

01

データに合った解析法を自動で構築します ～階層情報を用いた潜在変数モデルの自動生成～

どんな研究

与えられた**複雑なデータを解析するための潜在変数モデルを自動的に生成する**研究です。例えばブログ記事は、ユーザ、記事、文、単語などの複雑な階層構造を持ちます。提案法はデータの持つ階層構造の中から解析に必要な情報を抽出し、自動的にデータをモデル化します。

どこが凄い

これまで潜在変数モデルの設計は、データと機械学習の双方に詳しい専門家が時間をかけて設計、実装、テストを行っていました。提案手法はデータの持つ**階層構造を反映した多様なモデルを自動生成可能**であるため、これらの工程をすべて自動化できます。

目指す未来

現在、多様なデータが大量に蓄積されています。これらのデータを解析することで様々な情報や知見を得ることができます。本技術を発展させ、**蓄積データの解析、知見の抽出が自動化**できれば、データをただ蓄積するのではなく、**知識や知見を自動的に蓄積するシステム**が実現できると期待されます。

複雑化するデータ

- 多様なデータが大量に蓄積されている。
- 特に **階層構造** を持つデータが増えている。
- データに直接現れない **潜在構造** を自動抽出したい。

階層構造の例

ブログ : ユーザ - 記事 - 文 - 単語
 音楽 : 作曲家 - 曲 - 小節 - 音符
 購買 : 日時 - 客 - 店舗 - 商品
 センサ : 場所 - 日 - 時間 - 検出量

単語トピック
 ユーザクラスタ

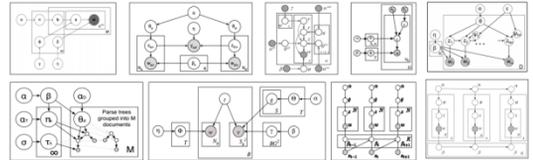
階層毎の隠れた構造を抽出したい

階層情報を活かした解析は困難 ☹

複雑化する潜在変数モデル

- データが持つ **潜在構造** を **潜在変数** を用いて抽出する。
- データの複雑化に伴い、様々なモデルが提案されている。
- データに合ったモデル設計は専門家しかできない。

潜在変数モデルの例



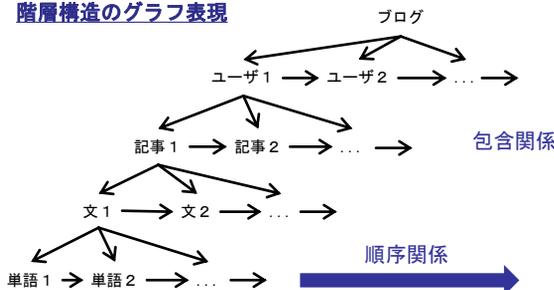
データと機械学習、双方の専門知識が必要 ☹

提案：階層構造を利用した潜在変数モデルの自動生成

- 適切な階層を自動抽出** することで様々なモデルを生成する。
- 単一の汎用アルゴリズム** で全モデルを学習できる。
- 動的計画法** により、効率的に計算できる。

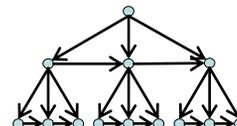
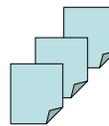
- データが変化しても、**再設計、再導出、再実装不要!**
- データ中の**潜在構造(クラスタ)**を**完全自動抽出可能!**

階層構造のグラフ表現



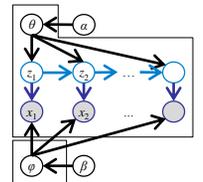
データの包含関係や順序関係を考慮したモデルを生成 ☺

入力



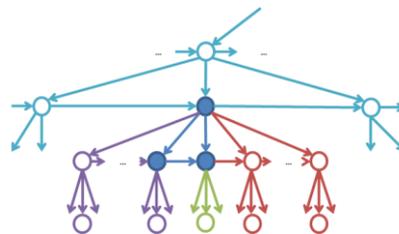
データ + 階層構造を表現するグラフ

出力



潜在変数モデル

動的計画法に基づく効率的な学習



外側確率
 内側確率
 前向き確率
 後向き確率

4種の確率を再帰的に計算することで効率的な学習を実現 ☺

関連文献

- [1] M. Ishihata, T. Iwata, "Generating structure of latent variable models for nested data," in *Proc. The 30th Conference on Uncertainty in Artificial Intelligence (UAI-2014)*, 2014.
- [2] 石島正和, 岩田具治, "データの階層構造を利用した潜在変数モデルの自動生成." 第28回人工知能学会全国大会, 2014.

連絡先

石島正和 (Masakazu Ishihata) 協創情報研究部 創発知能環境研究グループ
 E-mail: ishihata.masakazu(at)lab.ntt.co.jp

