事例集合に含まれないものをいう．

この問題の学習手法は，適切な分割そのものを示すことは容易だが，それを導く規則を明示することが困難な状況で有用である．今まで，このような状況ではクラスタリングの手法を無理に適用することが一般的だったが，この適用の問題点を指摘し，LCE 問題の手法でこれらの問題点がどのように解決されるかについて議論する．本研究では，LCE 問題を形式的に定義し，学習手法を開発する．さらに，ドット・パターンの分割とベクトル画像のセグメンテーションにこの手法を適用する実験によって，その有効性を示す．

1 Introduction

新たな学習タスク『クラスタ例からの学習（Learning from Cluster Examples; LCE）』を定式化し，学習方法を開発し，その有効性を示す．

LCE は，未知の（学習時の事例集合には含まれない）分類対象集合に対する真の分割を導く規則を，事例集合から獲得する問題である．ただし，真の分割とは全ての分割の中で利用者の目的に最も適した分割をいう．また，事例集合は，分類対象集合とその集合に対する真の分割の対である事例の集合である．

LCE は，クラスタリングと例からの学習という既存の二つの学習タスクを合成したものとも見なせる．クラスタリングとは，クラスタと呼ぶ部分集合に，事前に定めた基準や規則に基づいて，分類対象集合を分割するタスクである．一方，例からの学習とは，真のクラスへ分類する規則を訓練事例集合から獲得するタスクである．LCE は，事例集合から規則を獲得する点では例からの学習と同じだが，その規則の目的はクラスタリングのように分類対象集合を分割することである．よって，この学習タスクを『クラスタ例からの学習』と名付けた．

この LCE 手法は，真の分割そのものを示すことは容易だが，それを導く規則を明示することが困難な，画像のセグメンテーションといった状況で有用である．今まで，このような状況ではクラスタリング手法を無理に適用することが多かった．クラスタリング手法によって真の分割を導くためには，分割の基準や規則を事前に人手で試行錯誤によって調整しなくてはならない．この無理な調整の問題点を指摘し，LCE 手法の適用でこの問題点がどう解消されるかについて議論する．

この LCE タスクを定式化し，その解決手法を開発した．その手法で，ドット・パターンの分割とベクトル画像のセグメンテーションを対象に実験を行う．実験により，真の分割の獲得に必要な情報が十分に学習できているかを検証し，LCE 手法の特性を調査する．

2 An Overview of Learning from Cluster Examples

LCE には，事例集合から分割用の規則を獲得する学習段階 (図 1 左) と，その規則を未知の対象に適用する分割段階 (図 1 右) とがある．学習段階では，分類対象集合 O_I とこの集合に対する真の分割の具体例 π_I^* の対である事例の集合 EX から，分割用の規則を学習する．分割段階では，この規則を未知の分類対象集合 O_U に適用し，推定分割 $\hat{\pi}_U$ を得る．

この学習は，真の分割そのものを示すことは容易だが，その分割を得るための規則は容易には獲得できない状況で有用である．この状況の例として，画像を “意味のある” 領域に分割するというセグメンテーション問題がある．この問題には，今までは，分割の基準や規則を手作業で与えることで，クラスタリング手法が適用されてきたが，これには以下の短所がある．

- 真の分割の獲得に有用な特徴は，設計者が感覚的に発見しなくてはならない．そのため，対象領域への深い洞察と，多大な試行錯誤を要する．
- 有用な特徴を発見しても，その性質が多様で複雑なため，具体的な規則として表現で

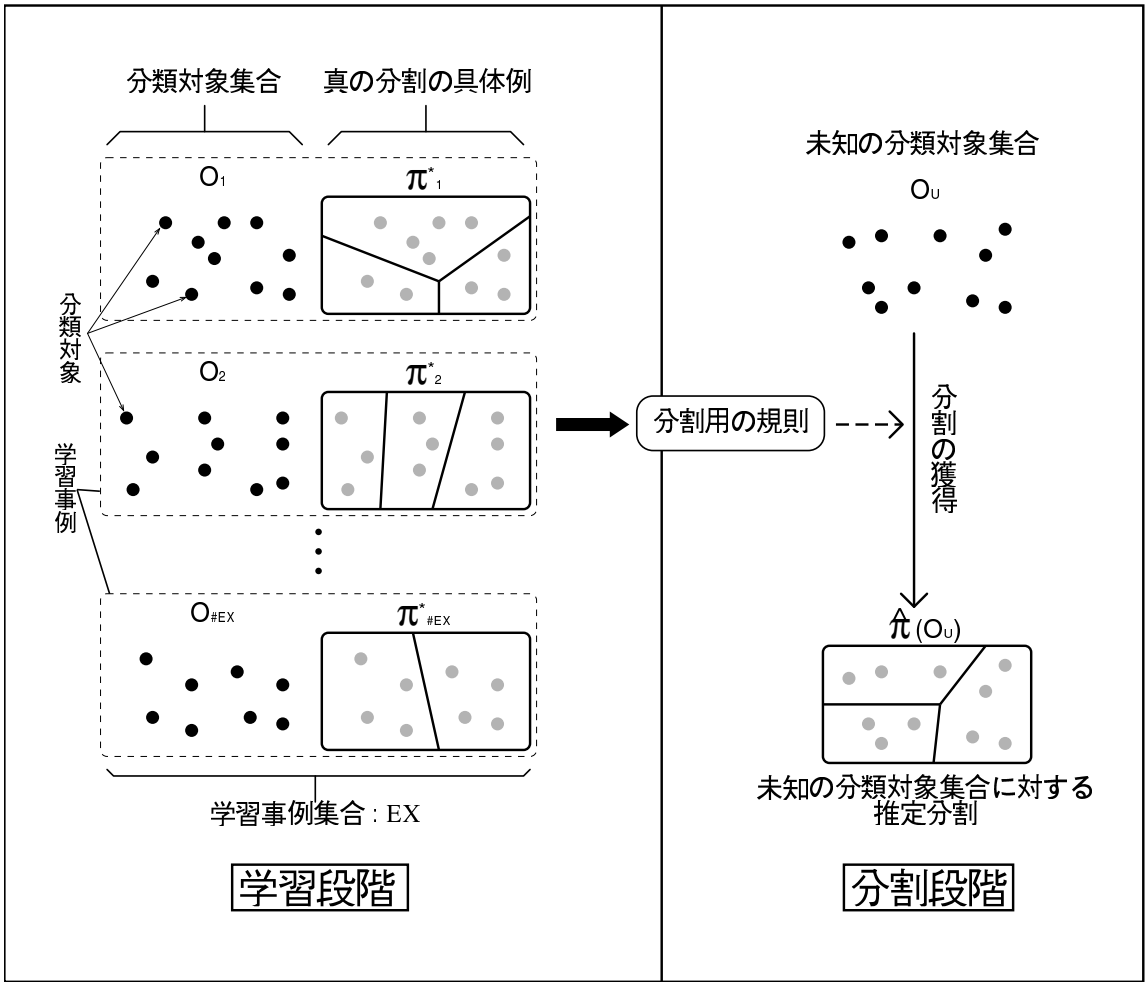

図 1: クラスタ例からの学習の概要

きない場合がある．

- 分割すべき分類対象集合には多様性があり，それに対応するため規則に調整できる部分が残こることが多い．そのため，規則の設計者だけでなく利用者にも対象領域に対する十分な知識が要求される．

- 分割結果が統計的に不安定である．理由は二つあり，一つは，規則の獲得が手作業であるため，そのときに参照する事例と，テスト事例とを明確に区別できないので，性能を公正に評価できないからである．もう一つは，人の認知能力によって規則の獲得に利用できる情報量が制限されるからである．

それに対し，LCE 手法を用いれば，以下のような改善が予測される．

- 感覚に依存することなく，具体的な分割の例が提示できれば規則は獲得可能である．

- 例からの学習では，複雑で多様な状況を具体化した規則が獲得されている．同様の効果が LCE にも期待される．
- 例からの学習では，利用者に調節の必要がない規則が獲得されており，LCE にも同様の効果が期待される．
- アルゴリズムが規則を自動的に獲得するので，訓練事例とテスト事例を明確に分離でき，性能が公正に評価される．また，計算資源の量によってのみ訓練事例の数は制限されるので，より多くの情報を考慮した規則を獲得できる．これらのことにより，統計的安定性が向上する．

画像認識の他にも，LCE は，多戦略学習，遺伝子予測，データマイニングなどの分野でも，有用であると考える．

3 Formalization of Learning from Cluster Examples

クラスタ例からの学習は以下のように定式化される．

事例集合 EX は $\#EX$ 個の事例からなる集合 $\{(O_1\,,\pi_1^*)\,,(O_2\,,\pi_2^*)\,,\ldots,(O_{\#EX}\,,\pi_{\#EX}^*)\}$ であり，各事例は分類対象集合 O_I と，これに対する真の分割の具体例 π_I^* の組である．分類対象集合 O は分類対象 o^i からなる集合であり，分割 π は，O の排他的で網羅的な部分集合であるクラスタの組 $\{C^1\,,C^2\,,\ldots,C^{\#\pi}\}$ である．また，O 中の分類対象 o^i と o^j の対を分類対象対と呼び，p^{ij} で表し，O の要素から作ることのできる全ての分類対象対の集合を P としておく．

分類対象集合は，その集合の一部分と属性ベクトルを関連付け，ベクトルにその部分の特徴を記述させることで表現する．これらの属性ベクトルは，どの部分と関連付けられるかによって以下の三種類に分けられる．

$A(o^i)$**:分類対象属性** 分類対象 o^i に関連付けられたベクトル．分割の状態には依存せず，分類対象集合のみに依存して値が決まる．ドットパターンでは点の位置などを表すために用いられる．

$A(p^{ij})$**:分類対象対属性** 分類対象対 p^{ij} に関連付けられたベクトル．$A(o^i)$ と同様，分類対象集合のみに依存．ドットパターンでは点の間の距離などを表すために用いられる．

$A(\pi)$**:分割全体の属性** 分割全体に一つだけ関連付けられるベクトル．分類対象集合のみに依存する他の属性ベクトルの値と，分割の状態から属性値が計算される．ドットパターンではクラスタの数などを表すために用いられる．

真の分割を導く規則は，これら三種類の属性を利用して記述される．

4 The Partitioning Method

与えられた分類対象集合に対する可能な分割の中から，分割の評価関数を最大にする分割を探索し，これを推定分割とする．この評価関数は，MAP 原理に基づくもので，未知の分類対象集合についてある分割 π を定めたとき，その分割が真の分割であるという事象と，この集合に関連付けられた全ての属性値ベクトルとの次の結合確率密度である．

$$\mathsf{Pr}[\pi{=}\pi^* \, , A(\pi); \{A(o)\} \, , \{A(p)\}] \tag{1}$$

ただし，$\pi{=}\pi^*$ は π が真の分割であるという事象で，$\{A(o)\}$ と $\{A(p)\}$ はそれぞれ，O と P の要素に関連付られた属性値ベクトル全体の集合である．また，$\{A(o)\}$ と $\{A(p)\}$ は O が与えられたときに値が決まっているので，$A(\pi)$ とは区別して前提条件やパラメータのように扱う．この結合確率は，属性値ベクトルの数が変動するため，従来の統計的推定手法で計算することが困難である．そこで，この結合確率は次の確率と確率密度の積に比例するものと考える．

$$\mathsf{Pr}[\pi{=}\pi^*; \{A(o)\} \, , \{A(p)\}] \tag{2}$$

$$\mathsf{Pr}[A(\pi)|\pi{=}\pi^*] \tag{3}$$

式 (2) は，次の関数を利用して計算する．

$$f_1(p^{ij}) = \mathsf{Pr}[\mathsf{in}(p^{ij} \, , \pi^*) = 1; A(p^{ij}) \, , A(o^i) \, , A(o^j)]$$

ただし，$\mathsf{in}(p \, , \pi)$ は，分類対象対 p が分割 π の同じクラスタの要素であるとき 1，そうでないとき 0 をとる関数である．

この $f_1(p)$ を学習段階で獲得しておき，Dempster&Shafer の確率結合則 (D&S 則と略) を用いて統合することで，式 (2) の定数倍が計算できる．D&S 則とは，異なる証拠に基づいた確率を統合する規則である．この D&S 則を適用するために，$\mathsf{in}(p^{ij} \, , \pi^*) = 1$ となる事象は，あらゆる可能な分割のうち分類対象対 p^{ij} が同じクラスタの要素となっている分割のいずれかが生じる事象の言い換えであることに注目する．この言い換えで，$f_1(p)$ は p が異なっていても，可能な全ての分割という同じ集合についての確率とみなせるようになる．一方，p^{ij} に関する前提条件 $A(p^{ij})$，$A(o^i)$ 及び，$A(o^j)$ を，確率分布を決定する証拠とみなす．前提条件と証拠は厳密には異なるが，どちらも，確率分布の決定に関わる点では同じであるので，同一視できると考える．また，p^{ij} に関する証拠に基づく確率を，分類対象対集合の全ての要素について求めて，それら統合した確率は，$\{A(o)\}$ と $\{A(p)\}$ の全ての要素を証拠とする確率と見なせる．これらの考えをもとにすると，式 (2) の定数倍が次式で計算できる．

$$\prod_{p \in P^+} f_1(p) \times \prod_{p \in P^-} \bar{f}_1(p))$$

ただし，P^+ は分類対象対集合の要素のうち $\mathsf{in}(p \, , \pi) = 1$ となるものからなる集合で，P^- はそれ以外の要素からなる集合であり，$\bar{f}_1(p) = 1 - f_1(p)$ である．

もう一方の，確率密度 (3) を簡単に $f_2(A(\pi))$ と表記する．これも学習段階で獲得しておく．

$f_1(p)$，$f_2(A(\pi))$ の獲得方法は次章で述べる．

次に，上記の評価関数を最大にするような分割を探索する方法について述べる．最初に，各クラスタに分類対象が一個ずつだけ含まれる初期分割を作成する．この分割に，次の二種類の操作を適用してできる分割の中で，評価関数を最大にするものを次の時点の分割とする．

1. 二個のクラスタを一個のクラスタに併合
2. あるクラスタの要素一つを別のクラスタに移動

そして，これらの操作では評価関数を大きくする分割を発見できなかったときに探索を終了し，そのときの分割を最終的な推定分割とする．

5 The Learning Methods

分割段階では関数 $f_1(p)$ と $f_2(A(\pi))$ が必要であるが，これらは学習段階で以下のように獲得する．

はじめに，$f_1(p)$ の推定方法について述べる．学習事例 $(O_I\,,\pi_I^*)$ の O_I の分類対象対 p_I^{ij} に対し，この対が分割 π_I^* において同じクラスタの要素となるかどうかを表す値 $c=\mathsf{in}(p_I\,,\pi_I^*)$ を計算する．さらに，この分類対象対が関連する三つの属性値ベクトル，$A(p_I^{ij})$，$A(o_I^i)$，及び $A(o_I^j)$ を合成した属性値ベクトル $A_C(p_I^{ij})$ も計算する．この c と $A_C(p_I^{ij})$ の対を新たな学習事例とする．この変換した学習事例を EX 中の全ての事例の，全ての分類対象対について求める．この変換した学習事例から，$\mathsf{in}(p_I\,,\pi_I^*)=1$ となる確率を求める問題は，通常の例からの学習問題とみなせる．本研究では，MDL 基準に基づき，この確率を推定するアルゴリズムを開発し，適用した．

次に，$f_2(A(\pi))$ の推定について述べる．各学習事例 $(O_I\,,\pi_I^*)$ について，O_I を π_I^* に分割したときの分割全体の属性値ベクトル $A(\pi_I^*)$ を計算する．この属性値ベクトルを，EX の全ての事例について計算する．この属性値ベクトルの集合から，確率密度関数 $\mathsf{Pr}[A(\pi)\,,\pi{=}\pi^*]$ を推定する．$\pi{=}\pi^*$ は均一分布に従うとすると，この密度関数は $f_2(A(\pi))$ に比例する．この密度関数は回帰木を用いて表され，MDL 基準に基づき適切な回帰木を推定するアルゴリズムを開発し，適用した．

6 Experimental Domains and Testing Methods

Experimental Domains: Dot Patterns

ドットパターンとは，平面にドットが分布している分類対象集合で，クラスタリング手法の実験的検証によく利用される．各分類対象集合は 50 個のドットから構成されており，2〜4 個のクラスタに分割されている．各クラスタ内ではドットは円形のガウス分布に従って分布している．EM アルゴリズムを用いてクラスタリングした結果との比較によって，未知の分類対象集合に対する真の分割を，LCE 手法で導くことができることを確認する．

Experimental Domains: Vector-Data Images

ベクトルデータ画像は，5 種類の図面部品で構成された論理回路の図面をスキャナーでビットマップ画像とし，それを細線化・ベクトル化といった処理によって，直線によって構成されるベクトルデータと呼ばれる画像に変換したものである．これらの画像について，一つの図面部品を表す直線の集合が各クラスタとなるような分割を求めた．ドットパターンより実用的な分割推定問題であるこの問題に LCE 手法を適用することによって，本手法の有効性を調査する．

A Testing Method

LCE によって，どれだけ真の分割に近い分割が推定されたかの評価方法について述べる．

この評価の方法には，厳密な交叉確認法である，leave-one-out 法を用いた．すなわち，K 個の事例を含む事例集合について，K 分割の交叉確認を行う方法である．各事例 $(O_I\,, \pi_I^*)$ について O_I の推定分割 $\hat{\pi}_I$ を獲得し，π_I^* との RIL を求める．この RIL を EX の全ての要素について求め，これらの平均で，真の分割を推定する性能を評価した．

RIL(Ratio of Information Loss) とは，真の分割が生じる事象を Π^*，推定分割が生じる事象を $\hat{\Pi}$ と記したとき，$RIL = H(\Pi^*|\hat{\Pi})/H(\Pi^*)$ で表される．ただし，$H(\Pi^*)$ は事象 Π^* の事前エントロピーで，$H(\Pi^*|\hat{\Pi})$ は推定分割を知ったあとの事後エントロピーである．RIL は 0 から 1 の間の値をとり，0 のときに完全に二つの分割が一致する．

7 Experimental Results and Discussions

Testing Using Dot Patterns

- EM アルゴリズムと LCE 手法による分割との比較
  EM アルゴリズムによる分割と対等な分割を導くことができる規則を LCE 手法で獲得できた．

- 事例数が LCE 手法の性能に与える影響の調査
  事例数の増加にともない，より正確な分割を導く規則を獲得できることを確認した．

- 分割全体の属性を採用する利点を確認する実験
  RIL に対する調査では差は見られなかったが，どれだけ正確なクラスタ数を推定できたかの検証では，分割全体の属性を採用した方が正確な推定ができた．

- 参考のため代表的な分割結果を表示
  EM アルゴリズムと LCE 手法それぞれの場合で，RIL が最大になる事例の結果を図示した．

Testing Using Vector-data Images

- 分割全体の属性を採用する利点を確認する実験
  分割全体の属性を採用すると推定分割の RIL が有意に減少することを確認した.
- 参考のため代表的な分割結果を表示
  推定分割の RIL が最小，中央値，および，最大である事例の分割結果を図示した.

Discussions

2 章にて，利用者の意図にそった分割をクラスタリング手法で無理に導くことの問題点を四つ指摘した．その四つの問題点が，LCE により以下のように解決された.

- ドットパターンの実験において，EM アルゴリズムによる分割と同等の精度で推定できる規則を，訓練事例から LCE 手法で獲得できた．よって，設計者の感覚に依存した特徴の発見は不要で，分割の具体例を与えることができれば分割用の規則を獲得できた.
- ベクトルデータ画像の分割問題では，“ヒゲ” や “スキマ” などの，人手では具体的な規則として表現することが困難な特徴を扱う必要がある．LCE 手法では事例集合から分割に必要な知識を得て，それらを具体的な規則へ自動的に変換できた.
- 分割用の規則を利用者が調節する必要がある場合が多かったが，LCE 手法では調節が不要な規則を獲得できた．そのため，利用者は対象領域について十分な知識がなくても，それらの規則を適用できる.
- 訓練事例とテスト事例を明確に区別して定量的で公正な性能評価を行った．また，手作業による手法では，LCE 手法のように多数の事例を扱うのは困難であった．これらによって，獲得された規則の統計的安定性が向上した.

8 Conclusions

『クラスタ例からの学習』という学習タスク新たに提案した．このタスクを定式化し，解決手法を開発し，ドットパターンとベクトルデータ画像という二つの問題で実験した．実験の結果，分割用の規則の獲得において，感覚に依存する必要がなく，複雑な特徴も具体化できた．獲得した規則には調整が不要で，その規則の適用結果は統計的に安定であった.