

## 時系列ブログデータからの知識発見に関する研究

### ～データマイニングの立場から～

田村 慶一

広島市立大学大学院情報科学研究科

**概要:** ウェブログ (ブログ) の登場によりウェブに関する深い知識を持たない人々も容易に情報を発信できるようになっている。ブログは個人の意見を反映したものが多く、世の中の動きを知る上でブログのデータから有益な知識を発見することが重要な課題となっている。特に、ブログは膨大なテキストデータだけではなく、データ同士がトラックバックやリンクなどにより“つながり”を持つことに特徴があり、この“つながり”に着目した解析が必要となる。

本研究では、時系列ブログデータの“つながり” (ブロガー同士のつながり) から作成されるグラフ集合に着目し、データマイニングの技術を応用して、グラフの集合から有益な知識を取り出すことを研究の目的としている。具体的には、ブログのトラックバックが形成するグラフ集合に焦点を当て、このグラフ集合から頻出かつ重複を許したコミュニティを発見する手法の開発を行ってきた。頻出なコミュニティとは、ある一定期間ごとに発生するグラフの中で、頻出する部分グラフであり、特定の話題を頻繁に扱っているブロガー群といえる。そのようなブロガー群を発見することは、ブログ検索、クチコミ情報の信頼性の向上やブロガーへの情報推薦などへの応用が期待することができる。

本発表では、時系列ブログデータから頻出するコミュニティを抽出する方法、重複を許すコミュニティ抽出法とコミュニティ抽出法の高速度手法を説明するとともに、評価実験の結果などを示す。