■ 戦略研97th ミーティング議事録

日 時：2014年4月5日(土) 14:00-17:00

場 所：東京／竹橋 ちよだプラットフォームスクウェア

テーマ：「ビッグデータによる経済・社会現象の解析」

発表者：大西立顕さん（東京大学大学院情報理工学系研究科准教授／
キヤノングローバル戦略研究所主任研究員）

参加者：参加者 15人（発表者を含まない）

（銀行印、財務コンサルタント、IT コンサルタント、会社経営、会社員、研究員、
NPO 法人理事長、社会保険労務士、行政書士、司法書士等）

目次；

- 0. イントロダクション
- 1. ビッグデータ解析
- 2. 不動産バブル分析
- 3. テキストデータ分析

発表；

0. イントロダクション

物理学や経済学などいろいろなところに所属していました。だからこそできる研究というものがあります。東京大学に、ソーシャルICT研究センター（以下、センター）が2013年4月、設置されました。従来はコンピューター、数理、ロボット、経済それぞれそれだけを研究していました。センターではこれらを一緒にやるのを趣旨としています。

1. ビッグデータ解析

スパコンの演算性能は急速に上がっています。いわゆるムーアの法則です。半導体の集積度は18ヵ月ごとに倍増しています。

日々の活動による多様で詳細な情報が高頻度に記録される時代になっています。センターではカードの履歴データを解析しています。解析は統計科学、経済物理学、複雑ネットワーク科学などで行っています。

社会経済現象は人々の感情が支配しています。たとえば、コインを6回投げたとき、表裏裏表裏裏となる確率と裏裏裏裏裏裏となる確率について質問すると前者のほうが高いとします（確率的には同じです）。また、松の木一本の値段はいくらでしょうか？。一本1万円もしないと考えますが、陸前高田の奇跡の一本松は1億円ぐらいの価値があります。人の心情は必ずしも数理で読み解くことはできません。現実のデータを入手すると、人の活動は物理に従わないことがわかります。

ビッグデータ利活用の難しさは下記のとおりです。①データの入手（個人情報）、②データの利用（解釈能力）、③データの分析・解析（数理的知識）、④データの処理・実践（計算処理技術）、⑤現場での利活用（実フィールドでの説明する能力）です。

2. 不動産バブルの分析

中古マンション売買の取引価格を分析しました。週刊住宅情報（現SUMO）の1986年1月～2013年11月のデータです。データは100万件ありました。不動産バブルの見える化、網羅ができました。

分析はフィールドで感じたことがけっこう大切です。価格と面積の関係がベースになりますが、平方メートル単価×面積ではありません。面積が大きくなるほど、高級マンションであるほど、平方メートル単価以上に高くなるのがわかりました。これにより、面積調整価格の累積分布にてバブル度を把握することができました。正規分布であれば桁違いものがないはずですが、不動産バブルが発生すると、桁違いのものがあります。周辺とかけ離れた価格が現れます。

近年、日米欧でみられる経済の混乱は、不動産バブルの崩壊がきっかけで起きています。不動産バブルにリアルタイムで干渉できることが大切です。

3. テキストデータ分析

いままでは数字だけで分析していました。現在は、トピックモデルと呼ばれるものです。ドキュメントの中の文字を分析します。いままでは単語の出てきた数で分析していましたが、英語の社名の「アップル」と果物の「アップル」のように同じ言葉でも意味を分析できるようになりました。文章から自動的にトピックと関連語を抽出します。この手法にて、ロイター社のニュース記事を分析しました。ニュースと株価の関係がわかりました。ニュースが多いと株式の取引が多くなるという相関があります。また、トピックを介した国のつながりもわかります。さらに時系列で見ると、つながりの変化がわかります。このような分析を客観的に自動的に行うことができます。しかし、解釈、解析できる人が必要になっています。

テキストデータ分析は政府による情報発信にも活用できます。世論をどう動かすために、大統領がいつ話すべきかということもわかります。金融危機などの混乱期にはとても大切です。

最後に。時系列データ（金融市場）、時空間データ（不動産市場）、ネットワーク（貿易ネットワーク）、テキストデータ（古文書）などを実証科学の視点から研究していきたいと考えています。

以上