

関係データとテンソル分解

講演者

林 浩平 (NAIST)、石黒 勝彦 (NTT)、ダヌシカ ボレガラ (東大)

セッションオーガナイザー

富岡 亮太 (東大)

関係データと行列

- 例: Netflix

映画

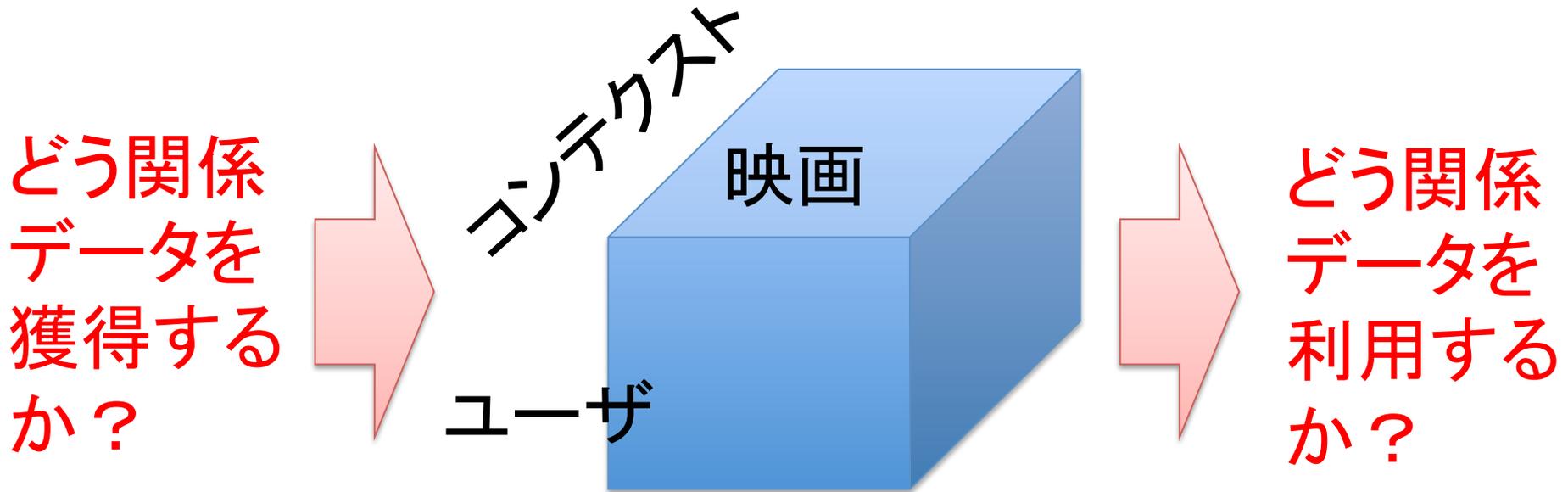
ユーザ

	Star Wars	Titanic	Blade Runner	...
User 1	5	2	4	
User 2	1	4	2	
...				

- 普通の多変量データも関係データと思えなくもない。
- さらにユーザ間の類似度、映画間の類似度も関係データ

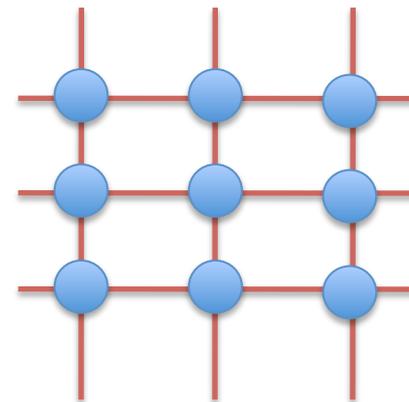
世の中は関係データで溢れている！

関係データとテンソル



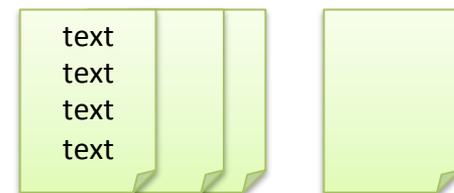
- 状況や時間に依存する関係
- 複数種類の関係
- 複数のオブジェクトの間関係

関係データをどう獲得するか

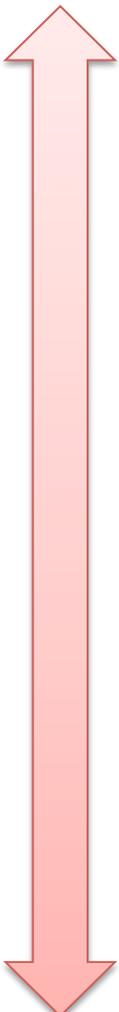


- 事前知識
 - 半教師付き学習
 - マルチタスク学習
- 関係が陽に得られている場合
 - Netflix など
- 大量のテキストデータから非自明な関係を獲得 (ボレガラさん)

	M1	M2	M3	...
U1	5	2	4	
U2	1	4	2	
...				



自明



非自明

関係データをどう利用するか

- テンソルの低ランク分解 (1960-)
 - 多重線形性に基づく数学的にきれいな世界
 - 凸最適化も適用可能 (Tomioaka et al 2010)
 - ランクの計算がNP完全だったり、最良ランク-R近似が存在しないなど数学的に非自明な性質も
- クラスタリング
 - ユーザや映画のグループ化 -> 石黒さん
 - 関係のグループ化、分類 -> ボレガラさん
- クエリー
 - 関係の意味で似ているペアを検索-> ボレガラさん

関係データとテンソルの間の接点は？

- 関係データの表現としてのテンソル
- 関係の関係は高次テンソルとして表現できるか？
- テンソル分解的な見方を関係データに持ち込むメリットは？

テンソル分解とクラスタリングの関係 (1/2)

- K-means クラスタリング

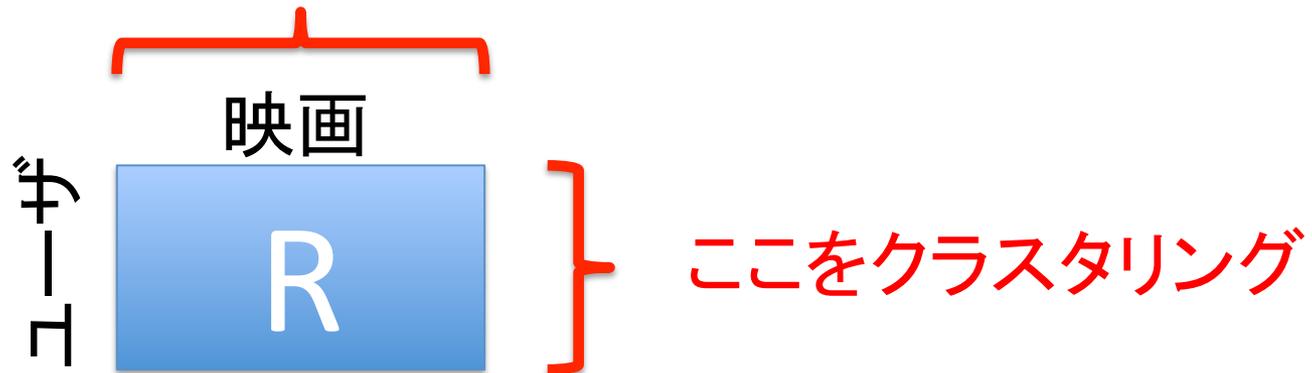
$$\begin{array}{c} N \\ D \quad \boxed{X} \\ \text{データ} \end{array} = \begin{array}{c} K \\ D \quad \boxed{M} \\ \text{平均} \end{array} \times \begin{array}{c} N \\ K \quad \boxed{Z} \\ \text{クラスタ割り当て} \\ (0 \text{ or } 1) \end{array}$$

- 無限関係モデル (IRM)

$$\begin{array}{c} N_2 \\ N_1 \quad \boxed{R} \\ \text{関係データ} \end{array} = \begin{array}{c} K_1 \\ N_1 \quad \boxed{Z_1} \\ \text{割当1} \end{array} \times \begin{array}{c} K_2 \\ K_1 \quad \boxed{\eta} \\ \text{相互作用} \end{array} \times \begin{array}{c} N_2 \\ K_2 \quad \boxed{Z_2} \\ \text{割当2} \end{array}$$

テンソル分解とクラスタリングの関係 (2/2)

- ユーザや映画をクラスタリングする方法 (IRM)



- 関係をクラスタリングする方法



テンソル
分解
(林さん)

関係の抽
出、関係
の検索
(ボレガラさん)

関係データの
テンソル表現

時間変化
する関係
モデル
(石黒さん)

