

第55回シンポジウムルポ



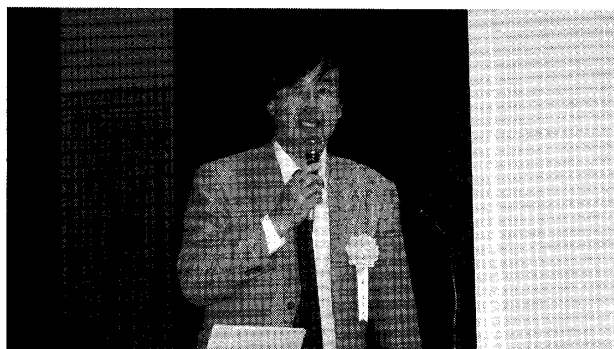
中原 孝信 (大阪府立大学)

平成18年3月13日、第55回シンポジウムが中央大学で開催された。テーマは「データ解析の最前線」であり、近年注目されているデータマイニングの企業活動における適用事例や最新動向の紹介ということから、多くの参加者が集まった。

実行委員長の加藤直樹氏(京都大学)は、冒頭の挨拶で、データマイニングとは膨大なデータから有効な知識を導き出す技術であり、データの部分的な特長の発見と、データから導き出せる仮説を生成することを目的としていると述べられた。そして、データマイニングは様々な分野で利用されており、その分野は、オペレーションズ・リサーチ(OR)の技術も適用できることから、今後、ORの技術に着手することで、新技術の開発が期待されるという挨拶から始まった。以下に4名の方々の講演を紹介する。

鷲尾隆氏(大阪大学産業科学研究所)の講演「生産プロセス品質管理におけるデータマイニングの現状」は、製造現場の品質管理だけではなく、原料の仕入れから加工、そして製品の販売まで、全てのプロセスにおいて統一した品質管理を念頭においたお話であった。従来行われている製造現場の品質管理は、正常状態からの逸脱を検査する監視、その原因を追究する診断に対して、統計解析などの技術が積極的に用いられているが、全てのプロセスで統一した品質管理を行うには、従来の技術に加え、データマイニング技術を適用することが有効であると述べられた。

データマイニング技術を用いた実際の応用例として、鷲尾氏らの研究室と、三菱電機(株)生産技術センターで共同開発された、顧客との関係を考慮した品質管理システムである「電気製品の市場品質監視システム」について解説された。これは、クレームのデータベースと製品の製造履歴情報を1つのトランザクションとして扱い、バスケット分析を適用することで、どんな製品やどんな部品を利用したときに、どのような不具合が起りやすいのか、その組み合わせ条件を把握するものである。そして、バスケット分析から得られた仮説を従来の品質管理手法を用いて検証することで品質



加藤実行委員長挨拶

管理を行う。このシステムは、現場では4週間のクレーム情報に対して適用され、そこから得られた仮説を検証し、製造現場にフィードバックすることで、製造プロセスの改良や、設計仕様の変更などに利用されてきたとのことである。氏は最後に、今後さらに膨大になるデータや、数100次元にも及ぶ変数を解析できるような手法の開発が重要になってくるとのお考えを示された。

小野潔氏(三菱東京UFJ銀行)の講演「データマイニングを用いた中小企業の信用リスクの推定モデル」は、企業ベンダーが提供する中小企業約42万件の予測倒産率と、業種、設立年数などの企業データ、経営者の年齢、年収などの個人データ、そして評点から成る企業情報から倒産モデルを再構築し、そのモデルに全ての企業情報を当てはめることで、中小企業約120万社の予測倒産率を算出するというお話であった。企業ベンダーが提供する予測倒産率は、どのような説明変数から得られた結果であるか、その内容が明らかにされていないため、モデルを再構築することでモデルの内容を明らかにし、他の中小企業の予測倒産率を算出することは与信フレームを考える上で重要であると述べられた。また、中小企業を対象にする場合、財務データの情報開示が進んでいないことや、その信憑性に疑問があることから、定量的アプローチではなく、定性的アプローチによりモデルを再構築する必要があることをご指摘された。小野氏は、決定木モデルを構築する際、存続企業の情報には倒産情報が含まれてい

ないため、ある閾値を設け、予測倒産率が閾値以上である企業を仮想倒産企業、閾値未満を正常企業と定義する工夫をされており、閾値の値は、バギング学習により複数の決定木の平均値を利用されていた。

再構築モデルにより明らかになった変数で注目する点としては、企業の情報公開の有無が優遇される加工変数が、ある程度の説明力を持っていることであり、中小企業では情報公開に義務はないものの、それを公開しているということで、説得性を持たすことが可能であると氏は強調された。モデルの精度もビジネス応用が十分期待できるものであり、120万社の予測倒産率は、中小企業の信用リスク管理に利用できるとのことであった。

大澤幸生氏（東京大学）の講演「『売る』前に『創る』と『作る』のチャンス発見」は、人が意思決定を行う際に重要となる事象をいかに発見するかというお話であり、チャンス発見ツールの1つである Key-Graph を用いた製品開発に関する事例などを紹介された。チャンス発見では、データから得られる客観的な判断だけではなく、人間の主観的な判断を大いに取り入れることで、意思決定者が納得してアクションを起こせるようになるということを強調されていた。KeyGraph は、アイテム間の共起関係を可視化できるツールであり、共起頻度の高いアイテムとの間に、稀に出現するようなアイテムを発見できるという特徴を持っている。チャンスの発見には人間の介在が必要不可欠であり、機械がチャンスを見つけるのではなく、KeyGraph により作成されるシナリオ・マップを人間が理解し、解釈することで、チャンスの発見につながるということであった。

チャンス発見の実用例として、朝食の献立データ、服の生地データ、そしてフィルムの傷をチェックする装置のクレームデータなどに KeyGraph を適用し、意思決定において重要なシナリオを導出されていた。朝食の献立データでは、ビタミン剤の潜在的な需要に関するシナリオの発見。生地データは、ビジネスウェアを好む顧客と、カジュアルな服を好む顧客との間を橋渡しできる素材の発見に関するシナリオ。装置のクレームデータからは最終的に5つの特許に結びつけるシナリオを得ることができたということであった。また、

これらのシナリオは、業務に精通している人達がグループディスカッションを行うことで初めて明らかになったシナリオであることを示された。

矢田勝俊氏（関西大学）の講演「流通業におけるデータ活用の実際」では、データマイニングのビジネス応用の観点から、他の分野で開発された技術や手法を流通業で蓄積される膨大なデータへ適用し、いかに知見を発見するかということについて事例を交えて解説していただいた。遺伝子解析手法のビジネス応用では、発泡酒の購買履歴データを対象に、各顧客の発泡酒の購入履歴を文字列で表現し、それに遺伝子解析手法を適用することでパターンを抽出する。結果としては、特定のブランドを3回連続で購入する顧客は、将来ある特定のブランドを好んで購入するようになるということであった。そしてその結果をビールメーカーのスタッフが活用し、販売促進のアイデアへとつなげ、実施することで、ある特定ブランドのシェアが販売促進後も増加したということである。別の内容では、債権の回収に関する問題にテキストマイニングと統計手法を適用し、効率良く債権を回収するオペレータの交渉技術を明らかにされていた。顧客とオペレータの会話の内容をテキストマイニング手法を用いて単語に分割し、因子分析により単語をグルーピングすることで特徴を抽出されていた。特徴としては、高回収率は支払い方法を表す単語や、債務者を励ますような単語、債務者への状況説明を表す単語などを低回収率者に比べて多く使っているなどの傾向が把握できているということであった。

矢田氏は、様々なデータを分析されてきた経験から、開発された手法が必ずしもビジネスの現場で適用することができない点、現場からのフィードバックが新たな課題をもたらす点など、現場の知識と、新手法を開発できる能力の重要性について強調され、両者を連携する仕組み作りの必要性を述べられていた。

4名の講師の方々は、企業で蓄積されているデータから、有益な知見を発見するために様々な分析手法を駆使されており、非常に興味深い講演であった。また、限られた時間の中、盛りだくさんの内容であったことから、研究者、実務家、双方の方々にとって大いに参考になったご講演であったと思われる。