修 士 論 文

題 目

周辺情報の関係性抽出による
在席情報推定に関する研究

指導教員

報 告 者

香西 英樹

岡山大学 大学院自然科学研究科 電子情報システム工学専攻

平成23年2月10日 提出

要約

在席管理はタイムカードやスケジュール管理ソフトなど，様々な手法によって実現されている．これらの手法は利用者本人からの直接的な情報のみに依存している．また，単一の情報源から利用者の在席情報を取得する在席管理システムは，精度とコストのバランスを取るのが難しいという問題がある．

この問題に対して，電子メールの送信履歴や道路の渋滞情報などといった，利用者の周囲の情報源を利用することにより推定精度を向上させる手法が存在する．さらに，複数の周辺情報源から在席情報を取得することで，より精度の高い在席情報を推定できることがわかっている．在席推定に利用する情報源は，目的に合わせてコストと精度のバランスを考え選択することができるので，利用者にとって有益である．しかし，どの周辺情報の関係性が強いかは直感的にわかりづらい．さらに，周辺情報の抽出は人間の主観に左右され，人によって関係性の強い周辺情報は異なる．このため，単に周辺情報を抽出しても有効に活用できるとは言い難い．以上より，周辺情報を在席推定に利用するには，客観的に周辺情報同士の関係性を抽出し，応用する必要がある．

本論文では，複数情報源を用いた在席推定の問題点を指摘し，解決法として周辺情報の関係性抽出に相互情報量を用いる手法を提案した．相互情報量は，2種類の確率変数の相互依存の尺度を表す．次に，関係性抽出システムを導入した在席管理システムを設計，実装した．そして，関係性抽出システムを使用するにあたって周辺情報の値の取り方や閾値による出力結果と正答率の関係性の発見など，課題の検討を行った．最後に，関係性抽出システムを用いた在席推定の実験を行い，関係性抽出システムの評価を行った．実験の結果から，関係性抽出システムは信頼できる周辺情報を抽出できることがわかった．また，関係性の強い周辺情報を抽出し，関係性の弱い周辺情報は無視できることがわかった．そして，主観により在席情報に無関係な周辺情報を使用しても，在席推定への影響を排除できることがわかった．

目次

図 目 次

表 目 次

第 1 章

はじめに

在席管理手法の多くは，利用者が自身の在席情報を直接入力することで実現されている．例えば，タイムカードやスケジュール管理システムがある．ここで，在席情報とは利用者の位置や何を行っているかの状態を表す情報である．在席情報を管理する手法(以降，在席管理手法と略す)は，利用者自身が在席情報を入力する手法 [1] と特殊な端末を利用する手法 [2] の2種類に分類できる．前者は導入コストが低いため導入が容易であるものの，精度は利用者が入力を怠らないかどうかに依存する．また，後者は高い精度の在席管理手法を実現できるものの，特殊な端末の導入コストは高いため，簡単に導入できない．つまり，単一の情報源から利用者の在席情報を取得する在席管理システムは，精度と導入コストのバランスを取るのが難しい．

在席管理の精度を向上させる手法として，従来の直接入力による情報に加えて，利用者の周辺情報を利用することが有効とされている [3] ．周辺情報の例として，電子メールの送信履歴や道路の渋滞情報がある．利用する情報源は目的に合わせて，精度と導入コストのバランスを考え選択することができる．しかし，どの周辺情報が利用者の在席情報と強い関係性を持つかは，直感的には判断できない．なぜなら，周辺情報の抽出は人間の主観によって左右され，どの周辺情報が利用者に影響を与えるかについては個人差があるからである．そして，在席判定と不在判定が重なった場合，判断が難しいという問題もある．

本論文では，複数情報源を用いた在席推定の問題点を指摘し，解決法として周辺情報の関係性抽出に相互情報量を用いる手法を提案する．相互情報量は，2種類の確率変数の相互依存の尺度を表す．次に，関係性抽出システムを導入した在席管理システムの設計，実装について述べる．そして，関係性抽出システムを使用するにあたっ

て周辺情報の値の取り方や閾値による出力結果と正答率の関係性の発見など，課題の検討を行う．最後に，関係性抽出システムを用いた在席推定の実験を行い，関係性抽出システムの評価を行う．

第 2 章

複数情報源を用いた在席管理システムの特徴

2.1 複数情報源の必要性

在席管理は様々な手法によって実現されている．これらの手法は，以下の3種類に大別される．

(1) 利用者の自発的操作による手法(手動タイプ)
例：在席表，タイムレコーダ

(2) 自動的に行う手法(自動タイプ)
例：ビーコン

(3) (1),(2)を併用する手法(併用タイプ)
例：RFID

これらの手法は利用者本人からの直接的な情報のみに依存しており，また利点と欠点が共存している．以下に各タイプの利点と欠点を示す.

(1) 手動タイプ

(利点1) ホワイトボードの表の設置，あるいは行き先を示したボードなど簡単な方法で実現できる

(利点2) 特殊な機器を必要としない場合が多く，導入コストが低い

(欠点 1) 常に利用者側に入力を意識させる必要があるため，利用者の負荷が大きい

(欠点 2) 入力が全て利用者側に依存するため，精度の保証がない

(2) 自動タイプ

(利点 1) 利用者側への負荷はほとんどないため，利便性が高い

(利点 2) 利用者の入力に対する意識に左右されないため，精度が高い

(欠点 1) 機器の単価が高く，各部屋に必要なため導入コストが大きい

(欠点 2) 部屋の入退室しかわからないため，出張などでその場にいないことを検知できない

(3) 併用タイプ

(利点 1) 手動タイプと比べて利用者の負荷が軽く，精度が比較的安定している

(欠点 1) 利用者の入力に依存するため自動タイプに比べると精度が安定していない

(欠点 2) 導入コストは手動タイプに比べると高価である

これらに対して,「コストと精度のバランスを取りたい」という利用者の要求がある．これを実現するためには,「利用者の直接的な情報だけでなく周辺の情報も利用する」方法がある．上記の要求を解決するため，複数情報源から在席情報を推定する在席管理システムが提案されている [4] ．在席管理システムの概略図を図 2.1 に示す．ここでは情報源として，例として以下の 3 点を挙げている．

(1) 利用者からの直接的な情報
例：手動入力

(2) スケジュール情報
例：スケジュール管理システム

(3) ネットワーク利用状況
例：メール送信状況，IP アドレスリース状況

以上のように，複数情報源から得た情報を用いることで，より正確な推定が可能となる．

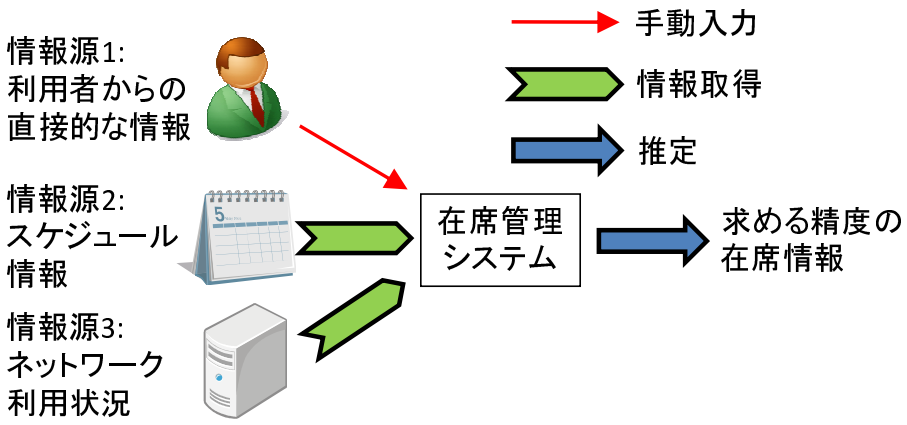

図 2.1 在席管理システムの概略図

2.2 複数情報源の有効性

複数情報源を用いた在席推定の例を図 2.2 に示す．これは被験者に自身の行動を記録してもらい，周辺情報としてスケジュール，DHCP の情報を使用した在席推定である．

出現する用語の意味は，以下のとおりである．

(1) 記録
被験者による行動記録である．実験では正解データとして扱う．

(2) スケジュール
被験者があらかじめ登録しておいたスケジュール表から得られた予定である．

(3) DHCP
被験者が使用している PC から，DHCP サーバへの IP アドレスリース要求が発生した時刻である．リース要求は 15 分毎に発生する．

図 2.2 からは，以下の 3 点の情報が得られる．

(1) 記録では，15 時 12 分で講義が終了している．

(2) スケジュールでは，15 時 50 分に講義が終了する予定となっている．

(3) DHCP は，15 時 12 分にリースされている．

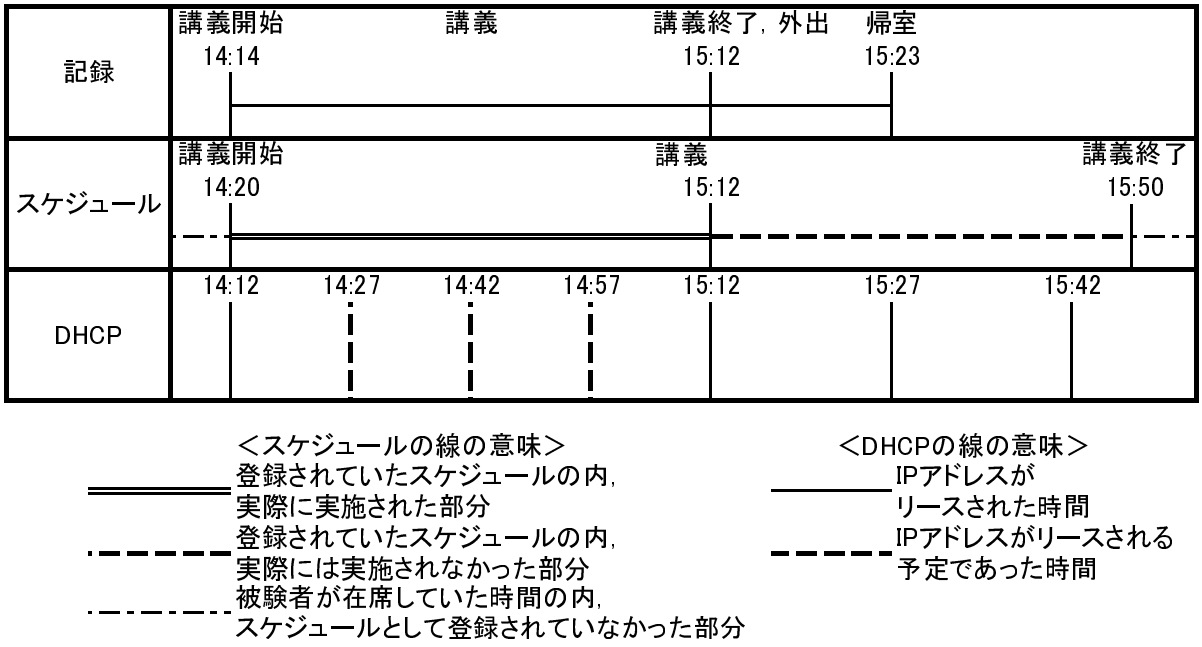

図 2.2 複数情報源を用いた在席推定例

この場合，スケジュールだけを用いると講義終了予定時刻と実際の講義終了時刻との間に38分の誤差が生じる．ここでIPアドレスリース情報を利用する．IPアドレスは15分毎にリースされているが，リースは14時12分を最後に途切れている．つまり，14時12分から14時27分の15分間に被験者がPCの電源を切ったと推定することができ，被験者は退室したと推定できる．また，リースは15時12分に再発生している．これにより，15時12分に被験者はPCの電源を入れた，つまり被験者は帰室したと判断できる．よって，この例では誤差を小さくできている．

また，人間の行動には，ある程度の法則性(以降，行動法則と呼ぶ)があると考えられる．例えば,「仕事の進捗状況が良ければ定時と共に帰宅するが，進捗状況が悪ければ遅くまで残業する」や,「重要な仕事が直近に控えていれば，早朝から出勤する」などである．このため，行動法則を発見することで在席推定の精度を改善できると思われる．

例えば，フレックスタイムで働く利用者に対して，以下の行動法則があるとする．

(1) 天気が雨だと午後から出勤する

(2) 前日に飲み会があれば，午前の遅くに出勤しがちである

(3) 現在の仕事の進捗状況が良ければ，早朝には出勤しない

これらの行動法則があった場合,「天気が雨で前日に飲み会があり，現在の仕事の進捗状況が良い」という情報が得られれば，午前中には出勤していないと推定できる．以上より，複数情報源から情報を収集することは重要であるといえる．

2.3 問題点

在席情報に関係する情報の抽出に関しては，以下の問題点がある．

(問題 1) どの周辺情報の関係性が強いかは直感的にわかりづらい

例えば,「遅くまで残業する」という結果に対して，何らかの法則性が考えられたとする．ここで「仕事が残っている」と「重要な役職を与えられている」という周辺情報が影響を与えていると推定できたとする．この 2 種類の周辺情報のいずれか，あるいは両方が在席情報に影響を与えているのか，それとも全く影響を与えていないのかは，人間の直感による判断に頼らざるを得ない．結果，周辺情報を利用しても在席推定結果は改善しない可能性がある．

(問題 2) 情報抽出は主観的である

(問題 1) の例で挙げた 2 種類の周辺情報は，在席推定において関係性が強いと容易に想像できる．ここで「近所のスーパーマーケットが特売日である」という周辺情報があるとする．ある利用者が頻繁にその店を利用していれば，特売日に生活必需品を買うために早めに帰宅することが考えられる．しかし，こうした情報は有効であるにも関わらず，無視されてしまう．

(問題 3) 人によって関係性の強い要素は異なる

「仕事が残っている」という周辺情報に対して,「残業をして片付ける」という人もいれば,「自宅に持ち帰る」という人もいる．このため，ある周辺情報が特定の利用者に有効であっても，他の人にも有効であるとは限らない．逆に，ある周辺情報が特定の利用者には有効でなくとも，別の利用者には当てはまる可能性もある．

(問題 4) 単体では関係性が小さい周辺情報の場合，見逃しやすい

例えば,「雨が降りそうである」という周辺情報があるとする．雨が降ってもほとんどの人は在席しているので，この周辺情報単体では関係性が小さいといえ

る．しかし,「布団を干している」という周辺情報があるとすると，早く帰宅して布団を取り込もうと考える人もいる．このように，他の周辺情報と組み合わせることで意味を持つ周辺情報の場合，単体で有効性を判断する方法を採用していると，周辺情報の候補から除外されてしまう．

(問題 5) 多数決で判断する場合，周辺情報個別の出現確率が考慮されていない

例えば,「雨が降っていれば午後から出勤する」「予定が特にない日は午後に出勤する」「月曜日の午前中は在席している」という行動法則があるとする．多数決で判断する場合,「予定が特になく，雨が降っている月曜日」という状況では,「午前中は不在である」と判断される．しかし，月曜日は週の始めであるため，午前中の在席率は高い．このため，周辺情報を用いることによって在席推定の結果が悪化してしまう．

2.4 問題の解決法

2.3 節で示した問題は,「周辺情報の関係性を客観的に判断することができない」ということに起因する．これらの問題に対し，入手した情報を単純なルールに基づいてそのまま活用することは困難である．このため，情報同士の関係性抽出に計算機による機械学習で対応する．これにより，前述の問題を以下のように解決できる．対処法の番号は，各問題点に対応している．

(対処 1) 関係性の数値化により，関係性の強さを判断する．

(対処 2) 計算機による客観的な判断で主観を交えることなく判断する．

(対処 3) 人物別に学習データを取得して判断する．

(対処 4) 周辺情報同士の関係性も判断する．

(対処 5) 在席結果に対する関係性の強さを数値化する．

第 3 章

相互情報量を用いた関係性抽出システム

3.1 相互情報量の性質

関係性抽出には，情報同士の関係性を偏見なく示すことが有用である．ここで，相互情報量を用いた関係性抽出システムを提案する．相互情報量とは，2種類の確率変数の相互依存の尺度を表す量である [5] ．その性質は以下の3点である．

(1) 値が大きければ異なる事象間の関係性は強い

(2) 事象が増えるにつれ値は単調に減少する
値が小さい組み合わせについては計算する必要はない．これにより，計算時間の短縮が見込まれる．

(3) 個別ではエントロピーが大きい周辺情報でも，関係性の強さを示すことができる
エントロピーが大きいということは，その周辺情報は発生するかどうか曖昧である．つまり，単体では有効性が小さいといえる．しかし，相互情報量を用いることで他の周辺情報との関係性を示すことができ，情報を有効活用することができる．

以上より，相互情報量は複数の周辺情報の関係性を示すのに役立つといえる．

3.2 提案手法で使用する用語

本論文で使用する用語は，以下のとおりである．

表 3.1 状態とその値の例

要素	ID	値
天気は晴れである	0	1(真)
天気は曇りである	1	0(偽)
天気は雨である	2	0(偽)
PC の電源が入っている	3	1(真)
スケジュールに記述された予定中である	4	0(偽)

(1) 要素

「対象者の身の回りの状況を述べた述語」を指す．例として,「天気は晴れである」や「現在 12 時である」などがある．また，類似した要素の集合を「要素群」とする．要素群の例として，要素群「天気」には「天気は晴れである」や「天気は雨である」などが存在する．要素は 0 か 1 を対応する値として取る．値が 1 であればその要素は真であり，0 であれば偽である．また，要素にはそれぞれ ID を付与し，他の要素と区別する．ID は 0 から始まる非負整数である．要素とその値の集合を「状態」とする．

状態の例を表 3.1 に示す．この状態中の要素群として「天気」と「PC の電源」,「スケジュール」がある．各要素群は以下の要素を持っているとする．

(A) 天気

「晴れ」と「曇り」と「雨」を持つ.「晴れ」の ID は「0」,「曇り」の ID は「1」,「雨」の ID は「2」とする．

(B) PC の電源

「入っている」を持つ.「入っている」の ID は「3」とする．

(C) スケジュール

「予定中である」を持つ.「予定中である」の ID は「4」とする．

このとき「晴れで PC の電源が入っており予定中でない」という状態は「10010」という値で表現される．

(2) 教師値

「状態を得た時の在席結果」を指す．教師値は「在席」と「不在」という 2 種類

の値を取る．これにより，教師値毎にデータを分類する．
具体的に取る値として，ある状態について「在席」であれば「y」，不在であれば「n」としている．

(3) 相関属性
「関係性の強い要素の集合」を指す．相関属性は複数の要素の集合として定義する．要素はハイフンで繋げて表現する．

例えば表 3.1 において相関属性「0-3」が出力されていれば，要素「晴れ」と「PC の電源」の関連性は強いことがわかる．なお，相関属性は在席と不在のそれぞれについて算出される．また，相関属性の抽出は相互情報量を用いて行われる．

(4) ルール
「相関属性に頻出する値」を指す．これを用いて，実際に取得した情報に対し在席情報の推定を行う．値はハイフンで繋げ，異なるルールはカンマで区切って表現する．
例えば表 3.1 において，相関属性「0-3」に対し在席ルールとして「1-1」が出力されていれば,「天気が晴れで PC の電源が入っていれば在席である」と解釈できる．ルールの抽出は，確率によって判断する．

(5) 計算回数
「相関属性に組み合わせる要素の数の上限」を示す．値は 2 以上の自然数を取る．
例えば表 3.1 において，計算回数が「2」であれば相関属性として「0-3」や「3-4」，計算回数が「3」であれば「1-4」や「0-3-4」が算出される可能性がある．

(6) 在席情報
「実際の推定に用いる情報に対し，ルールにマッチする情報があった場合，推定結果として返す値」を指す．値は「在席」か「不在」とする．

3.3 関係性抽出の流れ

周辺情報から関係性の強い要素を抽出するまでには，以下の手順が必要となる．

(1) データベースから要素群と教師値を取得する．

(2) 要素毎のエントロピーの上限値，相互情報量の下限値，計算回数，ルール出現確率の下限値を入力する．

(3) 教師値毎に相関属性を抽出する．

(4) 相関属性毎のルールを抽出する．

(5) (1)〜(4) を初回は対象者の人数分，以降は抽出要求が発行され次第，個人毎にルール抽出を行う．

第 4 章

在席管理システムの概要

4.1 関係性抽出システム導入前のシステム概要

4.1.1 各部構成

提案手法を導入する前の在席管理システムの構造を図 4.1 に示す. 図 4.1 の詳細は以下のようになる.

(A) 在席情報入力部
利用者に対して現在の在席情報を表示する.

(B) インタフェース部
利用者の在席状況を変更するインタフェースを提供する.

(C) 在席情報管理部
在席情報管理DBへのアクセスや正解情報管理DB, 関係性管理DBへのアクセスインタフェースを提供する他, 在席情報集約部に在席情報あるいは在席推定の問い合わせを行う.

(D) 在席情報管理DB
利用者の最新の在席情報を保存しておくDBである.

(E) 周辺情報
スケジュール管理システムやメールサーバなどの, 利用者による直接的な情報ではない情報の集合である.

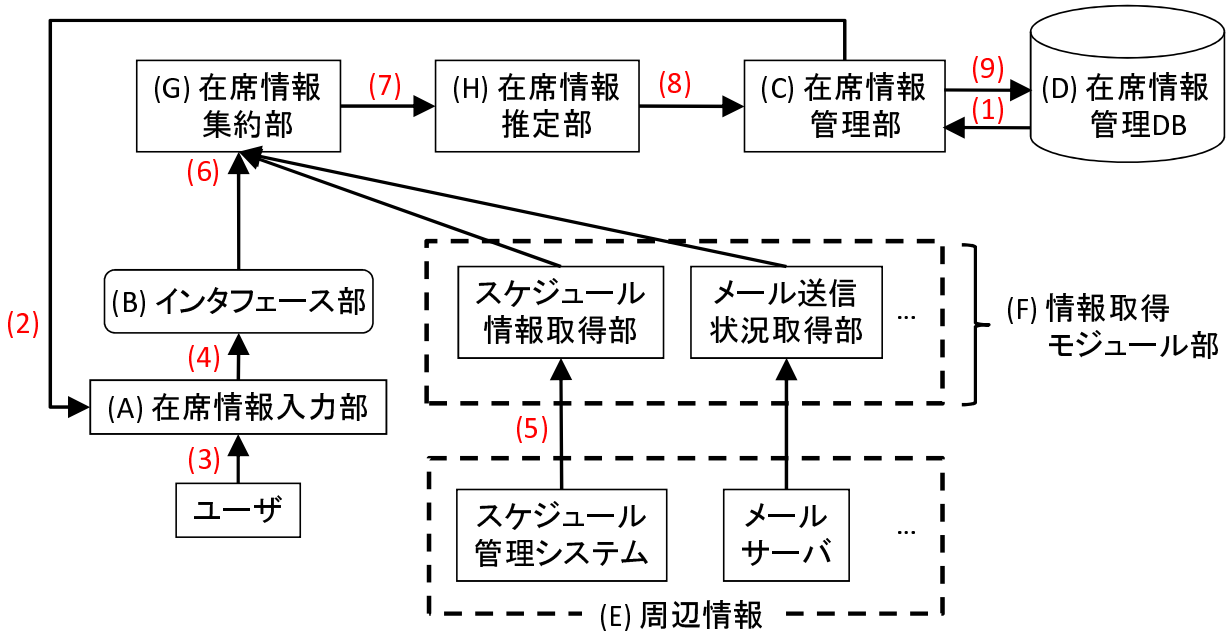

図 4.1 在席管理システムの概略図

(F) 情報取得モジュール部
周辺情報から得られた情報を在席管理に有用な値を取るように変換するモジュールである.

(G) 在席情報集約部
取得した周辺情報を一つにまとめる.

(H) 在席情報推定部
周辺情報から対象者の在席情報を推定する.

4.1.2 動作内容

処理の内容を以下に示す. 図 4.1 の矢印に付記している番号は, 以下で述べる動作内容の説明と対応している.

(1) 在席情報管理部 (C) は在席情報管理 DB(D) から在席情報を取得する.

(2) 在席情報入力部 (A) は在席情報管理部 (C) が取得した在席情報を反映させる.

(3) 在席情報入力部 (A) は利用者から在席情報の変更要求を受け取る.

(4) インタフェース部 (B) は，在席情報入力部 (A) を通して，利用者からの在席情報変更要求を受け取る.

(5) 情報取得モジュール部 (F) は対応する周辺情報から情報を個別に取得し，取得した情報毎に在席情報集約部 (G) が要求する形式に変換する.

(6) 在席情報集約部 (G) は，インタフェース部 (B) と情報取得モジュール部 (F) から在席情報を集約する.

(7) 在席情報推定部 (H) は在席情報集約部 (G) が集約した在席情報からより精度の高い在席情報の推定を行う.

(8) 在席情報管理部 (C) は在席情報推定部 (H) が推定した在席情報の反映を在席情報管理 DB(D) に要求する.

(9) 在席情報管理 DB(D) は在席情報管理部 (C) が要求した変更を反映する.

4.2 提案システム導入後の在席管理システム

4.2.1 各部構成

図 4.1 には在席情報推定部が存在する．これに周辺情報の関係性抽出システムを連携させる．提案手法を導入した後のシステムの構造を図 4.2 に示す．図 4.2 で新たに追加された箇所は以下のようになる.

(I) 正解情報管理 DB
今までの在席結果の出力数と，その正答数を記録している.

(J) 周辺情報管理 DB
利用者毎の学習データの集合を保存している DB である.

(K) 関係性抽出部
ある利用者について，どの周辺情報同士の関係性が強いか判断を行う.

(L) 関係性管理 DB
ある人物に対するルールを記録している DB である.

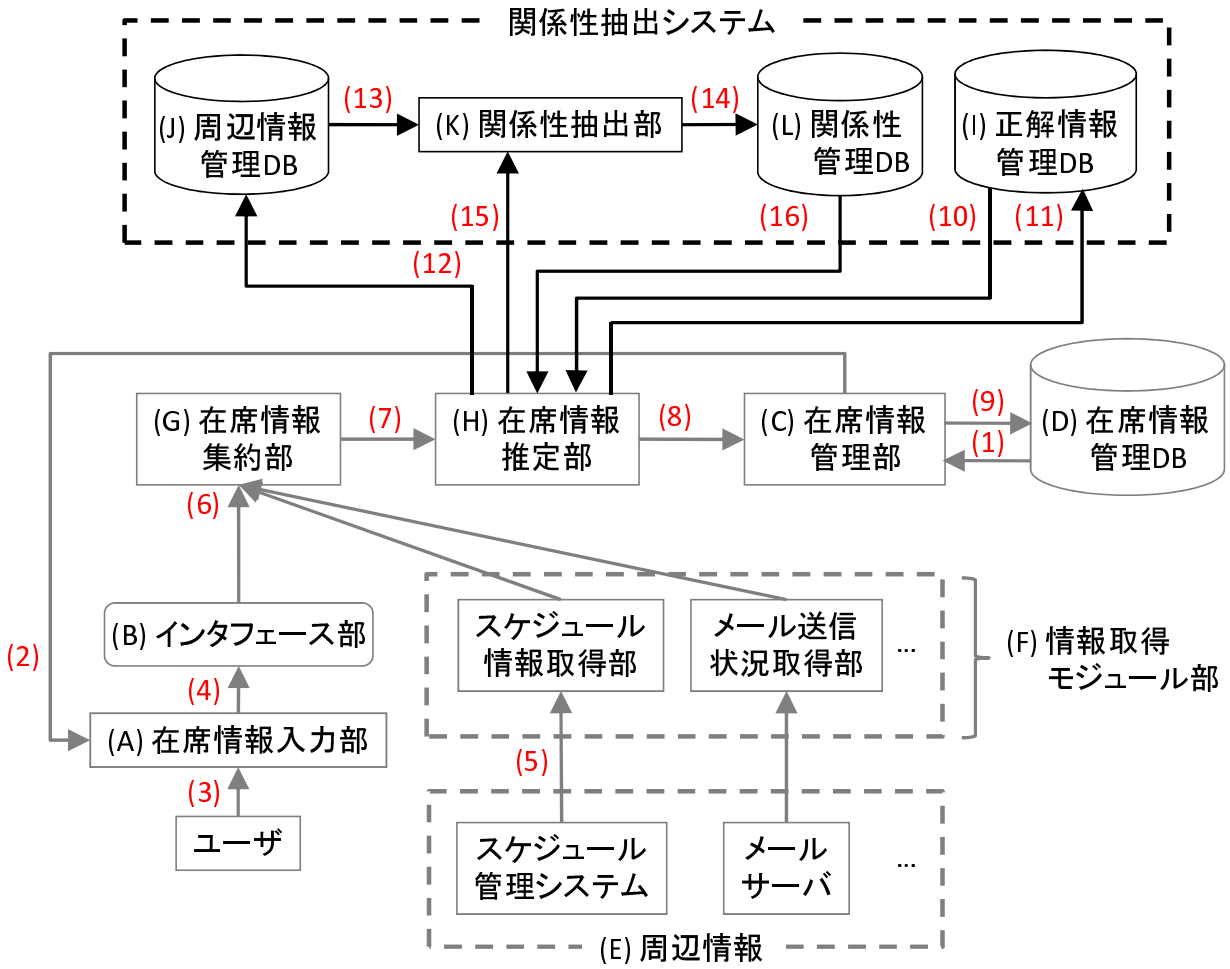

図 4.2 関係性抽出システムを導入した在席管理システムの概略

4.2.2 動作内容

図 4.2 の矢印に付記している番号は，以下で述べる動作内容の説明と対応している．

(10) 在席情報推定部 (H) は，正解情報管理 DB(I) から正答率を取得する．

(11) 在席情報推定部 (H) は，正解情報管理 DB(I) に正答率を記録する．

(12) 在席情報推定部 (H) は，周辺情報管理 DB(J) に要素群と教師値を記録する．

(13) 関係性抽出部 (K) は，周辺情報管理 DB(J) からある人物に関する全ての要素群とそれに対応する教師値を取得する．

(14) 関係性抽出部 (K) は，抽出結果を関係性管理 DB(L) に記録する．

(15) 在席情報推定部 (H) は，関係性抽出部 (K) にルール抽出の依頼を行う．

(16) 在席情報推定部 (H) は，関係性管理 DB(L) からルールを取得する．

システムの処理の概要は，以下の 3 点となる．

(1) 在席推定

(2) 学習データ取得

(3) ルール抽出処理

以降で，それぞれについて詳細を述べる．

4.2.3 在席推定の流れ

在席推定は学習データが多いほど精度が向上する．流れは以下のようになる．

(a) 在席情報集約部 (G) はインタフェース部 (B) と情報取得モジュール部 (F) から周辺情報を抽出する (6).

(b) 在席情報集約部 (G) は，在席情報推定部 (H) に抽出した要素群を渡す (7).

(c) 在席情報推定部 (H) は関係性管理 DB(L) からルールを参照し (16)，在席情報を推定する．

(d) 在席情報推定部 (H) は，在席情報管理部 (C) に推定結果を通知する (8).

4.2.4 学習データ取得の流れ

在席推定には多くの学習データを取得することが，より高い精度の推定を実現することにつながる．流れは以下のようになる．

(a) 在席情報の推定結果を利用者に入力してもらい (3)(4)，結果がインタフェース部 (B) から在席情報集約部 (G) を通じ (6)，在席情報推定部 (H) に通知される (7).

(b) 在席情報推定部(H)は評価結果を教師値として，現在の周辺情報を教師値と共に学習データとして周辺情報管理DB(J)に記録する(12).
また，在席情報推定部(H)は推定結果の正誤を判断し，結果を正解情報管理DB(I)に記録する(11).

4.2.5 ルール抽出処理の流れ

学習データからルールを抽出することで，初めてシステムは周辺情報から在席情報を推定することができる．流れは以下のようになる．

(a) 在席情報推定部(H)は関係性抽出部(K)に，対象者のルールを抽出するよう依頼する(15).

(b) 関係性抽出部(K)は周辺情報管理DB(J)からデータを取得する(13).

(c) 抽出結果を関係性管理DB(L)に記録する(14).

第 5 章

検討事項

関係性抽出システムを実装するには，以下の5点の課題が存在する．

(1) 要素の値の取り方

(2) 閾値による出力結果と正答率の関係性の発見

(3) 学習データの取得方法

(4) ルールの正答率の目標値

(5) ルール抽出時の学習時間の短縮

以降，それぞれの詳細と対処を述べる．

5.1 要素の値の取り方

要素の算出条件として，その性質上「0か1で表現できる」という条件が必要である．これに関する問題点として，何も考えず適用してしまうと情報の信頼性に影響を及ぼす可能性がある．以下にその例を挙げる．

(1) それ自体が主観的な要素
「真面目である」といった要素は，対象者がどれくらい真面目なのか判断する時点で主観が入ってしまっている．
これについては，要素として記録する前に他人にアンケート等で判断してもらうことで対応できる．

(2) 連続量を持つ要素群の量子化

「気温」や「湿度」等の連続量を持つ要素群は，情報取得に高性能な装置を用いれば，より細かな単位で情報を取得できる．しかし，それをそのまま適用してしまうと不都合が発生する．

例として,「気温」の場合を考えてみる．この場合要素は「20 ℃である」「21 ℃である」と 1 ℃刻みにすることが可能である．しかし，あまりにも要素の判定を細分化してしまうと各要素のエントロピーが高くなってしまう．その結果，信頼性のある情報として判断されなくなってしまう．

対処法として，ある程度の値で区切る方法がある.「気温」の例では,「20〜24 ℃である」「25〜29 ℃である」とすることで，各要素のエントロピーが不必要に高くなってしまうことを防ぐことができる．

5.2 閾値による出力結果と正答率の関係性の発見

相互情報量を計算するには，各々の事象のエントロピーを計算しなければならない．しかし，エントロピーという概念は直感的ではない．言い換えれば，エントロピーが大きくなればなるほどその事象は曖昧であるが，エントロピーが 0.1 ビット増えたから確実性がどれくらい低くなったとは明確にはいえないのである．これは，相互情報量も同様である．すなわち，エントロピー並びに相互情報量の値が推定結果にどう影響するかは，実際にシステムを運用してみなければわからない．このため，各パラメータに対する推定精度の相関を発見する必要がある．

本システムでは，エントロピーや相互情報量，ルール出現確率毎に数パターン準備しておき，計算時間と正答率を記録する．そして，正答率が最も高い閾値による学習結果を正式に記録する．この時，計算時間があまりにもかかるようであれば，他の結果を採用するものとする．計算時間の目安としては，一度関係性を抽出すればある程度の期間は有効であると思われるので，1 人あたり 10 分を目安とする．なお，閾値による推定精度の相関については 6.8.3 項，計算時間については 6.8.6 項で詳しく述べる．

5.3 学習データの取得方法

関係性抽出システムを使用するためには，学習データを入手する必要がある．また，ある一定量以上の学習データを保持している必要がある．しかし，その入手法

が問題となる．常に在席情報を全ての利用者に要求した場合，頻度の多さから推定する必要性がなくなってしまう．しかし，頻度を減らしてしまうと十分な学習データを取得できない．この問題に対し，段階的に在席情報問い合わせ回数を減らしていく．具体的には以下のようにする．

(1) システム導入後最初の1週間は，全員を対象に1時間に1度在席情報を問い合わせる．

(2) 次の1週間は，全員を対象に1日に2度問い合わせる．

(3) 以降は1日に1度利用者の誰かを対象に問い合わせる．

これにより，利用者の負担を極力減らしつつも，学習データを蓄積していくことが可能となる．

5.4 ルールの正答率の目標値

ルールを現実の事象に適用する場合，利用者の予期しない行動や事象が発生するため完全な在席推定は難しい．しかし，精度の低い推定は有用ではない．また，ルールの適用に失敗するたびにルールの再抽出を行うのは，計算機の負荷や最終的に残るルールの数から見ても効率が悪い．本システムでは，在席推定の正答率を80パーセントとすることを目標とし，ルール抽出を行う．再抽出の確率は，無作為に推定した場合より下回った場合とする．例えば，2種類の在席情報を扱う場合であれば，50パーセントが再抽出の確率となる．

5.5 ルール抽出時の学習時間の短縮

学習データが多ければ多いほど，抽出されるルールは信頼性が高いものになる．しかし，学習時間がその分長くなってしまうという欠点がある．このため，新しい学習データを入手する度にルール抽出を行うと，抽出時間に対する抽出結果の改善度は小さいと思われる．本システムではこれに対し，定期的あるいは利用者の要求があればルールの再抽出を行うことで対処する．

具体的には，以下の条件でルールの再抽出を行う．

(1) 推定結果の利用者への確認で，正答率が悪かった場合
システムは定期的に推定結果の確認を利用者に問い合わせ，その結果を正解情報管理DBに記録する．記録は利用者による任意の結果入力に対しても行われる．記録された正答率が一定値を下回った場合，システムは自動的にルール再抽出を行う．再抽出を行う正答率については，5.4 節のとおりである．

(2) 利用者の要求
利用者がシステムから直接再抽出要求を出すことで再推定を行う．不必要な再抽出を防ぐため，抽出状況は利用者にわかるように表示する．

(3) 一定データ量の追加
ルール抽出からある程度時間が経過すれば，相当の学習データが蓄積されている．このため，再抽出によるルールの精度向上の効果は高いと考えられる．

第 6 章

関係性抽出システムの評価実験

6.1 実験目的

本実験では，以下の 2 項目を評価することを目的とする．

(1) 性能評価
在席推定の正答率と情報の集約及び関係性抽出に必要な時間の評価を行う．在席推定の正答率は 6.7 節で，情報の集約及び関係性抽出に必要な時間の評価は 6.8.6 項で詳しく述べる．

(2) 閾値の適切な値の発見
5.2 節の問題に対して，正答率の高いルールを出力する閾値を発見，評価する．この評価は 6.8.3 項で詳しく述べる．

6.2 実験内容

(1) 実験期間
2011 年 1 月 8 日 (土)0 時 0 分 0 秒から 2011 年 1 月 21 日 (金)23 時 59 分 59 秒までの 14 日間の被験者の在席記録と周辺情報の記録を本実験で用いた．

(2) 被験者
同じ研究室に所属する 10 人を元に推定を行った．10 人は同じ部屋で活動している．

(3) 実験手順

1月8日から14日までの7日間の記録を学習データとし，ルールを抽出する．抽出されたルールを元にして，1月15日から21日までの7日間の記録を対象に1分間隔で在席推定を行い，その正答数等とルールについて調査，考察した．

(4) 判定内容

在席，不在の2種類の情報で判定を行った．

(5) 推定回数

推定回数は1人あたり10080回となる．これは，1分間隔での在席推定を7日間行ったからである．

6.3 使用した周辺情報

使用した周辺情報とその要素の区分は以下のとおりである．なお，時間の最小単位は1秒である．

(1) 曜日

その日の曜日を示す．要素の区分は，曜日毎の7種類である．情報源は関係性抽出システムを使用する計算機の時刻情報を使用した．

(2) メール

10分以内にメールを送信したかどうかの情報である．要素の区分は2種類である．情報源は研究室で運用しているメールサーバの送受信ログを使用した．

(3) スケジュール

現在時刻が，スケジュールに登録された予定中かどうかを示す．要素の区分は2種類である．情報源はGoogleカレンダー [6] に研究室用のカレンダーを作成し，それを使用した．

(4) 気温

現在の気温を示す．要素の区分は-5.0 ℃未満，-5.0 ℃～-0.1 ℃，0.0 ℃～4.9 ℃，5.0 ℃～9.9 ℃，10.0 ℃～14.9 ℃，15.0 ℃～19.9 ℃，20.0 ℃～24.9 ℃，25.0 ℃～29.9 ℃，30.0 ℃～34.9 ℃，35.0 ℃～39.9 ℃，40.0 ℃以上の計11種類である．情報源は気象庁の過去の気象データ [7] を使用した．

(5) 天気
現在の天気を示す．要素の区分は晴れ，曇り，雨，雪，その他の計5種類である．情報源は気象庁の過去の気象データを使用した．

(6) DHCP
DHCPによるIPアドレスのリースが15分以内に行われたかどうかを示す．要素の区分は2種類である．情報源は研究室で運用しているDHCPのシステムログを使用した．

(7) 時刻
現在の時刻を示す．要素の区分は毎時0分を起点とし，毎時59分59秒を終点とした1時間刻みの計24種類である．情報源は関係性抽出システムを使用する計算機の時刻情報を使用した．

6.4 閾値

実験で用いた各閾値を表6.1に示す．この値を選択した理由は以下のようになる．

(1) システムでは2種類の値しか扱わない．このため，エントロピーは最大で1ビットとなる．1ビットとなる場合，それぞれの値の出現確率は50パーセントとなる．この場合その要素は全く信頼できない値となる．

(2) 5.2節で述べたように，エントロピーと相互情報量は直感的な値でない．このため，確実にルールが出力されるような値として選択した．

(3) 組み合わせ回数は多くすればするほど計算時間が長くなる傾向にある．このため，ある程度の計算時間で終了するような値として選択した．

(4) ルール出現確率は低くし過ぎると在席情報に無関係なルールが多く出力されてしまう．しかし，高くし過ぎると全くルールが出力されなくなってしまう．このため，ある程度のルールが出力されるような値として選択した．

閾値の適性値については6.8.3項で述べる．

表 6.1 実験で用いた各閾値

エントロピー	0.99 ビット
相互情報量	0.5 ビット
組み合わせ回数	5 回
ルール出現確率	30 パーセント

6.5 用語定義

実験で使用する用語の説明を以下に記述する.

(1) 正答数
ある被験者に対して, 抽出されたルールから推定される在席情報が正しかった場合の回数である. 例として, 現在の状態に在席ルールが合致していて, 実際の在席結果が在席の場合である. 単位は回である.

(2) 正答率
ある被験者に対し正答した確率である. 単位はパーセントである.

(3) 誤答数
被験者に対して抽出されたルールから推定される在席情報が誤っていた場合の回数である. 例として, 現在の状態に在席ルールが合致していて, 実際の在席結果が不在の場合である. 単位は回である.

(4) 誤答率
ある被験者に対し誤答した確率である. 単位はパーセントである.

(5) 未判定数
被験者に対して在席とも不在とも判定できない場合であった回数である. 例として, 現在の状態に在席, 不在ルール共に合致しない場合や, 在席, 不在ルール共に同じ個数のルールが合致した場合である. 単位は回である.

(6) 期待正答率
未判定の場合に対し 50 パーセントの確率で正答するとした場合, 期待される正答率である. 単位はパーセントである.

表 6.2 学習データ例

在席データ	不在データ
100	011
100	011
101	010
111	010
100	000

表 6.3 状態例

000
001
010
011
100
101
110
111

6.6 関係性抽出の例

例となる学習データを表 6.2 に示す．表 6.2 において，在席データとは教師値が在席である学習データ，不在データとは教師値が不在である学習データである．また，要素の ID は左から順に 0，1，2 とする．この学習データから，ID が 0 の要素の値が「1」であれば必ず在席であり，ID が 0 の要素の値が「0」であれば必ず不在であるというルールが抽出できる．このルールを用いて表 6.3 で示される状態を対象に在席推定を行った結果を表 6.4 に示す． 表 6.4 より，関係性抽出システムは在席情報に関係する要素を発見することができているので，正常に動作しているといえる．

表 6.4 推定結果

正答数 (回)	正答率 (%)	誤答数 (回)	誤答率 (%)	未判定数 (回)	期待正答率 (%)
8	100.0	0	0.0	0	100.0

表 6.5 実験結果 (エントロピー:0.99 ビット, 相互情報量:0.5 ビット, 組み合わせ回数:5 回, ルール出現確率:30 パーセント)

被験者	正答数 (回)	正答率 (%)	誤答数 (回)	誤答率 (%)	未判定数 (回)	期待正答率 (%)
A	2343	23.2	7737	76.8	0	23.2
B	3458	34.3	6622	65.7	0	34.3
C	1412	14.0	8668	86.0	0	14.0
D	2224	22.1	7856	77.9	0	22.1
E	2497	24.8	7583	75.2	0	24.8
F	2407	23.9	7673	76.1	0	23.9
G	1215	12.1	8865	87.9	0	12.1
H	2476	24.6	7604	75.4	0	24.6
I	1469	14.6	8611	85.4	0	14.6
J	1685	16.7	8395	83.3	0	16.7
平均	2118.6	21.0	7961.4	79.0	0	21.0

6.7 実験結果

実験結果を表 6.5 に示す. 平均と確率は小数点第二位を四捨五入して表示している. 以降, 特に断りがない限り, 表 6.5 の結果を元に記述する.

6.8 考察

6.8.1 関係性の高い要素

関係性の高い要素は以下のようになる.

(1) 同一要素群内

(A) 気温
気温が0.0 ℃～4.9 ℃でなく，5.0 ℃～9.9 ℃である場合，つまり気温が5.0 ℃～9.9 ℃である場合に在席というルールが8人，気温が0.0 ℃～4.9 ℃であり，5.0 ℃～9.9 ℃でない場合，つまり0.0 ℃～4.9 ℃である場合に不在というルールが4人出力された．

(2) 異なる要素群間
見出すことができなかった．

これ以外にも，信頼できる要素として単独で出力されていたり，ルールが出力されていないことがあった．

6.8.2 抽出結果について

抽出結果についての検討項目は以下のとおりである．

(1) 期待正答率が良かった被験者の評価

(2) 期待正答率が悪かった被験者の評価

ここでは期待正答率が良かった被験者としてB，期待正答率が悪かった被験者としてGを対象に考察を行う．

期待正答率が良かった被験者

Bは土日以外は必ず在席しており，かつ長時間在席していた．このため，期待正答率が高かったものと思われる．

期待正答率が悪かった被験者

Gは学習データ中の平日は毎日在席していた．しかし，推定期間中は不在であることが多かった．これにより，在席ルールの数から在席と判定される状況で不在であった場合が多かったため，期待正答率が悪かったものと思われる．

表 6.6 エントロピーを0.97ビットに変更した場合

被験者	正答数(回)	正答率(%)	誤答数(回)	誤答率(%)	未判定数(回)	期待正答率(%)
A	2343	23.2	7737	76.8	0	23.2
B	3458	34.3	6622	65.7	0	34.3
C	1412	14.0	8668	86.0	0	14.0
D	2224	22.1	7856	77.9	0	22.1
E	2497	24.8	7583	75.2	0	24.8
F	2407	23.9	7673	76.1	0	23.9
G	1215	12.1	8865	87.9	0	12.1
H	2476	24.6	7604	75.4	0	24.6
I	1469	14.6	8611	85.4	0	14.6
J	1685	16.7	8395	83.3	0	16.7
平均	2118.6	21.0	7961.4	79.0	0	21.0

6.8.3 閾値の評価

エントロピー

要素は各列が0か1の数字列で表されている．このため，各列単体のエントロピーは最大で1ビットである．よって，エントロピーの閾値は1ビット未満が適切である．閾値を変更して実験して，列単体のエントロピーを0.97ビットに変更した結果を表6.6，0.95ビットに変更した結果を表6.7に示す． 表6.5と比較して，今回の実験中では変化は見られなかった．

相互情報量

相互情報量を0.4ビットに変更した結果を表6.8，0.6ビットに変更した結果を表6.9に示す． 表6.5と比較して，今回の実験中では変化は見られなかった．

表 6.7 エントロピーを0.95ビットに変更した場合

被験者	正答数(回)	正答率(%)	誤答数(回)	誤答率(%)	未判定数(回)	期待正答率(%)
A	2343	23.2	7737	76.8	0	23.2
B	3458	34.3	6622	65.7	0	34.3
C	1412	14.0	8668	86.0	0	14.0
D	2224	22.1	7856	77.9	0	22.1
E	2497	24.8	7583	75.2	0	24.8
F	2407	23.9	7673	76.1	0	23.9
G	1215	12.1	8865	87.9	0	12.1
H	2476	24.6	7604	75.4	0	24.6
I	1469	14.6	8611	85.4	0	14.6
J	1685	16.7	8395	83.3	0	16.7
平均	2118.6	21.0	7961.4	79.0	0	21.0

組み合わせ回数

抽出されたルールは合計413個であった．このうち，要素の組み合わせ回数が2回となるルール数は13個であった．組み合わせ回数は非常に多くの学習データが追加されたり，新たに要素を追加しない限り極端に増えることはないと思われる．以上より組み合わせ回数は，余裕を考え5回が適切と思われる．

ルール出現確率

ルール出現確率を20パーセントに変更した場合の結果を表6.10，40パーセントに変更した場合の結果を表6.11に示す． 表6.5と比較して，今回の実験中では変化は見られなかった．

6.8.4 相関属性のみを使用した場合

推定結果が悪い原因のひとつとして，単体でも信頼できる要素が非常に多く出力されている点がある．原因として，2週間という限られた期間では，要素群が出力す

表 6.8 相互情報量を0.4ビットに変更した場合

被験者	正答数(回)	正答率(%)	誤答数(回)	誤答率(%)	未判定数(回)	期待正答率(%)
A	2343	23.2	7737	76.8	0	23.2
B	3458	34.3	6622	65.7	0	34.3
C	1412	14.0	8668	86.0	0	14.0
D	2224	22.1	7856	77.9	0	22.1
E	2497	24.8	7583	75.2	0	24.8
F	2407	23.9	7673	76.1	0	23.9
G	1215	12.1	8865	87.9	0	12.1
H	2476	24.6	7604	75.4	0	24.6
I	1469	14.6	8611	85.4	0	14.6
J	1685	16.7	8395	83.3	0	16.7
平均	2118.6	21.0	7961.4	79.0	0	21.0

る値の範囲が限られてしまうからである．その結果，本来ならば信頼できない情報でも信頼できると判断したため，推定結果が悪くなったと考えられる．単体で有効とされたAの在席ルールと不在ルールを表6.12に示す.

在席ルールで出力されている要素の大半は要素群「気温」についての情報である．このため，14日間は実験期間としては短いものであったといえる．表6.12の場合，在席ルールと不在ルールのほとんどが共通している．さらに，在席ルールの方が不在ルールに比べて明らかに多い．これより現在の状態に対しほとんどの場合で在席という結果を返すことが考えられる．記録によると，Aは12時間以上在席していることはほとんどなかった．その結果，半分以上の推定において在席と推定したため推定結果が悪くなっていると思われる．ここで，相関属性のみを使用した場合の結果を表6.13に示す.

被験者全員の期待正答率は表6.5と比較して，明らかに期待正答率が良くなっていることがわかる．以上より，学習データが少ない場合は，相関属性のみを使用するべきといえる．

表 6.9 相互情報量を0.6ビットに変更した場合

被験者	正答数(回)	正答率(%)	誤答数(回)	誤答率(%)	未判定数(回)	期待正答率(%)
A	2343	23.2	7737	76.8	0	23.2
B	3458	34.3	6622	65.7	0	34.3
C	1412	14.0	8668	86.0	0	14.0
D	2224	22.1	7856	77.9	0	22.1
E	2497	24.8	7583	75.2	0	24.8
F	2407	23.9	7673	76.1	0	23.9
G	1215	12.1	8865	87.9	0	12.1
H	2476	24.6	7604	75.4	0	24.6
I	1469	14.6	8611	85.4	0	14.6
J	1685	16.7	8395	83.3	0	16.7
平均	2118.6	21.0	7961.4	79.0	0	21.0

6.8.5 主観的な要素群のみを使用した場合

被験者 A の 1 月 11 日 (火) の在席情報と要素群の値を図 6.1 に示す. 実線は正解データ及び各要素群が出力していた値を示す.

要素群の選択の問題として, 2.3 節で述べたように, 人間による情報抽出は主観的であるという問題がある. ここで, 先に述べた 7 種類の周辺情報から, 在席情報に関係すると思われる 4 種類のみを用いて実験を行う. 使用する要素群は, 以下の 4 種類である.

(1) 曜日

(2) メール

(3) スケジュール

(4) DHCP

以上の周辺情報を使用した場合の結果を表 6.14 に示す. 表 6.5 と比較して, 今回は結果が変化していないため, 実験環境下では気温, 天気, 時刻の 3 要素は在席情報に

表 6.10 ルール出現確率を20パーセントに変更した場合

被験者	正答数(回)	正答率(%)	誤答数(回)	誤答率(%)	未判定数(回)	期待正答率(%)
A	2343	23.2	7737	76.8	0	23.2
B	3458	34.3	6622	65.7	0	34.3
C	1412	14.0	8668	86.0	0	14.0
D	2224	22.1	7856	77.9	0	22.1
E	2497	24.8	7583	75.2	0	24.8
F	2407	23.9	7673	76.1	0	23.9
G	1215	12.1	8865	87.9	0	12.1
H	2476	24.6	7604	75.4	0	24.6
I	1469	14.6	8611	85.4	0	14.6
J	1685	16.7	8395	83.3	0	16.7
平均	2118.6	21.0	7961.4	79.0	0	21.0

関係がなかったといえる．このことから，関係性抽出システムは周辺情報として在席情報に無関係な情報を与えたとしても，その影響を排除できるといえる．

6.8.6 関係性抽出に必要な時間

測定環境を表 6.15 に示す．周辺情報の統合に必要な時間は10人合計で39分24秒であり，関係性抽出に必要な時間は10人合計で9分53秒であった．必要な時間は1人あたりそれぞれ約3分56秒と約59秒である．また，アルゴリズムから算出される最悪の計算時間は2^nである．在席管理システムを運用するにあたり，これらの処理は，計算機があまり使用されていない深夜，早朝等の時間帯に行えば，計算機の負荷を抑えることができると考えられる．以上より，計算時間の長さの問題は解決できると考えられる．

6.8.7 要素群の性質

実験の結果，今回使用した要素群の性質は以下のようになる．

表 6.11 ルール出現確率を 40 パーセントに変更した場合

被験者	正答数 (回)	正答率 (%)	誤答数 (回)	誤答率 (%)	未判定数 (回)	期待正答率 (%)
A	2343	23.2	7737	76.8	0	23.2
B	3458	34.3	6622	65.7	0	34.3
C	1412	14.0	8668	86.0	0	14.0
D	2224	22.1	7856	77.9	0	22.1
E	2497	24.8	7583	75.2	0	24.8
F	2407	23.9	7673	76.1	0	23.9
G	1215	12.1	8865	87.9	0	12.1
H	2476	24.6	7604	75.4	0	24.6
I	1469	14.6	8611	85.4	0	14.6
J	1685	16.7	8395	83.3	0	16.7
平均	2118.6	21.0	7961.4	79.0	0	21.0

(1) 曜日

被験者は平日は在席し，休日は不在であることがほとんどであった．ただし，仕事が残っている場合は休日でも在席しており，また平日でも病欠で不在の場合がある．

(2) メール

メール送信後 10 分間は在席していることがほとんどであった．ただし 12 時付近や 18 時以降になると，メールを送信すると同時に不在になる場合が見られた．

(3) スケジュール

スケジュール開始 5 分前には不在であることが多かった．また，スケジュール中で確実に不在であるのは予定時間の半分程度までで，それ以降は予定時間より早く終了したり，逆に延長するなどの誤差があった．

(4) 気温

今回の気温の区分は，実際に出力された値と比較すると非常に大きく幅を取っていた．このため，全く出力されない値があるなど情報としては不十分な点が

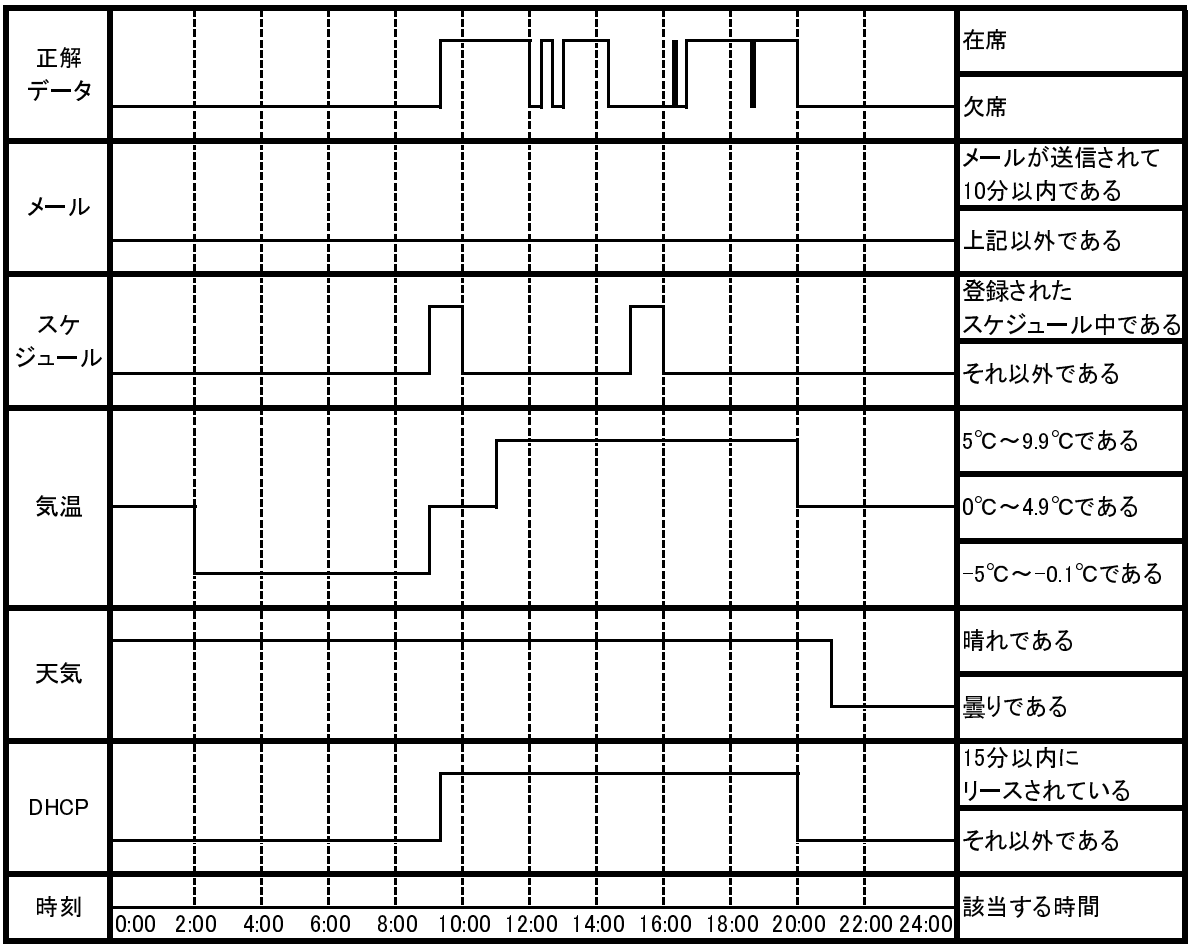

図 6.1 在席情報と周辺情報の関係

あった．よって，この要素群については情報を長期的に取得した上で評価したり，要素の決め方を検討する必要がある．

(5) 天気

実験期間中には雨や雪は降らなかった．これらの天気は在席情報に影響を与えることが容易に想像できる．このため，今回の実験期間中では有益な情報を入手することはできなかったといえる．また，雷や雪などは時期が限定される天気である．よって，この要素群については情報を長期的に取得した上で評価する必要がある．

(6) DHCP

この情報でわかることは，入室時間と退室時間がわかるだけである．このため，リース直後は在席しており，解放直後は不在であると推定できる．これ以外の状況では在席情報を推定することは難しいといえる．

(7) 時刻

被験者たちは 10 時頃には入室し，19 時頃には退席する傾向にあった．ただし，被験者により入退室時間には差がみられた．また，当日や翌日にスケジュールがある場合はその被験者の傾向に当てはまらない場合があった．

表 6.12 被験者 A の在席ルールと不在ルール

在席ルール	不在ルール
晴れである	雪でない
曇りでない	その他の気象でない
雨でない	-5.0 度未満でない
雪でない	10.0 ℃～14.9 ℃でない
その他の気象でない	15.0 ℃～19.9 ℃でない
-5.0 度未満でない	20.0 ℃～24.9 ℃でない
10.0 ℃～14.9 ℃でない	25.0 ℃～29.9 ℃でない
15.0 ℃～19.9 ℃でない	30.0 ℃～34.9 ℃でない
20.0 ℃～24.9 ℃でない	35.0 ℃～39.9 ℃でない
25.0 ℃～29.9 ℃でない	40.0 ℃以上でない
30.0 ℃～34.9 ℃でない	メール送信後 10 分以上経過している
35.0 ℃～39.9 ℃でない	
40.0 ℃以上でない	
0 時でない	
1 時でない	
2 時でない	
3 時でない	
4 時でない	
5 時でない	
6 時でない	
7 時でない	
20 時でない	
21 時でない	
22 時でない	
23 時でない	
日曜日でない	
月曜日でない	
土曜日でない	
0.0 ℃～4.9 ℃でなく，5.0 ℃～9.9 ℃である	

表 6.13 相関属性のみを使用した場合

被験者	正答数 (回)	正答率 (%)	誤答数 (回)	誤答率 (%)	未判定数 (回)	期待正答率 (%)
A	1667	16.5	1093	10.8	7320	52.8
B	0	0.0	0	0.0	10080	50.0
C	1154	11.4	1606	15.9	7320	47.8
D	674	6.7	5266	52.2	4140	27.2
E	941	9.3	4999	49.6	4140	29.9
F	1845	18.3	915	9.1	7320	54.6
G	998	9.9	1762	17.5	7320	46.2
H	2021	20.0	2359	23.4	5700	48.3
I	1244	12.3	1516	15.0	7320	48.7
J	6131	60.8	949	9.4	3000	75.7
平均	1667.5	16.5	2046.5	20.3	6366	48.1

表 6.14 主観的な要素のみを使用した場合

被験者	正答数 (回)	正答率 (%)	誤答数 (回)	誤答率 (%)	未判定数 (回)	期待正答率 (%)
A	2343	23.2	7737	76.8	0	23.2
B	3458	34.3	6622	65.7	0	34.3
C	1412	14.0	8668	86.0	0	14.0
D	2224	22.1	7856	77.9	0	22.1
E	2497	24.8	7583	75.2	0	24.8
F	2407	23.9	7673	76.1	0	23.9
G	1215	12.1	8865	87.9	0	12.1
H	2476	24.6	7604	75.4	0	24.6
I	1469	14.6	8611	85.4	0	14.6
J	1685	16.7	8395	83.3	0	16.7
平均	2118.6	21.0	7961.4	79.0	0	21.0

表 6.15 測定環境

OS	Ubuntu 10.04
CPU	Intel Core2 Duo
周波数	2.66GHz
使用言語	Ruby 1.9.1

第 7 章

おわりに

本論文では，在席管理において周辺情報を効果的に利用するため，複数情報源を用いた在席推定の問題点を指摘し，解決法として周辺情報の関係性抽出に相互情報量を用いる手法を提案した．次に，関係性抽出システムを導入した在席管理システムを設計，実装した．そして，関係性抽出システムを使用するにあたって周辺情報の値の取り方や閾値による出力結果と正答率の関係性の発見など，課題の検討を行った．最後に，関係性抽出システムを用いた在席推定の実験を行い，関係性抽出システムの評価を行った．実験の結果から，関係性抽出システムは信頼できる周辺情報を抽出できることがわかった．また，関係性の強い周辺情報を抽出し，関係性の弱い周辺情報は無視できることがわかった．そして，主観により在席情報に無関係な周辺情報を使用しても，在席推定への影響を排除できることがわかった．

今後の課題として，気温のような全ての値が出力されるまでに非常に長い期間を必要とする要素群の対処と，関係性抽出システムの性能改善，在席管理に関連すると思われる周辺情報の発見，主観的な要素の導入，在席管理システムの運用実験がある．また，在席推定の正答率を改善するため，関係性抽出に使用する学習方法の変更も検討する必要がある．

謝辞

本研究を進めるにあたり，懇切丁寧なご指導をして頂きました乃村能成准教授に心より感謝の意を表します．また，数々のご助言を頂きました谷口秀夫教授，山内利宏准教授，および後藤佑介助教に厚く御礼申し上げます．最後に，日頃の研究活動において，お世話になりました研究室の皆様ならびに本研究を行うにあたり，経済的，精神的な支えとなった家族に感謝いたします．

参考文献

[1] サイボウズ，“サイボウズ Office 8，” http://products.cybozu.co.jp/office/

[2] 平田敏之，國藤進，“複数情報源を用いた位置情報の補正手法の提案，” 人工知能学会全国大会論文集 (CD-ROM)，Vol.18，No.1H1-03，2004.

[3] 土持幸久，高橋伸，田中二郎，“プライバシを考慮しつつユーザの状況・状態を推定と提示を行うシステム，” 情報処理学会シンポジウム論文集，Vol.2006，No.6-1，pp.497-500，2006.

[4] 檀上正光，“複数情報源からプレゼンス情報を推定する在籍管理手法の提案，” 岡山大学情報工学科特別研究報告書，2009.

[5] 汐崎陽，“情報・符号理論の基礎，” 国民科学社，1991.

[6] Google Inc.，“Google カレンダー，” http://www.google.com/calendar/

[7] 気象庁，“過去の気象データ検索，” http://www.data.jma.go.jp/obd/stats/etrn/index.php