データマイニングを活用したモノづくりの意思決定支援

Decision Aiding of Manufacturing Using Data-Mining/Hiroyuki KODAMA

岡山大学 児玉紘幸

1. はじめに

近年のCAD (Computer Aided Design)/CAM (Computer Aided Manufacturing) システムの発達により、非熟練の技能者であっても、工作機械加工において必須なNC (Numerical Control) プログラムの構築が容易となってきた。しかしながら、工程設計を考慮した使用工具の設定や、切削条件の決定には、熟練技能者の暗黙知に頼っているのが現状である。特に中小企業の多い日本において、多品種少量生産が主流の現場では、熟練者の暗黙知を体系化することが困難であるため、非熟練技能者の習熟過程を円滑化し、多様な条件決定を支援するシステムの構築が必要となる。実験計画法や機械学習を応用した、最適な切削条件決定プロセスに関する研究は多々見られるが、すべての切削工具や被削材に対して、柔軟に適用可能な切削条件決定システムを提案する研究は少ないようである。一方で、近年の製造業において、モノのインターネットIoT (Internet of Things) 技術を製造現場に実装していくことが当たり前となりつつある。IoT 技術において、現場で取得したビッグデータを一元管理し、それらのデータから傾向を抽出する目的として、データマイニングの技術が注目されている。そこで本稿では、数あるデータマイニング手法の中で、代表的な手法の特徴や製造業分野への適用方法を解説するとともに、具体的な適用例として、工具カタログデータに対してデータマイニング手法を適用した意思決定支援システム (カタログマイニングシステム¹⁾) の紹介を行う。

2. データマイニングの定義と利用目的

データマイニングとは、有益なパターンやルールあるいはノイズを含むデータベースに蓄積された膨大なデータ (ビッグデータ) から、価値ある情報を発掘する方法のことである²⁾。図1に示したデータマイニングの概念図のように、データマイニングを行うことにより専門家も気づかない新しい仮説・知識を発掘することがデータマイニングの醍醐味である。

近年において、データマイニング技術は医学、化学、経営学のみならず、さまざまな分野で用いられており、一部で製造システムへの応用も試み始められている。データマイニング手法の発展により、数々の知識発見に基づいた意

思決定システム、すなわち KDD (Knowledge-Discovery in Database) システムが提案されてきた³⁾。KDD システムは、数々のパラメータが複雑な相関関係を有するような大規模データ (例えば POS データや株価、消費者アンケート結果など) から有益な知識発見を行うまでの全体のプロセスを示すのが一般的である。特に製造分野において、多くの研究成果により KDD システムの発展が推進されてきており、プロセスレベルでの応用例⁴⁾として、製造現場の機器メンテナンス頻度の予測や、欠陥の発見、設計、製造、品質保証、スケジューリングなどの意思決定支援システムに应用可能な知識を抽出することができるようになった。近年では、上記の技術を応用したトレンド解析手法を適用することにより、製造現場での IoT システムに実装されている。データマイニングの特徴は、解析の目的に合致した最適な解析手法を幾つか組み合わせることにより、データに忠実なルール生成が行えることである。これは未知なるルールの発見という利点をもつが、多数の解釈困難なルールを生成してしまうという問題点も内在しており、解析の特性をよく理解して、抽出された結果の解釈を行わなければ、正しい結果は出力されない。また、データマイニングの最終目的は意思決定、つまり問題解決であり単なる予測ではない。ここでデータから新知識獲得までの模範的なフローを図2に示す。複数のプロセスは独立して存在しており、一連の流れにより新知識が獲得できるが、さらに質の良い新知識を獲得するには、繰り返して精練していくことを示している。データマイニングの過程においては、データの傾向が見える化し、各プロセスで得られた知見を解析者がさまざまな角度で考察が行える点にある。

図1 データマイニングの概念図

図2 データマイニングフロー

図3 データマイニングの目的と解析手法

3. データマイニングにおける代用的な解析手法

データマイニングで取り扱うデータの形態は、量的データ（例えば、1, 2, 3...などの数値データ）や、そのままでは演算のできない質的データ（例えば、誕生日や血液型）、時系列データなどが存在する。それぞれのデータの形態に合わせて、使用可能な解析手法は異なる。

図3に、データマイニングをする上で目的に沿った代表的な解析手法を示す。これらの解析手法は、図1中の統計解析プロセスで使用される。解析の目的は、一般的に大きく三つに大別できる。すなわち、①データのグループ化、②データの特徴抽出、③データの予測である。図2では、それぞれの目的に最適な解析手法がいくつか示されているが、データマイニングに精通していないエンジニアでも、取り扱いのしやすい代表的な手法をそれぞれの目的について選択した。すなわち、目的①ではクラスタ分析、②では主成分分析、③では回帰分析について、それぞれの特徴や使い方を解説する。図4にそれぞれの解析の特徴概念図を示す。詳細な計算アルゴリズムなどは、それぞれ文献を参照していただきたい。次章において、各解析手法の具体的な適用例を紹介する。

3.1 データをグループ化する

データマイニングの用語として、「グループ化する」こ

① データのクラスタリングと見える化 非階層的的手法

階層的的手法

② データの特徴の抽出

③ データの予測

図4 データマイニングの代表的な解析手法概念

とは、「クラスタリング」と同義に扱われることが多い。クラスタリングは、属性間の距離などを基準にして、似通っている属性をもつデータをグループ化して、クラスタを生成するものである。クラスタリングの手法は、数値分類法とも呼ばれ、大きくは変数クラスタ分析などに代表される階層的的手法と、K-means法に代表される分割最適化手法（非階層的的手法とも呼ばれる）に分けられる。クラスタリングの目的としては、大規模データの相関関係を見える化することにより、データの全体の特徴を把握することや、似たような属性を有するデータの集合を集めることにより、次なる解析手法を適用した段階での新知識の抽出や解釈の精度を高め、その過程を解析者にとって少しでも容易にすることが挙げられる。階層的的手法の特徴は、クラスタ間の融合の順序とその類似度を表すデンドログラム（樹状図：図4）が形成されるという特徴がある。そのため、解析の初期段階として、全体のデータの構造を相関関係の観点から見える化を行う最適な手法であると考えられるが、デンドログラムの形状は、クラスタの融合を決定する幾つかの種類を有する評価関数に依存するため、その評価関数の選択には試行錯誤を有する。非階層的手法の特徴は、クラスタ間の融合の順序はランダムであるが、データの分布を多次元的にクラスタリングすることによって見える化し、全体のデータの分布の特徴が把握できる点にある。一般的には、両解析手法は組み合わせて適用し、データの構造を分析することが多い。

3.2 データの特徴を抽出する

ここでは、量的データ（数値データ）に対して適用可能な解析手法として、主成分分析を対象とする。主成分分析とは、多変量（多次元）が有する相関関係を取り除くことで、少数の次元に統合化し、見える化や特徴的なパターンの抽出を図る分析手法として用いられる。すなわち、多変量データの統合・縮約を目的とし、分析の効率化と多変量データの視覚化が行える便利な手法である。また、主成分分析によって抽出された特徴を定量化する手法として、主成分回帰が適用されている。また、相関のないデータ（例えば、一定値の数値データ）に対して、主成分分析の手法

は適用できないことに注意すべきである。近年では、主成分回帰手法よりも、データの特徴抽出精度の高い発展的な手法として、最大情報係数 MIC (Maximal Information Coefficient) や機械学習アルゴリズムの一種であるランダムフォレストを導入した結果が報告されている⁵⁾。多くの場合、データの特徴の抽出と、データの予測は組み合わせて解析に導入されることが多いため、ビッグデータの有する特徴を、より効率的に精度良く抽出可能な手法を適用することが重要である。

一方で、前節で解説したクラスタリングの手法も組み合わせて適用することにより、見える化されたデータの相関関係を補助的に解釈することにより、さらに有意な特徴の抽出が行える。

3.3 データを予測する

データマイニングの解析手順は、解析者の知見や解析の目的に合わせて適宜決定することが可能であると前述したが、データの予測プロセスは、データのクラスタリングおよび特徴抽出を行った後に、ある程度予測に必要な変数やクラスを特定した段階で行われることが多い。数ある予測手法の中でも、よく使われるのは回帰分析である。回帰分析は目的変数（量的変数データ）を、説明変数を用いて予測した回帰式であり、両者の相関関係を定量的に評価するための手法として使われる。回帰式中に説明変数が二つ以上存在する場合は、重回帰分析と定義されている。回帰式の予測精度を示すパラメータとして、0~1 の範囲の数値で決定係数（寄与率）が使われる。決定係数が1に近ければ近いほど、予測精度は高いと評価できる。一方で、決定係数が、0.5を下回る場合は、説明変数と目的変数の相関関係が小さいため、高い予測精度は保証できない。回帰式を構成する説明変数の、目的変数に対する相関関係の定量値としては、標準化偏回帰係数が使われる。標準化偏回帰係数の高い説明変数を選ぶことによって、回帰式の決定係数は高くなる傾向にある。データマイニングの観点からは、回帰式を構成する説明変数の標準化偏回帰係数を参考にして、どのような変数が、目的変数に対して影響度合いが強いのかを判別し、なぜそのような変数が影響度合いを有するのかを考察（解釈）することが重要である。考察にあたり、上記で紹介したデータをグループ化する手法や、特徴を抽出する手法によって得られた結果も併せて行うことにより、包括的なデータの傾向が俯瞰できるのがメリットとなる。特に重回帰分析では、互いに相関関係の強い説明変数を二つ以上使用して回帰式を作成した場合、多重共線性を示すことにより、予測される値のとり得る範囲が広くなり、結果が安定しなくなる。そのため、変数クラスタ分析と主成分回帰を併用することによって、重回帰分析の解析を行う前に、説明変数間で相関の高い変数を明らかにし、より目的変数に対して相関の強い説明変数を取捨選択してから、回帰式に用いることによって、予測精度をある程度保証することが可能となる。

4. 工具カタログデータマイニング

4.1 意思決定支援システムとしての役割

上記で述べた解析手法をそれぞれ適用した意思決定支援プロセスの適用事例として、著者が取り組んできた工具カタログデータマイニング技術について紹介する。

近年のCAM システムでは、非熟練技能者でも比較的簡単にNCプログラムとして工具パスが作成できるように進化してきたが、加工に使用する工具の切削条件（例えば、主軸回転数 S rpm, テーブル送り速度 F mm/min, 切り込み量 P fmm および A n mm）は自動決定できない。すなわち、NC 工作機械の運動学的な指令の決定は、熟練技能者の暗黙知的な経験や知見に頼っていることが多く、一般的に体系立てられていないため、非熟練技能者がそれらを受け継ぐのに長期間を要する。工具の切削条件の決定には、加工する被削材料の材料特性や加工形状、工具を保持するホルダーの突き出しの剛性や工具自体の材質、刃部の形状、コーティングの種類、工作機械の剛性など、多岐にわたる条件をそれぞれ考慮して決定する必要がある。最終的にその切削条件により、工具の摩耗状況が変化し、要求精度の確保が困難となり、製品の製作にかかるコストや納期に大きな悪影響を与える。特に、切削に関する知見の浅い非熟練技能者にとって、最適な切削条件の決定は難しく、それらの決定には試行錯誤的な切削実験を要する場合が多く、相当な時間や被削材料、工具、労働力を無駄にってしまう。これを解決することが不可欠である。したがって、工具や切削条件の決定を支援するシステムを構築していくことで、設計から製造までのプロセスを合理化することができると考えられる。

ここで、非熟練技能者は、工具カタログに記載されている切削条件推奨条件表を基に、条件設定を行うことが多い。しかしながら、カタログは工具メーカーの工具に対する膨大な知識を集めた加工技術に関するビッグデータであるため、工具形状や被削材によって条件が多岐にわたることから、条件設定に時間がかかり非効率である。また、そのようなデータベースを単に参照データとして用いるのではなく、熟練技能者の有する、暗黙知的な切削条件決定に関する知識が集約された情報源として扱うことで、そこから切削条件決定に有意な知見を抽出していくことが可能であると考えた。

4.2 カタログマイニングプロセス

切削条件の決定支援のためのカタログマイニングプロセスの概要図を図5に示す。データベースとして工具カタログデータを用いる。データマイニングにおいて特に重要なのは、データの獲得、選択、データクレンジングである。データを解析可能な形に変換するプロセスが、新知識発見のために必要な全プロセスに占める割合は70~80%であるといわれている⁶⁾。しかし、工具カタログデータにおいては、すでに工具メーカーで実験を繰り返した結果に基づく良質なデータが大量に存在していると考えられるた

図 5 工具カタログへのデータマイニング手法の適用例

め、解析者にとってデータの獲得、選択に時間を割く必要はない。また、カタログには、すでに数値で表示されている工具データが多くデータクレンジングの多くを省略できるものと考えられる。そのため、カタログマイニングでは、クレンジングの次のプロセスであるデータのクラスタリングが特に重要になると考えられる。そこで、データ構造解析手法（非階層型解析手法）である K-means 法によって工具形状によるビジュアル的なクラスタ分けを行った。次に統計解析手法として、階層型解析手法である変数クラスタ分析により、変数間の関係をビジュアル的に表現することでデータ構造の階層を把握する。そこで抽出された切削条件決定に有意な変数を用いた重回帰分析により切削条件式を導出し、工具形状（直径、全長、刃長、刃数）と切削条件について考察する。マイニングによって得られた知見の解釈のために、推奨切削条件が未知の材料に対して実際に導出された切削条件を用いて切削実験を行った。導出された切削条件決定式に工具パラメータを代入することによって、工具カタログに推奨切削条件が記載されていない被削材料であっても、指標となる切削条件決定のための試行錯誤的な実験を最小にし、ある程度実用的な切削条件を迅速に導出できることが示せた。

5. おわりに

本稿において、データマイニングの紹介とともに、代表

的な手法の特徴について概説した。また、製造分野におけるデータマイニングの具体的な応用例として、工具カタログデータマイニングについて紹介した。貴重なお時間を割いて本稿をご覧になられた方に、少しでもデータマイニングの面白さや、適用範囲の広さをご理解いただければ幸いである。デジタル化が進む社会でデータを有効に活用することの効果が認められることで、製造現場にデータマイニングを適用する研究が、世界的に今よりも発展していくことを期待する。

参考文献

- 1) H. Kodama, T. Hirogaki, E. Aoyama and K. Ogawa: Proposal of Ball End-milling Condition Decision Methodology Using Data-mining from Tool Catalog Data, *Journal of Precision Engineering*, **79**, 10 (2013) 964-969.
- 2) K.B. Irani, J. Cheng, U.M. Fayyad and Z. Qian: Applying Machine Learning to Semiconductor Manufacturing, *IEEE Expert*, **8**, 1 (1993) 41-47.
- 3) G. Christine and D. Alan: Knowledge Discovery from Industrial Database, *Journal of Intelligent Manufacturing*, **15** (2004) 29-37.
- 4) W. Sakarinto, H. Narazaki and K. Shirase: A Knowledge-based Product Model Data for Integrating CAM-CNC Operation, *LEM21*, (2009) 95-100.
- 5) T. Sakuma, T. Hirogaki, E. Aoyama and H. Kodama: Proposal of tool catalog data mining process with maximal information coefficient MIC, *Proc of JSPE*, (2017).
- 6) 元田, 山口, 津本, 沼尾: データマイニングの基礎, 情報処理学会, **10** (2006).