

## 論文内容の要旨

博士論文題目 Latent Variable Models for Discrete Data and the Learning  
Methods

(離散データのための潜在変数モデルとその学習法)

氏名 小西 卓哉

(論文内容の要旨)

Latent variable models are probabilistic models that are widely applied in modern data analysis. Latent variable models have unobserved random variables, which are beneficial for revealing hidden structures behind observations and giving meaningful interpretation to complex data.

While the latent variable models became known as popular methods, there are still research questions remaining. A practical issue is how to design latent variable models according to the properties of tasks. Unless suitable models are used, the users will fail to obtain desirable latent representation. Another issue is how to learn the models when observed data are given. While efficient learning algorithms have been proposed for many models, several models have not been explored enough. Such exploration is important for clarifying the characteristics of learning methods and finding better ones for the models.

On the basis of the above perspectives, this thesis studies two specific problems about the latent variable models. In Chapter 3, we focus on search queries on the Web search engines. A search query consists of a combination of terms and the possible number of them is enormous. However, the search queries can be represented as common low-dimensional patterns. We propose a probabilistic topic model that extracts such patterns as pairs of latent topics. Using two real query datasets, we demonstrate that the obtained topics are intelligible by humans and are highly accurate in keyword recommendation and query generation tasks.

In Chapter 4, we study the variational Bayesian inference methods of the infinite relational model for network data that have not attracted attention. We derive the collapsed variational Bayesian inference that we obtain by marginalizing out the parameters analytically. The collapsed variational Bayesian inference empirically outperforms the standard variational Bayesian inference in many real network datasets. The results imply the collapsed variational Bayesian inference indicates even better performances in dense networks.

氏名	小西 卓哉
----	-------

(論文審査結果の要旨)

潜在変数モデルは近年のデータ解析で広く用いられている手法であり、観測データに内在する構造を抽出し、データから関係性を見出すことができる。しかし与えられたデータに対してどの潜在変数モデルを用いればよいかを定める一般的な方法はないため、特に実用性の高いデータに対してはそれを表す適当なモデルを見つけることが重要になる。また、データからモデルのパラメータを推定する方法についても、計算量と精度の観点から適切なものを選ぶ必要がある。本研究はこれらの問題に対して新たな方法を提案し、その有効性を示したものである。

本研究ではまず、ウェブサーチエンジンにおける 2 語のクエリを対象とし、それらを低次元のパターンで表すトピックモデルを提案した。その結果、提案手法によりデータからトピックを抽出することに成功し、それがキーワード推薦などに有効であることを示した。

次に、潜在変数モデルの中でも利用度の高い無限関係モデルについて、従来の変分ベイズ法を改良した周辺化変分ベイズ法およびその近似解法を提案した。提案手法は、ネットワークに関する実データにおいて従来法よりもよい性能を発揮し、特に密なネットワークでその性能向上が大きいことを示した。

以上をまとめると、本論文は潜在変数モデルにおいて新たなモデルを提案するとともにパラメータ推定法をも提案した研究であり、ビッグデータ時代のデータ解析に資すると考えられる。よって、博士（工学）の学位に値するものと認められる。