

## 情報意味論 (第2回)

慶應義塾大学工学部  
櫻井 彰人

## エキスパートによるデータマイニングの例



### 概念学習 - 心理学から

概念とは

概念とは、ある特徴を共通にもつ、刺激の範疇である。

例: 左にある形はすべてある概念範疇, 矩形, に属する. その共通特徴は

(1) 4本の直線; (2) 対向する線が平行; (3) 線は端点で交わる; (4) 直線は4個の直角を作る.

色, 大きさ, そして方向は「矩形」という概念を定義する特徴ではない。

### 概念学習 - 心理学から

概念とは

正事例 (positive instance): ある概念に属する刺激のこと,

負事例 (negative instance): そうでない刺激のこと

下記のものとは矩形概念の負事例である:



ある刺激は、もし矩形を定める特徴の一つでも欠いていれば、矩形の負事例ということになる。

### 概念学習 - 心理学から

概念とは

概念の主要素2個:

属性: ある刺激が正事例であるか負事例であるかを定める, 刺激が有する属性.

規則: ある刺激がある属性の正事例であるためには、どの属性がなければいけないか、またはあってはいけないかの記述. 勿論、もっと複雑な記述である場合もある.

矩形の例でいえば、属性は、前述の4個の属性であり、規則は「すべての属性が存在すべき」となる。

### 概念学習 - 心理学から

概念とは

規則の中でもっとも単純なものとして、「某属性がある」「某属性がない」という形の規則がある。

例えば、脊椎動物は「脊椎がある」動物として定義される。

正事例・負事例の例:



(仮に)肯定命題: それで真であれば刺激は正事例となり、偽であれば、刺激は負事例となる規則である。

## 概念学習 - 心理学から

(仮に)否定命題: 刺激がある特定の属性を持たないという形の命題



否定命題の例: 「無脊椎動物は脊椎のない動物である」。

上図は、当該否定命題を用いた場合の、正事例と負事例である。

## 概念学習 - 心理学から

より複雑な規則、すなわち、複数の属性を用いて記述する規則もある。

比較的単純な例: 連言 (conjunction) がある。  
「ある刺激にA属性があり、かつB属性がある」。

矩形を定義するとき用いた、先ほどの記述方法は、これである。

## 概念学習 - 心理学から

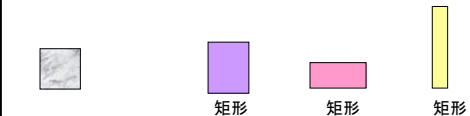
(行動学的に言えば) 概念の学習には、2つの過程がある:

汎化: ある呼び名 (例えば、対象の名称)・反応を、共通特徴に基づき、当該概念クラスの全メンバーに一般化して適用する。

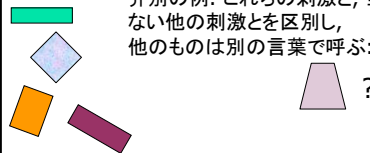
弁別: ある刺激が当該概念に属するか否かを、当該概念を定める属性がすべて存在するか、一つでも欠けているかに基づき、判定する。

## 概念学習 - 心理学から

汎化の例: 「矩形」という言葉を、定義属性をもつすべての刺激に一般化して用いる。



弁別の例: これらの刺激と、当該概念クラスに属さない他の刺激とを区別し、他のものは別の言葉で呼ぶ:



## 概念学習 - 心理学から

丸暗記は、学習対象を理解しないで学習することである。

場合によっては、ある未知刺激に対して、それが属する概念範疇を発見しなくても、正しい応答をするように学習することは可能である (すなわち、丸暗記するだけで、刺激に対応することが可能な場合がある)。

より技術的に言えば、

当該刺激が同一概念クラスの他の刺激と共有する属性を発見しなくても、正しい反応をすることは可能である。

## 概念学習 - 心理学から

丸暗記は次の場合である

ある単語・反応を、それに対応する事例とともに記憶している。しかし、

どうしてその対応があるのかを理解していない

すなわち、

次に同一事例をみれば、対応する正しい単語を言い当てることができる。

しかし、以前見たことのない事例には答えられない

概念学習ができたか否かのテスト:

同一概念に属する新事例を認識する (正しい応答をする) ことができるか?

## 概念学習 - 心理学から

暗記学習 vs.  
概念学習

チンパンジーに対するオペラント条件付け実験 (Kelleher, 1958) は、丸暗記学習と概念学習の差を明確に示している。

Roger T. Kelleher, Concept Formation in Chimpanzees, Science, Vol. 128, no. 3327, 777-778(3 October 1958).

手順は弁別訓練であり、

13 個の弁別刺激 (ある概念の正事例) と  
13 個のデルタ刺激 (ある概念の負事例)を用いた。

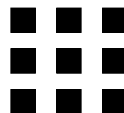
刺激はランダムな順で提示された。

正事例による刺激の間、ボタン応答が強化された (報酬が与えられた)(VR 100)。  
負事例による刺激の間、応答は強化されなかった。

## 概念学習 - 心理学から

暗記学習 vs.  
概念学習

各刺激は 3 行 3 列に配置された9個の小窓からなる:



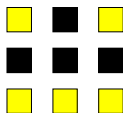
各小窓は点灯されるか暗いままかのいずれかである。例えば:



最近のチンパンジーはゲームもできる

## 概念学習 - 心理学から

暗記学習 vs.  
概念学習



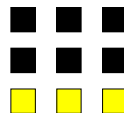
この配列は正事例の一つである。



最近のチンパンジーは自動販売機も使える

## 概念学習 - 心理学から

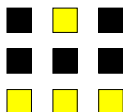
暗記学習 vs.  
概念学習



自然界で道具を使うのはおてのもの

## 概念学習 - 心理学から

暗記学習 vs.  
概念学習

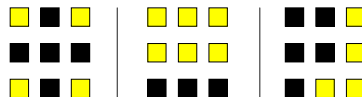


正事例はいずれもある共通の属性をもつ。負事例はその一つまたは複数をもたない。負事例を示そう:

## 概念学習 - 心理学から

暗記学習 vs.  
概念学習

### 負事例 (デルタ刺激)



正事例の共通属性が当該概念を定めている。それは...?

最下段の3個の小窓が点灯している。

## 概念学習 - 心理学から

暗記学習 vs.  
概念学習

13個の正事例と13個の負事例を、弁別が非常によくできるまで提示した:

正事例には一定速度の、素早い応答(ボタン押下);  
負事例には、ほとんどまたは全く応答なし。

その後、刺激の種類はそのままに、しかし提示順序を変更した。

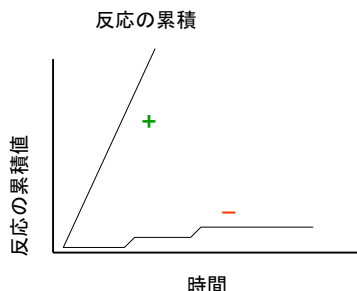
成績の低下は見られなかった。第一回目の典型的な、反応の累積を示す:



ヒト, チンパンジー, アカゲザルの脳

## 概念学習 - 心理学から

暗記学習 vs.  
概念学習



## 概念学習 - 心理学から

暗記学習 vs.  
概念学習

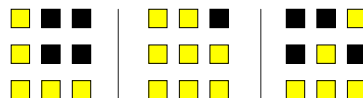
この弁別課題の第二フェーズでは、  
6個の正事例と6個の負事例を取り除き  
別の新しい、しかし、同じ概念の刺激を挿入して実験を行った。

例: 新正事例の3個と新負事例の3個である。

## 概念学習 - 心理学から

暗記学習 vs.  
概念学習

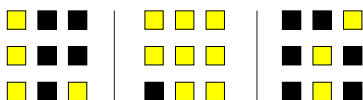
新正事例 (フェーズ 2 用)



## 概念学習 - 心理学から

暗記学習 vs.  
概念学習

新負事例 (フェーズ 2 用)



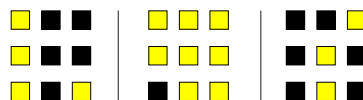
?

問: チンパンジーは、  
刺激がそれまで見たことがない新しい刺激であつても、  
以前同様正しく応答することができるだろうか?

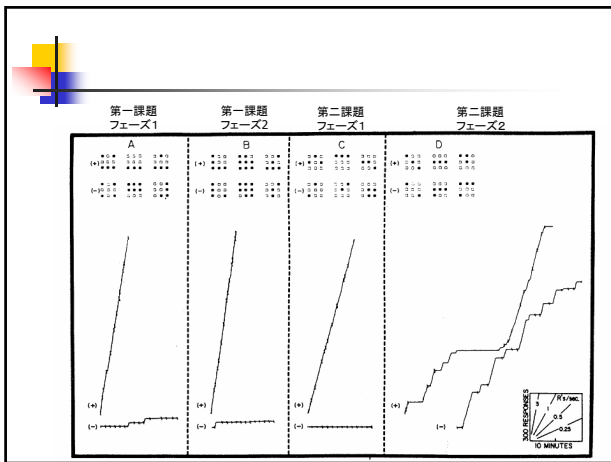
## 概念学習 - 心理学から

暗記学習 vs.  
概念学習

新負事例 (フェーズ 2 用)



結果: うまくできた、すなわち、チンパンジーは  
正事例に対して素早く応答し、  
負事例に対してはほとんどまたはまったく応答し  
なかった



暗記学習 vs. 概念学習

## 概念学習 – 心理学から

この結果は、チンパンジーはフェーズ1で**概念**を学習したことを示している。すなわち、パターンが何であれ、最下段の3つの小窓を見て、もしすべてが点灯していれば、ボタンを押下したと考えられる。

ボタン押下は、“これが正事例だ、”と言っているようなものである

この実験は「概念」を用いて問題を解くことの有利さを明らかにしている：新しい状況に正しく応答できるし、その際重要な属性のみに着目すればよい。

暗記学習 vs. 概念学習

## 概念学習 – 心理学から

### 暗記学習の例

チンパンジーに2番目の概念問題を与えた。  
13個の新しい正事例と13個の新しい負事例である。

これらの刺激の例を示す。概念は何か？

"Congo the Chimp" の1枚。1950年代中ごろ。3枚で20万円くらいの見取りでオークションに出された

暗記学習 vs. 概念学習

## 概念学習 – 心理学から

### 新正事例 (第二課題)

■ ■ ■	■ ■ ■	■ ■ ■
■ ■ ■	■ ■ ■	■ ■ ■
■ ■ ■	■ ■ ■	■ ■ ■

"The Bonhams" で、1万4,400ポンド(約295万円)で落札されたという。

暗記学習 vs. 概念学習

## 概念学習 – 心理学から

### 新負事例 (第二課題)

■ ■ ■	■ ■ ■	■ ■ ■
■ ■ ■	■ ■ ■	■ ■ ■
■ ■ ■	■ ■ ■	■ ■ ■

画筆を振るう Congo

暗記学習 vs. 概念学習

## 概念学習 – 心理学から

### 暗記学習の例

正解は“どれか3個の小窓が点灯している。”  
負事例は、点灯小窓の数は2または4である。

実験手順は課題1とのときと同じである。

13個の正事例と13個の負事例を、確実に弁別されるまで提示する：

正事例のとき素早く応答し、負事例のときにはほとんどまたはまったく応答がなくなるまで。

400枚の一枚

## 概念学習 - 心理学から

暗記学習 vs.  
概念学習

### 暗記学習の例

その後概念学習ができたか否かのテストを行う:  
6個の新正事例と6個の新負事例を提示する。  
チンパンジーは正しく答え続けるであろうか?

結果: 正事例, 負事例のときの応答の累積を記したグラフを示す。

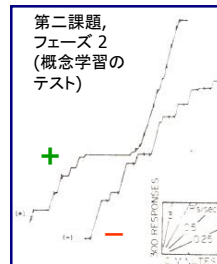


しかし馬でも掛ける

## 概念学習 - 心理学から

暗記学習 vs.  
概念学習

### 暗記学習の例



## 概念学習 - 心理学から

暗記学習 vs.  
概念学習

### 暗記学習の例

成績は, 新しい刺激に対しては, 悪化している,  
正事例に対しても, しばしば反応は途絶し  
負事例に対しても, しばしば高速で応答する。

チンパンジーは課題を“回答を記憶”し, 正事例が  
共有する特徴を発見しないで, 学習した。理由は  
たぶん当該概念が抽象的であったためであろう:

点灯する3個の小窓はある特定の場所と結びつい  
ているわけではない。

個数“3”に反応しなければならない

(なお、数の区別ができることは知られている)。

## 外延と内包

- (集合を) 定義する方法
- 外延 (extension)
  - その集合に属する要素をすべて挙げる
  - 例: {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
- 内包 (intension)
  - その集合に属する要素が持つべき性質を述べる
  - 例: 1以上10以下の整数

## 概念学習

- 概念:
  - 対象集合を(より大きな集合の部分集合として)特徴付ける記述
  - 特に、フル値関数で記述できるもの
- 概念学習:
  - あるフル値関数を、その入力値と出力値の対からなる訓練例から推定すること
- 例: EnjoySport

Sky	AirTemp	Humid	Wind	Water	Forecast	EnjoySport
Sunny	Warm	Normal	Strong	Warm	Same	Yes
Sunny	Warm	High	Strong	Warm	Same	Yes
Rainy	Cold	High	Strong	Warm	Change	No
Sunny	Warm	High	Strong	Cool	Change	Yes

- これらを一般化した概念は何だろう

## 仮説と仮説空間

- 機械学習では、「概念」を探索する
  - すなわち、与えられた正事例を含み、負事例を含まない概念記述を求める
  - この「概念の記述」を仮説という
  - 探索する仮説の集合を仮説空間という
    - 仮説空間は一般に無限集合である
- 注意すべきこと
  - 一般に、正事例も負事例も有限個しかない
  - 一般に、記述によって定義される集合は無限集合
    - 記述自体は、有限長
  - ということは、一般に、正事例を含み、負事例を含まない仮説は複数個あることになる。どれが正しいのだろうか? どれも同様に正しいのだろうか?

# 仮説の表現

- 仮説の表現はいろいろありうる
  - さらに、仮説空間自体もいろいろありうる
- (これから説明する概念学習では)各仮説  $h$  は属性値に対する制約の連言 (and)
- 各属性に対する制約は、
  - 特定値 (Water = Warm)
  - don't care (Water = ?)
  - どんな値も許さず (Water=?)
- 例

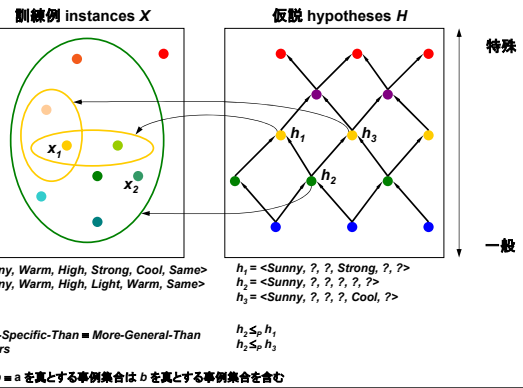
< Sky Sunny AirTemp ? Humid ? Wind Strong Water ? Forecast Same >

# 概念学習の基本

- 下記の所と条件のもと
  - 事例集合  $X$ : 次の属性で表現される Sky, AirTemp, Humidity, Wind, Water, Forecast
  - 目標関数  $c$ : EnjoySport:  $X \rightarrow \{0,1\}$
  - 仮説空間  $H$ : リテラルの連言。例えば、
    - $\langle \text{?}, \text{Cold}, \text{High}, \text{?}, \text{?}, \text{?} \rangle$
  - 訓練例  $D$ : 目標関数の正例及び負例
- 次を求める:
  - $H$  中の  $h$  で  $D$  中の全ての  $x$  に対して、 $f(x)=c(x)$  となるもの

# 暗黙の前提

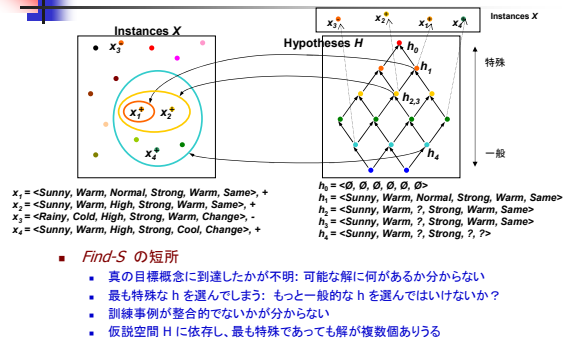
- 十分大きな訓練事例集合に対して目標関数を近似する仮説は、未観測の事例に対しても目標関数を近似する
  - 統計的枠組みのもとで、この言明は「高い確率で」正しいということが示せる。勿論、適当な確率分布を仮定する必要がある。
  - 確率が導入できない場合でも、「真の仮説は十分短い」という前提のもとで、正しいということが示せる。



# Find-S アルゴリズム

- 仮説  $h$  として仮説集合  $H$  の中で最も特殊なものを選ぶ
- 正の訓練事例の各々の  $x$  に対し
  - $x$  が  $h$  を満たす(成立させる)なら、何もしない
  - $x$  が  $h$  を満たさない(成立させない)なら、 $x$  も  $h$  も満たす(ちょっと言葉の誤用)仮説で最も特殊なもの(最も  $x$  と  $h$  に近いもの)を見出し、それを新たな  $h$  とする
- または
- $h$  中の属性制約  $a_i$  に対し
  - もし、 $x$  が  $a_i$  を満たすなら何もしない
  - そうでなければ、 $a_i$  を  $x$  によって満足される、より一般的な制約に改めよ
- $h$  を出力する

# Find-S の仮説空間探索



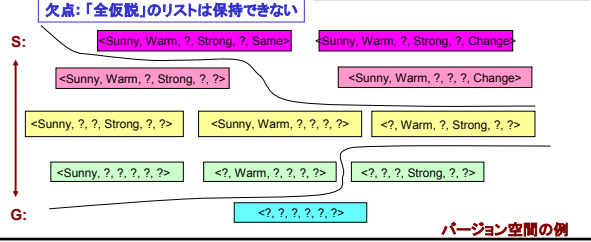
# Find-S の短所克服

- 候補除去 (candidate-elimination) アルゴリズム
  - 訓練事例に整合的な仮説を全て求めよう (しかし、全ての仮説を枚举することはしない。どうするか？それが知恵)
- 整合的な仮説
  - 仮説  $h$  が目標関数  $c$  の訓練事例集合  $D$  と整合する (consistent) であるというのは、 $D$  中の任意の訓練例  $\langle x, c(x) \rangle$  に対して、 $h(x) = c(x)$  となることを言う。
    - $h(x) = \text{真}$  であれば、 $h$  は  $x$  を cover する、である
    - $h(x) = \text{偽}$  であれば、 $h$  は  $x$  を cover しない、である
  - $\text{Consistent}(h, D) = (\forall \langle x, c(x) \rangle \in D) h(x) = c(x)$ 
    - $h$  は  $D$  中の正例を cover し、負例を cover しない
- バージョン空間
  - (仮説空間  $H$  と訓練例集合  $D$  に関する) バージョン空間  $VS_{H,D}$  とは、 $D$  中の全ての訓練例と整合する仮説  $h \in H$  の集合のことである
  - $VS_{H,D} = \{h \in H \mid \text{Consistent}(h, D)\}$ 
    - $D$  中の正例を cover し負例を cover しないような  $h$  の集合

# List-Then-Eliminate アルゴリズム

- $VS \leftarrow H$  の全仮説を要素とするリスト
- 各訓練事例  $\langle x, c(x) \rangle$  につき、 $VS$  から  $h(x) \neq c(x)$  となる  $h$  を除く
  - 正例を cover しない  $h$  を除く
  - 負例を cover する  $h$  を除く
- $VS$  を出力する

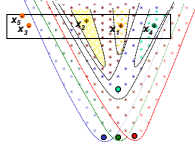
$x_1 = \langle \text{Sunny, Warm, Normal, Strong, Warm, Same} \rangle, +$   
 $x_2 = \langle \text{Sunny, Warm, High, Strong, Warm, Same} \rangle, +$   
 $x_3 = \langle \text{Rainy, Cold, High, Strong, Warm, Change} \rangle, -$   
 $x_4 = \langle \text{Sunny, Warm, High, Strong, Cool, Change} \rangle, +$



# バージョン空間の表現方法の工夫

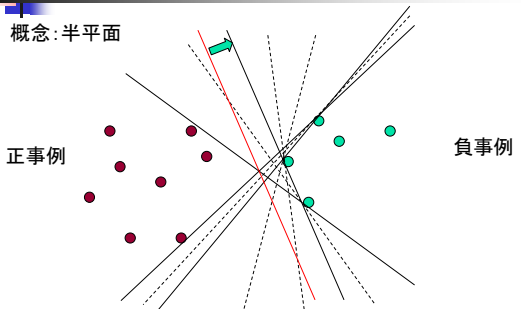
- 仮説空間 (一般の場合)  $H$ 
  - 有限交叉性を持つ半束 (半順序  $x \preceq y$  は、 $y$  はより特殊か等しい;  $\perp \equiv$  most general)
  - どの二つの仮説にも 最大下界 greatest lower bound (GLB) が存在する
  - $VS_{H,D} =$  整合的な仮説の部分集合
- 定義: 一般化側境界 General Boundary
  - $VS_{H,D}$  の一般化側境界 G: 最も一般な要素の集合
  - 最も一般な要素  $\equiv VS_{H,D}$  の極小要素  $\equiv$  「必要条件の集合」
  - 負例を cover しない範囲でできるだけ cover 範囲を大きくした仮説
    - どの負例も cover しない、極大の仮説
- 定義: 特殊化側境界 Specific Boundary
  - $VS_{H,D}$  の特殊化側境界 S: 最も特殊な要素の集合
  - 最も特殊  $\equiv VS_{H,D}$  の極大要素  $\equiv$  「十分条件の集合」
  - 正例を cover する範囲でできるだけ cover 範囲を小さくした仮説
    - 全正例を cover する、極小の仮説
- バージョン空間
  - バージョン空間内の全ての要素は S と G の間にある
  - $VS_{H,D} = \{h \in H \mid (\exists s \in S), (\exists g \in G), g \preceq h \preceq s\}$
  - 但し、 $x \preceq y \equiv x$  は  $y$  より一般的か等しい (正確には、より特殊ではない)

仮説を実効的には全部持っているが、記憶しないで済ませる方法はないか?



# 直感的な例 (一般化境界)

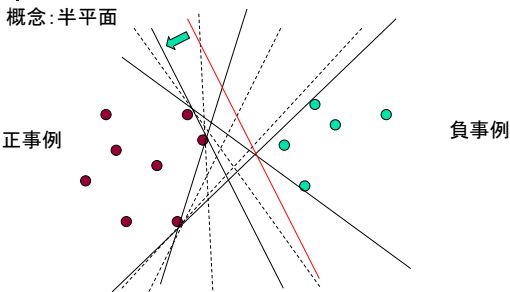
概念: 半平面



ある境界 (直線である) に平行な境界は、より一般かより特殊。平行でない境界は、どちらでもない

# 直感的な例 (特殊化境界)

概念: 半平面

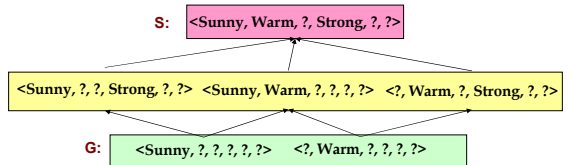


ある境界 (直線である) に平行な境界は、より一般かより特殊。平行でない境界は、どちらでもない

# バージョン空間の例

- 特殊化側境界 (S) と一般化側境界 (G) とに挟まれた領域

$x_1 = \langle \text{Sunny, Warm, Normal, Strong, Warm, Same} \rangle, +$   
 $x_2 = \langle \text{Sunny, Warm, High, Strong, Warm, Same} \rangle, +$   
 $x_3 = \langle \text{Rainy, Cold, High, Strong, Warm, Change} \rangle, -$   
 $x_4 = \langle \text{Sunny, Warm, High, Strong, Cool, Change} \rangle, +$



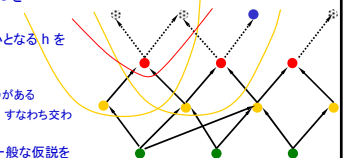
# 候補削除アルゴリズム(1/3)

- G ← H 中で「一般化が極大」である仮説の集合
  - {<?, ..., ?> } と記す
- S ← H 中で「特殊化が極大」である仮説の集合
  - {<0, ..., 0>} と記す
- 各訓練事例 d 毎に次のことを行なう

# 候補削除アルゴリズム(2/3)

- d が正例であるとき
  - d を覆わない仮説を G から除去
    - S 中で d を覆わない仮説 s につき
      - s を S から除去
      - s の一般化で、一般化が極小となる h を全て S に加える
        - h は d を覆い、
        - G の中に h より一般なものがある (これらは  $VS_{h,d}$  中の最大下界、すなわち  $s \vee d$  である)
    - S から、S 中の他の仮説より一般な仮説をすべて除去する

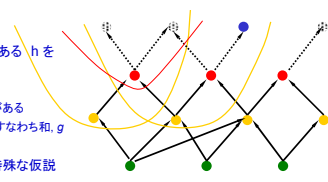
仮説はすべから一般であるべき  
Gはより特殊に、Sはより一般に



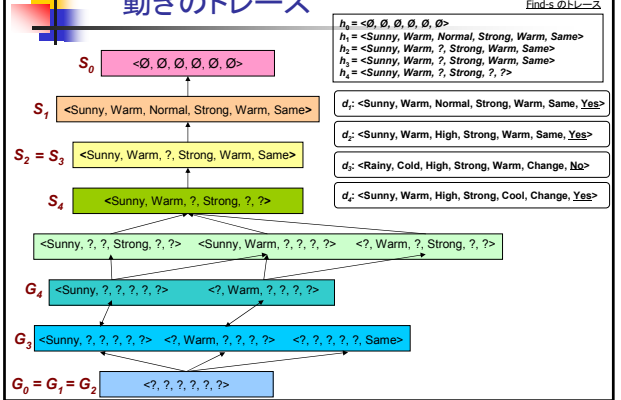
# 候補削除アルゴリズム(3/3)

- d が負例であるとき
  - d を覆う仮説を S から除去
  - G 中の d を覆う仮説 g につき
    - g を G から除去
    - g の特殊化で特殊化が極小である h を全て G に加える
      - h は d を覆わない、
      - S の中に h より特殊なものがある (これらは  $VS_{h,d}$  中の最小上界、すなわち  $g \wedge d$  である)
    - G から、G 中の他の仮説より特殊な仮説をすべて除去する

仮説はすべから一般であるべき  
Gはより特殊に、Sはより一般に



# 動きのトレース

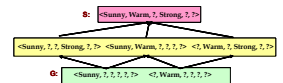


# 補足

- 真の解に収束するか
  - 訓練例に誤りなく、真の解が仮説空間内にあるならば、yes
  - 訓練例に誤りがあれば、バージョン空間は空集合となる
- 質問 query が許される場合
- 未知事例に対する予測

# 帰納的飛躍の正当化

- 例: 次の帰納的一般化は正当化できるか? どうして?
  - 正例: <Sunny, Warm, Normal, Strong, Cool, Change, Yes>
  - 正例: <Sunny, Warm, Normal, Light, Warm, Same, Yes>
  - 帰納結果 S: <Sunny, Warm, Normal, ?, ?, ?>
- 未知データが分類できると考えるのは何故?
  - (新しい事例に) どれだけ情報があれば、予測してもよいとするのか? 例えば
  - < Sunny, Warm, Normal, Strong, Cool, Change > - yes, 全員賛成
  - < Rainy, Cool, Normal, Light, Warm, Same > - no, 全員反対
  - < Sunny, Warm, Normal, Light, Warm, Same > - 半々
  - < Sunny, Cold, Normal, Strong, Warm, Same > - no, 4人がno, 2人がyes





## 概念学習のまとめ

- 仮説空間  $H$  における探索としての 概念学習
  - 学習: 正しい仮説を見出すこと
  - 複数あると: 最も特殊なもの or 全部
- 仮説空間  $H$  をより一般的なものからより特殊なものへと順序つける
  - 半順序
  - $H$  中に上限と下限がある → 束
- バージョン空間を用いた候補削除アルゴリズム
  - $S$  および  $G$  境界が、学習者が持つ不確かさ uncertainty を表現
  - バージョン空間を用いると、未知事例に対して予測を行うことができる
- (バージョン空間があると) 有用な質問が可能