

研究ノート

体育・スポーツ科学分野への決定木分析の応用事例： 分析方法の紹介と分析の注意点

Application of decision tree analysis in health and sports science: Introduction and instructions of analytic procedures

鈴木 宏哉¹⁾Koya SUZUKI¹⁾

Abstract

The purpose of this article was to introduce analytic procedures of decision tree analysis in the field of health and sports science. This article dealt with rule search of past sports experiences for establishing exercise habits as an example of decision tree analysis. There are some algorithms in decision tree analysis. A classification and regression tree (C&RT) was selected as an analytic algorithm. The samples were 200 (169 males and 31 females) Japanese university students. Survey items included current exercise habits (frequency, and duration), and past sports experiences (number of sports events, time spent in exercise per week, extent of perceived exercises, and enjoyment of exercise). From the result of a C&RT analysis, the rule of the highest ratio of exercise habit group (75.0%) was enjoyment of exercise > 3, extent of perceived exercises > 4, and number of sports events > 2.5 events (misclassification cost: 34%). It has been reported that a result of decision tree analysis varied depending on analytic algorithms and split criteria. Researchers also reported that samples more than 1,000 needed to show meaningful results. Therefore, the reliability of this result should be examined using different algorithms and split criteria, or different multivariate analysis (e.g., logistic regression analysis). However, it is important for the development of health and sports science to show interim criteria (association rule), although there are some limitations related in analytic procedures.

Key words : classification and regression trees, CART, C&RT, logistic regression analysis

[Received January 22, 2008 ; Accepted May 8, 2008]

1. はじめに

データマイニングは、大容量データをパターン認識テクニックや統計的手法によって処理することにより、意味のある新たなパターン及び傾向を発見するプロセスである(牛田ほか, 2003)。データマイニングはサンプル抽出, 探索, 加工・変換, モデリング(データマイニング手法の選択), 評価, が最善の結果を得るまで繰り返し実行される反復プロセスである(奥山, 1997)。データマイニングプロセスの中心となるモデリング手法は判別・予測, クラスタリング, 関連分析の3つに大別できる。代表的な手法として, 分類(classification), 回帰(regression), クラスタリング(clustering), 従属関係のモデリング(dependency modeling), 変化や偏差の検出(change and deviation detection), 決定木(decision tree), 連関ルール(association rule), ニューラルネッ

トワーク(neural network)などが挙げられる(Fayyad et al., 1996)。

決定木分析はその中で判別や予測を目的とするデータマイニング手法のひとつであり, データマイニング手法として最も注目され, 使用頻度が高いといわれている(岩崎, 1999)。決定木分析は従来の統計手法の判別分析やクラスター分析と同じく, 判別や分類を行う手法である(新村, 2002)。従属変数が二値データである場合には, ロジスティック回帰分析と同じように従属変数への独立変数の影響を検討したり, 予測したりする手法である(奥ほか, 2004)。従属変数が3件法以上の順序尺度データに対しても適用可能であるため重回帰分析とも類似した分析手法である。従来法と異なる点は, 従属変数に対する独立変数間の相互関係を階層的に捉えることが可能であり, 従属変数に影響する複数の独立変数によって標本を二股以上に枝分かれする形で均質なグループに多段階

1) 東北学院大学 教養学部 Faculty of Liberal Arts, Tohoku Gakuin University

別するという概念で解析が行われる。したがって、これまでの因果分析では、それぞれの独立変数が従属変数に対してどの程度影響しているのかを分析しているのに対して、決定木分析では複数の条件を組み合わせた従属変数に対する影響を検討することができることが特徴である。また、データの分布に依存せずに分析が可能であるため、判別分析と重回帰分析に変わるノンパラメトリックな多変量解析の一種ともいえる(大滝ほか, 1998)。加えて、判別分析と同じように従属変数を判別する基準を示すことができることから、従属変数を判別するための独立変数の複合的な基準を示すことが可能となる。したがって、どのような基準の組み合わせによって従属変数を判別する精度が向上していくのかを逐次的に確認することができる。これらの知見は、実践現場で活用する実務家にとって、従属変数を意図した状態に変化させるための意思決定を行う際に、複数ある独立変数のうち、どれを強調すべきか、どれくらいの基準を満たすべきか、そしていくつの基準を満たすことで従属変数が意図した状態になる確率がどの程度になるのかを知るために有用である。

次に、伝統的な統計解析とデータマイニングの相違点のひとつとして、分析に用いられるデータ数の多さとデータ収集法が挙げられる(岩崎, 1999; 奥山, 1997)。統計解析では研究者の設定した仮説を検証するためにデータをどのようにして収集するかはとても重要なことであるが、データマイニングでは商品の購買履歴などのように日々大量に、そして意図せずして蓄積されたデータを活用し、検証すべき仮説を導き出すことがその主たる目的となる。フィットネスクラブなどのスポーツ施設では利用者の利用履歴が蓄積される。決定木分析を用いれば、意図せず蓄積されたこれらのデータから利用者がドロップアウトせず継続して施設を訪れ、運動するための方法を発見することができるかもしれないし、文部科学省が毎年数万件規模で実施している体力・運動能力調査データから体力向上の手がかりを探索することができるかもしれない。

子どもの体力の長期的低下傾向は、Nishijima et al. (2003) が指摘するように、集団全体が体力低下の傾向を示しているのではなく、低体力者層がより低下傾向にあることから高体力者層との体力差が拡大し、集団の平均値が低下している。すなわち、集団の分布が維持されたまま低い値へシフトしているのではなく、集団の分布が変化しながら平均値が低下している。そして、Body Mass Index (BMI) の経年変化についても将来的には分布の変化を伴いながら平均値の増加(肥満者の増加)

が予想されている(Prentice, 2001)。決定木分析は前述したようにデータの分布に依存せずに分析が可能であるため、このようなデータに対しても有効である。

決定木分析を含むデータマイニング手法の特徴は統計学における検定論との立場の違いにも現れている。統計的仮説検定の理論的枠組みでは、標本数が増加するほど、有意性が検出されやすい仕組みになっている(出村, 2007)。したがって、仮説検証型の研究では、仮説を設定し、仮説を検証するために必要な測定項目を準備し、どの程度の差が現れた場合に差があると判断するのかをあらかじめ決め、それに見合った標本を抽出し、測定を行うことが必要となる。しかし、意図せずして大量にデータ化される購買履歴や顧客情報を活用して、有益な知見を得ようとする場合には、これまでの仮説検証型の分析ツールは適用できない。なぜならば、大量データに対しては統計的仮説検定の理論は有意であるという紋切り型の無意味な反応しか返せないからである(豊田, 2001)。決定木分析を含むデータマイニング手法が対象とするデータは、ネットワークやWorld Wide Webを通じて1日あたり極めて大量に蓄積されるデータであり、意図しない受動的なデータであり、データの蓄積とともに構造や分布が変化するデータに対して適用するツールとしてビジネスの分野で発展してきた。ビジネスの分野だけでなく、体育・スポーツあるいは健康関連分野においても同じ特徴を有するデータがある。医療機関や健康・福祉関連施設の個人データは受診や施設利用ごとに、データ利用の意図に関わらず大量に蓄積される。また、学校保健統計や体力・運動能力調査に代表される官公庁統計は実施主体者側の調査目的はあるものの、最近では総務省統計局のホームページから自由に閲覧でき、申請することで素データを利用することもできるため、調査目的外の利用者にとっては仮説検証ではなく、探索型のデータ解析となる。

以上のように体育・スポーツ科学分野における決定木分析の活用可能性は大いにあると思われる。実際に国内において既にいくつかの研究が報告されている。例えば、西嶋(2002a)は体力評価値を用いて体力の優劣を評価し、体力の優劣を規定する生活習慣のパターンを検討した。その結果、中学生女子において体力評価値の劣る層に含まれる割合が高まる生活習慣のパターンとして、運動実施時間が30分未満かつ運動部・クラブ所属をしていないというパターンを抽出した。同様の研究として、高校生では毎日運動し、過去の運動遊び経験が9種目以上のときに体力評価値が優れる層に含まれる割合が95%となることを報告した(西嶋, 2002b)。また、中田ほか

(2006) はメタボリックシンドローム危険因子の改善に求められる条件を探索し、8.4%以上の体重を減少させることが第一の条件であることを報告した。しかしこれらの報告は学会での発表であり論文にはなっていない。次に、学術雑誌に掲載された研究として、Nakano et al.(2007)は国内のトップレベルの競技者が長期にわたって日々記録した心身のコンディションと生活習慣のデータを用いて、心身のコンディションが良好になる生活習慣のパターンを検討した。その結果、ある対象者の自覚的体調が良好となる割合が高まる生活習慣のパターンとして、起床時刻が4時19分以上かつ睡眠時間が6時間39分以上であるというパターンを抽出した。この他にも Suzuki and Nishijima (2007) による過去の運動経験と現在の運動習慣に関する研究などがあるが、体育・スポーツ科学分野における適用例はごくわずかである。

様々な実践現場で収集されたデータに対して新しいデータ解析手法を適用するためには、解析手法の理論的な発展が前提にあることはもちろんであるが、それ以上に、理論家ではなく、実務家が新しい解析手法を活用するためには、簡易な分析プログラムの開発が重要となる。決定木分析に関しても簡易な分析プログラムの登場により多くの実務家が応用するようになった。すなわち、理論的な理解が不十分であっても分析の実行が可能となった。しかし決定木分析ではアルゴリズムの選択など、分析結果を左右する選択がいくつか存在する。いくつかの解説書ではアルゴリズムの選択基準などについて説明が示されているが、これらはあくまで、意図せず日々大量に蓄積されるデータを解析する上での選択基準である。フィットネスクラブの顧客データは、日々大量に蓄積されるデータの例であるが、このような体育・スポーツ科学分野の例題を用いている解説書は見あたらない。また、マーケティングにおける意思決定に活用されるような探索型の分析ツールとしてではなく、西嶋(2002a; 2002b)や Nakano et al (2007) の報告のように何らかの仮説を基に測定変数を決定し、意図してデータを収集することで、設定した仮説を検証しようとするような仮説検証型の分析ツールとして決定木分析を用いる場合の選択基準や解釈の注意点については、まとまった報告は見あたらない。

そこで本論文では、応用事例として、大学生における運動習慣獲得に関するルール探索を取り上げ、決定木分析の分析方法及び分析の注意点を紹介することを目的とした。決定木分析は分岐の仕方、扱える変数の型(量的データ、質的データ)、変数の選択と分岐の基準の違いによってさまざまなアルゴリズムが存在する。本論文で

は決定木分析において最も広く利用されており(Breiman et al., 1984)、二股に分岐する手法である2進木解析アルゴリズムのひとつである Classification And Regression Trees (CART 又は C&RT) を取り上げる。なお2進木解析は従属変数が質的データのときには分類2進木解析、量的データのときには回帰2進木解析と呼ばれる。

決定木分析を実行可能な汎用的プログラムには、SPSS 社製 AnswerTree, SAS 社製 Enterprise Miner, Salford Systems 社製 CART, 数理システム社製 Visual Mining Studio などがあるが、特にここでは AnswerTree 3.0 J を用いた分類2進木解析の応用事例を紹介する。

2. 大学生における運動習慣獲得に関するルール探索

2.1. データ

一般大学生男子 169 名(19.6 ± 1.4 歳), 女子 31 名(19.6 ± 1.2 歳), 合計 200 名(19.6 ± 1.4 歳)を対象として、現在の運動習慣と過去の運動経験に関するアンケートを集合調査法により実施した。分析に用いたデータは現在の運動習慣(週3日1日1時間以上実施者を1, それ以外を0: 質的データ), 過去の運動経験について、過去に経験したスポーツ種目数(量的データ), 過去に経験したスポーツ種目の週平均運動時間(量的データ), 過去の主観的運動量(5件法: リッカートスケール), 過去の運動に対する好意度(5件法: リッカートスケール)であった。ここで、過去の運動経験とは、大学入学前に経験した運動のことを指し、特に過去に経験したスポーツとは、クラブ活動やスポーツクラブ等に所属した経験を指す。分析に用いたデータセットの形式は表1に示す通りであった。なお、各項目の平均値及び標準偏差は、スポーツ種目数が2.2 ± 1.1 種目, 平均運動時間が5.9 ± 1.6 時間/週, 主観的運動量が3.9 ± 1.1, 運動好意度が4.3 ± 1.0 であった。

2.2. 分析手順

- 1) アルゴリズムの選択: 今回の分析では「C&RT」を用いた。この他に独立変数が質的データである場合に限り「QUEST」でも2進木解析を行うことができる。
- 2) 変数の選択: 従属変数に「運動習慣」, 独立変数に「ス

表 1. 分析に用いたデータ

被験者	運動習慣	スポーツ種目数	平均運動時間	主観的運動量	運動好意度
1	0	3	6	4	5
2	0	2	7	4	5
...
199	1	2	5	3	5
200	0	2	6	2	3

スポーツ種目数」「平均運動時間」「主観的運動量」「運動好意度」を選択する。選択する際には「変数の定義」によってあらかじめ各変数を「名義変数」「順序変数」「量的変数」のいずれかに定義しておく。

- 3) 樹木の検証方法の選択：樹木とは決定木分析における結果の図を指す。得られた樹木（法則）が一般化できる樹木であるかを異なる独立したサンプルを用いて確認する必要がある。このことを交差検証といい、一般には検証用のサンプルを別個に確保する必要があるが、AnswerTreeには代替法として1つのデータセットを学習用サンプルと検証用サンプルに分割するテストサンプル法とデータセットを例えば10個のサブグループに分割し、9個のサブグループで樹木を作成し、残りの1個のサブグループを用いて樹木を検証することをすべての組み合わせについて行う交差検証法がある。サンプルサイズが十分に大きければ「データをサブサンプルに分割」を指定すればよいが、今回はサンプルサイズが小さいので「交差検証を実行」を指定し、交差検証回数をデフォルトの「10」に指定する。ちなみに、Nakano et al. (2007) の論文では1013ケースのデータに対してテストサンプル法を用いており、全データの75%を学習用サンプル、残りの25%を検証用サンプルとして交差検証を実施している。
- 4) 詳細オプションの設定：分岐や樹木の成長を制御する基準を指定することができる。今回はサンプルサイズが小さいため、樹木の成長を停止させる基準（停止則）として、親ノードのサンプルサイズを「25」、子ノードのサンプルサイズを「10」に指定した。ここで、ノードとは分岐の節目を意味し、分岐の元となる節目を親ノード、分岐されてできる新しい節目を子ノードという。これらの値が小さいほど得られる法則は煩雑になり、一般化しにくい法則が導かれる。通常であればデフォルトのままでもよい。C&RTにおける分岐の基準には不純度という概念

が用いられる。不純度とは1つのノード内の応答が、従属変数のカテゴリのうちの1つに集中している度合いを表す。不純度が0のノードとは従属変数のカテゴリのうちの1つにすべてのケースが含まれる場合であり（SPSS, 2002）、今回の分析では、運動習慣あり（又は運動習慣なし）にすべてのケースが含まれる場合である。不純度は様々な方法で定義され、一般的な方法にGiniの分散測度がある（SPSS, 2002）。Nakano et al. (2007) や Suzuki and Nishijima (2007) の論文でもGiniの分散測度が用いられている。第1番目の分岐は不純度を最も減少させる独立変数によって行われ、親ノードと子ノードの不純度の変化（改善度と呼ばれる）が最も高くなるように順次独立変数が選択される。Giniの分散測度はAnswerTreeのデフォルトとして採用されている。

- 5) 樹木の成長：すべての設定が終了した後に「完了」ボタンをクリックすると、樹木ウィンドウが表示される。「樹木成長」ボタンをクリックすると分析が実行される。

2.3. 結果の解釈

図1はC&RTによる分類2進木解析の結果を示している。一番上のノードは分岐する前の状態であり、運動習慣のある者が44.00%（88名）、ない者が56.00%（112名）であることが分かる。決定木分析では従属変数に対して影響の強い順に分岐される。今回の結果では、過去の運動に対する好意度が最も現在の運動習慣に影響を及ぼし、続いて過去の主観的運動量と過去に経験したスポーツ種目数が影響を及ぼしていることが分かった。過去に経験したスポーツ種目の週平均運動時間については分岐項目として採用されなかったことから運動習慣の獲得にはあまり影響しないことが分かった。各項目の分岐値に注目すると、運動好意度が3より大きい（4：好き

だった、5：とても好きだった)と答えた者の中では、運動習慣のある者が50.92%に増加し、さらに主観的運動量が4より大きい(他者と比較して良く動いていた方だと思う)に対して、5：かなりそう思う)と答えた者の中では、運動習慣のある者が59.15%に増加、さらに経験したスポーツ種目数が2.5種目より多いと答えた者の中では、運動習慣のある者が75.00%に増加した。したがって、各項目同士の因果関係は不明だが、運動が好きで、他者と比較して良く動き、大学入学前までに3種目以上のスポーツを経験することが大学生において運動習慣を獲得するために必要であることが示唆された。このように決定木分析ではある法則を導くための基準値を階層的に示すことができるのが特徴である。

AnswerTreeでは、導出された分岐ルールを用いて予測された従属変数の分類精度を「推定誤差」として出力する(図2)。出力結果は樹木ウィンドウの「誤差」タブをクリックし、「推定誤差」の中の「誤差統計量」を確認する。推定誤差はモデルから予測される誤分類率のことであり、全標本数の中で誤って分類された標本数の割合によって算出される。今回は $(24+44)/200 = 0.34$ (誤分類率:34%)となった。また今回は交差検証を行っているので「推定誤差」の中に誤差統計量の他に「交差検証」が出力された。交差検証とは、交差検証回数を10回に指定した今回の場合、樹木作成に用いた9個のサブ

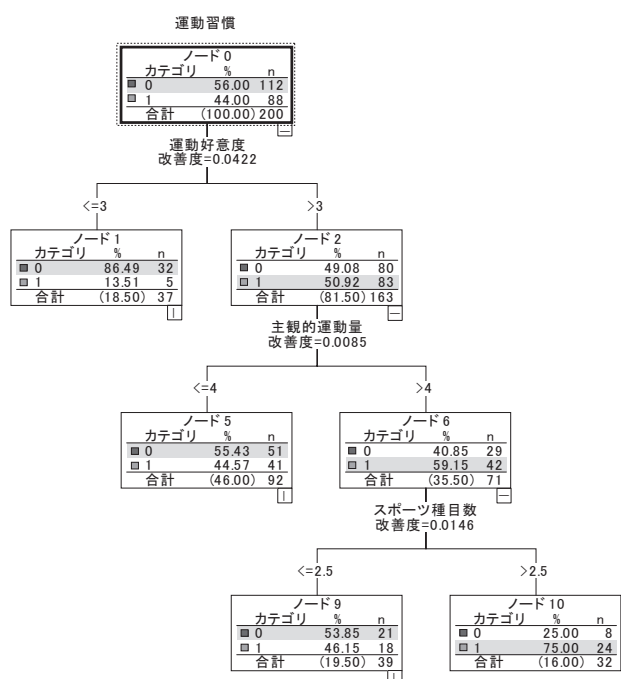


図1. C&RTによる分類2進木解析の結果：樹木図
注) カテゴリについて、0は運動習慣なし、1は運動習慣ありを意味する。

グループ以外のサブグループを用いた樹木の検証が10回行われ、そのときに得られる10個の推定誤差の平均値である。「推定誤差」の中の「交差検証」を確認すると0.4であり、誤差統計量との差はそれほど大きくないのでこの結果は一般化できそうである。しかし、推定誤差の絶対値や誤差統計量と交差検証の値の差から樹木の精度を判断するための基準はなく、適用する場面に応じて判断されなければならない。

分岐に採用された項目や分岐順序の信頼性については、従来の多変量解析手法との整合性を考慮して判断することもできる。例えば今回の例では従属変数が二値データであるため二項ロジスティック回帰分析によって得られる項目のオッズ比や有意性の結果と比較することによって樹木の精度を確認することができる。

3. 分析における注意点

決定木分析は購買履歴データのように大量に日々蓄積されるデータの中からパターンを発見するために用いられることが多く、意味のある結果を得るためには少なくとも1000ケース以上必要であるとされている(大滝ほか, 1998)。鈴木(2007)の報告では600ケース以下であると分岐項目や分岐順序が不安定になり、800以上であっても分岐の値には安定しないものもあるとされている。ここでは200ケースを用いて分析したが、あくまでも例題として捉えてほしい。ケース数が1000に満たないようなデータを扱う場合には、従来法と併用し、結果の信頼性を確認する必要がある。加えて、決定木分析の結果は、決定木分析の手法や分岐の基準によって変化することが知られている(新村, 2002)。Suzuki and

		実際のカテゴリ		合計
		0	1	
予測されたカテゴリ	0	68	24	92
	1	44	64	108
合計		112	88	200

	誤差統計量	交差検証
推定誤差	0.34	0.4
標準推定誤差	0.0334963	0.034641

図2. 樹木ウィンドウの誤差タブ：推定誤差の出力結果
注) カテゴリについて、0は運動習慣なし、1は運動習慣ありを意味する。

Nishijima (2007) の論文では、高校生が習慣的な運動実施を獲得するために必要な過去の運動経験に関する条件を解明するために、習慣的運動実施の有無を従属変数、過去の運動経験を独立変数とする決定木分析を行った。さらに、習慣的運動実施（二値：有り・無し）を従属変数、過去のスポーツ経験3項目（スポーツ種目数、スポーツ開始時期、1週間あたりの平均運動時間）を独立変数とする3項目強制投入法によるロジスティック回帰分析を実施し、決定木分析の結果とロジスティック回帰分析の結果の一致性を確認している。

分析結果(研究の結論)が標本あるいは分析手法によって異なることは決定木分析に限ったことではなく、従来の多変量解析手法においても認められることである。したがって、このことは決定木分析手法に特有な問題点ではなく、いかなる統計手法を用いる場合にも注意しなければならないことである。なお、分析アルゴリズムや停止則の違いによる結果への影響については、新村(2002)の論文の中で例題を用いて紹介されているので参照されたい。

4. まとめ

いかなる分析手法においても、分析に用いていない変数が従属変数にどの程度影響しているのかを実証することはできない。一般に、データマイニングの一手法である決定木分析は、インターネットや Point of Sales (POS) を通じて日々刻々とデータが蓄積されているデータベースから価値ある情報を掘り出すために利用される(豊田, 2001)。したがって、分析に用いるデータは何らかの仮説に従って計画的に用意されたものでないことが多い。しかし、本論文で紹介した研究のように、体育・スポーツ科学分野のデータに対して決定木分析を適用する場合には、理論的背景に裏付けされ、研究者が設定した仮説の範囲内から用意された変数であることが多い。このことは元来、決定木分析の適用を想定していたデータと体育・スポーツ科学研究におけるデータの相違点である。狩野(2002)はこれまでの多変量解析手法について、因果推論あるは因果関係の大きさを検討する場合、分析モデルが交絡変数をすべて取り込んでいるという検証不可能な前提をおいており、科学的精密実験のように興味のある原因変数を1つに絞り、その他の変数をできるだけ一定に保ったまま1つに絞った原因変数を動かすことをしない限り、因果の大きさを同定することは難しいと述べており、横断的に収集された調査データを用いる場合に、分析に投入する変数の重要性を強調している。この

ことを前提としながらも、決定木分析を用いて現時点で従属変数との関連性が想定される変数群を用いた高い分類ルール(基準値)を提示することは、体育・スポーツ科学分野の発展にとって重要であると思われる。近年では統計的検定論が十分機能しないほどの数ギガ単位のデータが利用できる環境にある。その中から意思決定に必要な何らかの情報を取り出すための手法として決定木分析を含むデータマイニングは重要であると思われる。しかしながら、豊田(2001)が指摘しているように、データマイニングにおけるデータ解析手法の特徴は得られた解の最適性と一意性の保証のなさにあり、発見された知識の持つ価値が、その知識を発見するのに要した手間を上回れば分析は成功と考えるのである。したがって、意思決定のツールとして実務家が用いる場合には、対費用効果を考慮した最適解を得ることになるが、普遍的な理論体系を説明するために用いる場合には、仮説探索ツールとして捉え、得られた仮説を従来の仮説検証型の方法論において検証することが必要ではないかと考えられる。

付記

本論文を作成するにあたり、SPSS社の研究助成は一切受けていません。

文献

- Breiman, L., Friedman J.H., Olshen, R.A., and Stone, C.J. (1984) Classification and Regression Trees. Belmont, CA: Wadsworth.
- 出村慎一(2007) 健康・スポーツ科学のための研究方法 (pp. 124-125). 東京: 杏林書院.
- Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. (1996) Advances in knowledge discovery and data mining (KDD-95) (pp. 1-34). Menlo Park, CA: AAAI/MIT Press.
- 岩崎学(1999) データマイニングと知識発見－統計学の視点から－. 行動計量学 26: 46-58.
- 狩野裕(2002) 第II部構造方程式モデリング, 因果推論, そして非正規性. 甘利俊一ほか(編) 多変量解析の展開: 隠れた構造と因果を推理する (pp. 65-129). 東京: 岩波書店.
- Nakano, T., Nishijima, T., and Suzuki, T. (2007) The Effectiveness of Decision Tree Analysis on Routine Athletic Conditioning Data. International Journal of Sport and Health Science 5: 12-20.

- 中田由夫, 大蔵倫博, 松尾知明, 片山靖富, 沼尾成晴,
大河原一憲, 田中喜代次 (2006) メタボリックシン
ドローム危険因子の改善に求められる条件とは:
The SMART Study. 第 27 回日本肥満学会プログラ
ム・抄録集: 174.
- 西嶋尚彦 (2002a) 青少年の体力低下の現状と改善策. 日
本体育学会第 53 回大会合同シンポジウム (2) 配付
資料.
- 西嶋尚彦 (2002b) 子どもの体力低下傾向とその要因. 第
1 回日本発育発達学会シンポジウム①配付資料.
- Nishijima, T., Nakano, T., Takahashi, S., Suzuki, K.,
Yamada, H., Kokudo, S., and Ohsawa, S. (2003)
Relationship between Changes over the Years in
Physical Ability and Exercise and Sports Activity
in Japanese Youth. *International Journal of Sport
and Health Science* 1: 110-118.
- Prentice, A.M. (2001) Obesity and its potential
mechanistic basis. *British medical bulletin* 60: 51-67.
- 奥山真一郎 (1997) 経営情報戦略における「データマイ
ニング」の役割. 第 16 回日本 SAS ユーザー会研究
発表会論文集: 461-473.
- 奥喜正, 本村猛能, 前鶴政和, 内桶誠二 (2004) データ
マイニングにおける二値データ解析-決定木とロジ
スティック回帰分析-. *物流問題研究* 44: 1-14.
- 大滝厚, 堀江宥治, Dan Steinberg (1998) 応用 2 進木解
析法-CART による-(p.175). 東京: 日科技連.
- 新村秀一 (2002) 数理計画法を用いた最適線形判別関数
(5) -決定木分析との比較-. *オペレーションズ・
リサーチ* 5月号: 315-321.
- SPSS (2002) 第 5 章 C & RT 分類樹木. SPSS (編)
AnswerTree コーステキスト (p.3). 東京: SPSS
Japan.
- 鈴木宏哉 (2007) 分類 2 進木解析のサンプルサイズに対
する頑健性. 日本体育測定評価学会第 6 回大会抄録
集: 34.
- Suzuki, K., and Nishijima, T. (2007) The Influence of
Past Sports Experience on Determining Current
Exercise Habit in Japanese Youth. *School Health* 3:
22-29.
- 豊田秀樹 (2001) 金鉱を掘り当てる統計学: データマイ
ニング入門 (pp. 5-39). 東京: 講談社.
- 牛田一雄, 高井勉, 木暮大輔 (2003) SPSS クレメンタイ
ンによるデータマイニング (p.9). 東京: 東京図書.