

相関ルールデータマイニングにGPを応用したグローバル経営 人材育成のための海外派遣適性評価システム

松村幸輝† 小谷悠太†† 白木三秀†††
東京福祉大学† 鳥取大学†† 早稲田大学†††

1. はじめに

日本企業の海外展開の活発化に伴って、現地法人への人材派遣を成功させることが課題となっている。そのアプローチとして、海外派遣で必要となる要因を分析し明確にすることで、より効率のよい派遣人材育成を可能にすると考えられる [1]。そのため海外派遣者に必要な能力や資質について種々検討されているが、その詳細は未だ不明な点も多い [2]。

これに基づき本研究は、海外派遣者の能力・資質を明確にするとともに派遣要員としての適性度を評価することを目的として、データマイニングの代表的手法である相関ルールによって決定木を実装し、これに遺伝的プログラミング (GP) を応用することによって最適化する新しい方法を提案する [3]。

本報告は、決定木の精度を向上させるために決定木作成方法の改善について検討した結果を報告する。

2. 決定木作成方法

2.1 派遣者へのアンケート調査

海外派遣要員として求められる能力や要件を調査するために、現在海外派遣されている駐在員にアンケートを採取した。調査地域は中国、ASEAN、インド、ヨーロッパである。アンケートで実施した質問は業務遂行能力やパーソナリティを問う 64 項目と派遣の達成度を問う質問 7 項目からなる。

2.2 相関ルール

決定木の非終端子として相関ルールにより抽出したルールを用いる。実験では相関ルール分析はトランザクションデータの中から、各質問項目と達成度の相関を分析するために利用する。

相関ルールは $\{1-4, 13-5\} \Rightarrow \{3\}$ で表す。論理包含を表す右矢印の左側が条件部で、右側が結論部である。条件部の「 $\{1-4, 13-5\}$ 」は「質問番号 1 の回答値が 4」かつ「同 13 の同 4」であることを表し、結論部の「3」は達成度が 3 であることを意味する。全ての相関ルールの組み合わせを探索するのは難しいため、バックトラック法を用いて支持度、信頼度、リフト値に閾値を設定しルールの抽出を行った。

支持度、信頼度、リフト値は以下の 3 つの式で与えられる。X が条件部のアイテム集合、Y が帰結部のアイテム集合、|D| がデータベース中の全トランザクション数を表す。

$$support(X \Rightarrow Y) = \frac{count(X \cup Y)}{|D|} \quad (1)$$

$$confidence(X \Rightarrow Y) = \frac{count(X \cup Y)}{count(X)} \quad (2)$$

$$lift(X \Rightarrow Y) = \frac{confidence(X \Rightarrow Y)}{support(Y)} \quad (3)$$

2.3 決定木

決定木は一定の規則により自動的にデータの分類が可能であり、分析結果の解釈が容易であるため本目的に適した構造である。図 1 に決定木の例を示す。

決定木の非終端子に抽出した相関ルールを、終端子に達成度の値をそれぞれ設定することでアンケートデータの分類と派遣候補者の適性評価が可能な決定木を得ることができる。

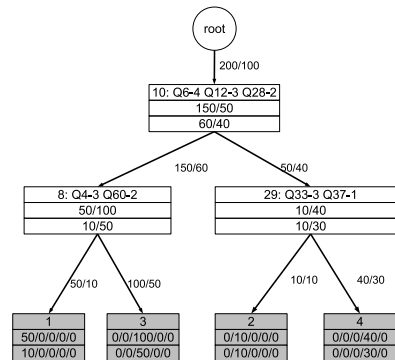


図 1 決定木の例

2.4 遺伝的プログラミング (GP)

決定木の最適化は GP により行う。抽出した相関ルールと達成度の値をランダムに設定した決定木を集団として、遺伝操作 (交叉、突然変異等) を繰り返すことでデータを分類する木を最適化する。

3. 評価ツールの実装

決定木の生成は、相関ルールを達成度ごとに抽出してそのルールを用いて非終端子と終端子を設定し、GP により進化 (成長と最適化) させる。

決定木は達成度 4 のデータとそれ以外のデータによる 2 値分類の木を作成した後、達成度 1, 2, 3, 5 の 4 段階により分類を行う。この操作はそれぞれの場合で 50 回行っている。実験に用いたデータは日本人派遣者からとったアンケートデータ 866 個である。

Expat Assessment System for Global Management
Personnel Training Utilizing Association-Rules-
Data-Mining Applied by GP

† Tokyo University of Social Welfare
†† Tottori University
††† Waseda University

表 1 GP のパラメータ

個体数	500
世代数 (2 段階/4 段階)	100/300
エリート保存率	5%
交叉率 (2 段階/4 段階)	80%/95%
突然変異率 (2 段階/4 段階)	10%/20%

GP の実験条件を表 1 に示す．非終端子には抽出した 300 個の相関ルールを使用している．またデータを 2 : 1 の比率で訓練データとテストデータに分け，最適化を訓練データで行い，テストデータで精度の検証を行った．

評価関数は誤分類率として以下の式 (4) を用いた． F はデータの誤分類数， ND はデータ数をそれぞれ表す．式 E の値が小さいほど優秀な木である．

$$E = \frac{F}{ND} \quad (4)$$

4. 結果と考察

表 2 に，非終端子にアンケートの質問項目を用いた方法と，相関ルールを用いた本手法との精度の比較結果を示す．50 回実験を行い誤分類率の平均値を，帰無仮説「平均値に差はない」として t-検定を行ったところ，P 値が 0.01 を下回り信頼度 99% で有意差ありという結果となった．相関ルールを用いた手法の方が良い結果であることが分かった．

決定木の終端子をアンケートと同じ達成度 5 段階で分類を行う方法で実験を行うと，作成された最良木の終端子の多くが達成度 4 または達成度 3 となり，分類精度が向上しない．

そこで，5 段階の分類を行うために，達成度 4 とそれ以外の 2 値分類を行う木と達成度 1,2,3,5 の 4 値分類の木を作成した．図 2 に 2 値分類の木と，達成度 4 段階の木の最適化の過程を示す．どちらも世代を経るごとに最適化されていることが分かる．

達成度 5 段階の実験結果と達成度 4 とそれ以外の実験結果を表 3 に示す．表 3 の適応度は GP による訓練時における誤分類率であり，誤分類率はテスト時における精度である．2 値分類で実験を行っても達成度 5 段階の場合と精度が変化していないことが

表 2 誤分類率の t-検定

誤分類率	平均値	標準偏差	P 値
相関ルール有	0.467	0.007	6.19E-03
相関ルール無	0.473	0.015	

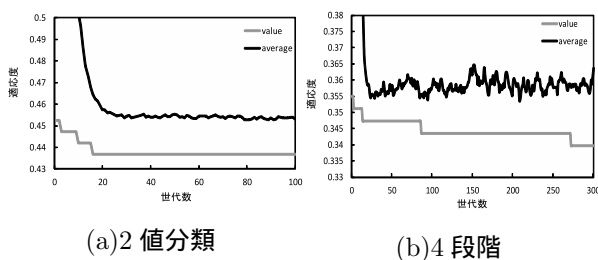


図 2 決定木の最適化過程

表 3 5 段階と 2 値分類の結果

実験条件	5 段階		2 値分類	
	適応度	誤分類率	適応度	誤分類率
平均	0.442	0.463	0.434	0.461
標準偏差	0.004	0.005	0.004	0.016

表 4 4 段階の結果

精度	適応度	誤分類率
平均	0.344	0.347
標準偏差	0.004	0.005

分かる．このことから達成度 4 のデータは分類が難しいことが分かる．

次に達成度 4 のデータを除き，終端子を 1,2,3,5 の 4 段階で分類を行った場合の結果を表 4 に示す．これまでで最も良い精度となっており，また 5 段階評価，2 値分類の結果と比較して訓練データとテストデータの分類精度に差がないこともわかる．終端子も 5 段階の結果ほどの偏りは出なかった．4 段階の決定木はこれまでと比較して最も良い木であった．

派遣者から採取したアンケートデータを用いた実験結果から，海外派遣候補者の適性を判断する場合，まず達成度 4 の木とそれ以外の木で適性評価を行った後，適性が 4 ではないと分類された場合に達成度 1,2,3,5 の木により適性を決定する方法がより高精度であることがわかった．また，作成した決定木を分析した結果，特定の達成度の分類を行うには，その達成度で相関を取ったルールが必要であることが分かった．

このようにして全データから必要な能力・資質について検討した結果，アンケートの質問 64 項目の中で，分類に重要なものとして「戦略立案ができる」，「数字分析が強い」などの項目が重要であることが示唆された．

5. おわりに

本研究では，海外派遣者から採取したアンケートを用いて決定木の作成を行い，派遣候補者の適性評価を行うシステムの作成を行った．

新たな分析手法として相関ルールを決定木の非終端子に用いて GP による決定木の最適化を行った結果，従来の手法である非終端子に個別の質問項目を用いる方法よりも安定した最適化を行うことが可能となった．さらに，従来の手法では難しかった 5 段階での評価をより行いやすい分類手法であることがわかった．これによって，派遣候補者にとって重要と考えられる要件についても分析することができた．

参考文献

- [1] 労働政策研究・研修機構：第 7 回海外派遣勤務者の職業と生活に関する調査結果 (2008)
- [2] 白木：日本企業のグローバル化とグローバル人材マネジメント日本在外企業協会のアンケート調査からの考察，グローバル経済 2009 年 6 月号，pp.4-11(2009)
- [3] 松村他：“進化型計算手法を用いたグローバル経営人材育成のための適性評価システム”，経営情報学会誌，Vol.19,No.2,pp121-145(2010).