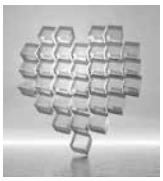

潜在変数モデルに基づく知識発見

【受賞タイトル】確率的潜在変数モデルに基づくデータマイニングに関する研究

岩田具治　NTT コミュニケーション科学基礎研究所

このたび，長尾真記念特別賞をいただくことができ，たいへん光栄に思う．このような栄誉ある賞をいただけたのは，これまでお世話になった先生，先輩，同僚，共同研究者の方々のおかげであり，皆様に感謝したい．

受賞の対象となったのは確率的潜在変数モデルを用いたデータマイニングに関する研究である．従来のマイニング技術は，頻出パターン抽出に代表される列挙に基づく手法であるため，複雑な構造を持つパターンの発見やノイズが多いデータへの適用は困難であった．私は「潜在変数モデル」と呼ばれる隠れ変数を持つ確率モデルを土台とするアプローチにより，さまざまなタスクに対して従来法の問題を克服する手法を提案した．確率モデルを用いることにより，ノイズなどの不確実性を扱うことができる．インターネットやソーシャルメディアの普及に伴いノイズを含む膨大なデータが容易に入手可能になり，確率モデルの重要性は急速に高まっている．

潜在変数モデルは，k 平均法，主成分分析，正準相関分析，次元削減，潜在意味解析など，統計などの分野で長年利用されてきた手法を確率モデルの枠組みで一般化したものと見なせる．確率モデルの利点は異種モデル・異種情報を確率論の枠組みで統合できることである．この利点を活かし，たとえば，潜在意味解析と次元削減の統合によるトピックを忠実に反映した文書群の可視化，非線形次元削減とクラスタリングを組み合わせた複雑な形を持つクラスタの抽出，画像とアノテーションの統合解析による関連するアノテーションの抽出などの研究を行った．

多言語コーパスからの普遍文法の学習や，ファッション雑誌の写真を参考にしたコーディネートの推薦，会話データからの影響力推定などの新しいタスクにも取り組んだ．新しいタスクの着想は，新しいデータに触れたり，さまざまな分野の人たちと議論するなかで得られた．潜在変数モデルを用いることで，興味を持った多様な課題に対して挑戦できよかったと思う．

最近取り組んでいるテーマは教師なしオブジェクトマッチングである．オブジェクトマッチングとは，異なるデータ集合の間で関連するオブジェクトを見つけるタスクである．たとえば，意味が同じ英語と日本語の単語の対応付け（辞書自動作成），画像と文の対応付け（説明文作成），複数のデータベースの ID の対応付け（名寄せ）に利用できる．このタスクに対し，複数のデータ集合が 1 つの潜在空間を共有する潜在変数モデルを提案した．提案法は，人手で対応付けたデータや異種データ集合間の類似尺度が不要であり，オブジェクト間に多対多の対応がある場合や，2 つ以上のデータ集合が与えられた場合でも適用可能である．現在，データは多様な用途で収集・管理されている．そのため異なるデータ集合の間の関係を解析することが困難であった．マッチング技術により複数のデータ集合の統一的解析が可能となり，これまで知り得なかった因果関係や知識の発見につながる．

タスクやデータの本質を理解してモデルを設計することは，この研究における醍醐味であった．しかし，モデル設計には専門知識が必要となる．より多くの場面で利用される技術とするために，モデル自動設計の研究が今後重要になると思われる．この受賞を励みにして社会に貢献できるよう研究に励んでいきたい．

（2014 年 5 月 14 日受付）

岩田具治（正会員） iwata.tomoharu@lab.ntt.co.jp
2001 年慶大・環境情報卒業．2003 年東大大学院・総合文化修士課程修了．同年 NTT 入社．2008 年京大大学院・情報学博士課程修了．博士（情報学）．機械学習，データマイニング，情報可視化の研究に従事．