閉集合を基礎とした機械学習

京都大学 情報学研究科
山本　章博

山本章博

略歴

- 1985年 京都大学理学部卒業 (当時は学科なし)
- 1990年 九州大学総合理工学研究科修了
- 1990年 北海道大学工学部電気工学科講師
- 1994年 北海道大学工学部助教授
- 2003年 京都大学情報研究科教授

生息場所

人工知能学会　SIG-FPAI

論理プログラミング，記号論理，自動証明
計算論的学習，機械学習
知識発見，データマイニング

北大時代に教わったこと

- Booleは 0, 1を $x\,x = x$ の解と説明している

津田一郎先生より

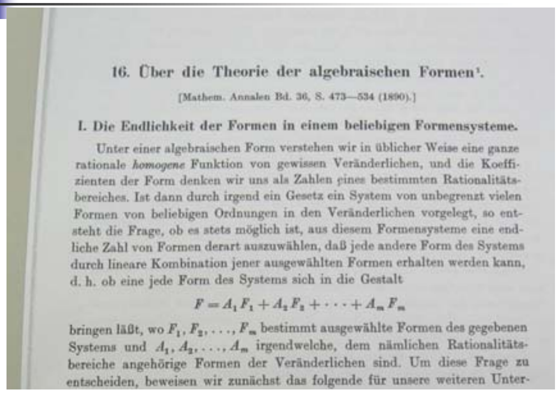
Boole: An Investigation of the Laws of Thought
on Which are Founded the Mathematical Theories
of Logic and Probabilities

- 多項式イデアルは順序イデアルが本質らしい

↓　辻下徹先生のセミナー

代数や数理論理の中には学習が埋め込まれている
学習理論をつくるにはその学習性を引き出すべき

Hilbert's original paper



16. Über die Theorie der algebraischen Formen[1].

[Mathem. Annalen Bd. 36, S. 473—534 (1890).]

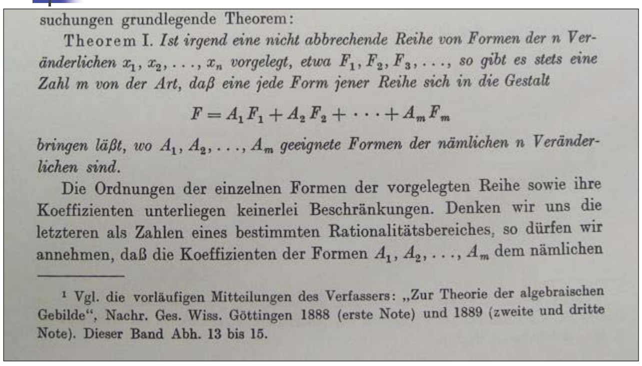
I. Die Endlichkeit der Formen in einem beliebigen Formensysteme.

Unter einer algebraischen Form verstehen wir in üblicher Weise eine ganze rationale *homogene* Funktion von gewissen Veränderlichen, und die Koeffizienten der Form denken wir uns als Zahlen eines bestimmten Rationalitätsbereiches. Ist dann durch irgend ein Gesetz ein System von unbegrenzt vielen Formen von beliebigen Ordnungen in den Veränderlichen vorgelegt, so entsteht die Frage, ob es stets möglich ist, aus diesem Formensysteme eine endliche Zahl von Formen derart auszuwählen, daß jede andere Form des Systems durch lineare Kombination jener ausgewählten Formen erhalten werden kann, d. h. ob eine jede Form des Systems sich in die Gestalt

$$F = A_1F_1 + A_2F_2 + \cdots + A_mF_m$$

bringen läßt, wo $F_1, F_2, \ldots, F_m$ bestimmt ausgewählte Formen des gegebenen Systems und $A_1, A_2, \ldots, A_m$ irgendwelche, dem nämlichen Rationalitätsbereiche angehörige Formen der Veränderlichen sind. Um diese Frage zu entscheiden, beweisen wir zunächst das folgende für unsere weiteren Unter-

Hilbert's original paper



suchungen grundlegende Theorem:

Theorem I. *Ist irgend eine nicht abbrechende Reihe von Formen der n Veränderlichen $x_1, x_2, \ldots, x_n$ vorgelegt, etwa $F_1, F_2, F_3, \ldots$, so gibt es stets eine Zahl m von der Art, daß eine jede Form jener Reihe sich in die Gestalt*

$$F = A_1F_1 + A_2F_2 + \cdots + A_mF_m$$

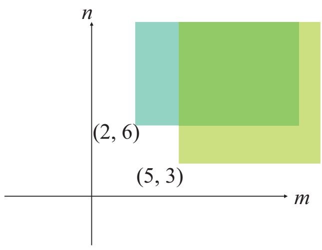
bringen läßt, wo $A_1, A_2, \ldots, A_m$ geeignete Formen der nämlichen n Veränderlichen sind.

Die Ordnungen der einzelnen Formen der vorgelegten Reihe sowie ihre Koeffizienten unterliegen keinerlei Beschränkungen. Denken wir uns die letzteren als Zahlen eines bestimmten Rationalitätsbereiches, so dürfen wir annehmen, daß die Koeffizienten der Formen $A_1, A_2, \ldots, A_m$ dem nämlichen

[1] Vgl. die vorläufigen Mitteilungen des Verfassers: „Zur Theorie der algebraischen Gebilde", Nachr. Ges. Wiss. Göttingen 1888 (erste Note) und 1889 (zweite und dritte Note). Dieser Band Abh. 13 bis 15.

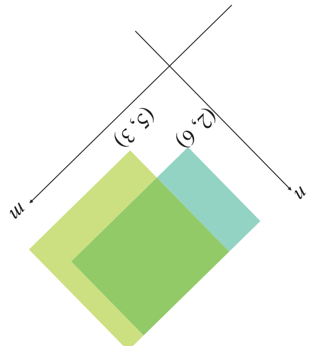
Dicksonの補題(1)

$$x^m y^n z^k \Rightarrow (m, n, k)$$

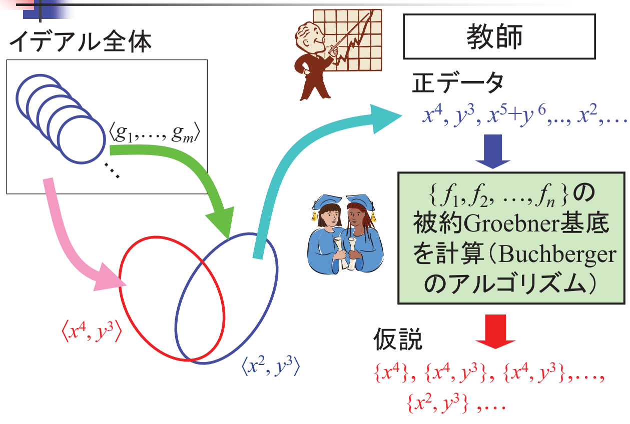


Dicksonの補題(2)

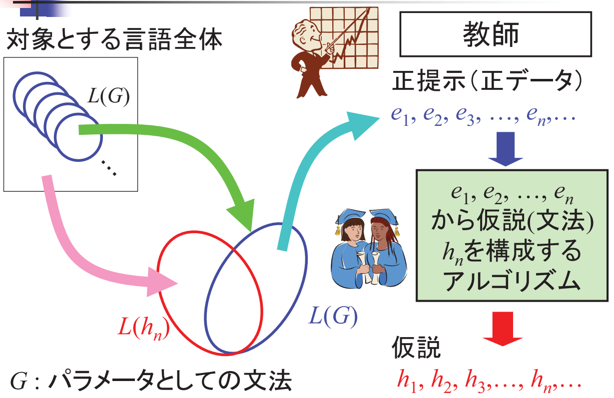
$x^m y^n z^k \Rightarrow (m, n, k)$



Q$[X_1,\ldots, X_n]$のイデアルの学習



形式言語の正提示からの極限同定学習



前後10年程度の研究テーマ

機械学習・データマイニングのための
代数的手法
数値の離散化

- 共通のキーワード:「閉集合」
- 分担・協力いただいている方々
  - 徳永浩雄(代数学, 暗号)
  - 平田耕一(データマイニング, 計算論的学習)
  - 立木秀樹(計算可能性解析, フラクタル)
  - 廣渡栄寿(計算論的学習, 計算可能性解析)
  - 杉山麿人(計算論的学習, 統計)

閉集合

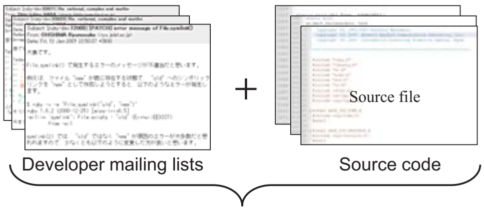
- 代数: 環, 体, イデアル,...
- 位相: 実数の閉区間, ...
- 記号論理: 論理的帰結,...
- データマイニング: 飽和集合
- グラフ理論: クリーク

...

話題1
代数的閉集合の応用

Nguen, Doi, Yamamoto
DS2009

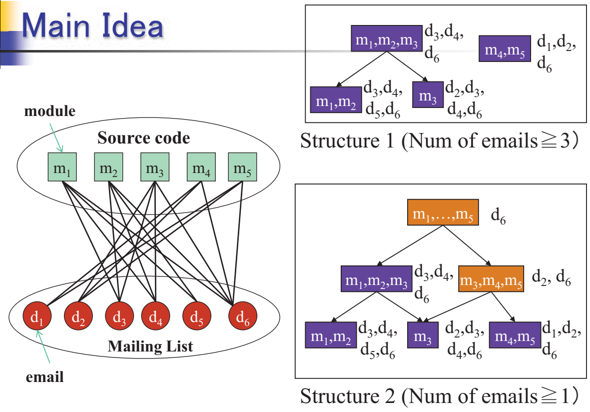
メーリング・リストからのプログラム構造抽出



メールとモジュールの間の関連性を抽出し，データマイニング手法を使い，プログラムの構造を抽出する

現在のところ日本語で書かれたメールを対象

Main Idea



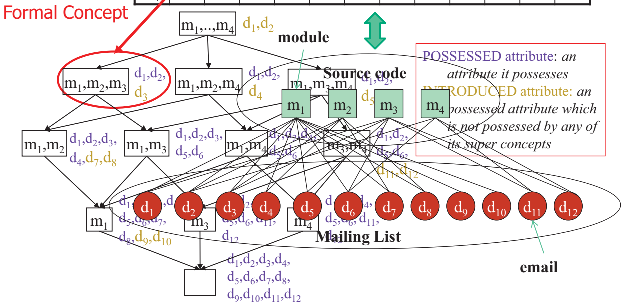
Assumption: a set of modules relating to same emails constitutes a subsystem

[Wille 84]: *Formal Concept Analysis*

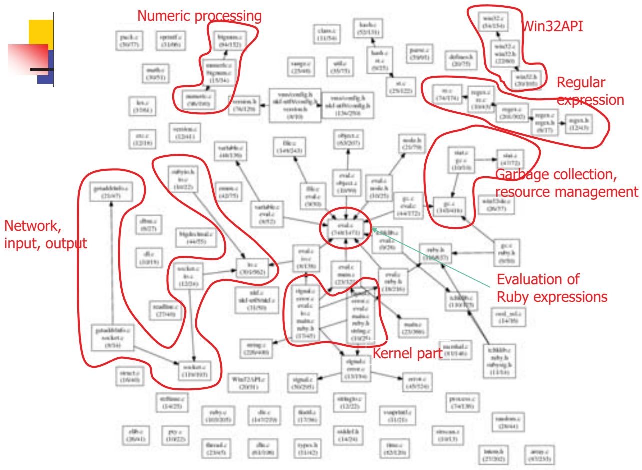
Generating Formal Concept Lattice

	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}
m_1	●	●	●	●	●	●	●	●	●	●		
m_2	●	●	●	●			●	●				
m_3	●	●	●		●	●					●	●
m_4	●	●		●	●	●					●	●

Formal Concept

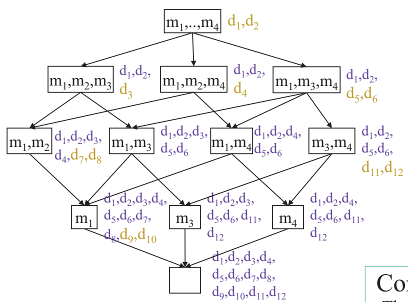


POSSESSED attribute: *an attribute it possesses*
INTRODUCED attribute: *an possessed attribute which is not possessed by any of its super concepts*



An extracted structure of the Ruby processor (σ=10,τ=8)

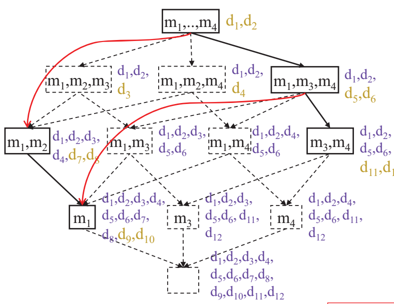
An Example of Applying PRUNIA (σ=2,τ=2)



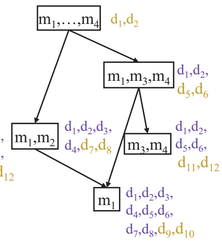
Lattice before pruning

Constraints:
- The number of its possessed attributes ≧σ
- The number of its introduced attributes≧τ

An Example of Applying PRUNIA (σ=2,τ=2)



Lattice during pruning



Lattice after pruning
（Extracted structure）

Properties of PRUNIA:
1. The representation of super-sub concept order is persevered
2. No redundant path

代数的閉集合

- A mapping $c: 2^U \to 2^U$ is a closure operator if it satisfies:

  $X \subset c(X)$

  $c(c(X)) = c(X)$

  $X \subset Y \Rightarrow c(X) \subset c(Y) \quad (\forall X, Y \subset U)$

- $X \subset U$ is closed if $c(X) = X$.

Galois対応による閉集合

- A binary relation R ⊆ O×A
- f : O → A and V : A → O

  f(S) = { a∈A | (o, a)∈R for some o∈S}

  g(T)= { o∈O | (o, a)∈R for some o∈T}

- Then both h(S) = g(f(S)) and k(T) = g(f(T)) are closure operators.
- A pair of a closed set S in O and f(S) (T in A and g(T)) is called a formal concept.

話題2
閉集合の意味論

望ましい閉集合

- コンパクト性(有限基底)

  閉集合 X = C($\{x_1, \ldots, x_n\}$)

  - 整数イデアル，多項式イデアル
  - 帰結集合
  - フラクタル(近似的に)

    コラージュ定理