社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

複雑ネットワークのトポロジ構造を利用した ネットワークマイニングフレームワークとその応用

杉山　浩平† 大崎　博之† 今瀬　真† 八木　毅†† 村山　純一††

† 大阪大学 大学院情報科学研究科
〒 565-0871 大阪府吹田市山田丘 1-5
†† 日本電信電話株式会社 NTT 情報流通プラットフォーム研究所
〒 180-8585 東京都武蔵野市緑町 3-9-11
E-mail: †{k-sugi,oosaki,imase}@ist.osaka-u.ac.jp, ††{yagi.takeshi,murayama.junichi}@lab.ntt.co.jp

あらまし　本稿では、さまざまなリンクマイニング手法や従来のデータマイニング手法を統合的に扱うことができるフレームワーク NMF (Network Mining Framework) を提案する。近年、リンクマイニングに代表される、複雑ネットワークのトポロジ構造を利用するさまざまな取り組みが活発に行われている。これまで数多くのリンクマイニング手法が提案されているが、これらは異なる分野でそれぞれ独立に研究されてきた。このため現状では、それらのリンクマイニング手法を相互に連携させて利用することは困難である。そこで本稿では、さまざまなリンクマイニング手法や従来のデータマイニング手法を統合的に扱うことができる、「ネットワークマイニング」のためのフレームワーク NMF を提案する。NMF は、入出力の形式が統一された 7 種類の機能ブロックによって構成され、これらの機能ブロックを自由に接続することにより、さまざまなネットワークマイニング手法の実現を可能とする。さらに本稿では、NMF の利用例を示し、NMF によって高度なネットワークマイニングが可能となることを示す。
キーワード　複雑ネットワーク、トポロジ構造、フレームワーク、ネットワークマイニング、リンクマイニング

NMF: Network Mining Framework using Topological Structure of Complex Networks

Kouhei SUGIYAMA†, Hiroyuki OHSAKI†, Makoto IMASE†, Takeshi YAGI††, and Junichi MURAYAMA††

† Graduate School of Information Science and Technology, Osaka University
1-5 Yamadaoka, Suita, Osaka 565-0871, Japan
†† NTT Information Sharing Platform Laboratories, NTT Corporation
3-9-11 Midoricho, Musashino, Tokyo 180-8585, Japan
E-mail: †{k-sugi,oosaki,imase}@ist.osaka-u.ac.jp, ††{yagi.takeshi,murayama.junichi}@lab.ntt.co.jp

Abstract　In this paper, we propose a network mining framework called NMF that supports integration of several link mining methods and the conventional data mining methods. In recent years, several researches using topological structure of complex networks represented by link mining have been performed actively. In the literature, although there many link mining methods exits, these methods have been independently developed in different fields. It is therefore difficult to cooperatively integrate utilize these link mining methods. In this paper, we propose a framework for network mining called NMF that supports integration of several link mining methods and the conventional data mining methods. NMF is composed of seven functional blocks with unified input and output format. By connecting these functional blocks flexibly, the advanced network mining methods are realizable. In this paper, we also present examples of NMF applications, and show how advanced network mining methods can be realized using NMF.
Key words　Complex Networks, Topological Structure, Framework, Network Mining, Link Mining

1 はじめに

近年、複雑ネットワークのトポロジ構造を利用するさまざまな取り組みが活発に行われている。例えば、複雑ネットワークのトポロジ構造が有する性質を分析する、ネットワーク分析と呼ばれる研究がさかんに行なわれている [1]。他にも、Web の検索エンジンのように、複雑ネットワークのトポロジ構造を利用した情報検索手法 [2] も広く用いられている。

このような複雑ネットワークのトポロジ構造を利用したデータマイニングは「リンクマイニング」と呼ばれる [3, 4]。リンクマイニングでは、データの構造 (データを「ノード」と捉えた時の、ノード間のリンク) を利用することによりデータマイニングを行う。リンクマイニングは、情報検索 / 分類・需要予測・マーケティング・テロや犯罪対策 [5, 6] など、さまざまな用途への応用が期待されている。現在も活発に研究が行われており、これまでさまざまなリンクマイニング手法が提案されている [7]。

数多くのリンクマイニング手法が提案されているが、これらは異なる分野でそれぞれ独立に研究されてきた [7]。このため現状では、それらのリンクマイニング手法を相互に連携させて利用することは困難である。例えば、リンクマイニングのためのツールもいくつか実用化が進んでいる [8, 9] が、それぞれ異なる分野への適用を前提として設計されている。このため、これらのツールではリンクマイニング手法のごく一部を利用できるにすぎない。

また、リンクマイニング手法と既存のデータマイニング手法を相互に連携させることも容易ではない。リンクマイニングではデータの構造を利用するが、従来のデータマイニングでは主にデータの量や意味を利用する。例えば、データの構造と意味の両方を利用するデータマイニング手法もいくつか提案はされている [10, 11] が、特定の用途に限定されている。

さまざまな分野における異なる要求に対応できる、より高度なデータマイニングを実現するためには、このようなリンクマイニング手法や従来のデータマイニング手法を統合的に扱うことのできる枠組が必要である。本稿では、ネットワークのトポロジ構造を利用するさまざまなリンクマイニング手法に加えて、従来のデータマイニング手法をも統合したデータマイニングを「ネットワークマイニング」と呼ぶ。

本稿では、まず、さまざまなリンクマイニング手法や従来のデータマイニング手法を統合的に扱うことができる、ネットワークマイニングのためのフレームワーク NMF (Network Mining Framework) を提案する。NMF は、入出力の形式が統一された 7 種類の機能ブロック (ネットワーク分析ブロック・定量分析ブロック・意味分析ブロック・データ加工ブロック・データ変換ブロック・データ可視化ブロック・外部入出力ブロック) によって構成される。これらの機能ブロックを自由に接続することにより、さまざまなネットワークマイニング手法の実現が可能となる。

さらに本稿では、NMF の利用例を示し、NMF を用いることにより、どのように高度なネットワークマイニングが可能となるかを議論する。具体的な例として、ノード特性の再帰的推定やネットワーククラスタの特徴語抽出が、NMF の機能ブロックの組み合せにより実現できることを示す。

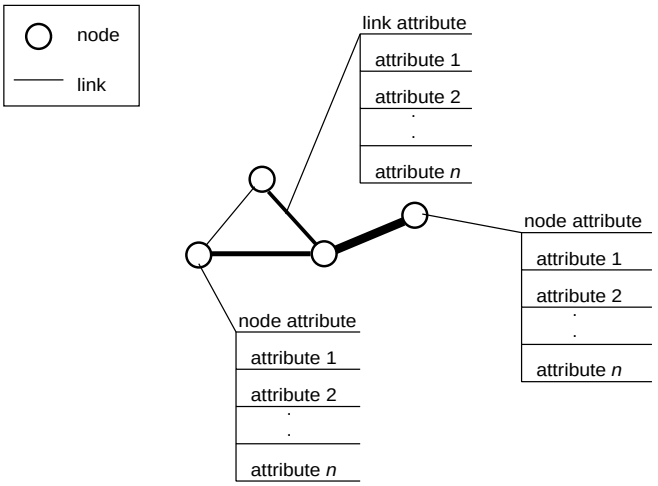

図 1: ネットワークマイニングで対象となるデータの例

本稿の構成は以下の通りである。2 章では、NMF の概要を説明し、7 種類の機能ブロックおよびそれらの入出力について述べる。NMF の利点もあわせて議論する。3 章では、ノード特性の再帰的推定およびネットワーククラスタの特徴語抽出を例に取り、また、NMF のの利用例を述べる。最後に、4 章において、本稿のまとめと今後の課題を述べる。

2 ネットワークマイニングフレームワーク NMF

2.1 概　　要

NMF (Network Mining Framework) とは、さまざまなリンクマイニング手法やデータマイニング手法を統合的に扱うことのできる、ネットワークマイニングのためのフレームワークである。ネットワークマイニングでは、データの構造だけでなく、データの量や意味もデータマイングの対象となる。つまり、構造を有するデータをネットワークとして見た場合に、ネットワークのトポロジ構造だけでなく、ネットワーク中のノードやリンクに付随するデータ (ノード属性やリンク属性) もネットワークマイニングの対象となる (図 1)。

NMF の概要を図 2 に示す。NMF は入出力の形式が統一された 7 種類の機能ブロック (ネットワーク分析ブロック・定量分析ブロック・意味分析ブロック・データ加工ブロック・データ変換ブロック・データ可視化ブロック・外部入出力ブロック) によって構成される。これらの機能ブロックを自由に接続することにより、さまざまなネットワークマイニング手法の実現が可能となる。各機能ブロックの入出力の形式が統一されているため、それぞれの機能ブロックを自由に組み合せることが可能である。

2.2 入出力の形式

さまざまな機能ブロックの柔軟な接続を可能とするために、NMF の機能ブロックの入出力は以下のいずれかの形式を用

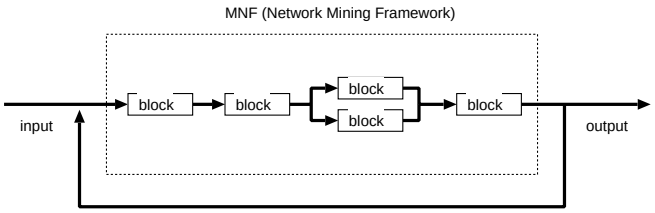

図 2: ネットワークマイニングフレームワーク NMF の概要

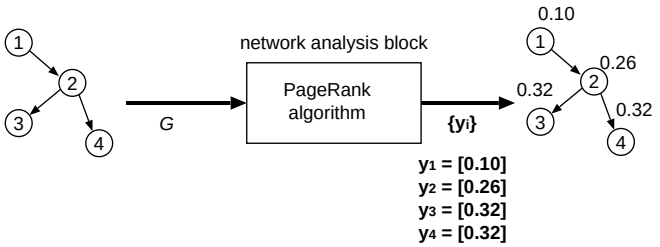

図 3: ネットワーク分析ブロックの例

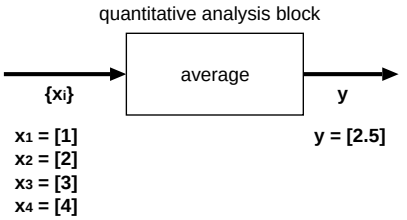

図 4: 定量分析ブロックの例

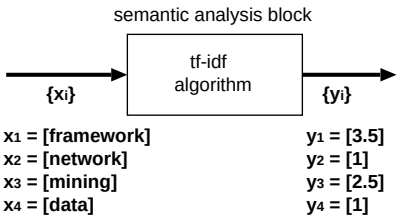

図 5: 意味分析ブロックの例

いる。

- ネットワークのトポロジ構造を表現するためのグラフ $G = (V, E)$

ネットワークのトポロジ構造を表現するためのデータ形式であり、ノードの集合 V およびリンクの集合 E によって構成される。

- ノードの属性を表現するためのベクトル $\mathbf{x}$ の集合

ネットワークマイニングでは、ネットワークのトポロジ構造だけでなく、ネットワーク中のノードやリンクに付随するデータもデータマイニングの対象となる。ノード属性を表現するためのデータ形式であり、単数もしくは複数のノード属性の集合によって構成される。

- リンクの属性を表現するためのベクトル $\mathbf{x}$ の集合

ネットワークマイニングでは、ネットワークのトポロジ構造だけでなく、ネットワーク中のノードやリンクに付随するデータもデータマイニングの対象となる。リンク属性を表現するためのデータ形式であり、単数もしくは複数のリンク属性の集合によって構成される。

- 一般的な値を表現するためのベクトル $\mathbf{x}$

主に、リンクマイングや従来のデータマイニングの結果を格納するために用いられるデータ形式である。

2.3 機能ブロック

NMF は、入出力の形式が統一された 7 種類の機能ブロック (ネットワーク分析ブロック・定量分析ブロック・意味分析ブロック・データ加工ブロック・データ変換ブロック・データ可視化ブロック・外部入出力ブロック) によって構成される。

- ネットワーク分析ブロック

ネットワーク分析ブロックは、ネットワークのトポロジ構造を利用したデータマイニング (リンクマイニング) を行うためのブロックである。一般に、ネットワーク分析ブロックは、ネットワークのトポロジ構造 G を入力とし、ネットワーク分析の結果 $\mathbf{x}$ を出力する。

ネットワーク分析ブロックは、ネットワークの基本的な特性の分析 (ノード数・リンク数・平均次数・平均距離・クラスタリング係数・ネットワーク密度など) に用いることができる。また、ネットワーク分析ブロックは、データのランキング [2, 12, 13]・データのクラスタ検出 [14]・データの分類 [15]・ダイナミクス分析 [16] のような、さまざまなリンクマイニング手法に用いることもできる。

例として、ネットワークのトポロジ構造から、各ノードの PageRank 値 [2] を求めるネットワーク分析ブロックを図 3 に示す。ネットワークのトポロジ構造 G を入力とし、各ノードの PageRank 値をノード属性の集合 $\mathbf{y}_i$ として出力する。

- 定量分析ブロック

定量分析ブロックは、データの量に着目したデータマイニングを行うためのブロックである。さらに、定量分析ブロックは、データの統計値 (平均・分散・分布など) に用いることができる。また、データ量の正規化に用いることができる。

例として、ノード属性の平均値を求める定量分析ブロックを図 4 に示す。ノード属性の集合 $\mathbf{x}_i$ を入力とし、その平均値 $\mathbf{y}$ を出力する。

- 意味分析ブロック

意味分析ブロックは、データの意味に着目したデータマイニングを行うためのブロックである。一般に、意味分析ブロックは、特徴語の抽出や、形態素解析に用いることができる。

例として、tf-idf 法 [17] によってリンク属性から特徴語を抽出する意味分析ブロックを図 5 に示す。ノード属性の集合 $\mathbf{x}_i$ を入力とし、各語の tf-idf 値をリンク属性の集合 $\mathbf{y}_i$ として出力する。

- データ加工ブロック

データ加工ブロックは、ノード属性、リンク属性や、データ間の構造 (リンク) の加工を行うためのブロックである。

- データ変換ブロック

データ変換ブロックは、外部のデータを NMF が取り扱える形式に変換する、もしくは NMF が取り扱う内部データを他の形式に変換するためのブロックである。

- データ可視化ブロック

データ可視化ブロックは、データの構造・量・意味の可視化を行うためのブロックである。

データ可視化ブロックでは、データの構造を 2 次元もしくは 3 次元でプロットすることが可能であり、可視化されたデータの表現形式の変更 (回転・ズームイン / アウト・フォーカス・選択・再配置など) に用いることができる。

- 外部入出力ブロック

外部入出力ブロックは、外部からリアルタイムでデータを入出力するためのブロックである。

2.4 特　　徴

NMF は、さまざまなリンクマイニング手法や既存のデータマイニング手法を統一的に扱うことが可能であり、以下のような特徴を有する。

- 柔軟性

NMF では、機能ブロックを自由に交換することが可能であるため、同じデータに対して異なるリンクマイニング手法や既存のデータマイニング手法を柔軟に適用することができる。

- 機能性

NMF では、機能ブロックの組み合わせが可能であるため、リンクマイニング手法や既存のデータマイニング手法を自由に組み合わせることができる。これにより、きわめて高度なネットワークマイニングになると考えられる。

- 拡張性

NMF では、機能ブロックの入出力を汎用性のある形で統一しているため、今後登場する新たなリンクマイニング手法やデータマイニング手法への拡張も可能である。

- 汎用性

NMF では、データ変換ブロックにおいてさまざまな種類のデータの違いを吸収する。このため、さまざまなデータを体系的かつ統一的な手法で分析することが可能である。

- 親和性

NMF では、外部入出力ブロックにより、例えば、外部のソフトウェアやネットワークサービスとの接続が可能である。これにより、既存のソフトウェアやネットワークサービスとの高い親和性を持たせることが可能である。

3 NMF フレームワークの利用例

3.1 ノード特性の再帰的推定

我々は文献 [18] において、既存のネットワーク分析手法を再帰的に適用することにより、データに含まれるノイズの影響を除去しながらノード特性を推定する手法を提案した。

文献 [18] において提案した、ノード特性の再帰的推定手法 RENC (Recursive Estimation of Node Characteristic) の概要は以下の通りである。従来のノード特性推定手法 [2, 12, 13] では、ネットワークのトポロジ構造をすべて利用してノード特性を推定している。このため、ノード特性の推定において重要でない (ノード特性の推定の上ではノイズとなる) ノードやリンクを含んだままノード特性を推定していると考えられる。ノード特性の再帰的推定手法 RENC では、ネットワークのトポロジ構造をすべて利用するのではなく、ノード特性の推定にとって意味のあるノードやリンクのみを用いることにより、ノード特性を推定する。まず、ネットワークのトポロジ構造すべてを用いてノード特性の推定を行う。この結果より、ノード特性を推定する上でノイズとなると考えられるノードと、そのノードに接続されているすべてのリンクを削除する。このようにして得られた部分ネットワークに対して、再度、ノード特性の推定を再帰的に適用する。

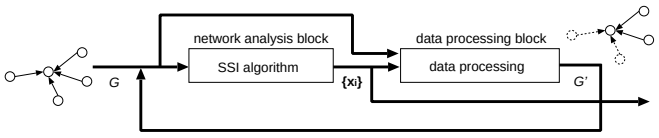

図 7: NMF の利用例 (ノード特性の再帰的推定)

上記のようなリンクマイニング手法は、ノード特性を推定するネットワーク分析ブロックと、重要でないと考えられるノードを削除するネットワーク分析ブロックを接続することによって実現が可能である (図 7)。

まず、分析対象となるネットワークのトポロジ情報 G がノード特性を推定するネットワーク分析ブロックに入力され、このブロックは推定したノード特性の集合 $\mathbf{x}$ を出力する。次に、分析対象となるネットワークのトポロジ構造 G と、推定されたノード特性のベクトル $\mathbf{x}$ が、重要でないと考えられるノードを削除するデータ加工ブロックに入力され、G の部分ネットワーク G' が出力される。このような手順が繰り返され、ノード特性の再帰的推定が可能となる。

このように、NMF では、機能ブロックおよびそれらの入出力を自由に接続することが可能である。図 7 に示すように、NMF では機能ブロック間のフィードバックも記述可能であり、さまざまなネットワークマイニング手法を実現できると考えられる。

3.2 ネットワーククラスタの特徴語抽出

文献 [19] において、ブログ記事をノード、トラックバックをリンクとするブログのトラックバックネットワークにおいて、ネットワーククラスタの特徴語を抽出する手法が提案されている。

本稿で提案する NMF の機能ブロックを組み合わせることにより、このようなネットワーククラスタの特徴語抽出も実現可能である。

文献 [19] において提案している、ネットワーククラスタの特徴語抽出手法の概要は以下の通りである。まず、ブログのトラックバックネットワークに対し、ネットワークのトポロジ構造を利用したクラスタ検出手法により、ブログの記事のクラス

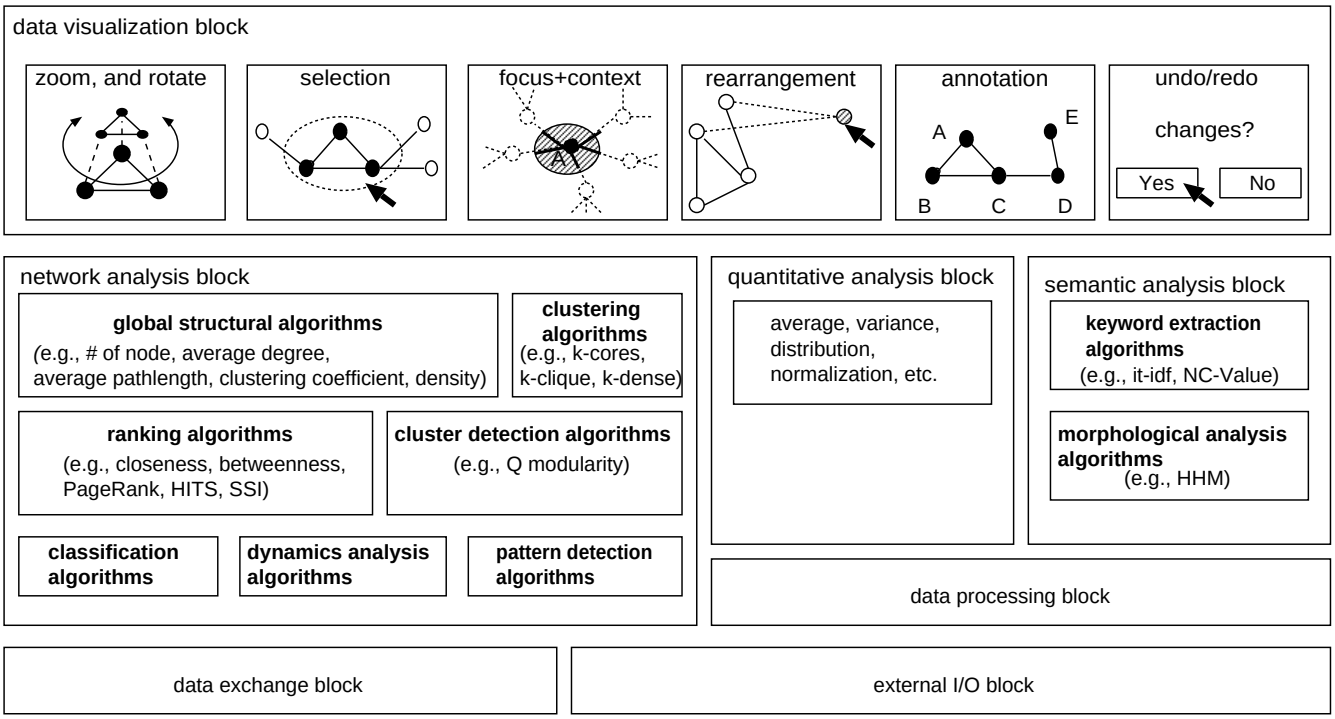

図 6: NMF を構成する 7 種の機能ブロック

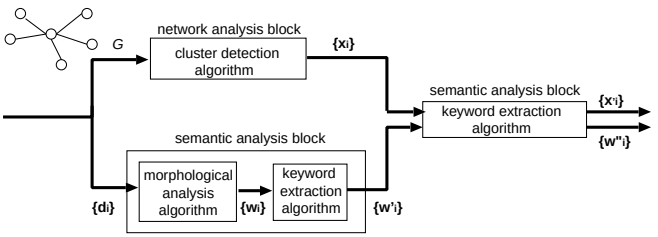

図 8: NMF の利用例 (ネットワーククラスタの特徴語抽出)

タを抽出する。そして、記事の本文から自然言語処理の手法によって、クラスタごとの特徴語を抽出する。

上記のようなリンクマイニング手法、およびデータマイニング手法は、ネットワークのクラスタ検出を行うネットワーク分析ブロックと、自然言語から特徴語を抽出する意味分析ブロックを並列に接続することにより実現できる (図 8)。この時、自然言語から特徴語を抽出する意味分析ブロックは、形態素解析を行う意味分析ブロックと特徴語を抽出する意味分析ブロックを連結することにより、実現できる。さらに、上記のように並列に接続したブロックと特徴語を抽出する意味分析ブロックを連結することによって実現が可能である (図 8)。

まず、分析対象となるブログのトラックバックネットワークのトポロジ構造 G がクラスタ検出するネットワーク分析ブロックに入力され、このブロックは推定したクラスタ ID をノード属性の集合 $\mathbf{x}_i$ を出力する。また、上記の処理と並行に、分析対象となるブログ記事の本文がノード属性の集合 $\mathbf{d}_i$ として、形態素解析する意味ブロックに入力され、語の集合 $\mathbf{w}_i$ を出力する。さらに、語の集合 $\mathbf{w}_i$ が特徴語を抽出する意味分析ブロックに入力され、このブロックは推定された特徴語の集合 $\mathbf{w}'_i$ を出力する。

次に、分析対象となる推定されたクラスタ ID の集合 $\mathbf{x}_i$ と、推定された特徴語の集合 $\mathbf{w}_i$ が、特徴語を推定する意味分析ブロックに入力され、それぞれクラスタ ID 集合と推定された特徴語の集合 $\mathbf{x}'_i$、$\mathbf{w}''_i$ が出力される。

図 8 に示すように、NMF では、機能ブロックの並列処理も記述可能であり、さまざまなネットワークマイニング手法を実現できると考えられる。

4 まとめと今後の課題

本稿では、さまざまなリンクマイニング手法やデータマイニング手法を統合的に扱うためのフレームワーク NMF (Network Mining Framework) を提案した。さらに本稿では、NMF の利用例を示し、NMF によってどのように高度なデータマイニングが可能かを議論した。具体的には、ノード特性の再帰的推定やネットワーククラスタの特徴語抽出が、NMF の機能ブロックの組み合せにより実現できることを示した。

今後の課題として、NMF を実装することにより、NMF の有効性を実証することが重要であると考えられる。また、NMF を用いることにより、より高度なネットワークマイニング手法を考案することも重要な課題であると考えられる。

文　献

[1] R. Albert and A.-L. Barabasi, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, June 2002.

[2] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the Web," tech. rep., Stanford Digital Library Technologies Project, 1998.

[3] D. J. Cook and L. B. Holder, "Graph-based data mining," *IEEE Intelligent Systems*, vol. 15, pp. 32–41, Mar. 2000.

[4] C. Ding, X. He, P. Husbands, H. Zha, and H. D. Simon, “PageRank, HITS and a unified framework for link analysis,” in *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 353–354, Aug. 2002.

[5] V. Latora and M. Marchiori, “How the science of complex networks can help developing strategies against terrorism,” *CHAOS, SOLITONS and FRACTALS*, vol. 20, pp. 69–75, Apr. 2004.

[6] H. Chen, W. Chung, J. J. Xu, G. Wang, and M. Chau, “Crime data mining: A general framework and some examples,” *IEEE Computer Society*, vol. 37, pp. 50–56, Apr. 2004.

[7] L. Getoor and C. P. Diehl, “Link mining: A survey,” *SIGKDD Explorations*, vol. 7, pp. 3–12, Dec. 2005.

[8] A. Perer and B. Shneiderman, “Balancing systematic and flexible exploration of social networks,” in *Proceedings of IEEE Transactions on Visualization and Comput Graphics (InfoVis)*, vol. 12, pp. 693–700, Sept. 2006.

[9] H.-J. Schulz, T. Nocke, and H. Schumann, “A framework for visual data mining of structures,” in *Proceedings of the 29th Australasian Computer Science Conference (ACSC)*, vol. 48, pp. 157–166, Jan. 2006.

[10] A. McCallum, A. Corrada-Emmanuel, and X. Wang, “Topic and role discovery in social netwrok,” in *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 786–791, Aug. 2005.

[11] M. Alexander and S. Lorenz, “Reconceptualizing latent semantic analysis in terms of complex network theory. a corpus-linguistic approach,” in *Proceedings of the 2nd International Conference of the German Cognitive Linguistics Association*, pp. 23–26, Oct. 2006.

[12] J. M. Kleinberg, “Authoritative sources in a hyperlinked environment,” *Journal of the ACM*, vol. 46, no. 5, pp. 604–632, 1999.

[13] K. Sugiyama, O. Honda, H. Ohsaki, and M. Imase, “Application of network analysis techniques for Japanese corporate transaction network,” in *Proceedings of 6th Asia-Pacific Symposium on Information and Telecommunication Technologies (APSITT 2005)*, no. 9–10, pp. 387–392, Nov. 2005.

[14] M. E. J. Newman, “Fast algorithm for detecting community structure in networks,” *Physical Review E*, vol. 69, p. 066133, June 2004.

[15] R. Kondor and J. Lafferty, “Diffusion kernels on graphs and other discrete structures,” in *Proceedings of the 19th International Conference on Machine Learning (ICML)*, pp. 315–322, May 2002.

[16] A. Ben-Hur and W. S. Noble, “Kernel methods for predicting protein-protein interactions,” *Bioinformatics*, vol. 21 Suppl 1, pp. 38–46, June 2005.

[17] G. Salton and C. Buckley, “Term-weighting approaches in automatic text retrieval,” *International Journal on Digital Libraries*, vol. 24, pp. 513–523, Jan. 1988.

[18] 杉山 浩平, 大崎 博之, 今瀬 真, “複雑ネットワークのトポロジ構造を用いたノード特性の再帰的推定手法,” 電子情報通信学会技術研究報告 (IN2006-175), pp. 77–82, Feb. 2007.

[19] M. Uchida, N. Shibata, and S. Shirayama, “Identification and visualization of emerging trends from Blogosphere,” in *Proceedings of International Conference on Weblogs and Social Media*, pp. 305–306, Mar. 2007.