

D\_046

駅社員向け情報配信システムのための情報フィルタリングルールの獲得  
Acquisition of information filtering rules for information delivery system for station staffs

野末 道子†      尾崎 尚也†      土屋 隆司†  
Michiko Nozue      Naoya Ozaki      Ryuji Tsuchiya

## 1. まえがき

鉄道のダイヤ乱れ時において、時々刻々と発生する列車の運転整理情報（列車の運休や順序変更、臨時列車運行など、列車運行計画のリスケジュールリングに関する情報）や列車遅延情報をどのように駅社員に対して伝達するかが課題となっている。駅社員は、伝達される膨大な情報の中から、各駅、各場面において、必要な情報を抽出して、旅客に対して的確な案内を行うことが求められる。しかし、この作業の巧拙は駅員の経験、技量に依存する部分が大きく、必ずしもスムーズに行われていないのが実情である。そこで、本研究では、各駅の特性や到着列車・後続列車の列車種別、各列車の停車駅情報等に基づいて、配信情報の重要度を決定し、適切にフィルタリングするための一般的なルールを獲得することをめざす。これにより、ダイヤ乱れ時における駅社員の案内業務を的確に支援するツールの構築が可能になると考える。

このような情報フィルタリングルール（情報の重要度判定基準）は、専門家（ベテラン駅員）の経験知識として暗黙裡に存在するものであり、さまざまな駅、場面に適用可能な一般化されたルールとして獲得するのが難しいという課題があった。また、抽出する重要度判定基準に関しては、高い可読性も要求される。情報の重要度判定の基準は、支援システムにとって最も重要な要件であり、当該システムの導入・運用・利用を行なう関係者間で、あらかじめ合意形成を行なっておく必要があるからである。

本研究では、上記の課題を解決するため、決定木学習を用いた重要度判定基準の獲得手法について提案する。一般に、獲得された決定木による判定性能は、考慮される属性集合に大きく依存する。しかし、本課題に関しては、属性集合は自明ではなく、情報の重要度決定にあたり、そもそもどのような属性を考慮すべきか、を決定することが最初の課題となる。また、可読性の高い重要度判定基準を構築するために、最終的にはできるだけ小さな決定木を獲得することも同時に求められる。

我々は、属性抽出段階では、複雑に繁る木構造が生成されるものの、初期段階で獲得された属性

の妥当性評価に適した木を出力することが可能であった random tree 学習と専門家へのヒアリングを組み合わせた手法を適用、また、最終的な決定木生成段階（システムに組み込まれる重要度判定基準の獲得フェーズ）では、コンパクトで可読性の高い決定木が生成される C4.5 による学習を用いた手法を適用し、上記の要件を満たす結果を得たので報告する。

## 2. 決定木学習における属性の抽出

本研究では、ダイヤ乱れ時において駅社員が受け取る各情報に対し、重要度を付与し、その重要度に応じた表示・非表示制御あるいは表示方法制御を行なうものである。情報の重要度は「最重要」「重要」「普通」「不要」の4段階の離散値を取るものと仮定した。具体的には「最重要」と判断された情報は強調表示で提示される一方、「不要」と判断された情報は、システム画面上は非表示となる、といったような表示制御が行なわれる。

本研究で取り扱う上記のような分類問題に関する課題として挙げられるのは、主に以下の二点である。

- 1) 与えられた課題に対する分類知識（属性集合）が自明ではない
- 2) 与えられた情報に対する専門家間の判断に揺れがある

これらは、特に本研究で扱う運転整理情報配信に固有の課題ではなく、各種の研究事例がある<sup>1)</sup>。鳥居らは、1)の課題に対し参加型モデリングのモデリングプロセスにおいて分類学習を組み込むことで、ロールプレイングゲーム(RPG)ログから、参加者の意思決定を一般的に説明する分類知識を獲得する方法を示した。分類知識を獲得する過程で、決定木(C4.5)から得られた情報を専門家に示し妥当な知識であるかどうかを判定させている<sup>2)</sup>。

本研究においても、鳥居らのRPG同様に、事前に実施した駅社員とのディスカッションから、重要度判定に関わる幾つかの属性を獲得した。しかし、それらの属性集合を用いた重要度判定ルールを組み込んだ情報配信システムによる伝達情報を専門家に評価してもらった結果、本来考慮すべ

†(財)鉄道総合技術研究所

きいくつかの属性が不足していることが判明した。一例としては、ある駅での調査において「下り列車ホームでの案内の場合には、〇〇駅（主要ターミナル駅のひとつ）までの情報があれば十分であり、〇〇駅以遠の情報は〇〇駅で最新情報を得るよう案内すればよい」という意見が出された。これは、多くの旅客が移動する駅までの情報が最優先で欲しい、と解釈できるものであり、「旅客の流動パターン」や「ターミナル駅情報」が新たに考慮すべき属性として抽出された。これらは、最初に属性集合を獲得するために実施したヒアリングの際には獲得できなかった属性である。このような属性の抽出のためには、専門家に対して一般的な重要度判断基準を提示し、可否を問うだけでは不十分であり、個別具体的な事例に沿ってのヒアリングが不可欠である。実際、重要度判定基準に関して一般論を議論している際には指摘されなかった条件や属性が、個別の事例に関する議論の中で指摘されるケースが多く見られた。

以上を踏まえて、我々は、鳥居らの研究とは逆に、小さな属性集合から出発し、専門家へのヒアリングを通して新たな属性を漸次獲得していくアプローチを取った。具体的には以下の通りである。

- (1) 専門家へのヒアリングとその結果に基づく決定木構築を繰り返し、相互にフィードバックをかける。
- (2) 個別具体的な事例に関する判断の自然なモデル化が可能であった random tree 学習を用いて決定木を構築する (C4.5 などでは、決定木の抽象度が高くなり過ぎ、専門家による属性情報取得のための判断情報 (未知の属性) の獲得が困難となるため)

また、与えられた専門家間の判断には揺れがあり、基準となる判定結果を得る必要が生じた。そこで、判断の揺れを解消するために、それぞれの判定結果を相手に見せ、議論を行う中で判定結果を集約する方法をとった。以下、具体的な実験のプロセスを示す。

### 3. 属性抽出実験

本実験においては、最初の属性抽出プロセスでは、自分のいるホームに到着する直近10本の列車に関連して配信された仮想の運転整理情報、列車運行状況 (遅延情報) に対して、各情報の重要度判定を行ってもらった。調査の被験者ならびに調査方法は表1の通りである。

表1 重要度判定調査

調査被験者	駅における旅客案内を経験したことのある社員 (30代前半一名、30代後半一名)
調査方法	ある駅のホームに立つ駅社員であると想定して、表2に示すような各情報に対する重要度の判定を1~4の4段階で行う。判定は、端末画面ではなく、印刷物を見て実施する。

表2の情報は一行毎が、被験者が判定の対象とする一件の情報である。ID 2、ID 3のように、同じ列車に複数の情報が出るものもある。

「ID2: 935 Mは現在遅延していない」、「ID3: 935 Mは西明石で列車番号5073の後に順序変更となる」はそれぞれで単独で判定を行う情報である。

判定のランクである4段階のうち、段階1は非表示となる最も重要度が低い情報であり、4が最重要である。駅社員の情報表示画面には「通常表示」「詳細表示」の2種類があり、非表示の情報は「通常表示」では閲覧することはできず、「詳細表示」に切り替えて閲覧する。

本実験において、PDA、PC等端末画面を見て判定を実施しなかったが、その理由は以下の通りである。一点目は、PDA画面などを見て判断する場合には、画面の大きさ、見栄え、フォントサイズなどの情報が、判定に対する心理的な影響要因となり、個別の情報の重要度のみを判定することが難しくなる点がある。もう一点は、紙の情報であれば、被験者が自由に書き込みなどを行うことができることから、判断根拠となる属性を抽出

表2 判定した情報の例

ID	列車番号	種別	行き先	現在駅	伝達情報	着	発	番線
1	209C	普通	西明石	塩屋	神戸-西明石:運休	17:43	17:44	3
2	935M	快速	網干	西明石	遅延なし	17:43	17:44	4
3	→				西明石:5073の後へ			
4	→				大久保:793Tの後へ			
5	3483M	新快速	姫路	明石	遅延なし	17:50	17:51	4
6	→				大久保:795Tの後へ			

しやすいという点が挙げられる。判定に際しては特に時間制限は設けず、見直し等も自由に行ってもらった。判定実験の流れについて、図1に示す。

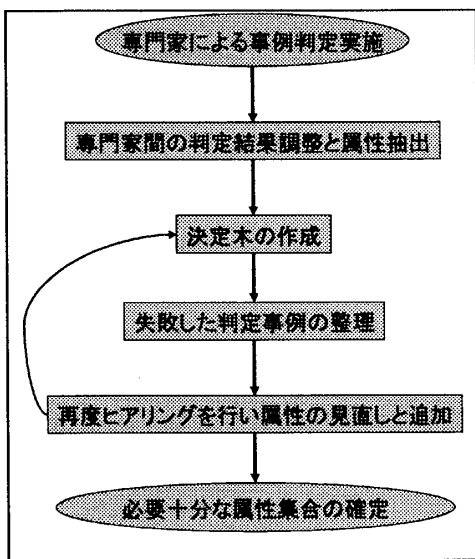


図1 判定実験の流れ

判定後、被験者の両名に同席してもらい、各情報に対する重要度の判定根拠（判定に使用した属性）を尋ねた。今回の実験では、双方の判定者から同じ属性集合が得られた。また、各属性が判定に及ぼす影響の方向（重要度が増加する方向か減少する方向か）については矛盾は見られなかった。

得られた判定結果集合（表3）に対し、判定実験の際の正解値を決定するために、2~4の判定については低い方の判定値を採用し、1と2で表示の有無に関わる場合については、再度判定の理由に関して尋ね、修正した方がよいと被験者が判断した場合には判定結果の修正を行い、修正の希望が無かった場合については2を採用した。

表3 判定結果の例

ID	判定者 A	判定者 B	調整結果
1	2	2	2 (普通)
2	1	1	1 (不要)
3	4	4	4 (最重要)
4	3	4	3 (重要)
5	1	2	2 (普通)

情報の重要度を判定するための属性を抽出する手順は次の通りである。まず、18件のプリテスト集合に対し、最初のヒアリングの結果得られた属性による決定木を用いた学習を行い、同一集合に対する判定を行った。この際、初期段階では、C4.5(WEKA<sup>3</sup>上の J48)、Random Tree、ReptTree などの複数の決定木手法を用いた。各手法について、判定結果が専門家の判断と異なっ

た場合には、専門家の意見を求めた。決定木生成試験の結果、Random Tree は複雑な木を構成したが、個々の事例の分類判断の過程がより詳細に決定木上に再現されるため、不足属性を専門家に指摘してもらうことが容易であった。この点で他の手法に比べて優位であったため、以降は、Random Tree による決定木に基づいて不足属性抽出のヒアリングを継続した。

以上のプロセスを経た結果、重要度判定のための属性情報として、以下のものを得た。

1. 列車到着順序：当該ホーム番線に何番目に到着する列車か（整数）
2. 客扱い有無：当該列車が当該駅での客扱いに関係する列車かどうか（Yes、No）
3. 優等列車：優等列車かどうか（Yes、No）
4. 新規列車種：これから来る列車の中での最初に到着する種別の列車かどうか（Yes、No）
5. 最遠終着列車：これから来る列車の中で一番遠くまで行く列車かどうか（Yes、No）
6. 新規到着駅有り：これから来る列車の中で、最初に到着予定の駅がある列車かどうか（Yes、No）
7. 行き先変更：行き先変更に関係するか（Yes、No）
8. 拠点駅での先着順変更：拠点駅までの先着案内の変更に関係するか（Yes、No）
9. 遅延情報：遅延情報かどうか（Yes、No）
10. 自駅番線変更：自駅の着発番線変更に関係するか+臨時列車もここに含む（Yes、No）
11. 自駅以降整理情報：自駅以降での運転整理内容かどうか（Yes、No）

上記の属性集合を用いて、専門家の判定済みの新たな事例 151 件に対して、Random Tree による決定木を生成したところ、専門家の判定と整合する判定結果を得られることが確認された。この結果から、上記のプロセスによって得られる属性集合を用いることにより十分な判定性能を有する決定木を生成できることがわかった。

一方、前述したように、フィルタリングシステムに組み込まれて使用される重要度判定ルール（決定木）に関しては、高い可読性も要求される。そこで、上記プロセスで得られた属性集合と C4.5 による学習を用いて決定木を構築し直した（図2）。構築した決定木による判定結果を、再現率と精度で評価した（表4）。判定結果は、再現率、精度のいずれの観点からも高い性能を示しており、十分実用に耐えるものと考えられる。特に、非表示カテゴリの情報に対して高い再現率と精度が得られたこと、最重要の情報に対する高再

現率の得られたことから、本手法の有効性を確認することができた。

4. まとめ

本研究では、鉄道ダイヤ乱れ時において大量に発生する情報を、決定木学習を用いて的確にフィルタリングして各駅の社員に伝達する手法について提案した。情報フィルタリングの性能を左右する属性集合の獲得プロセスに Random Tree 学習と専門家へのインタビューを組み合わせた手法を適用、また、最終的な決定木生成段階では、コンパクトで可読性の高い決定木が生成される C4.5 による決定木学習を用いた手法を適用し、良好な判定性能を得た。

参考文献

1) 秋葉 泰弘,フセイン・アルモアリム. “解説:例からの学習技術の応用に向けて 1.基本技術とその応用上課題”. 情報処理. Vol.39, No.2, p.145-

151(1998)

2) 鳥居大祐, ブスケ・フランソワ, 石田亨. “分類学習を用いた参加型モデリング手法の拡張”. 人口知能学会論文誌. Vol. 20, No.6, C, p.379-386(2005)

3) Weka 3: Data Mining Software in Java <http://www.cs.waikato.ac.nz/ml/weka/>

表4 C4.5による判定結果

	再現率 (決定木が正しく分類した数 / 学習集合中の当該要素の数)	精度 (決定木が正しく分類した数 / 決定木が当該要素と判定した数)
非表示	0.866 (39/45)	0.951(39/41)
普通	0.745 (38/51)	0.826(38/46)
要	0.778 (21/27)	0.656(21/32)
最重要	0.962 (25/26)	0.781(25/32)
平均	0.828	0.803

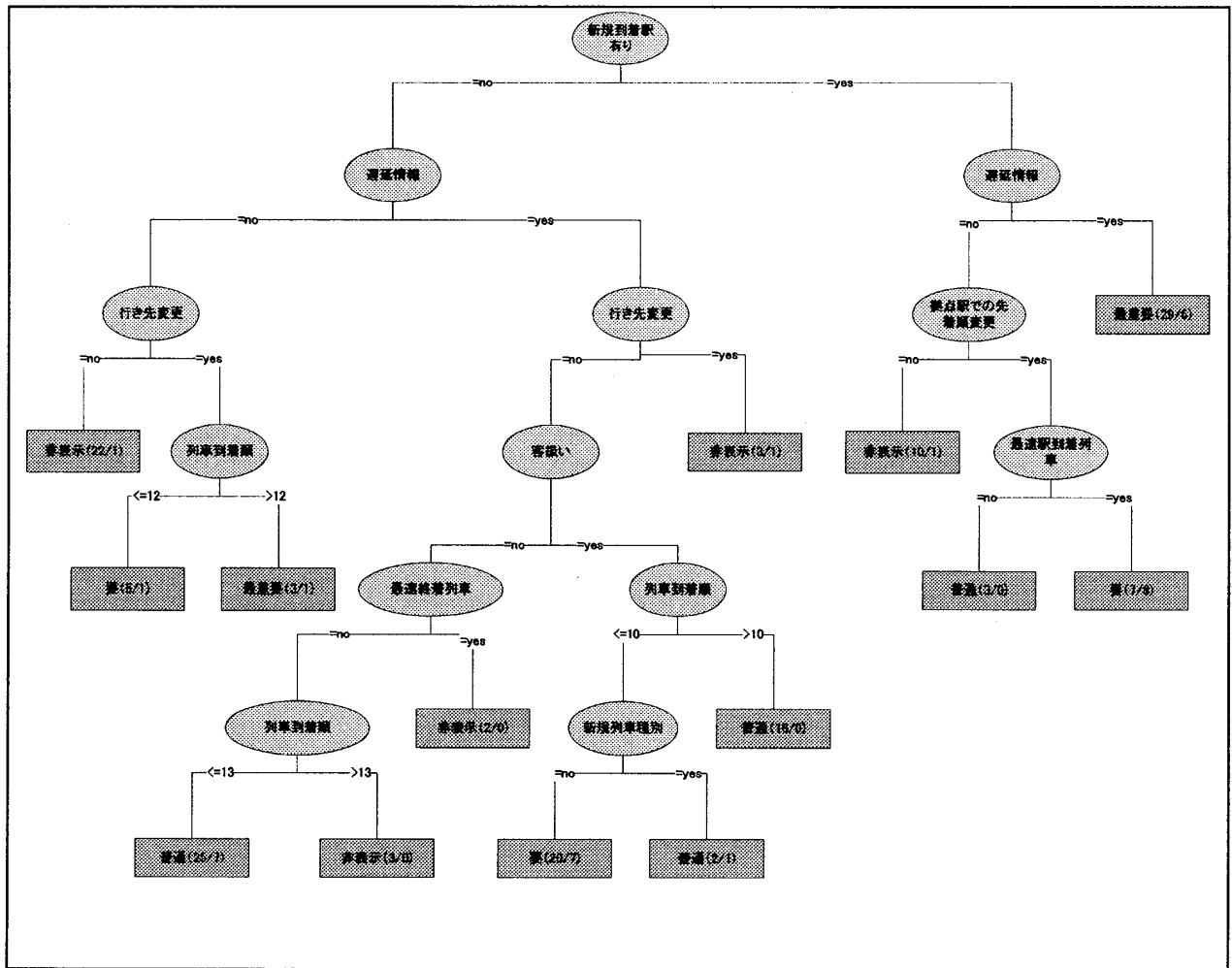


図2 C4.5による決定木 (学習事例数 151)