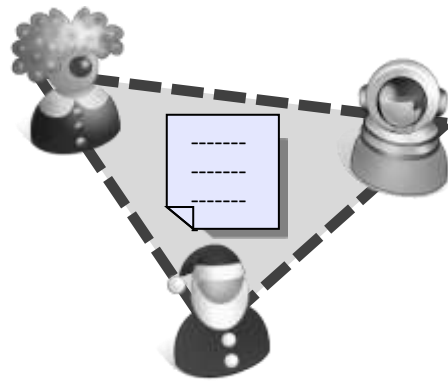


高精度でスケーラブルな多項関係予測の実現



鹿島 久嗣

東京大学

情報理工学系研究科

数理情報学専攻

$\text{Prob}(\text{co-author}(\text{Alice}, \text{Bob}, \text{Carol}, 2010, \text{Journal})) = 0.7$



概要：多数のデータ間の関係を予測する方法の開発と応用を通じて、実社会／学術的インパクトを目指します

- IT技術や計測技術の進歩などに支えられた情報基盤が急速に整いつつある現在、集められたデータの解析、とりわけ直接的に意思決定に結び付く予測技術の重要性が増している
- 従来の予測的データ解析では、個々のデータが対象とされてきたが、近年では一層高度で有用な予測を目指し、データの間の高次関係の予測が重要であると認識されつつある
- 高次関係の予測は、経営、マーケティング、バイオ、創薬など実世界の様々な分野の重要な場面における本質的な役割をもつが、現状ではこれを高精度で、効率良く実現する方法手法は存在しない
- 本研究では、近年著しい発展を遂げているテンソル分析、情報統合、最適化アルゴリズム等の先端技術に基づき、汎用、高精度でかつスケーラブルな高次関係予測機構を実現し、さらにこれを実問題に应用することで実社会へのインパクトを目指す

研究提案：高精度でスケーラブルな多項関係予測の実現

① 研究の背景：

「個別の予測」から「関係の予測」へ

② 本研究のフォーカス：

汎用／高精度／大規模な高次関係予測

③ 技術的チャレンジ：

「データ過疎対処」「スケーラビリティ向上」「予測精度向上」

④ 課題解決のためのキーアイデア：

「テンソル構造」「情報統合」「凸最適化」

① 背景：データの飽和により一層高まる解析技術 とりわけ予測技術は重要性が高い

- 近年、IT技術や計測技術の進歩などに支えられ電子的な情報基盤が急速に整いつつある
- 集められたデータを有効に活用するためのデータ解析技術の重要性がますます増している
 - 統計／機械学習／データマイニング／データ工学／...
- データ解析手法の大別：
 - 現状分析的な解析「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？」
- 直接的に競争力のある意思決定に結び付く予測的データ解析
 - 人が病気にかかりやすいかどうかの予測 → 生命や健康
 - 顧客が商品を購入するかどうかの予測 → 経済的利益

① 背景：データ解析技術の潮流は、個々のデータの解析からデータ間の関係の解析に移行しつつある

- 従来：「個々のデータを対象とした解析」から

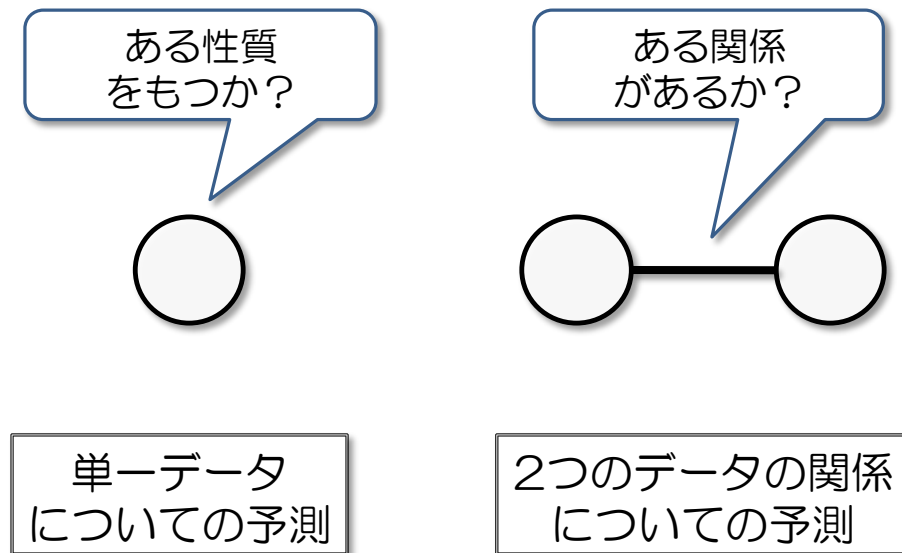


近年：「**データ間の関係の解析**」へと移行しつつある

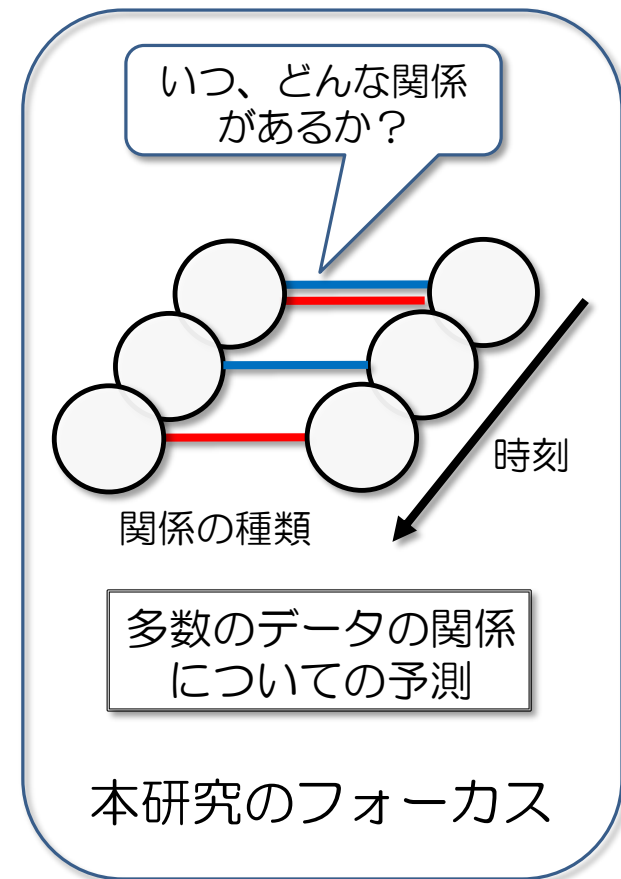
- 関係の分析は**様々な領域において盛ん**に行われつつある
 - ソーシャルネットワーク分析：人間関係
 - オンラインショッピング：顧客と商品の間の関係
 - 創薬スクリーニング：薬剤と標的の関係
- データ間の関係に注目することで、**個々のデータに注目しているだけでは見えない性質**が見えてくる
 - コンピュータネットワーク上のプロセス依存関係から異常を予測
 - 複数の脳波時系列の相関関係から思考を読みとる

② 本研究のフォーカス： 汎用／高精度／大規模 な データ間の高次関係予測

- 本研究では、現状の関係の予測をさらに推し進め、
3次～4次～さらに高次の関係を予測する汎用的な方法論を展開する
 - 現状ではせいぜい2～3のデータ間の関係を扱うのが限界
 - スケーラビリティとデータ過疎の壁



従来研究



本研究のフォーカス

③ 技術的チャレンジ： データ過疎の解決、スケーラビリティと予測精度の向上

1. データ過疎問題

- 高次の関係になるほど、データの組み合わせ数は指数的に増える
- 予測モデルを学習するのに使える事例数は相対的に小さくなる

2. スケーラビリティ問題への対処

- 関係に含まれるデータの数が多くなるにつれ、これを予測するのに用いる組み合わせ特徴のパラメータ数は爆発的に増加する

3. 予測精度の向上

- 多くの手法では非凸最適化もしくは組み合わせ爆発により準最適解に留まる

④ 課題解決のためのキーアイデア： テンソル構造の導入、情報統合、凸最適化問題としての定式化

- 立ちふさがる3つの困難を解決する鍵：パラメータへの
テンソル構造導入、情報の自動選択と統合、凸最適化としての定式化

