

和田研究室 研究グループ

機械学習と データマイニング

<http://www-wada.elcom.nitech.ac.jp/study/MLweb/>



キーワード
機械学習
データマイニング
論理
人工知能

人の思考や論理を参考に、計算機に学習や発見をさせるための方法を探る。

機械学習とデータマイニング

機械学習 = 人が経験から学習する仕組みを研究し工学応用する**人工知能**の一分野。

データマイニング = 大量のデータから役立ちそうな知識(規則性)を発見する方法を探る**データベース**の応用技術分野。



データベースからルール を見つける

ジャンル	価格	装丁	ページ数	言語	買う?
解説書	2000円	ソフトカバー	220ページ	英語	no
小説	650円	ソフトカバー	310ページ	日本語	yes
エッセイ	1200円	ハードカバー	180ページ	日本語	yes
小説	1000円	ハードカバー	450ページ	日本語	yes
学術書	2800円	ソフトカバー	320ページ	日本語	no
小説	950円	ハードカバー	210ページ	英語	no
...

{ (ジャンル = 小説) (ジャンル = エッセイ) }
(言語 = 日本語) (価格 < 1500円)
買う = yes

:かつ
:または

論理と知識がキーワード

経験、データ

知識、論理、ルール

環境、行動、予測

環境の意識的な理解。
高い可読性。

本当は学習に知識など必要ないのかも。
生物進化の模倣(自然淘汰)、神経回路網の自己組織化、統計学習



論理と知識

論理(logic)は人が自然に正しいと考えるステップを決める仕掛けです。

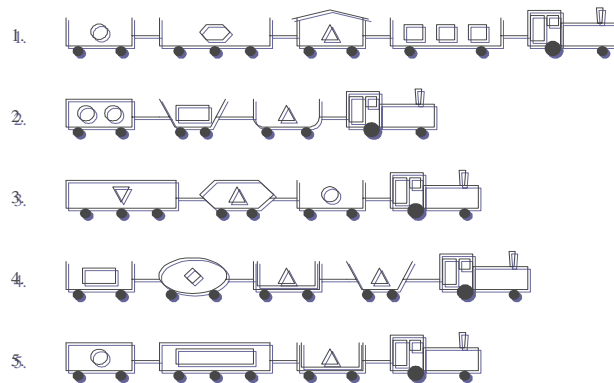
(論理と論理回路は、見掛け以外はほとんど無関係)

論理を使うと、

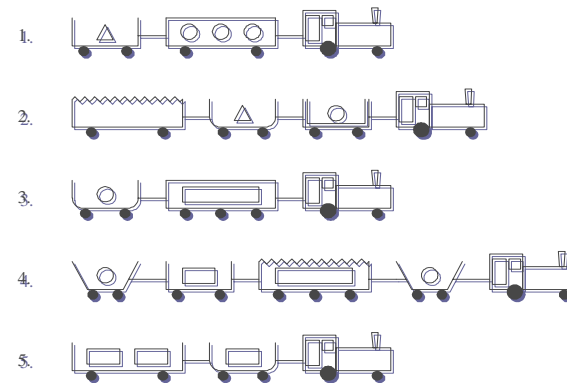
- 人が知識として考える規則性や事実をそのまま記述できる。
- 知識から導かれる結論を、推論(=計算)できる。
- コンピューターでそれらを実現できる。

東行きはどんな列車？

1. TRAINS GOING EAST



2. TRAINS GOING WEST



東行き

西行き

○ 事例

東行き(列車1)

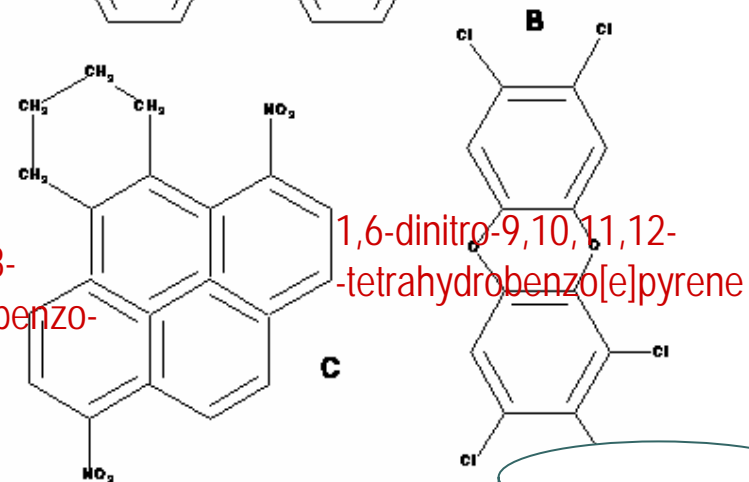
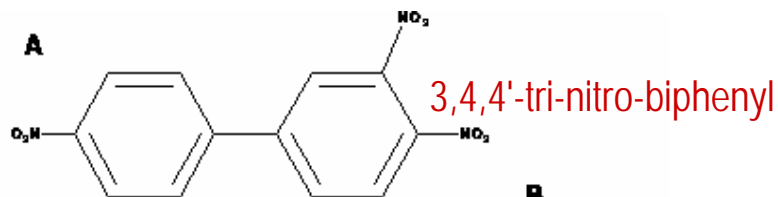
貨車(列車1,車1),方形(車1),短い(車1),屋根なし(車1),車輪数(車1,2対),荷物(車1,荷1),円形(荷1),荷物数(荷1,1),
 貨車(列車1,車2),方形(車2),長い(車2),屋根なし(車2),車輪数(車2,3対),荷物(車2,荷2),六角形(荷2),荷物数(荷2,1),
 貨車(列車1,車3),方形(車3),短い(車3),山屋根(車3),車輪数(車3,2対),荷物(車3,荷3),三角形(荷3),荷物数(荷3,1),
 貨車(列車1,車4),方形(車4),長い(車4),屋根なし(車4),車輪数(車4,2対),荷物(車4,荷4),方形(荷4),荷物数(荷4,3).

○ 仮説

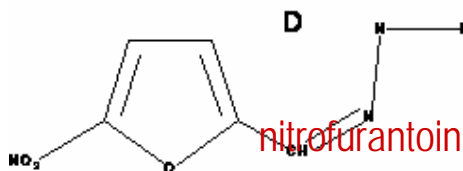
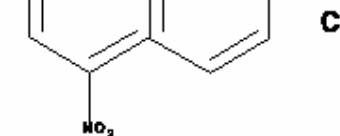
東行き(列車):-貨車(列車,車),短い(車),not 屋根なし(車).

短くて屋根なしではない貨車をもつ列車が東行き

物質の突然変異性予測



2-nitro-1,3,7,8-
-tetrachlorodibenzo-
-1,4-dioxin



化学構造等に
関する知識

化学物質は、列車が貨車を接続しているのと同じように、原子を接続し、化学的性質を乗せている。

突然変異性あり(分子1)

原子(分子1, 原子1)

元素(原子1, c),

電荷(原子1, 0.12),

...

結合(分子1, 結合1)

結合タイプ(結合1, 2重結合)

...

このグループの最近の課題

- **時間系列**を含むデータからの知識発見。
- 大量の**文書データ**からの知識発見。
- **一階述語論理** (高級な論理) を用いたデータマイニング。
- **強化学習** (試行錯誤学習) の結果から知識を獲得して、将来の試行錯誤に活かす研究。

