樹状モデル（決定木）とは？

株式会社フュージョンシス
http://fusionsys.com/
2005 年 6 月

樹状モデル（決定木）は、データの中にあるパターンや構造を抽出するための技術です(CHAID, CART, C4.5 などの手法が有名です)。あらかじめ何らかのモデルや仮説を用意しなくてもよいのが特徴です。樹状モデルは、データの分類、パターン認識、予測に使われます。

樹状モデル（決定木）の考え方の近いものに線形回帰があります。ガソリン消費量と気筒数、車体重量、車体の色の明るさの間に成り立つ関係を調べたいとします。多くのサンプルがあれば、

ガソリン消費量＝A＊気筒数　＋　B＊車体重量　＋　C＊車体の色の明るさ

という式の中の係数 A,B,C を決めることができます。気筒数や車体の重量はガソリンの消費量に密接に関係してくると考えられる一方、車体の色の明るさは、ガソリン消費量に関係してきません。したがって係数は

ガソリン消費量＝2＊気筒数　＋　3＊車体重量　＋　0.1＊車体の色の明るさ

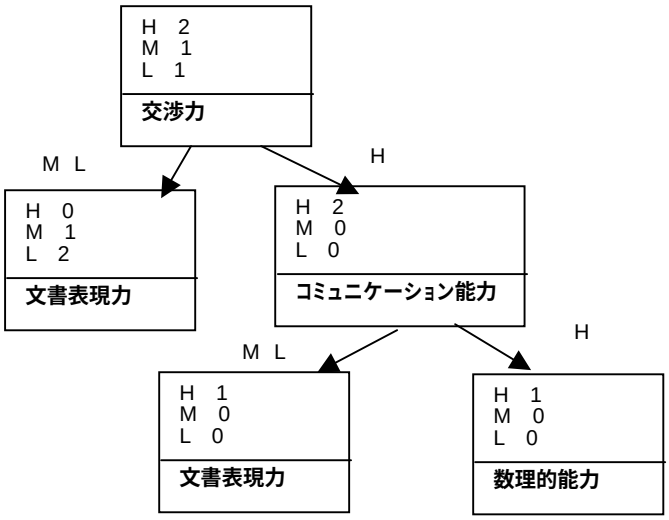
のような感じになるでしょう。係数をみることによって、車体重量や気筒数は、ガソリン消費量と関係が深いことを知ることができます。

決定木は、このような簡単な式で書けそうもない現象への拡張です。

社員	上司の評価	数理的能力	コミュニケーション能力	交渉力	文章表現力
田中	M	3	1	2	3
佐藤	L	2	2	1	1
鈴木	H	1	3	3	2
山田	H	2	1	3	1
佐々木	L	2	1	1	1

上司の評価は1年後に上司が3段階で決め、他の能力は入社時にテストで計ったものであるとします。上司の評価と、数理的能力、コミュニケーション能力、交渉力、文章表現力の中で関係があるものを知りたいとします。

人間の能力などは、線形でないと考えられるので、樹状モデルを使うことが適当だと考えられます。



上のような木ができたとすると、上司の評価と交渉力がもっとも関係が強いということがいえます（ただし、CHAID 以外の方法では、別の解釈になります）。この木は下の方にのばしていくことができます。

上の木はどのような項目（数理的能力、コミュニケーション能力、交渉力、文章表現力）で高得点をとれば、上司の評価が高くなる可能性が高くなるかといったルールを述べているとも考えられます。

さて、次に新入社員、高橋が入ってきたとしましょう。

社員	上司の評価	数理的能力	コミュニケーション能力	交渉力	文章表現力
田中	M	3	1	2	3
佐藤	L	2	2	1	1
鈴木	H	1	3	3	2
山田	H	2	1	3	1
佐々木	L	2	1	1	1
高橋	??????	1	2	3	3

高橋にはまだ上司の評価はありませんが、入社時にテストによって数理的能力、コミュニケーション能力、交渉力、文章表現力は分かっています。高橋が1年後にどのような上司の評価を売るかどうかは、上の木を使って予測できます。線形回帰をつかって予測ができるように、樹状モデル（決定木）も予測に使うことができます。

弊社では樹状モデル（決定木）を基にしたアプリケーションを昨年開発し、マーケティングの会社に納品させて頂きました。このアプリケーションでは、上で述べたような樹状モデルの予測能力を使って新入社員の評価を予測します。樹状モデルは広範な応用の範囲を持っていますが、様々なご要望に応えてカスタマイズを施し、お客様にさらに高い満足感をもたらすことができると確信しております。

■