

平成 18 年度 流通データ分析 データマイニングの活用事例 レポート

● データマイニングについて

出玉イニングとは小売店の販売データや電話の通話履歴、クレジットカードの利用履歴など、企業に大量に蓄積されるデータを解析し、その中に潜む項目間の相関関係やパターンなどを探し出す技術である。例えば、スーパーの販売データをデータマイニングで分析することにより、「ビールを買う客は一緒に紙おむつを買うことが多い」「雨の日は肉の売り上げが良い」など、項目間の相関関係を見つけることができる。

データマイニングを利用し多彩な事象の分析が可能であるが、事象が多様化し選好が細分化した「サブカルチャー」と呼ばれる現代日本文化の各ジャンルに対し、社会調査サイトで得たデータを基に、〈服飾ブランド〉を分析した一つの結果が以下である。

服飾ブランドにおいて女性のファッションブランドの認知が自身の理想形を求め、レイヤーを下るにしたがって、スタイルとしての個性を発揮するようになるが、最終的には身体性に帰着するという特徴を導き出した。

(引用) 2003 年度 森泰吉郎記念研究振興基金 成果報告書
データマイニングを援用したサブカルチャー事象解釈
<http://www.mag.keio.ac.jp/~ond/mori/2003/>

感想

インターネットで調べる中で様々な事象をデータマイニングによって分析し理解を深められることがわかったが、それと同時にデータマイニングによって得られた結果を一途に信用しがちだという意見も垣間見え、今の自分にはそれを判断する知識があまりにも不十分であることがわかったので、もっと学びたいと思った。また自分が興味を持っている分野などにデータマイニングを活かされるようにしたい。

=====

夜間気温の上昇に着目した地震予測方法に関する考察

はじめに

兵庫県南部地震いわゆる阪神淡路大震災は 1995 年 1 月 17 日に発生したが、その 14 日前である 3 日から 4 日にかけて気温変化異常があった。2002 年・地球惑星科学関連学会ではこの事象について述べた。[1] その後 1990 年から 10 年間の神戸データを対象にデータマイニング調査を行った。ここではこの調査についての報告を行うとともに夜間気温の上昇に着目した本方法の可能性についても報告する。

10 年間の神戸の気温データから得られた知見

分布を調べると 20 時から翌 4 時までの間に 2.0 度以上となったのは 3651 ケース中 37 ケースにとどまった。更に、この 37 ケースについて低気圧の通過状況などをもとに分類を実施した。その結果 9 ケースについては気象学的な乱れが無いものと考えられる。そこで更にこの 9 ケースについて調査した。その結果、2 ケースについては地震の前兆としての可能性がある。

地震予測方法としての可能性

ピックアップされた 2 ケースでは、一方は気温の日変化が目立たなくなるもので、もう一方は夜間気温が上昇するものである。これをもってただちに「夜間気温の上昇に着目した地震予測方法が有望である」とは言えないが、今後相関性について更なる検証ができれば実用的な「地震予測方法のひとつ（特に震源エリア予測に強い）」と言える可能性が出てきた。

また本手法を 2000 年・鳥取県西部地震 M7.3 において適用し地震の 21 日前に夜間の気温が上昇している事象を見つけている。

参考 URL : www-jm.eps.s.u-tokyo.ac.jp/2003cd-rom/pdf/s049/s049-012.pdf

=====

データマイニングは、さまざまな分野で活用されている。多くの場合は、商品のマーケティングなどの市場調査に活用されている。パソコンの普及によって以前よりも大量のデータを収集できるようになったので、さまざまなツールを利用してデータマイニングを行っているようだ。

さらに、最近では市場調査のみならず、データマイニングを利用して、ハッカーから WEB サイトを守るためのセキュリティ対策にも利用されている。データマイニングをどのように利用すれば、ハッカーから WEB サイトを守ることができるのだろうか。

大手企業の WEB サイトには、一日に数千件を超えるアクセスがある。その中には、悪意のある者から攻撃が多々ある。その攻撃は場合によっては、大きな損害につながる場合がある。

数千件のアクセスの中からデータマイニングを利用することにより、その WEB サイトにアクセスする人達のサイトの回り方を調査する。その中から不振なアクセスをするものを割り出すことで、攻撃の予兆を見つけることができるのだ。

このシステムは未完成らしいが、マーケティングなど以外でデータマイニングを利用する方法として着目してされているらしい。

参考にしたサイトの URL

<http://www.attention.jp/marke.html>

<http://enterprise.watch.impress.co.jp/cda/security/2006/12/21/9316.html>

<http://ja.wikipedia.org/wiki/%E3%83%87%E3%83%BC%E3%82%BF%E3%83%9E%E3%82%A4%E3%83%8B%E3%83%B3%E3%82%B0>

=====
時系列相関ルールマイニングに基づく人工衛星テレメトリデータからの異常検出法

概要：なにやら小難しい単語並んでおり、内容がよくわからないタイトルですが、これは町田・矢入研究室（東京大学 先端科学技術研究センター 知能・情報系 宇宙環境システム学）の論文で、端的に言えば人工衛星に発生するトラブルを事前に予測できないものかというものです。もう少し詳しく言えば、人工衛星から送られてくるデータをデータマイニング技術によって処理することで、トラブル発生の兆候をつかもうという研究です。そういった兆候は過去の人工衛星等のトラブル発生件数の少なさからデータの蓄積がなく、こういったデータがトラブル発生兆候となりうるのかはいまだ手探り状態であるため、データマイニングを適用することでそういった兆候が発見できるのではないかというものです。

<http://www.ai.rcast.u-tokyo.ac.jp/~yairi/papers/jsai-conf2001-yairi.pdf>

東京大学 先端科学技術研究センター 知能工学研究室

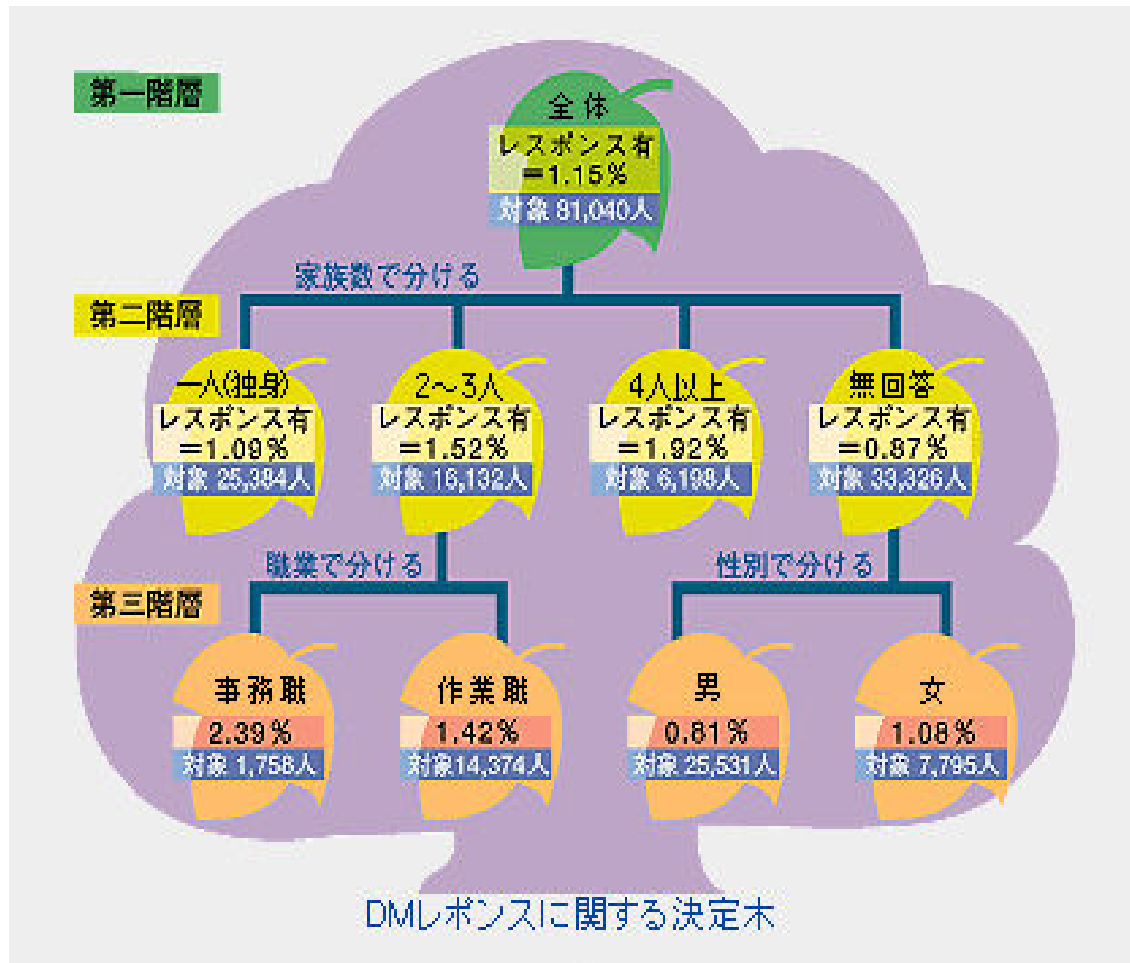
<http://www.ai.rcast.u-tokyo.ac.jp/index-jp.html>

感想：宇宙開発事業団から実際の人工衛星のデータをもらっているあたり、さすがに東大だと思います。（偏見？）

提案手法・実験評価のあたりはさっぱり理解できませんでした。

=====

雑誌の購買勧誘のためのダイレクトメール(DM)のレスポンス率(回答率)の事例



<http://www.ne.jp/asahi/itoh/seiichi/paper/DecisionTree/>

まず全体を家族数で分けると、家族数が2~3人と4人以上の家族の場合にレスポンス率が向上することがわかる。特に家族数が4人以上の場合ではレスポンス率が1.15%から1.92%となるため、無差別にDMを発送した場合と比べて4人以上の家庭に絞り込んで発送した場合、レスポンス率が大幅に向上すると考えられる。

次に家族数が2~3人のグループを職業別に分けると、事務職の場合、レスポンス率が大幅に上昇することがわかる。つまり、家族数が2~3人の事務職にDMを発送した場合にはレスポンス率が1.15%から2.39%に上昇するため、無差別に送った場合と比べて、倍以上の効果があると考えられる。

=====

近年のハードウェア性能の向上やネットワーク環境の整備等に伴って、データベースは急速に大規模化している。ところが、そのようなデータを分析し、利用するには統計の専門知識が必要であったり、統計では発見しにくい知識があるなどの問題があった。一方で、企業は生データから得られた情報を活用して、迅速な意思決定を行う必要がある。データマイニン

グにより、大量のデータから有効な知識を発見し、経営戦略作成等に活かすことができる。
例えば、

- ・有力な市場はなにか？
- ・新しいキャンペーンはどうすれば効果的か？
- ・どんな商品が望まれているか？
- ・売上に影響を与えているのはなにか？

といった知識を得ることができる。

またデータマイニングによって得られる情報には、通常予測とルールがある。予測は未来の事柄を推測することであり、ルールは予測のための規則を人間がわかりやすい形で表したものである。

具体例をあげると、

- ・商品 A は来週 20000 個売れるだろう
- ・来月の総売上高は 1 億 5000 万円くらいだろう

などは予測の例であり、

- ・商品 A は 20 代の男性によく売れている
- ・商品 B と商品 C は同時に売れることが多い

などはルールの例である。

予測とルールは表裏一体の関係にあり、両方同時に行われることが求められている。

現在用いられている主な手法としては、統計、決定木、ニューラルネットワーク等がある。

参考 URL : http://www.toshiba.co.jp/rdc/datamining/mining_j.htm#top

～感想～

現代の経済社会において、企業や市場でデータマイニングが行われることは必須であると思う。実際に分析が行われることで、最適な経営戦略を考えることができたり、利益にもつながると考えられる。そのような意味では、データマイニングは重要かつ実用性のある分析手法の一つであるといえる。

=====

適用例

粒納豆と豆腐

イワシだんごと揚げかまぼこ

参考 URL

<http://www.yamaguti.comp.ae.keio.ac.jp/japanese/2006-e%E3%83%93%E3%82%B8%E3%83%8D%E3%82%B9SW/%E7%9B%B8%E9%96%A2%E3%83%AB%E3%83%BC%E3%83%AB.pdf>

感想

探してはみたものの、意外に思えるような購買パターンは、いまだ紙おむつとビール以外には見つかっていない。上記のような、“おでんの具”，“定番の朝食”，“スパゲッティとソース”など、想像に難くないパターンを調べだすことしかできなかった。

ただ、記載した URL にある、もう少し視野を広げたデータマイニングに興味を持った。ドレッシングの購買パターンを時系列から考えたもので、ある商品を買った人が、次に来たときにも同じ商品を買う頻度や、他の同種の商品に乗り換える確率などを調べたものようだ。こういった調査は、“商品の中毒性”のデータマイニングとでも言うのだろうか、非常におもしろい手法だと感じさせられた。

＝＝

私は、今回の課題で、「女性が選ぶ海外ブランドの人気を探る」というデータマイニングに事例を選びました。

あらゆる年代の女性が海外ブランドを好み、こだわりを持っている。そこで女性への海外ブランド調査データから海外ブランドはどのような基準で選ばれるのか、どんなブランドが人気なのか、年代により好みに違いがあるのかを統計手法により探っている。

アンケート調査の結果

表1は96年8月31日の日経流通新聞に載っていた首都圏在住の15歳以上の女性モニター454人に聞いたアンケート調査データである。

	人気度	認知度	所有率	高級感	誇らしさ	品質の信頼性	センスのよさ	親しみやすさ	広告が魅力的
シャネル	159	377	209	318	136	150	123	36	86
エルメス	145	327	136	245	104	154	127	27	41
ティファニー	145	327	136	182	86	136	136	77	59
ルイ・ヴィトン	136	359	186	177	77	186	82	109	18
グッチ	123	350	154	163	73	141	114	68	32
ラルフローレン	114	295	200	54	27	114	91	154	36
カルテ	109	291	109	232	95	150	95	14	23

イエ									
フェラ ガモ	109	286	68	159	64	109	77	32	18
ブラダ	104	245	45	104	50	77	82	59	18
カルバ ン・クラ イン	100	263	123	32	23	64	118	132	54
ベネト ン	86	327	241	18	5	54	59	227	95

(表 1) 454 人回答 (複数回答)

評価項目の相関

相関係数を求め、9つの評価項目同士で相関 (関連) があるかを調べてみる。

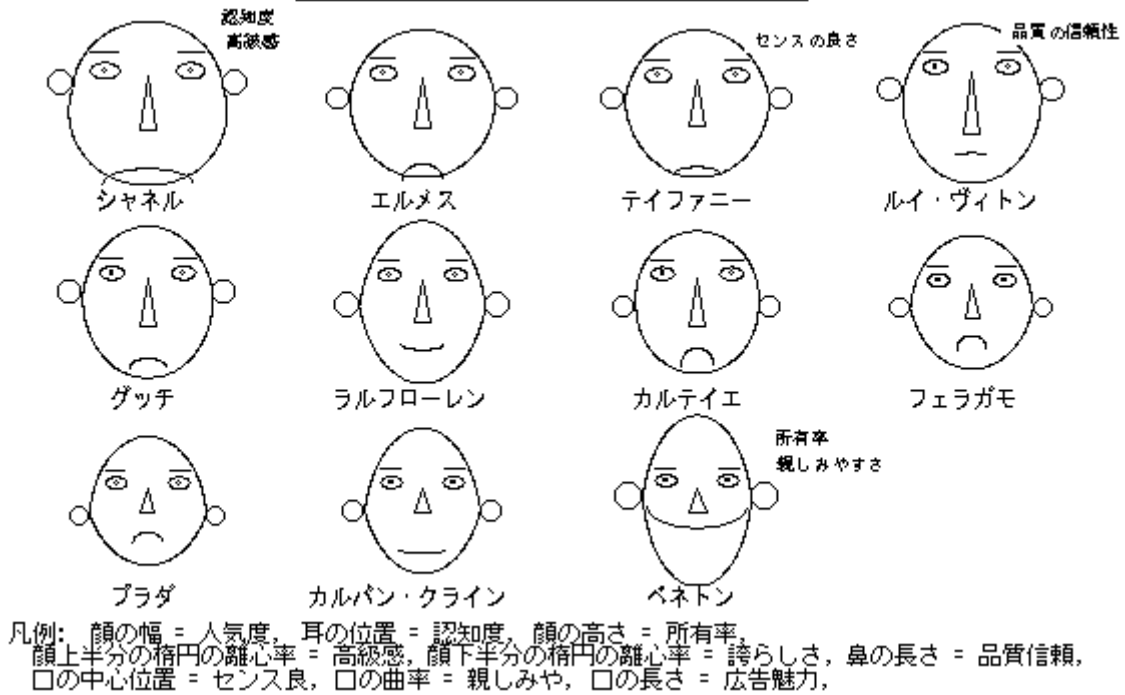
	人気度	認知度	所有率	高級感	誇らし さ	品質の信頼性	センスの 良さ	親しみや すさ	広告が魅 力的
人気度	1								
認知度	0.679	1							
所有率	0.177	0.699	1						
高級感	0.814	0.547	-0.064	1					
誇らしさ	0.853	0.537	-0.081	0.994	1				
品質の信頼性	0.783	0.640	0.119	0.788	0.789	1			
センスの良さ	0.703	0.266	-0.053	0.503	0.578	0.381	1		
親しみやすさ	-0.518	-0.016	0.613	-0.810	-0.813	-0.558	-0.451	1	
広告が魅力的	0.087	0.405	0.676	-0.059	-0.058	-0.323	0.159	0.469	1

(表 2) 相関係数行列

人気度と相関があるのは相関係数 0.679 の人気度、0.814 の高級感、0.853 の誇らしさ、0.783 の品質の信頼性、0.703 のセンスの良さである。人気のあるブランドは高級感、誇らしさがあり、品質に信頼性があり、センスが良いブランドであるということがわかる。認知度と相関があるのは 0.699 の所有率、0.640 の品質の信頼性である。認められるには品質の信頼性が必要であると思われる。所有率と相関があるのは 0.613 の親しみやすさ、0.676 の広告が魅力的である。高級感と相関があるのは 0.994 誇らしさ、0.788 品質の信頼性、-0.810 の親しみやすさである。高級感と誇らしさは相関係数がほとんど 1 に近い値となっており高級感と誇らしさは関係深いように思われる。また、高級感と親しみやすさとは相関係数が負となっている。これは、高級感と親しみやすさとは相反する評価項目だということを示している。

次に海外ブランドの特徴を顔グラフで表わしてみる (図 2)。

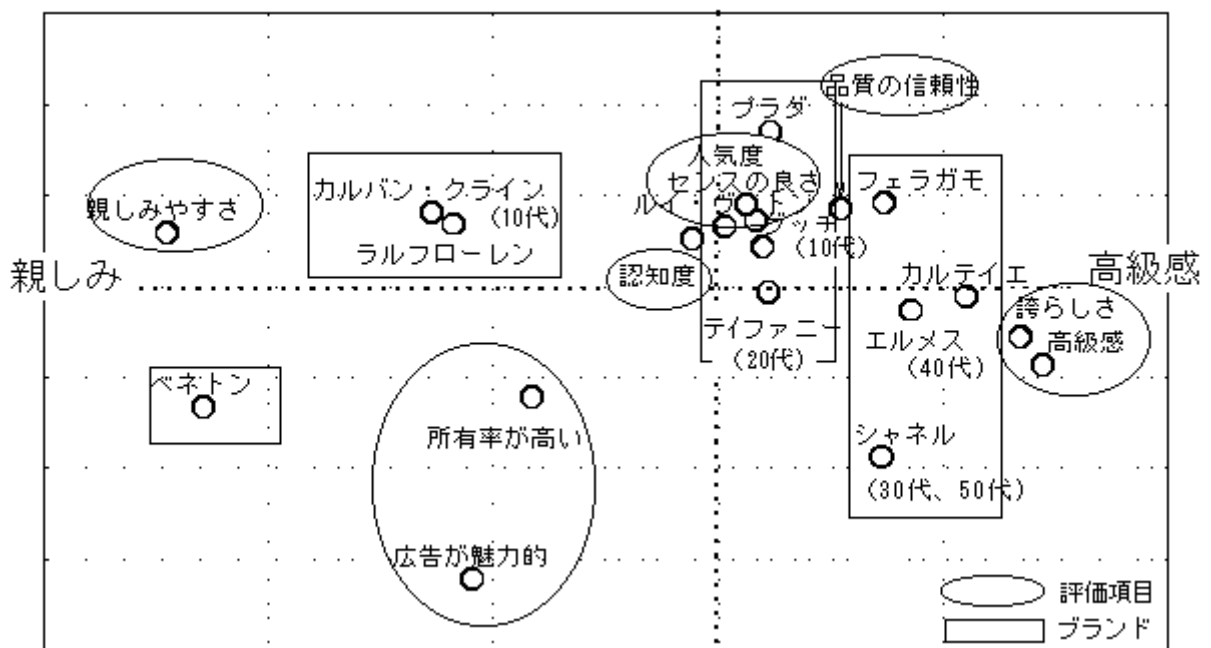
図2 海外ブランドを顔で表わすと



例えば、シャネルの顔を見ると耳が高いところにあり、頭が比較的大きいので認知度が高く、高級感があることがわかる。鼻が長いルイ・ヴィトンは品質の信頼性が高いことを表わしている。ベネトンの顔は長いので所有率が多く、口に特徴があり親しみやすさを表わしている。つまり、似た顔（ブランド）同士は同じような評価をされているということである。

次に、双対尺度法を用いてブランドおよび評価項目を同時に平面上にプロットした（図3）。似た評価をされているブランドは□でくくり、似た評価項目は○でくくっている。

図3 海外ブランドと評価関連図



全体の傾向として左に行くほど親しみを感じ、右へ行くほど高級感を感じるブランドがきている。年代も左から10代、・・・、50代となっており、若い年代ほど親しみを感じるブランドを好み、年を経るにつれて高級感志向となっています。

筆者が注目しているのはベネトンである。ベネトンは若い年代の女性に好まれている。そして広告が魅力的との評価も受けている。ブランドはベネトンからはじまり、年を経るとベネトンから離れていくのだろうか、それともベネトンにとっては40,50代の女性の心をとらえることができないのだろうか、あるいは、戦略的に若い女性のみを狙っているのだろうか、などということが考えられる。

以上概括すると、高級感のあるブランドはシャネル、エルメス、カルティエ、フェラガモが、センスの良さはプラダ、ルイ・ヴィトン、グッチ、ティファニー、品質の信頼性ではルイ・ヴィトンが高い評価を得ており、若い女性は親しみを感じるベネトンを好んでいることがわかる。このように、ブランドの評価項目同士の相関係数を調べることにより人気のあるブランドはどのように評価されているかわかる。

感想

今回の課題で、データマイニングのことについて、いくつかサイトをみて調べたので勉強になった。今回私が選んだ事例で、特に面白いと感じたのは、顔グラフである。各ブランドの特徴をブランドごとに顔の形で表していたので、視覚的に見やすかったと思う。

<参考>上田データマイニング塾 (<http://www.datamining.jp/index.html>)

=====

例として、ぶたまんをあげる。

概要

とあるお店でPOSシステムのデータにより、ぶたまんと併せて買われる商品について調べたところ、缶の十六茶という情報を得た。

では本当に併せ買いされているのか。缶の十六茶とぶたまんは、同時に特売されていたため、主婦はまとめ買いをしたのかも知れないのである。それを調べるため指標としてさらに、広告の有無、商品の陳列を近さ、値下げの有無、曜日（土日か平日か）、天候（晴れか雨か）を調べることにした。

そして相関係数と散布図により、ぶたまんと相関があるのは、広告、近距離、天候となったのである。さらに詳しく調べるために回帰分析を用いることとなった。

その結果のうち、缶の十六茶の売上数に影響しているのは、近距離と値下げだということがわかった。さらに細かく調べると、近距離と値下げのどちらが売上数に影響しているかは、近距離の方だということがわかったのである。

ぶたまんの売上数に影響しているのは、広告、近距離、天候であった。広告をし、十六茶と近距離におき、天候が晴れのときに売上数に対する効き方は、（回帰）係数を棒グラフにすると、広告、天候、近距離の順であることがわかるのである。

(<http://www.datamining.jp/index.html> 参照)

感想

データマイニングをインターネットで調べることで、データマイニングとはどのようなものであるか、より理解をすることが出来たので更に興味や意欲がわいた。

自分の普段の行動でもデータマイニングの相関となりうる行動をしているにではないかと考えることも出来、非常に意識が高まった。

=====

データマイニング導入事例～医療分野への導入～

アメリカ、シカゴにある Children's Memorial Hospital の脳腫瘍研究所長、エリック・ブレマー博士は小児脳腫瘍の治療にデータマイニングの技術を応用している。現在、小児脳腫瘍に対する有効な治療法は手術以外には確立されていない。このため、医師が実験的な研究や手法に基づいて治療法を作り出していかなければならないのである。そこで、博士は小児脳腫瘍に関する遺伝子発現のデータベースを構築し、このデータと、過去から現在までの有効な治療法との相関関係を発見しようとしている。

この小児脳腫瘍治療のデータベース作成においてまず必要とされるのは、腫瘍の種類を正確に分類することである。小児脳腫瘍は外観から判断することが困難であり、その分類は極めて主観的である。また白血病等の他の小児ガンで使用されている客観的な判断指標（マーカー）もほとんど存在していない。腫瘍の種類だけでなく、腫瘍の臨床病期分類（Stage）および病理学的組織分類（Grade）を判断することも重要である。しかし、神経病理学者にとって、その中間段階の区別は容易では無く、さらに隣接する段階における治療法が著しく異なることがあるので、段階を

間違って判断した場合、患者に大きなダメージを与えてしまう危険性がある。

ブレマー博士は、公表されている Pomeroy らのデータセット (Nature vol. 415, pp. 436-442, 2002) と自らのデータを組み合わせることで、主要な 6 種類の小児脳腫瘍から、合計 133 個の腫瘍サンプルを得ることが出来た。博士はデータマイニングを行うことにより、これらの腫瘍を 95% 以上の精度で分類した。このサンプルは病理学的にすでに詳細に記述されているもので、これが将来の小児脳腫瘍分類を予言するようなテストケースとなりそうである。特にこの方法は、診断が難しい腫瘍に対して有効であるとされている。

今後ブレマー博士のデータベースに臨床的、病理的、生化学的な情報も統合して、より完全かつ正確なものを提供できるようにしていく計画がある。また患者情報をデータベースに加えることで、遺伝子的また病理的な特徴を共有する脳腫瘍に対して最適な治療法を発見できる可能性もある。データベースをより大きなものにしていくと、腫瘍の発生に関与する遺伝子を複数明らかにすることも出来るだろう。『Nature』誌によると、悪性腫瘍の分類、診断、治療に関する記事は、2001 年だけでも 21,000 件以上もあった。ブレマー博士はデータマイニングにより、これらの記事を選別し、使用できそうな記事を抽出していく予定である。

感想

データマイニングの導入事例のほとんどはマーケティングに関するものであり、それ以外の事例は珍しい。今回は小児ガンの治療に大きな可能性を見出したが、反面、データを有効に生かしていないようなケースも多い。これはデータマイニングを行うためには専門的な知識がいるためであり、今後、有効にデータマイニングを行っていくためには、知識を持たない人でも使うことの出来るデータの解析ソフト等の開発が必要であると感じた。

参考：<http://www.spss.co.jp/labocase/files/c039.html>

=====

データマイニングとは、データの集合の中から、知識を発見しよう、というものです。ここでいう知識とは、データの中に見られるルールとか法則のことです。最近データマイニングの紹介でよく使われるものに、ビールと紙おむつという言葉があります。これは、スーパーで客の買った物を分析したら、紙おむつを買う人はビールを買うことが多い、という傾向が出てきた、ということです。この、「紙おむつを買う人はビールを買うことが多い」というのがある種の知識なのです。このような知識をデータの中から機械的に、自動的に見つけよう、というのがデータマイニングです。

この、ビールと紙おむつのような関係でしたら、単純にお客さんの買い物リストのデータを見るだけで、発見できます。しかし、例えば、5000 円以上の買い物をしてくれる人は、ビールと紙おむつを買うか、トイレットペーパーとティシュペーパーと洗剤を買う（あくまで適当に考えた例ですが）、というような、複雑な形をした知識を見つけ出そうと思ったら、単純にデータを見るだけでは、とんでもなく時間がかかります。例えば、1000 品目の中から 5 品目を組合せて、5000 円以上の買い物をする客がその組合せの買い物をしているかどうかをチェックするとしますと、およそ 1000 の 5 乗、1000 兆回もデータを見なければいけません。たいていのコンピューターでは 1

日や2日で終わる計算ではありません。こういった知識の発見をするためには、[数理計画](#)や[アルゴリズム](#)の技術を使って、効率的な方法を考えないといけないのです。

もう1つ例を見てみましょう。車を買った人のデータを分析して、車を買う人の年齢と、買った車の金額についてどんな関係があるか調べるとしましょう。絵で書くと図のようになります。

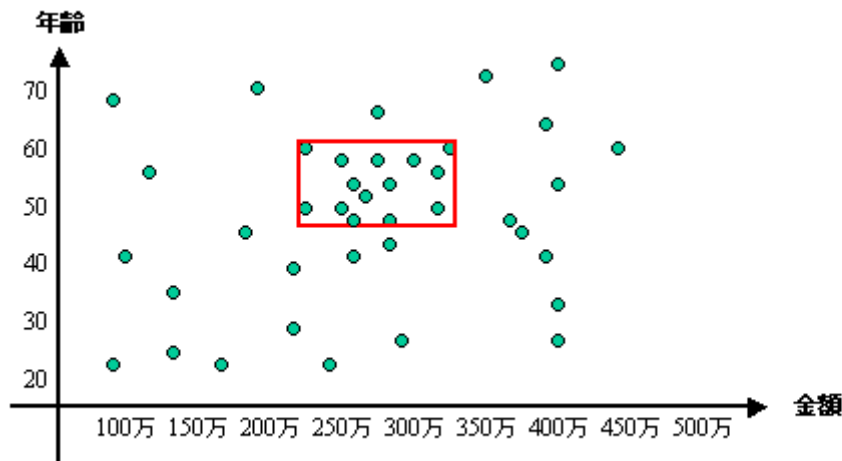


図1: 年齢と買った車の金額の関係

普通に考えて、年をとるほど高い車を買いたいという人はいるんですが、例えば、図の赤線で囲った部分が示すように、48歳から60歳の人が230万から320万の車を買う可能性が高い、というような具体的な年齢層と購入価格帯がわかれば、効果的な宣伝ができるわけです。年齢を1歳刻みで20から70まで、車の価格を10万刻みで100万から500万まで調べるとすると、 $50 \times 50 \times 40 \times 40$ の組合せを全てについてデータを検索する必要がありますが、これは、[アルゴリズム](#)を工夫すると、 50×40 の組合せを調べるだけですむようになります。

昔から、こういったデータの分析には統計的な方法がとられていました。主成分分析や回帰分析などです。これらの方法は、まず、データがどのように分布しているか、その分布の仕方を推定して、将来予測や分析をしようというものです。例えば、消費者金融会社が、お金を借りた人の年齢と年収と自己破産の関係を調べたいとします。集まったデータが図のようになっているとしましょう。

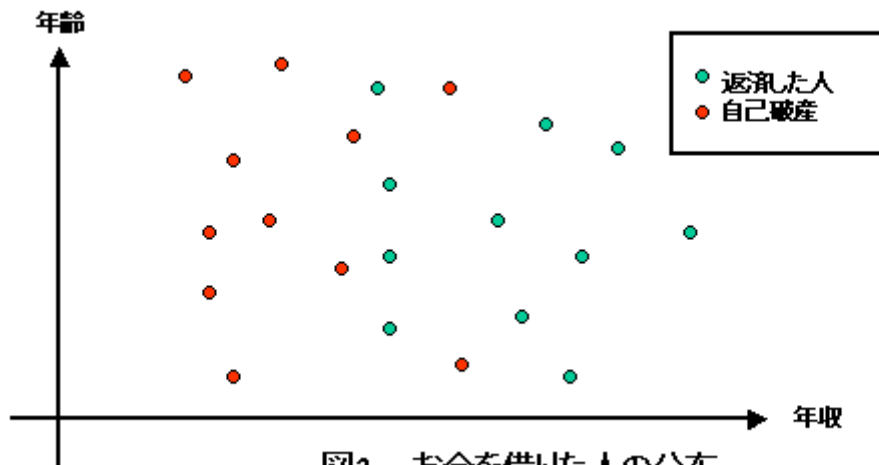


図2: お金を借りた人の分布

これはお金を借りた人それぞれを、年齢と年収に対応する位置に、破産した人なら赤で、そうでなければ緑で書いたものです。これを、赤の点、緑の点それぞれ、正規分布に従って分布しているのだ、と思って分布を推定すると、例えばこのようになります。

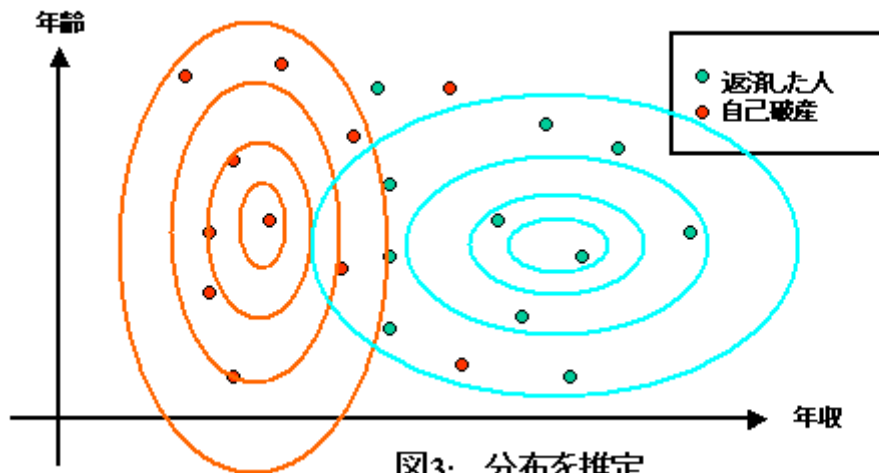


図3: 分布を推定

このように分布が推定できると、新しくお金を借りに来た人の年齢と年収がわかると、自己破産の確率はどれくらいかな、と推定できるわけです。データが多くなれば、それだけ推定も正確になります。ただ、「こういうふう分布しているだろう」という分布の形の仮定が間違っていると、推定は全然当たらなくなります。この、分布の推定、というのがやっかいです。

対して、最近開発された、データマイニングで研究されている方法の1つ、サポートベクターマシンという方法は、ちょっと発想が違います。赤い点のグループと緑の点のグループを、直線を使って分割してしまおうというものです。この図のように。

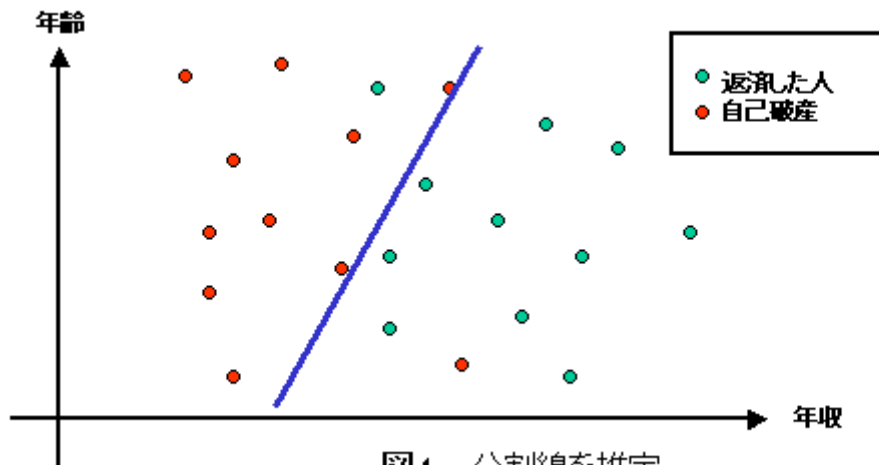


図4: 分割線を推定

このとき、直線の片方側にはなるべく赤い点が多くなるように、もう片方の側には緑の点が多くなるように、直線を決めよう、というものです。なるべく上手にグループ分けをする、最適化問題を解くわけです。新しいお客さんが来たときに、その年齢と年収を聞いて、この線のどっち側にお客さんが来るか、で自己破産しそうな人かどうかを判定しよう、というものです。この方法だと、データがどういう形の分布になっているか、さっぱりわからなくても、データから直接、こういった直線を得られる、というのが強みです。データの分布がわからないから上手に分析・予測ができなかった問題に対して、解決法を提示できる可能性があるわけです。

<http://research.nii.ac.jp/~uno/datamine.htm> 参照

感想・・・非常に多いデータを網羅的に適用することでこのような知識を得ることができた。データマイニングは使いようによっては物凄い力を発揮すると思った。

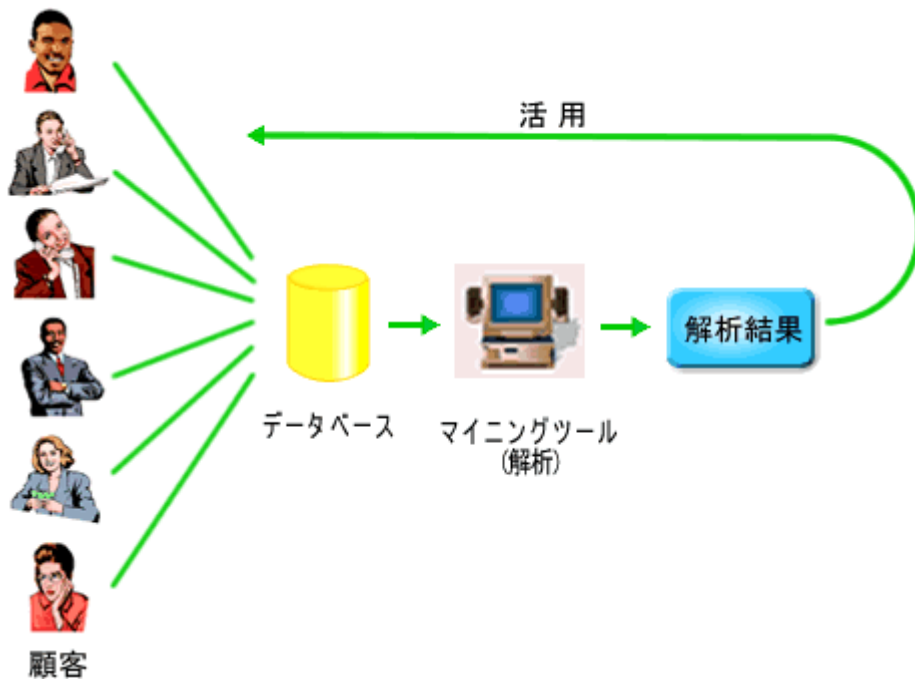
=====

「データマイニングの具体例」

小売店の販売データや電話の通話履歴、クレジットカードの利用履歴など、企業に大量に蓄積されるデータを解析し、その中に潜む項目間の相関関係やパターンなどを探し出す技術。従来は、こうした取引の「生データ」は、経理処理に必要なだけで活用されていなかったが、情報技術の向上により、潜在的な顧客ニーズが眠る「鉱山」として「採掘(mining)」されるようになった。例えば、スーパーの販売データをデータマイニングで分析することにより、「ビールを買う客は一緒に紙オムツを買うことが多い」「雨の日は肉の売上が良い」など、項目間の相関関係を見つけることができる。また、クレジットカードの利用履歴を解析することにより、不正使用時に特徴的なパターンを見つけ出し、あやしい取引を検出するなどの応用も考えられる。

営業データマイニングとは

マイニングとは「採鉱」を意味する。鉱山から鉱石を掘り出すことがマイニングである。企業の膨大なデータから意味ある関係性やビジネスパターンを探り出すのがデータマイニングだ。営業データマイニングでは、顧客属性～購買データ、営業活動～商談データなどの営業活動全般にわたるデータを分析し、ピンポイントのプロモーション展開などに結びつける。営業活動高度化の必須ツールが営業データマイニングだ。



詳細は「[5分でわかる IT](#)」をご覧ください

補足説明

- ◆ 図は営業データマイニングの流れを示したものである。データベースに蓄積された情報をマイニングツールが解析し、洗い出してきた解析結果を顧客にフィードバックし、さらなるメリットを提供する。
- ◆ データマイニングの活用例
 - ・ 製造・・・ 顧客満足度、品質管理、顧客ニーズの発見など
 - ・ 小売・流通・・・ 顧客維持、リコメンデーションサービス、マーケットバスケット分析 エリアマーケティングなど
 - ・ 通信サービス・・・ クロスセリング、解約分析など
 - ・ 金融サービス・・・ 顧客獲得、保険請求データ不正行為の発見と予測など
 - ・ 全般・その他・・・ 今後の売上げ予測、ダイレクトメールの返信率向上、成長傾向の予測、苦情分析による品質向上、優良顧客の発見～生涯価値判断、Web アクセスログの分析による Web サイト改善など

＝＝

製造業界の事例として、株式会社住友金属システムソリューションズをあげる。

株式会社住友金属システムソリューションズは、製品の検査不良率が悪化し、把握が困難であった工程でデータマイニングをおこなった。このときは不良要因分析を用いて検査をおこなった。その結果、不良の原因と無関係であると思われていた工程が不良の原因となっていることが発見された。そこで、歩留まり向上策として改善点を提案し、実行することで問題の改善をすることができた。

歩留まりの改善が必要な理由として、製造コストを押し上げる直接原因となっている製造不良を解消し、コスト削減を図ることと、増産に比例して増加する機会損失を最小限に食い止めるという2つの理由がある。

例えば、データマイニングが導入される以前では不良発生の要因の特定について、長く現場で作業をしている人の経験と勘に頼り、誰もが納得できるような理由や改善策を考えることができなかった。しかし、各工程でデータを取得し、分析していくことで、今まで経験と勘でしかわからなかった部分がデータやグラフになって視覚化され、問題が見えるようになった。そこで、問題に対して改善案を提案することで、問題の解消をおこなうことができる。

データマイニングを利用して、歩留まりが1改善されることでコストの削減ができる。このコストの削減により反対に利益として発生する場合も考えられる。

<http://www.setsunan.ac.jp/kubolabo/files/research/03/007142.pdf#search='%E3%83%87%E3%83%BC%E3%82%BF%E3%83%9E%E3%82%A4%E3%83%8B%E3%83%B3%E3%82%B0%20%E4%BE%8B'>

感想：上の事例でわかるように製造業界では、売上の向上や品質分野においての不良要因の取り除き、歩留まりの向上でもデータマイニングは利用されている。実際データマイニングはさまざまな分野においても使われているが、データマイニングの導入により、有益な情報を得るためにはデータベースの存在が大きいと思われる。これは、データベースを新たに構築したり、再構築するときに必要なデータの定義をおこなうときに経験の要素を含めることで、分析に必要なデータの収集や、最適な分析手法の選択をすること考えられるからである。

＝＝

●適用例の概要

「馬券をいかなる形でも良いので、データマイニングによる解析を行って儲けることができるかどうか確かめる」というもの。

・使ったデータ

JRA-VANより発売されているデータのうち、1997年～2001年6月までの中山競馬場開催分のデータを用いています（CD-ROMで書店に売っているもの）。

該当するデータ数は27,829件

・解析方法について

Cross-Validationを使う。

データ20,000件を決定樹によるモデル作成用データとし、残りのデータ7,829件をモデル

検証用データとする。

- ・ 結果について

「出力」×「単勝オッズ（但し単勝オッズは考えられる最低のオッズ）」が2以上である条件に適合するデータを抽出し、検証でも20以上あったものは次の通り。尚、2以上であっても中数が10未満だったものは除外している。

単勝オッズは15～35倍程度と、穴馬ばかりが出る結論になった。

個人的には馬体重は軽すぎず重すぎず、かつ枠は内目が良いということは何となくではあるが理解できるものの、その他の結論については、これがどういうことなのか現在のところ理由付けがし難い。

・・・という分析結果でした。

●情報提供元の URL http://www.h5.dion.ne.jp/~v_mt/mining01.html

●感想

自分は競馬をやらないので、詳しいことはわからないのだけれど、こういうことにデータマイニングをつかい分析をすることは、とても興味がわきやすいことだと感じました。

=====

バスケット分析の例

スーパーでの併買をバスケット分析で行った例を示す

● 牛肉（国産）との併買ベスト5

1	16.4	国産牛	佃煮、こんにゃく
2	15.6	国産牛	野菜類、こんにゃく
3	14.9	国産牛	牛乳、中華の素
4	14.8	国産牛	こんにゃく、(日曜日)
5	14.1	国産牛	豆腐、こんにゃく

● 鶏肉との併買ベスト5

1	30.9	鶏肉	唐揚げ粉
2	18.4	鶏肉	野菜類、こんにゃく
3	18.4	鶏肉	ベーコン、豚肉
4	17.8	鶏肉	佃煮、納豆、豚肉
5	17.8	鶏肉	牛乳、シチューの素

● パン粉との併買ベスト5

1	4.6	パン粉	サラダ油
2	4.0	パン粉	貝類
3	3.8	パン粉	豚肉、漬物
4	3.7	パン粉	豚肉、調味料
5	3.4	パン粉	豚肉、牛乳

● 牛肉（輸入）との併買ベスト5

1	9.0	輸入牛	焼肉のたれ
2	3.4	輸入牛	野菜類
3	3.1	輸入牛	カレールー
4	3.0	輸入牛	スープの素
5	2.8	輸入牛	香辛料

● 鮮魚との併買ベスト5

1	26.9	鮮魚	刺身
2	24.7	鮮魚	納豆、貝類
3	24.2	鮮魚	和惣菜
4	24.1	鮮魚	中華惣菜
5	23.7	鮮魚	魚卵

● サラダ油との併買ベスト5

1	24.9	サラダ油	醤油、(木曜日)
2	21.7	サラダ油	ティッシュ、(木曜日)
3	16.2	サラダ油	砂糖、(木曜日)
4	11.0	サラダ油	小麦粉
5	10.6	サラダ油	カップ麺、(木曜日)

● 豚肉との併買ベスト5

1	61.7	豚肉	牛乳、中華の素、調味料
2	59.5	豚肉	ハム、中華の素
3	58.4	豚肉	食パン、牛乳、中華の素
4	57.7	豚肉	野菜類、中華の素
5	56.7	豚肉	牛乳、中華の素

● 焼肉のたれとの併買ベスト5

1	13.9	焼肉たれ	その他精肉
2	9.2	焼肉たれ	輸入牛
3	4.0	焼肉たれ	ドレッシング
4	3.5	焼肉たれ	調味料、(日曜日)
5	3.5	焼肉たれ	国産牛

● 木曜日の併買ベスト5

1	14.6	サラダ油、醤油	木曜日
2	12.9	サラダ油、カップ麺	木曜日
3	12.8	サラダ油、ティッシュ	木曜日
4	12.5	ティッシュ、カップ麺	木曜日
5	11.7	ティッシュ、トイレット	木曜日

こういうデータをまとめたことによって、併買商品をわかり、よりお客さんへの対応がスムーズにできるようになる。まだ、価格の調整もやりやすくなる。売り上げに貢献。

木曜日の併買ベスト5には、サラダ油とティッシュがあまり関係ないのに一緒に買われるのは驚いた！

参考：「顧客分析とデータマイニングの動向」 Database Marketing and Data Mining
松田芳雄 http://www.unisys.co.jp/tec_info/tr68/6814.pdf

家電製品の価格データ（2001年1月時点）をマイニングした。一見何でもないような数字の並びであるデータをマイニングして知見・仮説を見つける、というもの。

電子レンジの価格は安いと感じる製品（メーカー）は三菱電機、東芝、三洋電機で、高いと感じるのはシャープ、松下。同一製品でも東京よりも大阪が安くなっているのは三洋電機で逆に東京が安くなっているのは東芝、三菱電機、シャープ。

冷蔵庫 400 リットル以上の価格は安いと感じるのは東芝、三菱電機で、高いと感じるのは松下冷機、日立。同一製品では大阪が安いのは三菱電機、日立で、東京が安いのは東芝、松下冷機。

全自動洗濯機の価格安く感じるのは三菱電機、日立、東芝、三洋電機で高く感じるのはシャープ、松下。総じて大阪が高くなっている。

エアコンの価格安く感じるのは三菱電機、高く感じるのは松下。総じて大阪が安くなっている。

また、これらの製品価格が三菱電機は安く松下は高く感じることから、この2つを価格比較している。その結果4つの仮説が立てられた。

仮説 戦略的に価格を設定している（松下は高く、三菱電機は安く）

1

仮説 松下製品は機能が豊富なので高い

2

仮説 三菱電機は技術力が高く、したがって安いコストで開発・製造できその結果安くなっている

3

仮説 松下は販売力、PR力があり、したがって高く設定している

4

・情報提供元 URL : http://www.datamining.jp/jirei/j007_frame.htm

・感想

ネットで具体例を調べましたが、分析によって今まで気付かなかったことが発見できれば分析のやり甲斐があるような気がしました。今まではあまり家電製品を購入したことがなかったので、購入する機会があればこの結果をどおりになっているかを確認したいと思います。また、実際に卒業研究をする中でデータマイニングをするならば「新たな発見」がたくさん見つけられればよいなと思いました。

=====

・適用例の概要

: クレジットカード

応用例として利用履歴を解析することにより、不正使用時に特徴的なパターンを見つけ出し、あやしい取引を検出するということも考えられる。

・情報提供元のURL

: IT用語辞典 e-WRODS

<http://e-words.jp/w/E38387E383BCE382BFE3839EE382A4E3838BE383B3E382B0.ht>

・感想

データマイニングは「お酒を買うお客は、雑誌も買う」や「雨の日は、練り製品が売れる」など、いろいろな項目間の相関関係を調べるものだが、それが、クレジットカードにも応用できるかもしれないと、知って驚きました。ただ、膨大な量のデータが必要になるので、実用化は難しいだろうと思いました。しかし、この考え方はその他の犯罪にも使えるかもしれないと思いました。犯罪の種類とその傾向がわかれば、犯人の割り出しにも使えると思います。ただ、過信できるものでもないと思うので、注意は必要ですが……。

=====

データマイニングを利用した具体例としてテキストマイニングを紹介します。テキストマイニングとは形式化されていないテキストデータ（通常は自然文）を単語などに分割し、その出現頻度や相関関係などをデータマイニングの手法を使って解析することで、一定の知見や発想を得る分析手法の総称です。従来、単語ごとに分かち書きする英語に比べて、日本語のテキストマイニングにおいては文章を単語ごとに切り分ける処理（形態素解析）が必要であることから“遅れている”とされていたが、近年技術的な進化により、ある程度ビジネスでも実用レベルに達するようになってきた。例えば、アンケートの自由回答、コールセンターやWebサイトなどに寄せられるユーザーからの定性情報、掲示板などへの書き込みなどを解析することによって、顧客や市場のニーズを抽出したり、自社製品への不満点を分析するといった使い方が期待されている。また電子メールによる顧客からの問い合わせに対して、これを自動的に解釈して返答したり、関係担当者へ転送するといったソリューションにも応用されています。

出典

情報マネジメント用語辞典

URL: <http://www.atmarkit.co.jp/aig/04biz/textmining.html>

感想

苦情のメールの文章から自社製品に対する不満を分類していけることに驚いた。私としてはメールの内容ごとに分類してそれが何件くらいきているかで分類していると思っていたので、今回調べてみてとても興味深かった。しかし、これによってメールを読む時間が短縮されてコストダウンを図れるのは分かるのだが、メールを出しても読まれないというのは消費者の立場から考えると少し納得がいかないなと感じた。

=====

・ 適用例の概要

とある百貨店の、ハウスカード全会員のプロフィール、販売データ、DM 反応データを統合する DWH を構築し、有料顧客の行動をデータマイニングを利用して分析してキャンペーンに活用することで、ダイレクトメールの反応率が上昇し、DM とキャンペーンコストの削減と売り上げ増につながった。

・ 参考 URL

<http://www-06.ibm.com/jp/software/data/bi/solution/datamining.html>

・ 感想

一般にはあまり知られていない技術であると同時に、適用するものによっては大きな費用対効果を秘めているものだと思った。まったく関連性のなさそうに思えるもの同士に何らかの関連を見つけ出すのは非常に面白いものであると思われる。

=====

ヤフー、データ・マイニングを活用してオークションの不正利用を防止

適用例の概要

ネット・オークション「Yahoo!オークション」の不正利用対策に不正利用検知モデルの導入。不正利用検知モデルは、データ・マイニングの手法を活用して、オークションの利用履歴を分析し、不正利用をパターン化する。不正利用のパターンは、データベースにまとめ、ID 登録や出品といった時点で、データベースと照合する。不正行為のパターンと合致する行為が検出されたら、不正行為を監視するパトロール・チームに通知し、不正と判断した場合に、ID の利用を停止したり、出品を削除したりする。

モデルを構築するに当たり、約 1 億円を投資する。

情報提供元の URL

<http://itpro.nikkeibp.co.jp/article/NEWS/20051114/224528/>

感想

古いニュースだが、データマイニングを犯罪プロファイリングとビジネスの両方に利用しているよい例だと思った。ヤフーオークションに参加していると盗品販売や架空出品だと思われる怪しいものは同じようなジャンルでたくさん見受けられたり、オークション詐欺被害者が利用する掲示板などをみても同じ人間や同じ手口で何人もの人が被害にあっているようなので、投資金額の 1 億円が高いのかどうなのかはよくわからないが不正対策としてはいいのではないかと思った。

=====

下の動物たちは、ある百貨店の顧客である。いつも苦情を言ってくる顧客もいれば、素直で友好的な関係を築いている得意先もいる。まさに十人十色、巨大な動物園の中のようなのである。

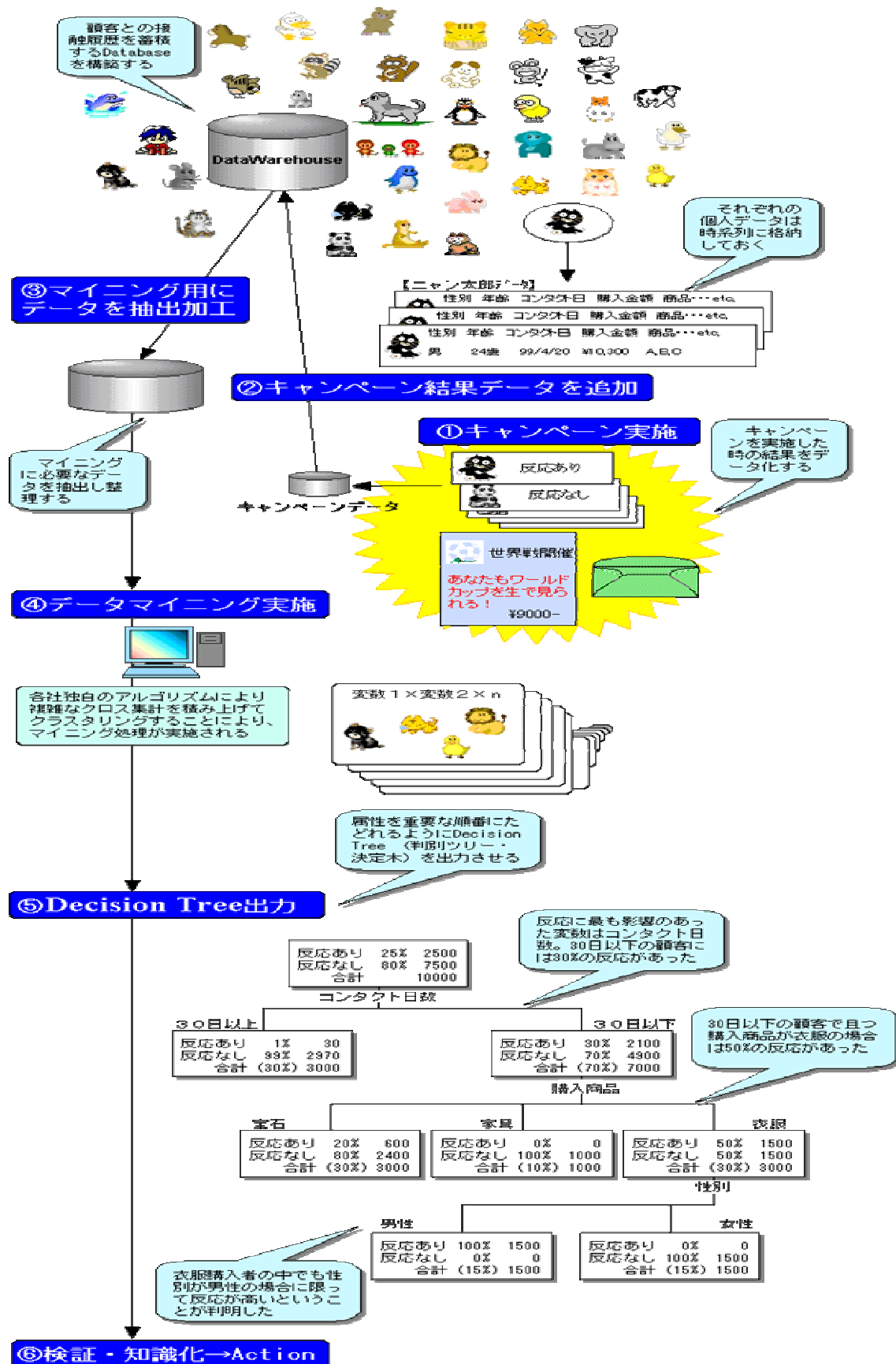
この百貨店では、夏に行われるワールドカップサッカーを利用して既存の顧客にチケットを販売するキャンペーンを実施した。運良く社内のデータウェアハウスには顧客の住所も保存されていたので、とりあえず1万人にダイレクトメールを発送し、全体の25%にあたる顧客から申込みをもらった。担当者はまずまずの結果であったと満足している。

そこで、秋に行われるサッカーの試合についてもキャンペーンを行おうと思っていた矢先、社長から「こんなに費用をかけて25%ならやめてしまえ！」と予期せぬ叱りを受けてしまったのである。

そんなとき、担当者の知り合いから、「データマイニングを利用して、前回の25%を分析してみたらどうだ」と提案されたのである。

そして、担当者はデータマイニングを実行して見ることになった。





担当者がデータマイニングを実施した結果、今回の反応に一番関係が深かったのは、

顧客とのコンタクト日数であることが分かった。

そして、次ページの下の図が Decision Tree というものである。特徴を発見するために利用されるデータマイニングの手法の1つである。これをみると、全体の25%の人たちの特徴がよくわかる。

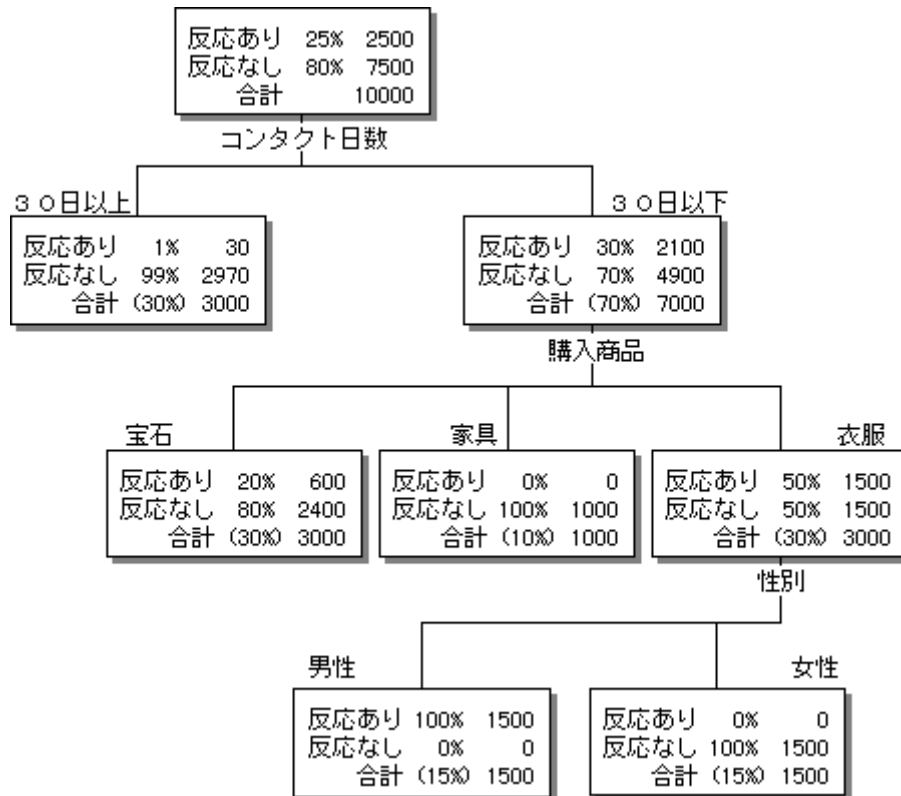
ツリーの下に判別効果をあらわすグラフがある。これは、反応のあった顧客をベースに、判別効果の高い順番に Node (分岐点) を並べて、どれくらいセグメントすれば何%の効果を得られるかを見たものである。

赤い線は、対象10000件を使ってアプローチした場合の理論線である。つまり、10000件を無作為に処理すれば該当者は直線的に積み上がっていくわけである。青い線は、今回のデータマイニングによって得られた Node を利用した場合のもので、この場合、Node5 の段階で90% ($2500 \times 0.9 = 2250$) をつかまえられるという事である。

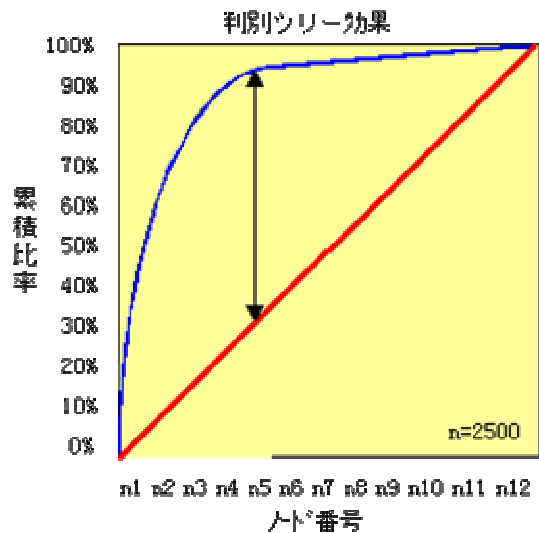
つまり、仮にこの Node5 までに該当する顧客が全体の50% ($10000 \times 0.5 = 5000$) だった場合、半分の経費で90%レベルの効果が得られる事になる。したがって、Node5 までに該当する顧客にアプローチすれば良い。それだけでなく、この情報は新規の顧客にも反映させることができるので、他の顧客情報に当てはめて見込度の大きい顧客を優先的にアプローチできるのである。

データマイニングは1つの意志決定支援ツールなので、この結果だけに頼ってはいけない。次回のアプローチでの検証も必要であり、残りの50%に対するアプローチをどうするのかも考える必要がある。単純に可能性の高い50%の対象者だけのアプローチで終わらせると、他の対象者の受注の可能性は0%になってしまう。今回の場合、可能性の高い対象者に対しては前回と同じアプローチを使い、それ以外の対象者には他の手段を使ってアプローチをするマーケティングも考えられるのである。大事なことは受注率を上げるだけでなく、契約しなかった顧客の原因分析を行い、その顧客へのコミュニケーション戦略を別に練っていかなくてはならない。データマイニングを使ったアクションプログラムは、通常トライ&エラーを繰り返しながら正しい法則を導き出していくのである。

【Decision Tree】



【グラフ】



参照URL

「データマイニングって何」

http://www2u.biglobe.ne.jp/~kihachi/kdd/kdd_00.htm

感想

データマイニングという言葉は以前から聞いたことがあったが、詳しい意味は知らなかった。しかし、授業を聞くことでなんとなくわかり、さらにレポートとして具体例を考えることによって理解を深められたと思う。

データマイニングは、マーケティングの分野はもちろん、物流や交通などの様々な分野で使える手法であると思ったので、より理解を深めてこれからは利用できればいいと思う。

=====

1. 流通サービス業：マーケットに関する情報を採取

- ・販売予測

天気や催し物、過去の実績といったデータから商品ごとの特性や購買数を分析したうえで、仕入れ量を決定する。

- ・販売分析

顧客（個客）の購買パターンや「紙おむつを買った顧客はビールを同時に買う」というような隠れた併売パターンを見つけ、収益性をより高める。

2. 製造業：コストや品質管理に関わる原因を分析

- ・品質管理

製造工程の不良原因を分析し、不良多発原因を排除する。

3. 金融、保険業：リスクや顧客特性の情報を採取

- ・リスク分析

過去の実績や契約者の属性データから、契約のリスクを分析し、上限値等を決定する。

- ・特性分析

顧客属性と契約商品の関係を分析し、顧客ごとに効果のある商品を紹介する。また、顧客離れを防止する。

4. 医療、バイオ産業：要因や効果に関する情報を採取

- ・要因分析：病気の原因を分析したり、薬の改善効果を調査する

※参照：http://www5.ocn.ne.jp/~shinya91/csm/12csm_tenbo.html

=====

適用例； am/pm にて 発注制度の改善

コンビニの主力商品は、おにぎりや、お弁当など購買頻度の高い商品である。これらは、

一日に複数回納品される商品である。しかし、品切れによる機会損失や在庫過多による廃棄ロスも毎日発生しているはずである。これを改善すれば売上と粗利益高も改善できる。

しかし、適切な発注量を決めることは難しい。本部では店舗に対して発注量を最適化するための多くの情報を提供しようとするが、立地環境や顧客層などが店により違うことで販売動向も変わってくるので適切な発注数量を求めるとなると個店別のデータ分析が必要となる。

そのようなことからコンビニ各社では、データマイニングと呼ばれる大量に蓄積されたデータの中から、その中に潜む項目間の相関関係やパターンなどを探し出す技術として、am/pmだけでなく、多くのコンビニで取り組まれている。

このデータマイニングを使って、以下のような分析を行うことができる。

- ・ 顧客属性や購買履歴から店舗タイプを算出できる
- ・ 店舗タイプに基づき店舗の売上、利益を予測できる
- ・ 商品別の販売予測を行う
- ・ 販売予測を元に発注数量の適切さを回答してくれる

情報源；小売業の業務改善コンサルタント URL <http://opedesign.exblog.jp/>

この記事では、コンビニ各社にて、このデータマイニングが使われており、発注量を予測することができるだけでなく、商品戦略や出店戦略にも使えると述べている。コンビニなどは商品などの量が多く、データも膨大になると思われる。それらをすべて分析し、店舗タイプ、販売予測・発注量などを出すことができるのであれば、力強いシステムが出来ると思う。

=====

リクルートで発行されている R25 は基本的に、駅前におかれているが、近日発売された女性向け L25 は、様々なところにおかれている。これはマーケティングの調査結果、男性と違い、女性は帰宅後の生活様式が非常に多様化していることが分かったからだ。

主に、実際お稽古事意、エステ、百貨店等女性の集まる場所に置かれている。

- ・ 情報提供元の URL

<http://curry.livedoor.biz/archives/50687102.html>

- ・ 感想

世界 GDP TOP10 の対マーケティングの平均割合は 0.7%であるが、日本はその中で 0.3%しか使用していない。この数値から分かるように、日本のマーケティング市場はまだ未成熟だと考えられる。実際、バンダイなどの会社の新製品の成功率は 1%だと言われている。また、今回の例であげたフリペーパーを発行しているリクルートだが、ネットの顧客をモニターとして使う、日本で最も優れた企業、マクミルは、リクルート出身の人たちで作られた会社だ。これらを考えた上でも、これから日本は、マーケット市場が活性化すると考えられる。

=====

適用例

⇒ コーヒー豆のブルーマウンテンを買う客は高額な輸入商品も併せて買う

情報提供元の URL

⇒ http://www.nikkei-bookdirect.com/science/page/magazine/0205/sp_2.html

(機械はアインシュタインになれるか：日経サイエンス)

感想

⇒ この適用例はコーヒーと輸入商品という異色商品を一緒に買うことが多いと結果付けられていることがユニークであると思う。更にコーヒーは銘柄がブルーマウンテンに、輸入商品は高額なものと、適用範囲がかなり狭められていることが分析の細かさを汲み取ることが出来ると思う。

また、今回の情報提供元は、雑誌に連載された記事の一部であると推測される。このなかで『ブルーマウンテンを大幅に値引きして客を集め、高額商品の販売増につなげるといった巧妙な販促手法も登場した。』と、この分析によって販売方法を工夫させられているところが、この分析における精度の高さを表しているのではないかと思う。

=====

POSデータに表れる文化の違い（北海道と沖縄の例）

北海道のスーパーで、生魚の分類で一番売れるのは「魚卵」である。特に8月、9月は、日によっては生魚の総売り上げの8割にまで、魚卵、特に生イクラが売れる。逆に、沖縄ではサンマと鮭が生魚分類の売り上げトップである。

沖縄の米の販売額について、日単位の米の売上額を見ていると、年末とお盆に明らかな売り上げ低下期がある。通常期の半分くらいまで米の売り上げが落ち、それが1ヶ月くらいかけて回復するという、妙なグラフを描く。それは沖縄では米の分類が2つあることが原因である。普通の米と「米ギフト」である。米ギフトの集計を取ると、一般米の売り上げが低下する直前に、急激なピークを描いている。その理由は、沖縄では知人宅を訪問するときの手土産として「米」を持っていくことが多く、そのためにお米のギフトパッケージまで売られているということである。だから、「米ギフト」という分類があるわけである。つまり、盆暮れにお土産でもらった米が自宅に在庫し、それが消費されるまでスーパーで米を買わなくなるということである。

情報提供元の URL

[http://www.venture.nict.go.jp/ezp/index.php/venture/content/view/full/119/\(offset\)/10](http://www.venture.nict.go.jp/ezp/index.php/venture/content/view/full/119/(offset)/10)

感想

こういった文化を、私はまったく知らなかった。2ヶ月もの間、8割も魚卵だけで占めるとは本当に驚く。メーカー側として、こういった文化を知らずにいたら、きっと

このPOSデータに戸惑うことになるのだろう。

今回調べたことはデータマイニングではなく、POSのデータを単に集計するだけでわかることかもしれない。ただ、POSデータから、普通に生活するうえでは知ることのできない情報を知ることができるということに興味を持った。

=====

・適用例の概要

データマイニングの適用例

アメリカのスーパーマーケットが実際にPOSデータに対して、データマイニングを適用し、抽出したルールとして以下のようなものがある・

～「おむつを買いに来る客はビールも買っていく」

～「チェンソーを購入した客はその後でマットレスを買いに来る」

これらの情報を考慮し、店内の棚の再配置を行うことによって、効果をもたらす。

・情報提供元の URL

http://gc.sfc.keio.ac.jp/class/2005_14283/slides/07/19.html

・感想

データマイニングはさまざまな業界で取り込まれている。流通サービス業におけるバスケット分析や顧客セグメント分析、金融業界におけるポートフォリオ分析やリスク分析、通信分野におけるアクセスログ分析などがある。

大型システムの構築によりに企業や政府機関に多量のデータが蓄積されるようになったことや、さらには、インターネットの普及にともない誰もが利用可能なデータが爆発的に増大したことによって、人手をかけた検索技法のみでは、多量なデータから意味ある情報を取り出すことは困難である。よって、こういった技法によって多量のデータから価値ある情報を掘り出せるといったことが分かった。

=====

・適用例の概要

具体的適用例について書かれていること、

『「地震予知はできない」と割り切るのはまだまだ早い。本チャレンジは、電子的に蓄えられた地震活動の履歴データだけから計算機を用いて未来の地震の情報を得る「地震データマイニング」の手法を人工知能から打ち出すことをめざす。チャレンジの目標としては危険区域を推定することに限定して短期決着をめざすが、伏在断層の発見や地震学研究者との議論の支援など、人工知能の研究者が力を発揮できる研究は適宜含める計画である。

われわれのチャレンジが成功し、要注意活断層の位置を絞り込むことができれば、廉価かつ早期に危険な地域を知ることができるようになるだろう』

と、HP で記載されている時点で実験段階のようだがデータマイニングを利用することで地震の予知をできないかという研究がされているようだ。

- ・ **感想**

本文にもあるように、HP に記載されている時点では研究段階のようだが活断層をはじめさまざまな地層を調査し、データマイニングを利用することで「地震の予知」を可能にしようとしている。これ実現できれば素晴らしいと思う。

しかし、このHP 自体が6年前のものなので今はどの程度まで研究されているのか…。もしかしたら研究そのものが今行われているのかも…どうなのか。

- ・ **情報提供元の URL**

<http://ultimavi.arc.net.my/ave/CREP/result99.html>

=====