クラスタ例からの学習 (Learning from Cluster Examples)

神嶌 敏弘

目的

クラスタ例からの学習 (Learning from Cluster Examples; LCE)という新たな学習タスクを提起する

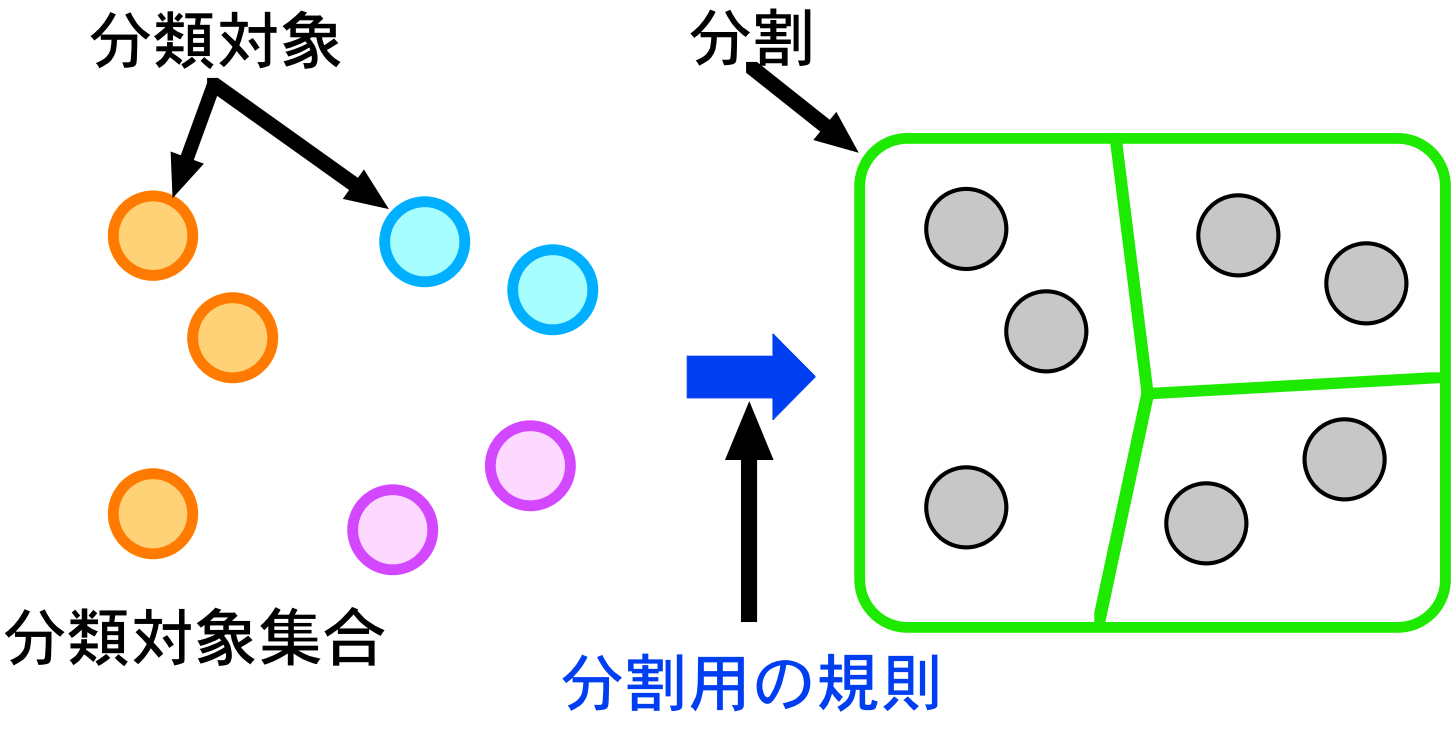
- 分類対象集合に対して適切な分割を導く規則を，事例集合から獲得するタスク
- クラスタリングと例からの学習の二つの学習タスクを組み合わせたもの

LCE問題を定式化し，その学習方法と分割結果の評価方法を提案．実験によりその有効性を示す

目次

- LCEとは何か
  - クラスタリング，例からの学習，LCE
- LCEはなぜ必要か
  - 適用範囲，従来手法の問題点，LCEによる問題点の解消
- LCEの定式化
- LCEでの分割の獲得と規則の学習
  - 分割の獲得方法，規則の学習方法
- LCEの結果の評価方法
- 実験
  - 人工的なドットパターン，実用的なベクトルデータ
- 考察
- まとめ

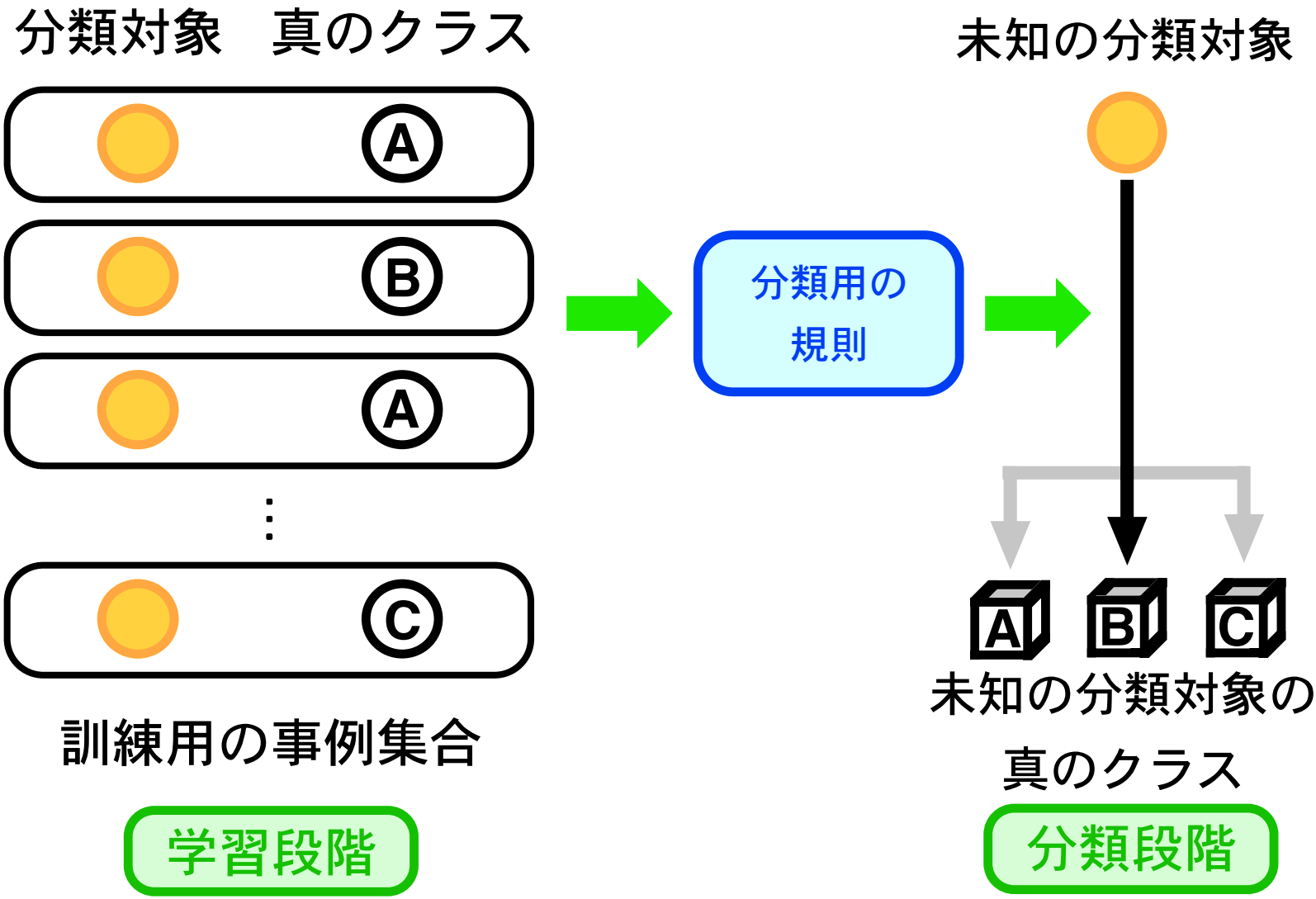
クラスタリング



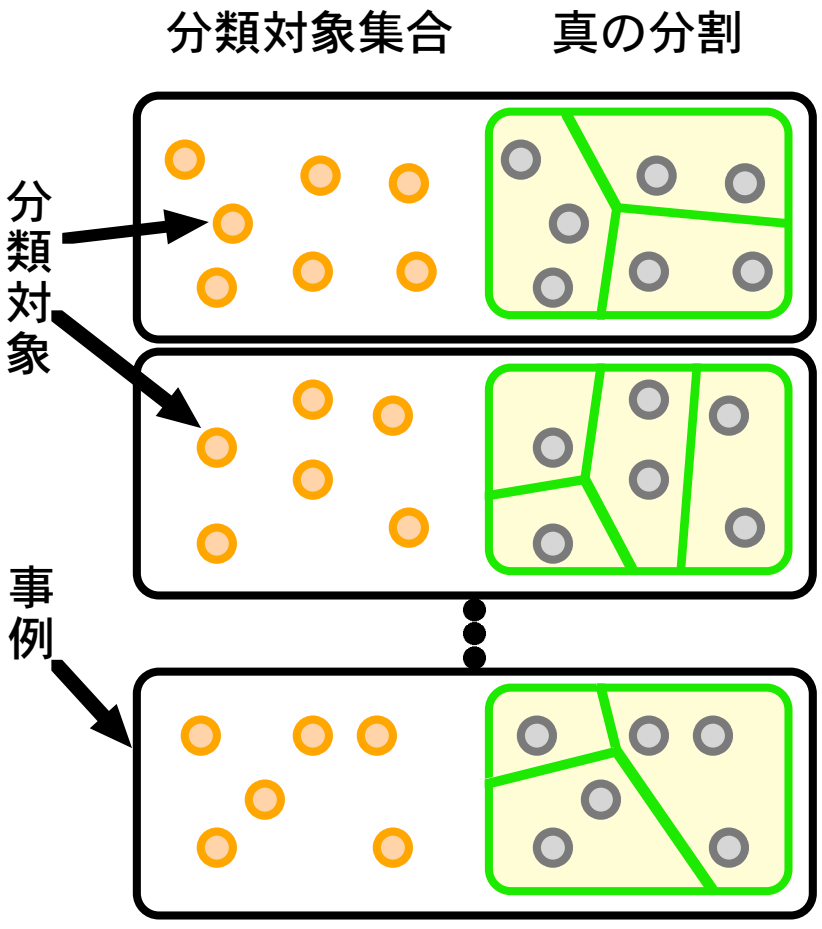
事前に定めた規則に基づき「似ているもの」を集めた部分集合（クラスタ）に分類対象集合を分割

（“似ている”ということは，評価関数や手続き自体によって定義）

例からの学習


分類対象　真のクラス
A
B
A
C
訓練用の事例集合
学習段階
分類用の
規則
未知の分類対象
A
B
C
未知の分類対象の
真のクラス
分類段階

クラスタ例からの学習(学習段階)



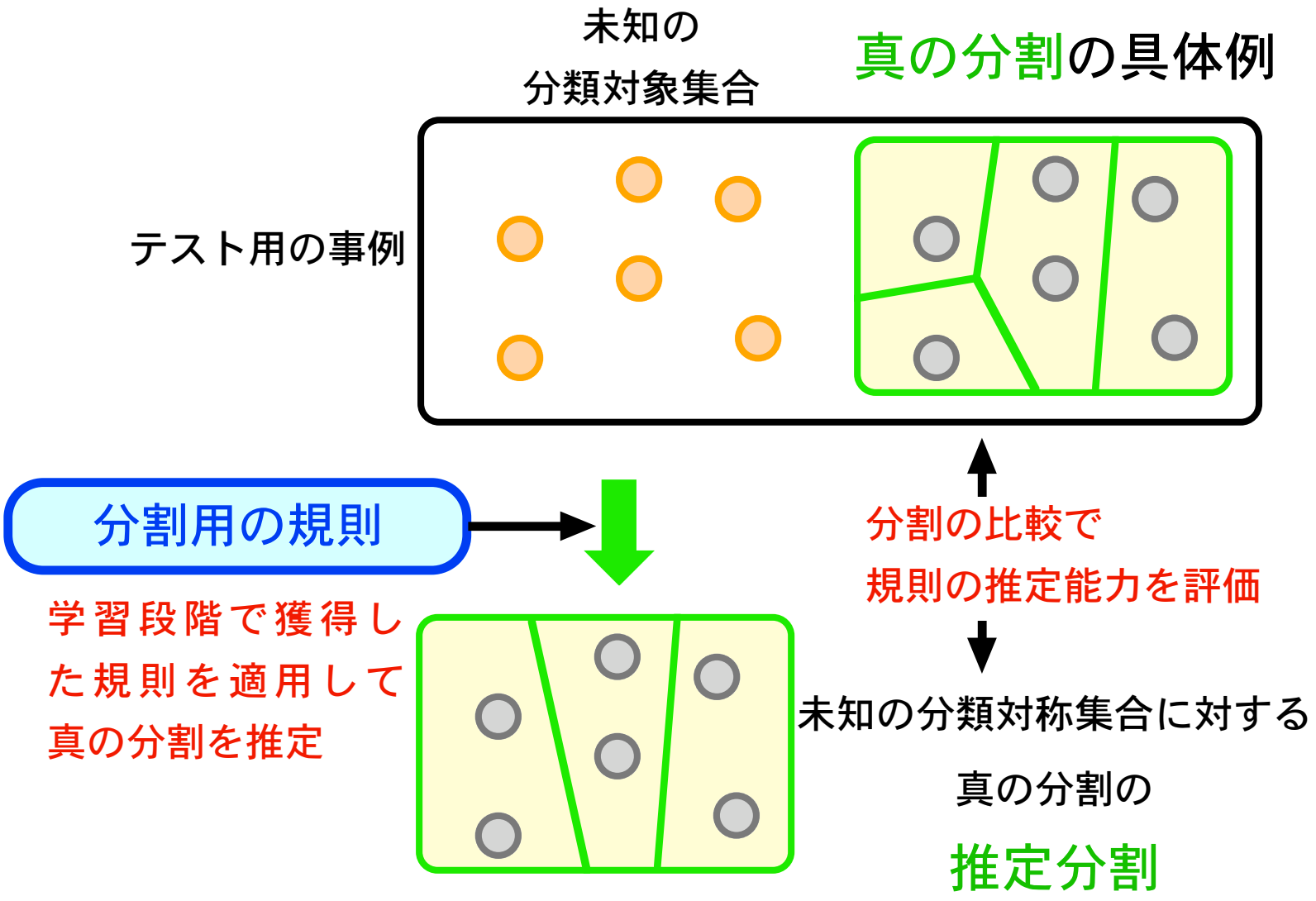
分割用の規則

クラスタ例からの学習

(Learning from Cluster Examples)

真の分割とは，利用目的にとって最も適切な分割

クラスタ例からの学習(分割段階と検証)



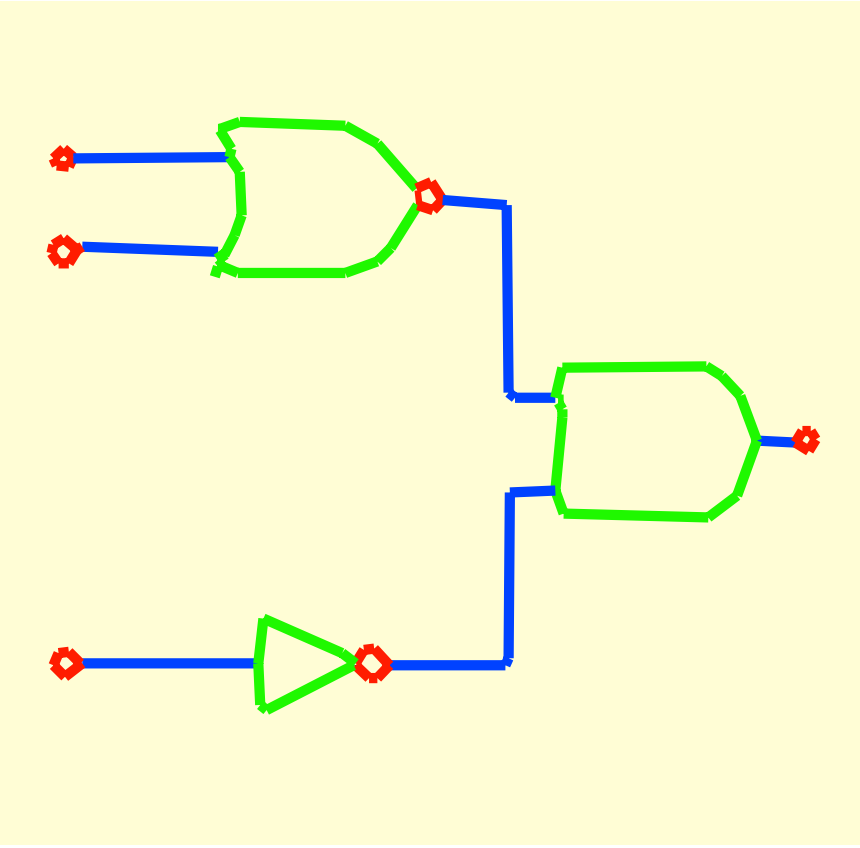
各学習タスクの関係

	演繹的タスク 規則の適用	帰納的タスク 規則の獲得
分類対象のクラスへの分類	クラス分類用の規則の適用 判別関数の適用 決定木や決定リストの適用 訓練済みニューラルネットの適用	例からの学習 線形判別分析 ニューラルネットによるクラス分類 決定木学習
分類対象集合の分割の獲得	クラスタリング 最小近隣法，*k*-means法 観察による学習	クラスタ例からの学習 本論文の手法

LCEを適用すべき問題の例

画像のセグメンテーション……画像の構成要素を何らかの意味をもつ集団ごとにまとめる操作

※画像認識の過程で利用される手法



・線分の集合で対象を表現したベクトル画像
・図面部品ごとに分割する例

LCE適用の条件

分類対象集合に対する真の分割が存在

・線分を図面の部品ごとにまとめた分割が真の分割に相当

真の分割の具体例を提示可能

・図面部品を認知するのは容易なので，真の分割は提示可能

真の分割を導く具体的記述は困難

・図面部品をまとめる評価関数や手続きを利用者が記述するのは困難

従来手法によるセグメンテーション

LCEとみなさずにこの種の問題を解決する手法

クラスタリングを利用したセグメンテーション

↓

分割の規則を具体的に定める必要

↓

しかし，分割の規則を提示するのは困難

（信号レベルと目標との概念的隔たりなどのため）

↓

利用者が試行錯誤で調整して適当に分割の規則を定める

（パラーメタや例外手続きの追加などによる）

↓

調節できればばよいが，現実にはうまくいかない

具体的記述獲得の困難さ

利用者が認知している部品の概念 

線分を表現した数値

試行錯誤で隔たりを埋めるのは困難

異なる立場での画像のセグメンテーションの定義

クラスタリングのJainの定義(獲得可能な分割)

an exhaustive partitioning of an input image into regions, each of which is considered to be homogeneous with respect to some image property of interest (e.g., intensity, color, or texture).

画像認識のKanadeの定義(獲得したい分割)

the ultimate goal of image analysis is to obtain a segmentation which separates out semantically meaningful objects or part of objects.

従来手法の例

従来手法でのセグメンテーションの具体的問題点を指摘するため，美濃の方法について説明する

美濃の方法：

美濃らは，認知的な観察をもとに，ベクトルデータ画像をシンボル候補(≒部品)に分割するために次の規則を採用

・ループに囲まれた領域
・短い線分を数多く含んだ領域

従来手法の問題点とLCEによる解決(1)

分割獲得に役立つ特徴の発見は感覚に依存

- 真の分割の獲得に役立つ特徴は感覚的に見つけられる．そのため，多くの対象領域の知識と試行錯誤が必用.
- 美濃の例では，認知学的見地の知識をもとにしているが，それを具体的に画像の特徴と結びつける点は感覚的.

LCEでは具体例の提示ができればよい

- LCE具体例の提示から規則を獲得し，感覚に依存した特徴の発見は不要

従来手法の問題点とLCEによる解決(2)

分割獲得に役立つ特徴を見つけても，その特徴を具体化できない

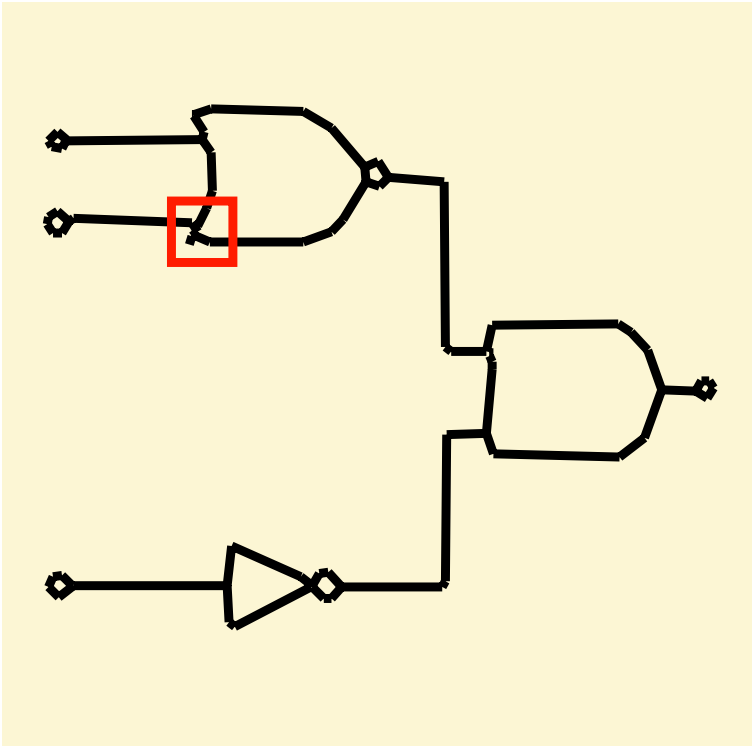
・美濃の例では，画像では例外的事象(ヒゲやスキマなど)が発生する．これらの存在を認知はしているが，いつどのように発生するかという特徴は複雑で，その特徴を具体化するのは困難.

LCEではアルゴリズムが規則を自動的に具体化

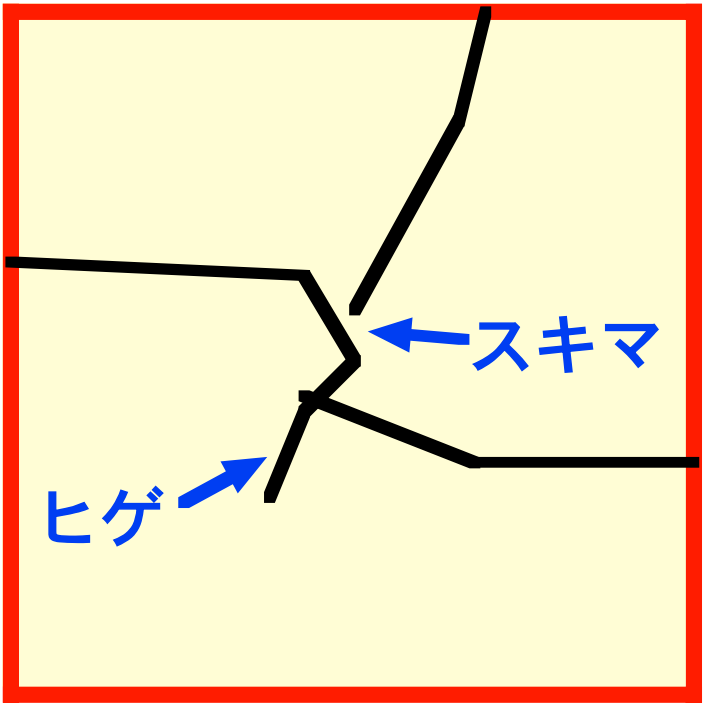
・例からの学習では，複雑な特徴を表した規則でも，事例から自動的に獲得できている．同様の効果をLCEに期待.

ベクトル画像のヒゲやスキマの例

ベクトル画像の全体図



赤枠部分の拡大図



従来手法の問題点とLCEによる解決(3)

獲得した規則を適用するときに調整が必要

- 設計者は画像の多様性に対応するために規則に調整できる部分を残す．よって，規則の設計者だけでなく，利用者も手法についての知識が必要.
- 美濃の例では“短い線分”を求める必要があるが，その短さはデータに応じて利用者が調整しなくてはならない.

LCEでは獲得した規則は調整が不要

- 多様性をもった事例集合を与えられるので，これらに対応した規則が獲得される．よって，その規則の適用時には調整は不要.

従来手法の問題点とLCEによる解決(4)

分割の結果は統計的に不安定

- 手作業での規則獲得ではテスト事例と訓練事例を厳密に分離できないので，規則の性能を公正に評価できない.
- 参照できる情報の量が人の認知能力で制限される.
- 美濃の例では，2～30程度の事例を参考に規則を獲得し，評価は結果を図示して定性的な評価を行うのみ.

分割の結果は統計的に安定

- テストと訓練事例を厳密に区別して検証可能.
- 学習アルゴリズムは，人間が扱うことが困難な多数の事例も扱える.

LCEの定式化

入力: 訓練事例 $EX = \{(O_1, \pi_1^*), (O_2, \pi_2^*), \ldots, (O_K, \pi_K^*)\}$

訓練事例(O_I, π_I^*)は，分類対象集合 $O_I = \{o_I^1, o_I^2, \ldots, o_I^{\#O_I}\}$ と真の分割の具体例π_I^*の対

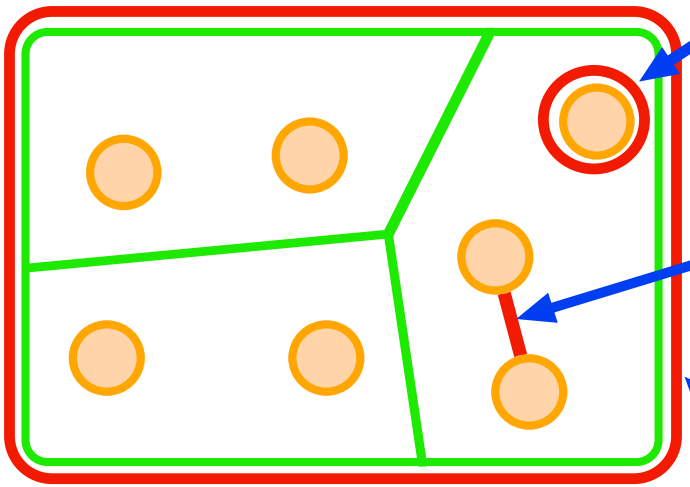
出力: 未知の分類対象集合 O_U に対する真の分割 $\hat{\pi}_U$ を推定できる規則

表記: 分類対象対 p^{ij} は分類対象 o^i と o^j の対

分類対象対全体の集合を P

分類対象集合の表現方法

分類対象集合 O を3種類の多数のベクトルで表現



A(o)…分類対象属性

各分類対象の特徴

ex. 位置

A(p)…分類対象対属性

各分類対象対の特徴

ex. 距離

A(π)…分割全体の属性

π に分割したときの分類対象集合全体の特徴

ex. クラスタ数

※ A(π) の値は，A(o) と A(p) の値，及び，分割 π の状態から計算

分割用の規則

分割用の規則は，分割の評価関数とこれを最大化する分割の探索手法で構成

分割πの評価関数: 事後確率最大(MAP)原理に基づき，次の結合確率を最大化

結合確率：$\Pr[\pi = \pi^*, A(\pi); \{A(o)\}, \{A(p)\}]$

$\pi = \pi^*$ …πが真の分割であるという事象
$A(\pi)$ …分割全体の属性
$\{A(o)\}$ …全ての分類対象属性の集合
$\{A(p)\}$ …全ての分類対象対属性の集合

評価関数を最大にする分割の探索

greedyな探索によって準最適な分割を探索

初期分割：各クラスタが分類対象を1個だけ含む

次の二つの操作を適用してみて評価関数を最も大きくする操作を適用し，次世代の分割とする

- 二つのクラスタを結合
- クラスタの分類対象1個を別のクラスタへ移動

この操作を，評価関数を大きくする分割が見つからなくなるまで反復

結合確率の分解・簡略化

結合確率：$\Pr[\pi = \pi^*, A(\pi); \{A(o)\}, \{A(p)\}]$

{A(p)} と {A(o)}の要素数は変化

➡ 計算が困難

➡ 次の二つの式に分解して計算

A(π)の尤度：$\Pr[A(\pi) \mid \pi = \pi^*]$

学習した関数 $f_2(A(\pi))$ を用いて計算

$\pi=\pi^*$の事前確率：$\Pr[\pi = \pi^*; \{A(o)\}, \{A(p)\}]$

学習した関数 $f_1(p)$ を用いて計算

$f_2(A(\pi))$の獲得

$$f_2(A(\pi))=\Pr[A(\pi) \mid \pi = \pi^*]$$

ex_2: EXを変換した属性値ベクトル集合

EX の各事例について，分割全体の属性ベクトルの値 $A(\pi_I^*)$ を計算

ex_2の要素ベクトルは確率密度$\Pr(A(\pi), \pi = \pi^*)$に従う

↓

仮定：$\Pr[\pi = \pi^*]$は均一分布

↓

ex_2の密度関数を$f_2(A(\pi))$の代用に

π=π*の事前確率の計算（1）

$\pi=\pi^*$の事前確率：$\Pr[\pi = \pi^*;\{A(o)\},\{A(p)\}]$

{A(p)}と{A(o)}の要素数は変化するので計算は困難

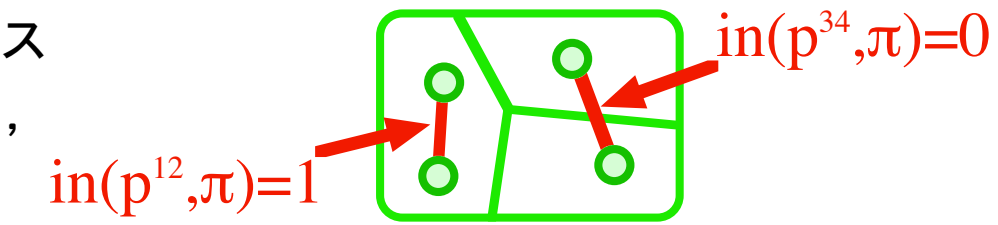
↓

分類対象対 p^{ij} に注目

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

注目した分類対象対に関連する属性値ベクトルがわかっているとき，その対が，真の分割で同じクラスタの要素である確率

$in(p^{ij},\pi)$: p^{ij}がπの同じクラスタの要素のとき1，それ以外0の関数



$\pi=\pi^*$の事前確率の計算 (2)

$f_1(p)$ を用いた次式は$\pi = \pi^*$の事前確率に比例

$$\Pr[\pi = \pi^*;\{A(o)\},\{A(p)\}] \propto \prod_{p\in P^+} f_1(p) \times \prod_{p\in P^-} (1 - f_1(p))$$

P^+　対の両方の分類対象が，πの同じクラスタの要素である分類対象対全体の集合

P^-　差集合：$P - P^+$

Dempster-Shaferの確率結合則を用いて導出

確率結合則(D&S則)：異なる証拠に基づく確率を統合

D&S則の適用(全体方針)

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

事象: $in(p^{ij}, \pi^*) = 1$

全ての可能な分割の中で，p^{ij}が同じクラスタの要素である分割のいずれかが真の分割

前提条件: $A(p^{ij}), A(o^i), A(o^j)$

確率分布を決定する証拠

$\pi=\pi^*$の事前確率：$\Pr[\pi = \pi^*; \{A(o)\}, \{A(p)\}]$

証拠 $\{A(o)\}, \{A(p)\}$ に基づく事象$\pi = \pi^*$の確率

D&S則の適用(D&S則の直観的説明)

D&S則の直観的説明：

$$\frac{\Sigma(\text{事象}\pi=\pi^{*}\text{を導く場合に割り当てられた確率})}{1-\Sigma(\text{矛盾を導く場合に割り当てられた確率})}$$

分子の計算方法：

- ◇全ての $p \in P$ について，事象$\pi=\pi^{*}$と矛盾しない事象が発生する確率を表す方を選ぶ
- ◇Σ記号の中身は，これら #P 個の確率の積

D&S則の適用(分子の計算)

事象$\pi = \pi^*$を導く組み合わせの確率の計算

ある分割 π が，真の分割 π^* と一致するには？

全ての $p \in P$ について，

◇真の分割 π^* で p が同一クラスタの要素ならば，

π でも同一クラスタの要素

[この事象の発生確率は $f_1(p)$]

◇真の分割 π^* で p が異なるクラスタの要素ならば，

π でも異なるクラスタの要素

[この事象の発生確率は $1-f_1(p)$]

これらの積は …… $\prod_{p \in P^+} f_1(p) \times \prod_{p \in P^-} (1 - f_1(p))$

D&S則の適用(分母の計算)

矛盾を導く場合に割り当てられたの確率の計算

この組み合わせは

・いかなる分割も導かない場合

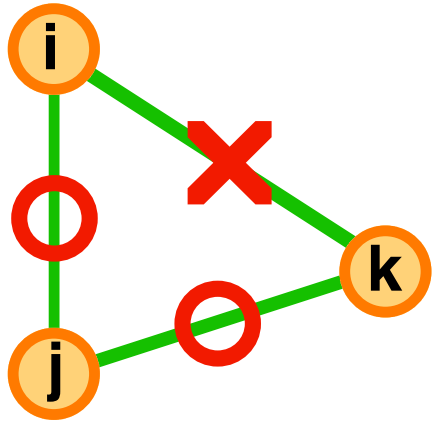
・分割 π の選択に依存しない

↓

分母は定数

↓

例: 矛盾する場合



$$\Pr[\pi = \pi^{*};\{A(o)\},\{A(p)\}] \propto \prod_{p \in P^{+}} f_1(p) \times \prod_{p \in P^{-}} (1 - f_1(p))$$

関数 f_1 の学習

訓練事例 EX の変換

全ての訓練事例の，全ての分類対象について

$$in(p^{ij}, \pi^*)$$

を計算し，さらに

$A(p^{ij})$ ， $A(o^i)$ ， および $A(o^j)$ の合成ベクトル $A_C(p^{ij})$

も求める

変換後の事例：$(in(p^{ij}, \pi^*), A_C(p^{ij}))$

この事例から事後確率 $f_1(p^{ij})$ は例からの学習の手法で計算可能

獲得した規則の評価

Leave-One-Outテスト (LVOテスト)

- K 分割の交叉確認法 (Kは事例数)

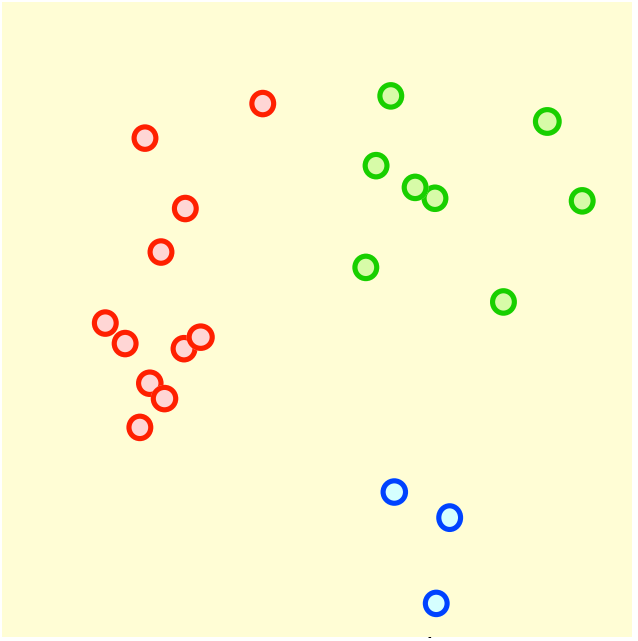
情報損失量 (Ratio of Information Loss; RIL)

- 真の分割と推定分割の類似性の尺度として利用
- 事後エントロピーを事前エントロピーで割った値
  ＝獲得すべき情報のうち獲得できなかった情報の割合
- 0 ならば完全に一致し，1 が最大で最も不一致

実験（ドットパターン）

ドットパターン

- 中心と分散が異なる複数の正規分布に従ってドットを発生させた人工データ
- 同じ分布から発生したドットが同じクラスタになる分割を真の分割とする



実験条件の詳細

クラスタ数：2～4個，ドット数：50個，総事例数：100個

正規分布の中心は一定の範囲でランダム

分布の標準偏差を変えた，クラスタの重なりが違う３種類の事例集合

EMアルゴリズム

EMアルゴリズム

・ドットパターンのクラスタリング代表的手法

・分布が既知のときには非常に有効

比較実験の目的：クラスタリングとLCEは違うタスク

データの発生源の情報を十分用いたEMアルゴリズムで推定した分割

対等

LCEの学習段階で獲得した規則の適用により推定した分割

結論：LCE手法で真の分割を導く規則を獲得できた

EMアルゴリズムとの比較実験

	LCE	EMアルゴリズム
分離	0.067	0.089
接触	0.161	0.161
重複	0.369	0.389

推定分割と真の分割の間のRILのLVOテストによる平均値

いずれの場合も検定では有意な差はなかった

↓

LCEでEMと対等な分割の推定ができた

↓

結論：LCE手法で真の分割を導く規則を獲得できた

ドットパターンについての実験結果

1. EMアルゴリズムと対等な分割をLCE手法で獲得

➡LCEで真の分割を導く規則が獲得できた

2. 事例数が増加するとより正確な推定ができた

➡事例を多く準備できれば，より優れた規則を獲得できる

3.分割全体の属性をより正確にクラスタ数が推定できた

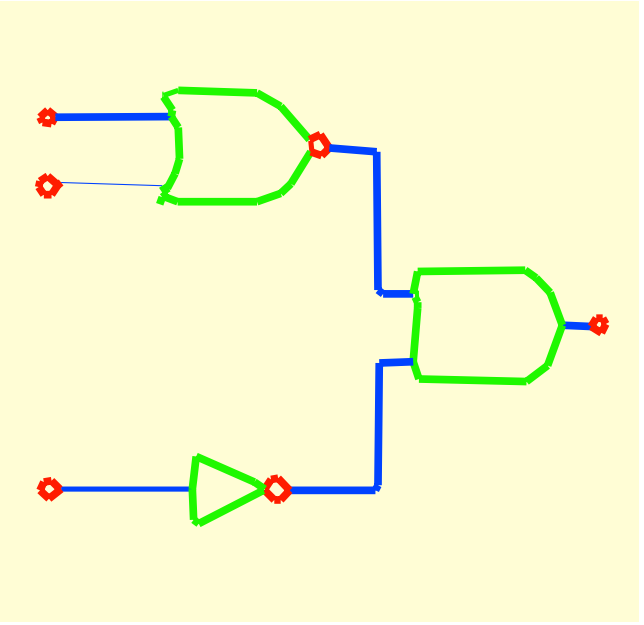
➡分割全体の属性の有効性を確認

4. 分割の具体例をいくつか図示 (参考)

実験（論理回路図画像）

論理回路図画像

- 線分の集合を表現するベクトルデータ形式の画像
- 分類対象を各線分とし，同一部品を表す線分をクラスタにまとめる問題



実験条件の詳細

平均クラスタ数：16.7個，平均線分数：102.9個，総事例数：100個
部品の種類は AND, OR, バッファ，端子，そして 接続線の５種類
手書きした図面をスキャナで取り込み処理した画像

論理回路図画像では比較実験をしない理由

- 従来のセグメンテーション手法では，テスト事例と訓練事例は本質的に分離不可能
- 論理回路図面は人工データではないので，発生源の特徴は不明
- ベクトルデータのセグメンテーション手法は対象画像に依存し，汎用性に乏しい
- パラメータ調節が多く，恣意性のない実験は難しい

分割全体の属性の有効性

分割全体の属性あり	分割全体の属性なし
0.430	0.442

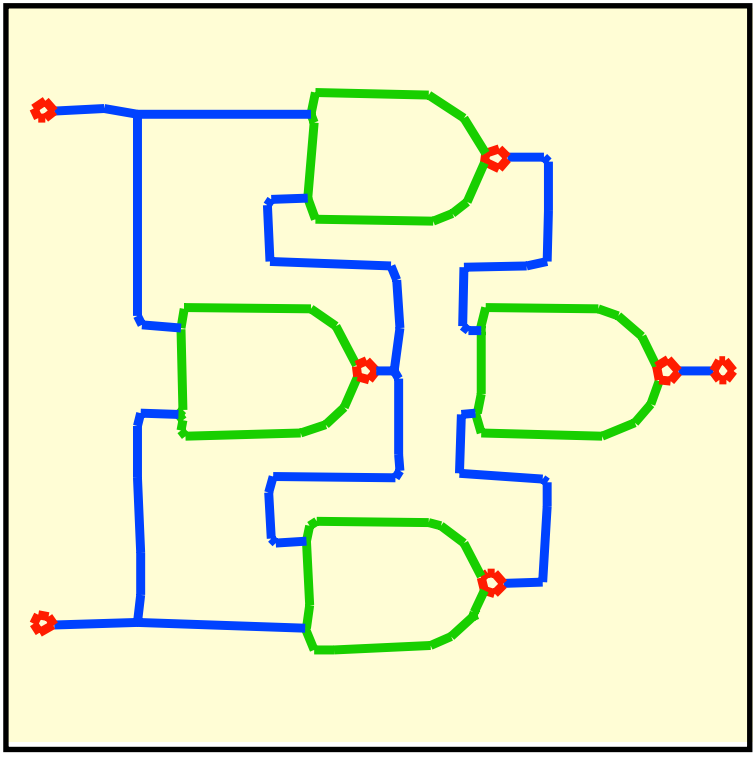
推定分割と真の分割の間のRILのLVOテストによる平均値
分割全体の属性を利用した場合とそうでない場合の値

分割全体の属性を利用した場合の方が，そうでない場合より統計的に有意にRILが小さい

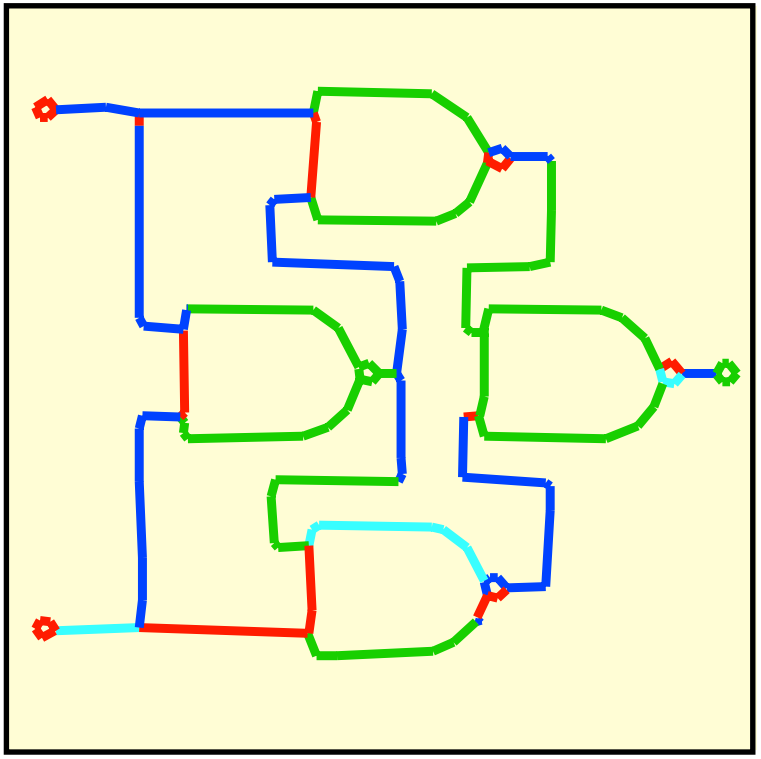
⬇

分割全体の属性の利用は有効

推定分割の具体例 (最も悪い結果)



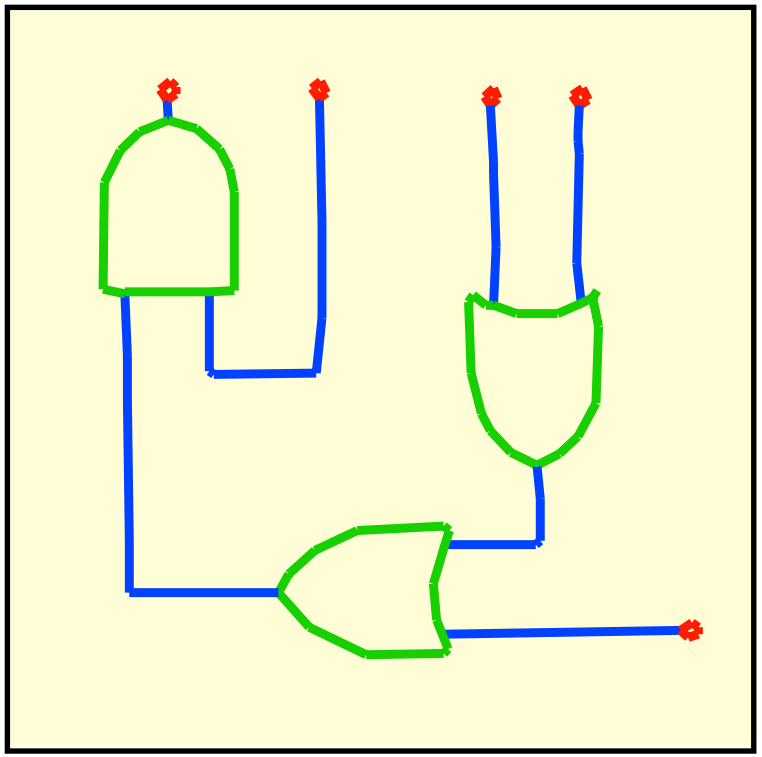
真の分割



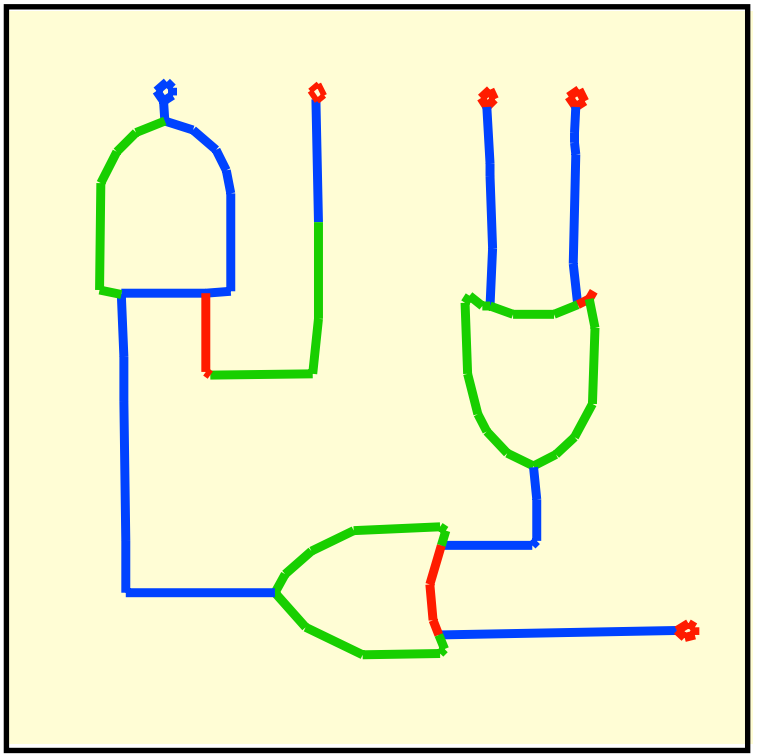
LCEによる推定分割

(RIL : 0.701)

推定分割の具体例（中間の結果）

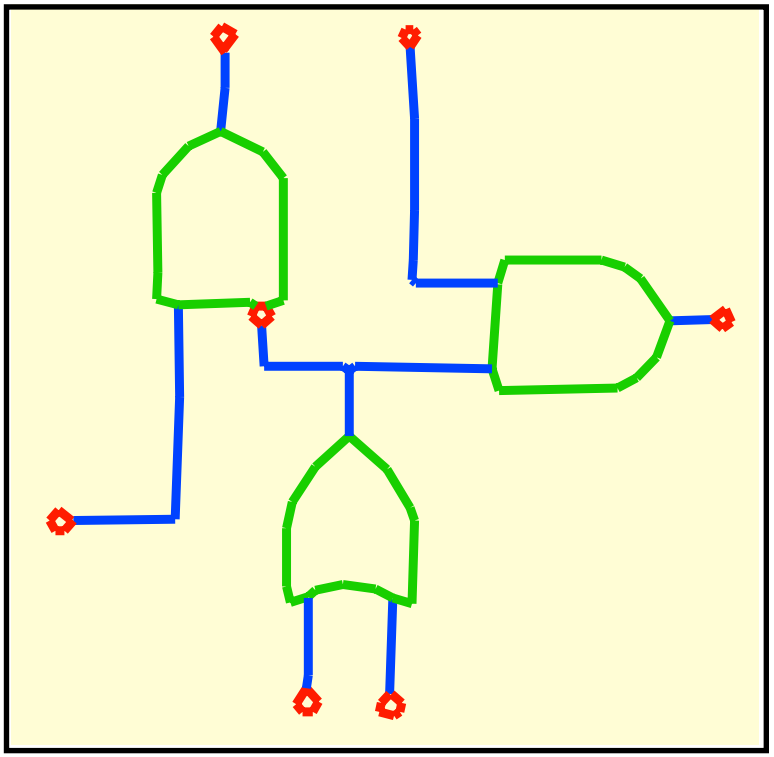


真の分割

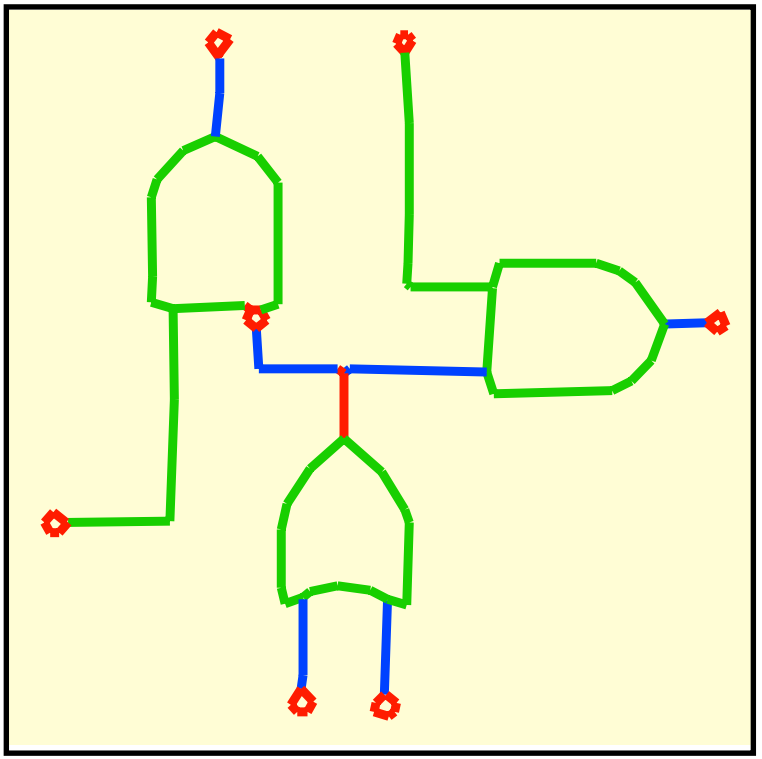


LCEによる推定分割
(RIL : 0.412)

推定分割の具体例 (最も良い結果)



真の分割



LCEによる推定分割

(RIL : 0.145)

論理回路図画像についての実験結果

1. 分割全体の属性を用いるとより正確な推定が可能

➡分割全体の属性の有効性を確認

2. 分割の具体例をいくつか図示（参考）

考察 (1)

分割獲得に役立つ特徴の発見は感覚に依存

・真の分割の具体例の提示で規則は，感覚に依存せず獲得できた

分割獲得に役立つ特徴を見つけても，その特徴を具体化できない

・アルゴリズムは事例から自動的に規則を具体化

獲得した規則を適用するときに調整が必要

・アルゴリズムは，適用時に調整が不要な規則を獲得

考察 (2)

分割の結果は統計的に不安定

- 手作業ではテスト事例と訓練事例を厳密に分離できないので，公正な性能評価ができない
- 参照できる情報の量が人の認知能力で制限

分割の結果は統計的に安定

- アルゴリズムが規則を獲得するため，訓練とテスト事例は厳密に分離でき，公正に性能評価
- 人の認知能力ではなく，計算機資源にしか制限されないので，多数の訓練事例を参照可能

まとめ

- クラスタ例からの学習という新たな学習タスクを定式化し，その学習方法と結果の評価方法を提案

- 分割用の規則を，感覚に依存せず，自動的に具体化して獲得できた

- 適用時の調整が不要な規則を獲得でき，その適用結果は統計的に安定あった