

# 「ネットワークと機械学習」入門

～ 標準タスクと基本モデル ～

SNEAK PREVIEW

鹿島 久嗣



DEPARTMENT OF MATHEMATICAL INFORMATICS

2012/03/13 NLP 2012 チュートリアル

# 本チュートリアルの概要：

## ネットワーク構造をもつデータを扱う問題と手法を概観

---

- ネットワーク構造をもったデータを扱う機械学習問題を、  

$\{\text{内部}, \text{外部}\} \text{ネットワーク} \times \{\text{ノード}, \text{リンク}\} \text{推論}$

  
の4通りに分類
- それぞれに対する基本的なアプローチとモデルを解説
  - グラフカーネル、グラフマイニング、条件付き確率場  
ラベル伝播、行列／テンソル分解  
マルコフネットワーク、確率的ブロックモデル

# 本講演の想定する聴衆：

機械学習は知っているがネットワークの扱いは想像つかない

- 機械学習の基本的なところは何となく知っている
  - 教師付き／なし学習、最尤／MAP／ベイズ推定
- ネットワーク構造（グラフ）を扱う解析といわれてもピンとこない

ひじょうにピンポイント！

# モチベーション

# 機械学習：

## データにもとづく分析や予測のための道具

---

- IT技術や計測技術の進歩などに支えられた情報基盤
- データを有効に活用するためのデータ解析技術への注目
  - 機械学習、統計、データマイニング、データ工学、...
  - 研究者だけではなく、産業界もその可能性に注目
- データ解析手法の大別：
  - 発見（分析）的な解析「いま何が起きているのか？」
  - 予測的な解析「これから何が起こるのか？」

# 統計的機械学習の成功： さまざまな分野でうまくいった

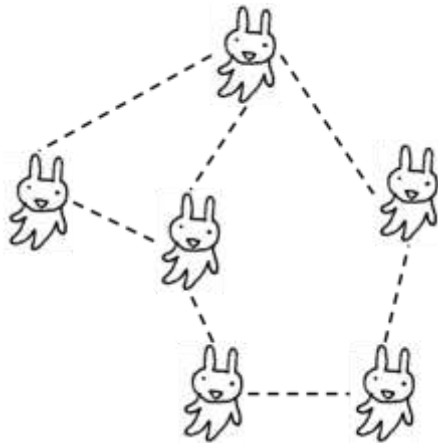
---

- 前世期半ばに始まった機械学習研究
  - 述語論理を基礎にした厳密な推論
  - ニューラルネットワーク
- 近年の機械学習の成功
  - ～ とりわけ「統計的」機械学習の各種応用における成功
  - 自然言語処理：文書分類、機械翻訳、評判分析など
  - ほか、バイオ、マーケティング、金融、Webなど

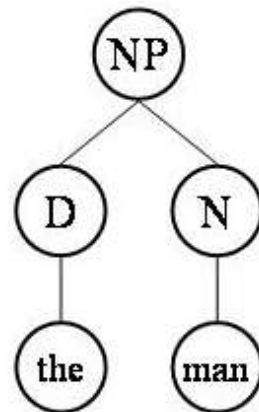
# ネットワーク構造をもったデータ:

自然言語処理、バイオ、Web等さまざまな分野で登場

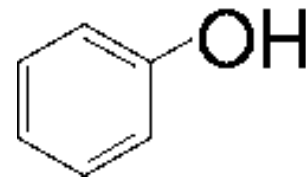
- ネットワーク構造をもったデータ
  - = グラフ（木、配列）によって表現されるデータ
  - 文書、構文木、Web、XML/HTML、化合物
  - ソーシャルネットワーク、DNA／タンパク質配列、RNA
  - 生体ネットワーク、引用関係、企業間取引、...



ソーシャルネットワーク



構文木



化合物

AGCTCGAA...

DNA配列

# ネットワークデータ解析における困難： 非ベクトル形式データの扱いは自明でない

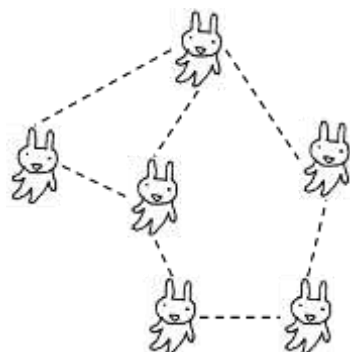
- 多くの手法ではベクトル（表）形式のデータを仮定

顧客番号	顧客氏名	年齢	性別	住所	...
0001	○○	40代	男性	東京都	...
0002	××	30代	女性	大阪府	...

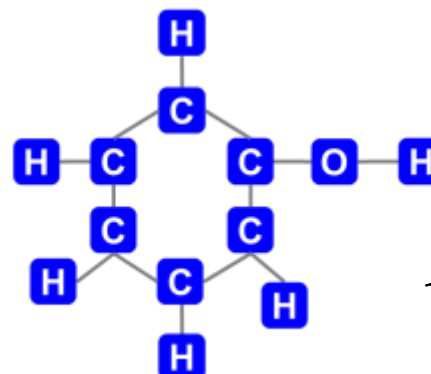
ベクトル

- 一方、非ベクトル形式データの扱いは自明ではない

## — ネットワーク構造はその最たるもの



隠れた人間関係  
を知りたい

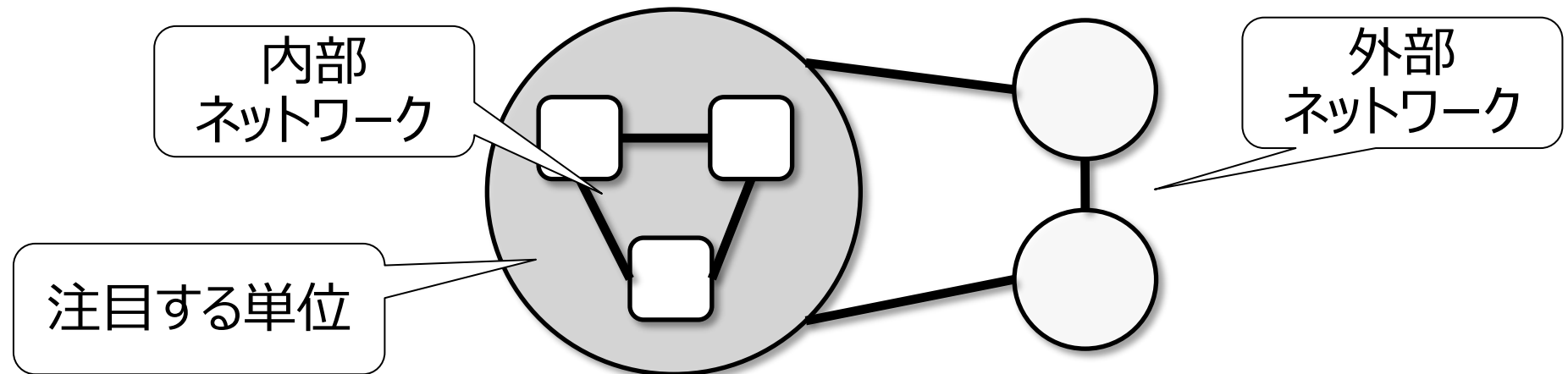




# ネットワーク構造の分類：

## 注目する単位の内側のネットワーク構造の2種類

- ネットワーク構造には 2 つの種類が存在する：
  1. 内部ネットワーク：  
注目するデータ単位の内側にあるネットワーク
  2. 外部ネットワーク：  
注目するデータ単位の外側にあるネットワーク

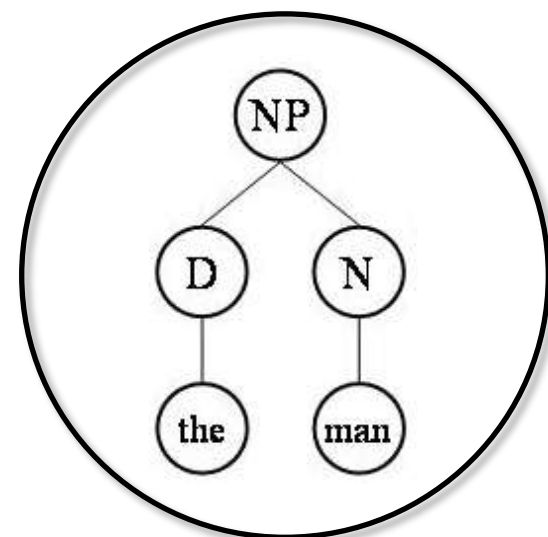


# 内部ネットワーク：

## 注目するデータ単位の内側にあるネットワーク構造

### ■ 内部ネットワークの例：

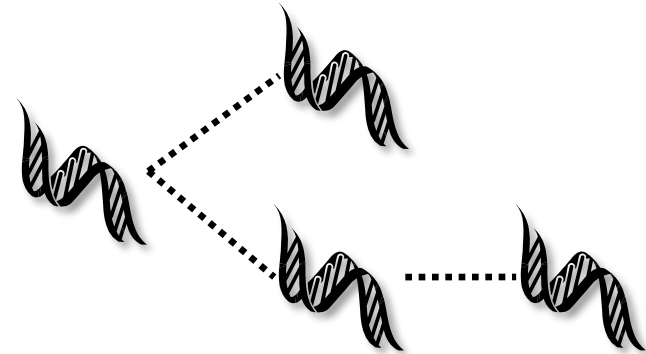
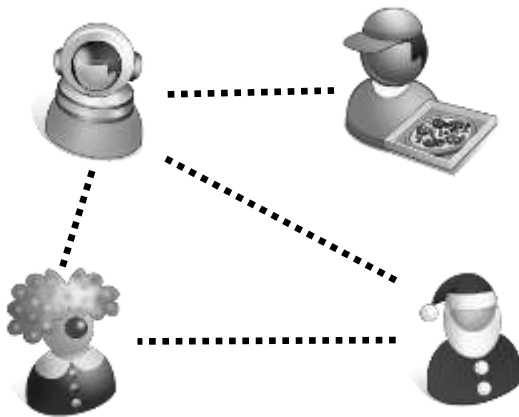
- 文書分類：文書に注目すると  
文書は内部に配列構造をもったデータ
- 構文解析：文に注目すると  
文は内部に木構造をもったデータ
- 活性予測：化合物に注目すると  
化合物は内部にグラフ構造をもったデータ



# 外部ネットワーク： 注目するデータ単位の外側にあるネットワーク構造

## ■ 外部ネットワークの例：

- 友人推薦： 人に注目すると  
ソーシャルネットワークは外部にグラフ構造をもったデータ
- 系統樹推定： 遺伝子に注目すると  
系統樹は外部に木構造をもったデータ

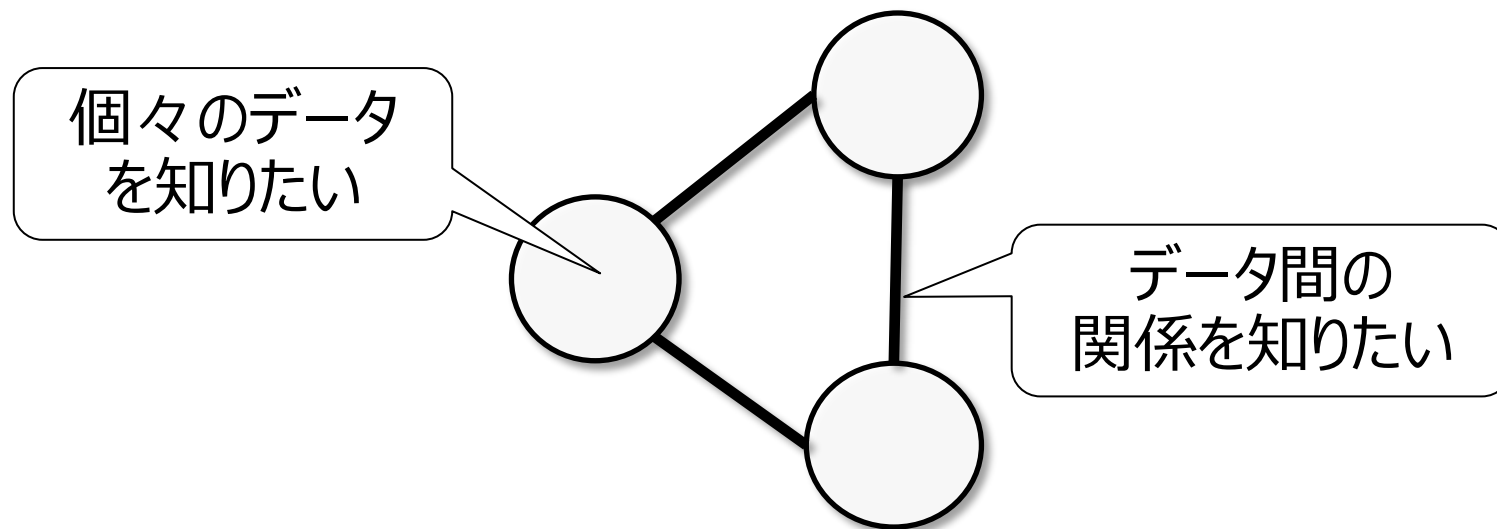


# 解析のフォーカス：

個々のデータに興味がある  $\Leftrightarrow$  関係に興味があるのか

---

- 解析のフォーカスとしても2種類ある
  1. 個々のデータの性質に興味がある
  2. データ間の関係について興味がある



# ネットワーク構造解析の世界観：

## 2 × 2 の 4 通りの分類

- {内部, 外部}ネットワーク  
× {個々のデータ, データ間関係}についての推論 の4通り

		ネットワーク構造の種類	
		内部ネットワーク	外部ネットワーク
フォーカス 解析の	個々のデータ	予測 クラスタリング 構造ラベリング パターン発見	予測 クラスタリング ランキング
	データ間の関係	(内部) 構造予測	リンク予測 構造変化解析

# ネットワーク構造解析のモデル： それぞれの分類にそれぞれのモデル

		ネットワーク構造の種類	
		内部ネットワーク	外部ネットワーク
フォ ー カ ス の 解 析	個々の データ	カーネル法 パターンマイニング	ラベル伝播 行列／テンソル分解 マルコフネットワーク 確率的ブロックモデル
	データ間 の関係	構造学習器 (HMM、CRF等)	リンク指標 ペアワイズ予測 行列／テンソル分解 マルコフネットワーク 確率的ブロックモデル

こんた



かんじ