

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

岡山大学大学院自然科学研究科

香西 英樹

はじめに

在席管理は様々な手法がある

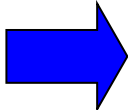
＜手法例＞

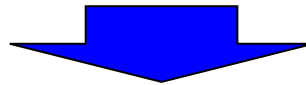
- (1) 利用者の自発的操作による手法 ... 在席表, タイムレコーダ
- (2) 自動的に行う手法 ... ビーコン
- (3) (1),(2)を併用する手法 ... RFID

どの手法も利点と欠点が共存

＜在席管理手法に対する要求＞

コストと精度のバランスを取りたい

 利用者の直接的な情報だけでなく**周辺の情報も利用したい**



複数情報源を用いた在席管理システム

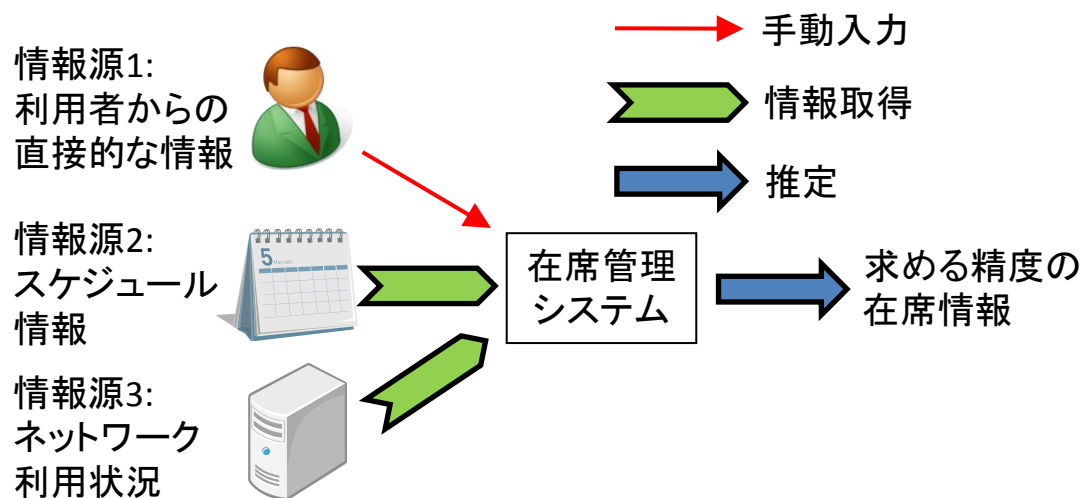
複数情報源を用いた在席管理システム

<特徴>

- (1) 複数情報源から在席情報を推定
- (2) 周囲からの情報を有効利用
- (3) 情報源を目的に合わせて選択

<情報源の例>

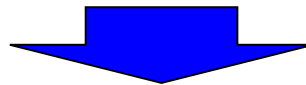
- (1) 利用者からの直接的な情報 ... 手動入力
- (2) スケジュール情報 ... スケジュール管理システム
- (3) ネットワーク利用状況 ... メール送信状況, IPアドレスリース状況



複数情報源の有効性

記録 (正答)	講義開始 14:14	講義	講義終了, 外出 (1) 15:12	帰宅 15:23			
スケジュール	講義開始 14:20	講義 15:12	講義終了 (2) 15:50				
DHCP	14:12	14:27	14:42	14:57	(3) 15:12	15:27	15:42

- (1) 被験者による記録(正答) ... 15:12講義終了
- (2) スケジュール情報 ... 15:50講義終了予定
- (3) DHCP利用情報 ... 15:12でリリースされている



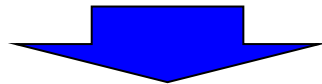
周辺情報によりスケジュール情報の誤差を軽減可能

情報抽出時の問題点

- (1) どの周辺情報の関係性が強いかは直感的に分かりづらい
在席情報に影響を与えているのかわからない
- (2) 情報抽出は主観的である
情報が無視されやすい
- (3) 人によって関係性の強い周辺情報は異なる
ある人に有効な周辺情報でも、他人にも有効とは限らない
- (4) 単体では関係性が弱い周辺情報の場合、見逃しやすい
単体で判定する場合除外されてしまう
- (5) 周辺情報個別の出現確率が考慮されていない
在席推定の結果が悪化してしまう

<問題が発生する原因>

周辺情報の関係性を客観的に判断することができない



計算機による機械学習を用いて情報抽出を行う

相互情報量を用いた情報抽出

<相互情報量>

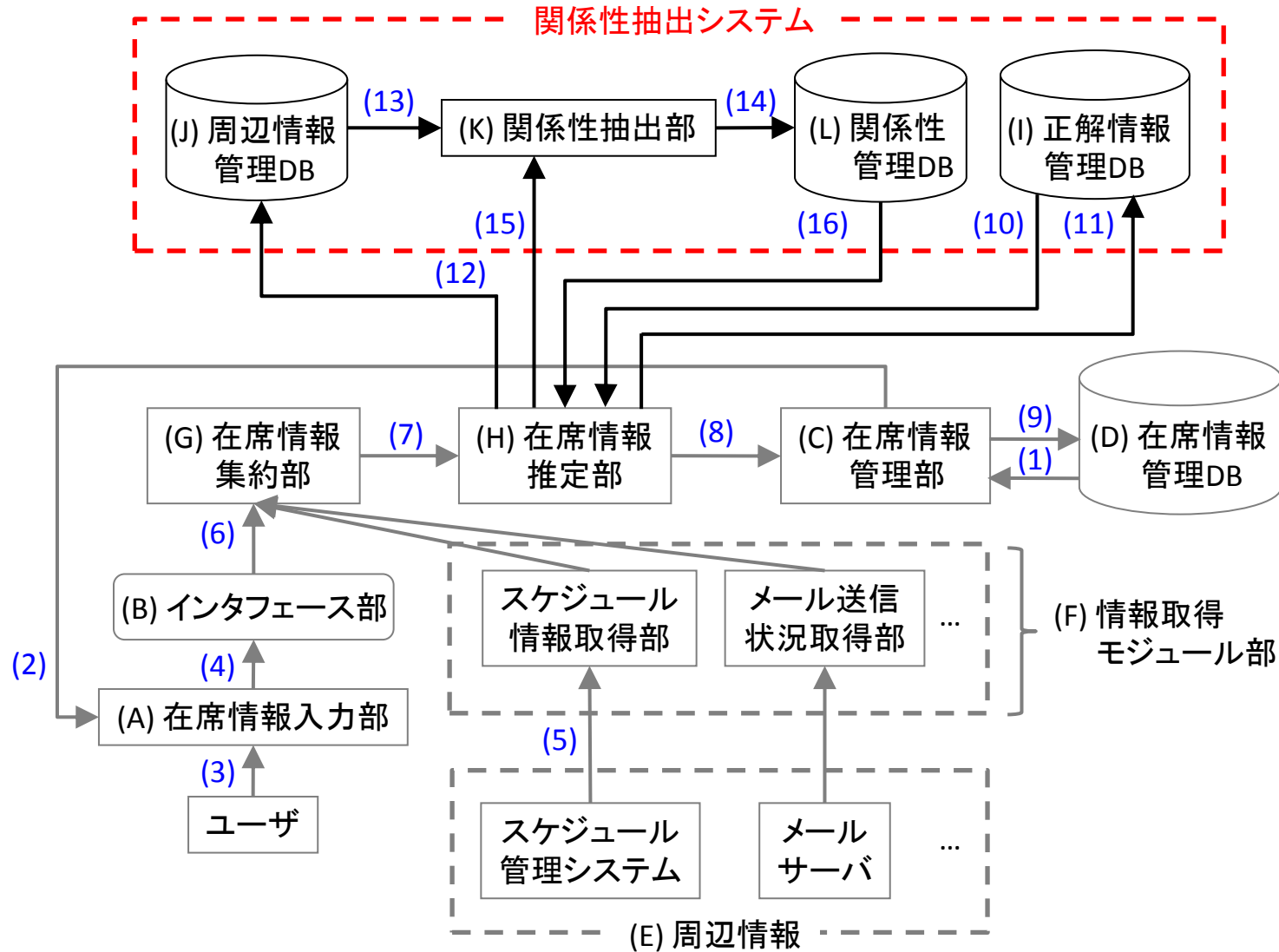
- (1) 2種類の確率変数の相互依存の尺度を表す量
- (2) 関係性を偏見なく示すことが可能

<問題の解決法>

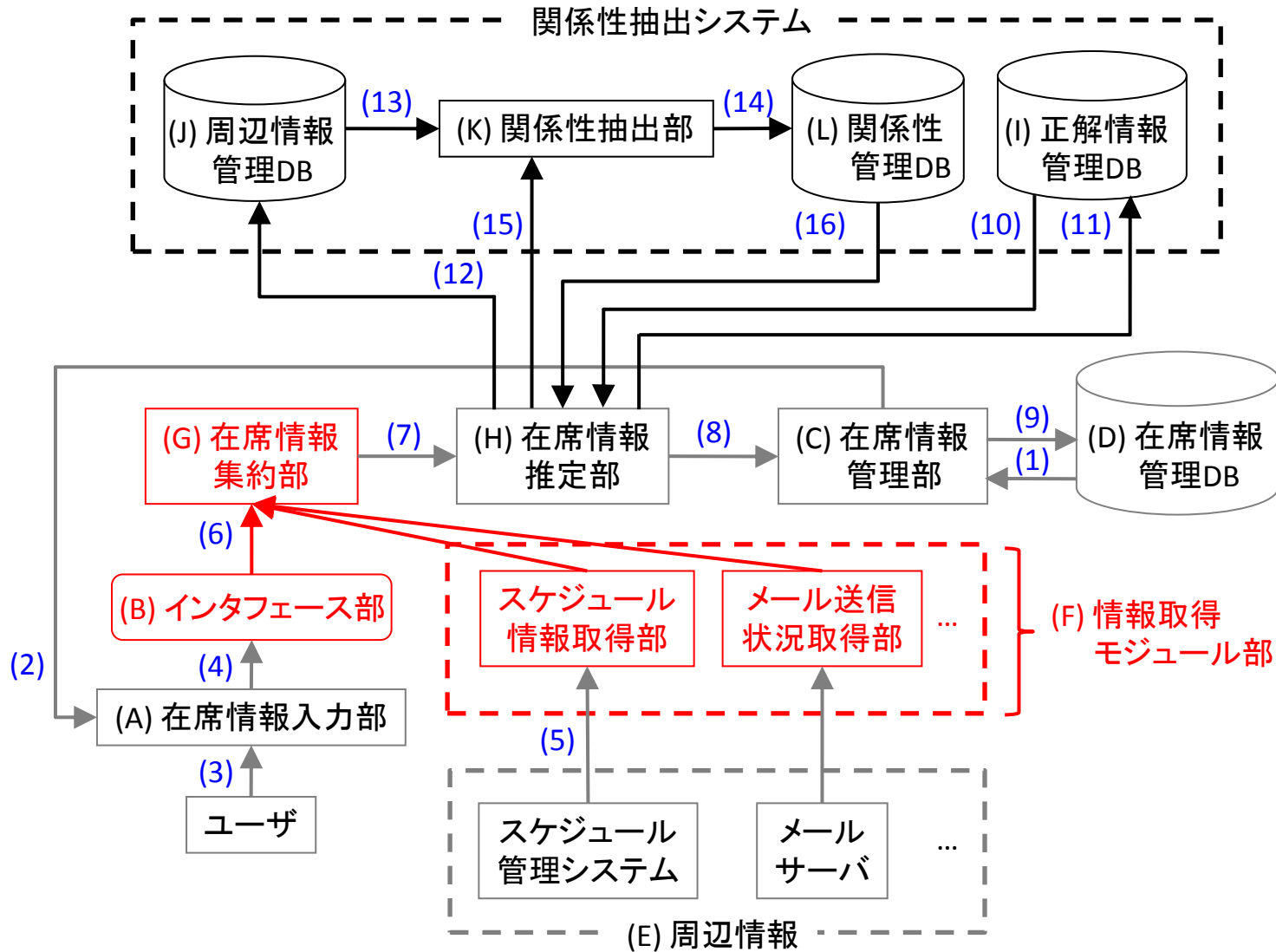
- (1) どの周辺情報の関係性が強いかは直感的に分かりづらい
関係性の数値化により, 関係性の強さを判断する
- (2) 情報抽出は主観的である
計算機による客観的な判断で主観を交えることなく判断する
- (3) 人によって関係性の強い要素は異なる
人物別に学習データを取得して判断する
- (4) 単体では関係性が弱い周辺情報の場合, 見逃しやすい
周辺情報同士の関係性も判断する
- (5) 周辺情報個別の出現確率が考慮されていない
在席結果に対する関係性の強さを数値化する

関係性抽出システムを在席管理システムに導入

関係性抽出システムを導入した在席管理システム

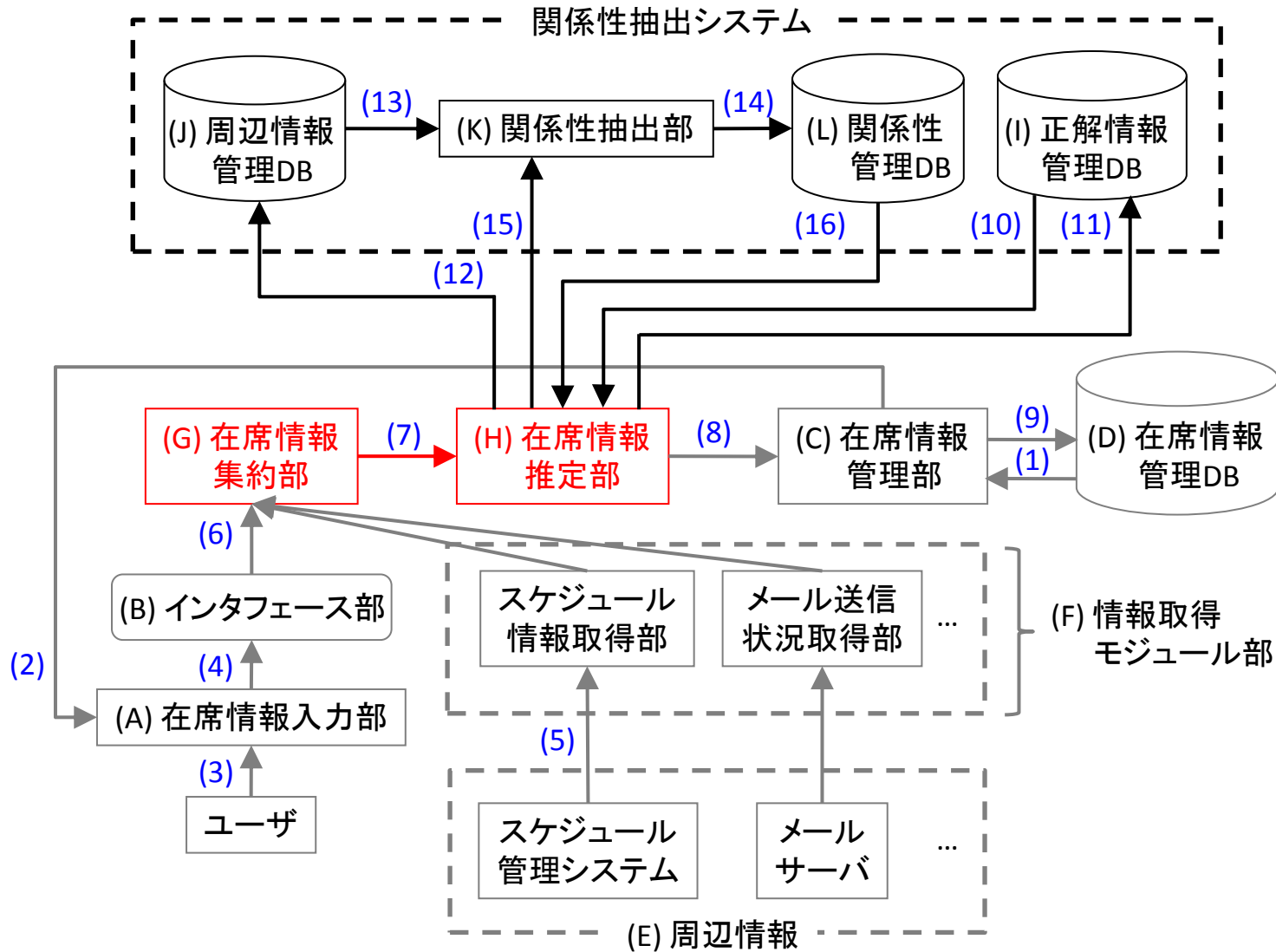


在席推定の流れ



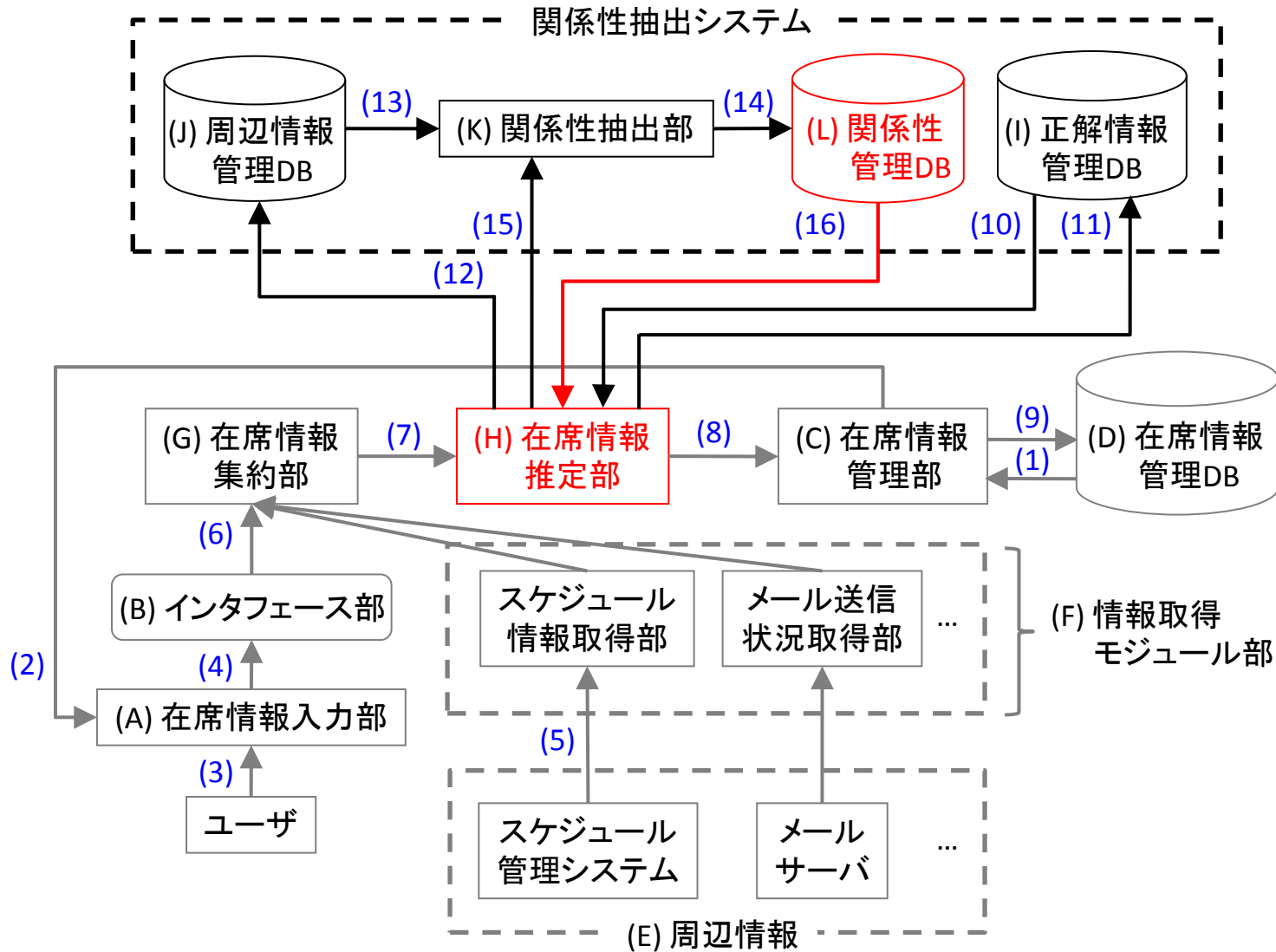
(a) 周辺情報を抽出する

在席推定の流れ



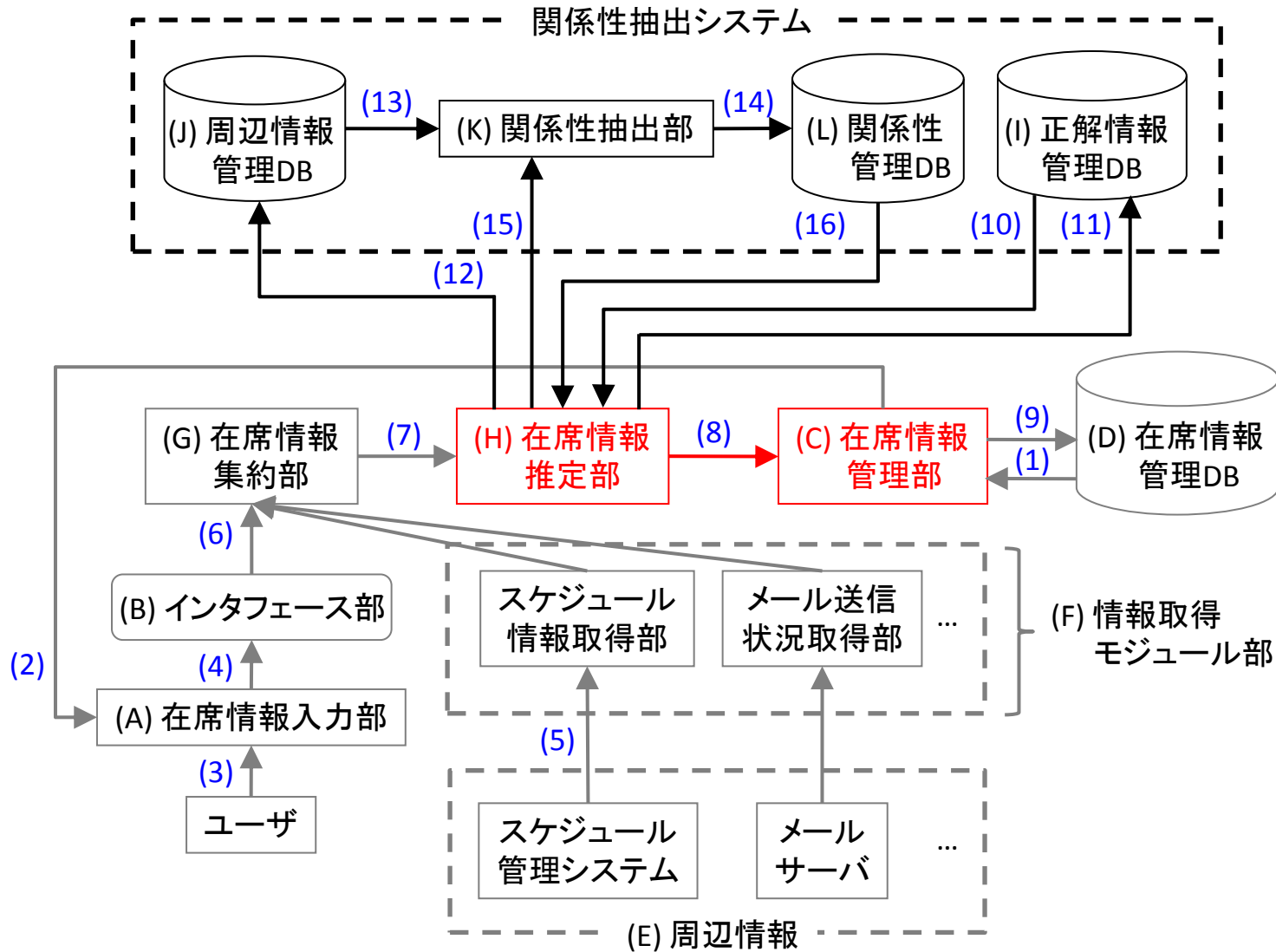
(b) 在席情報推定部にデータを渡す

在席推定の流れ



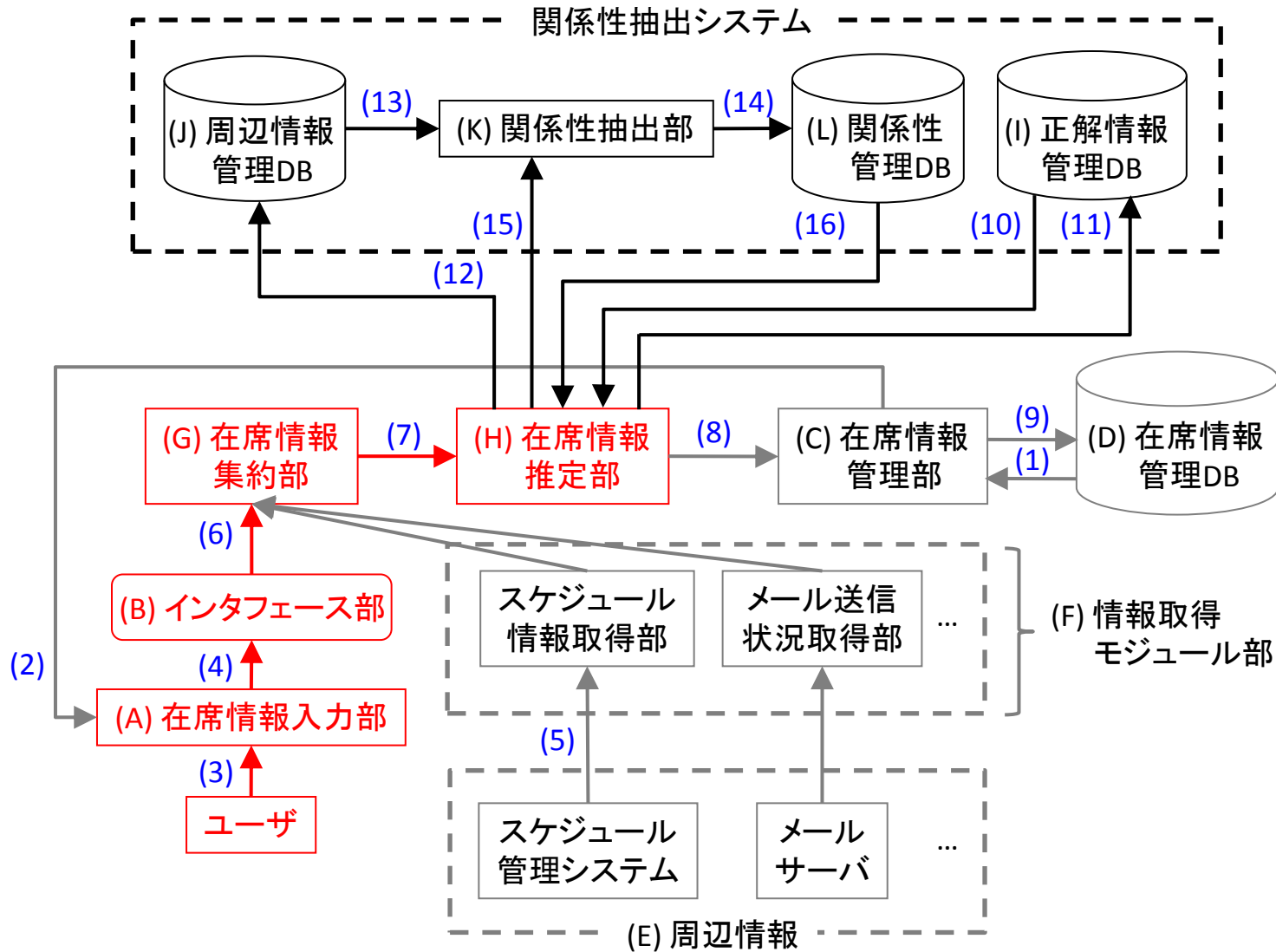
(c) ルールを参照し在席状態を推定する

在席推定の流れ



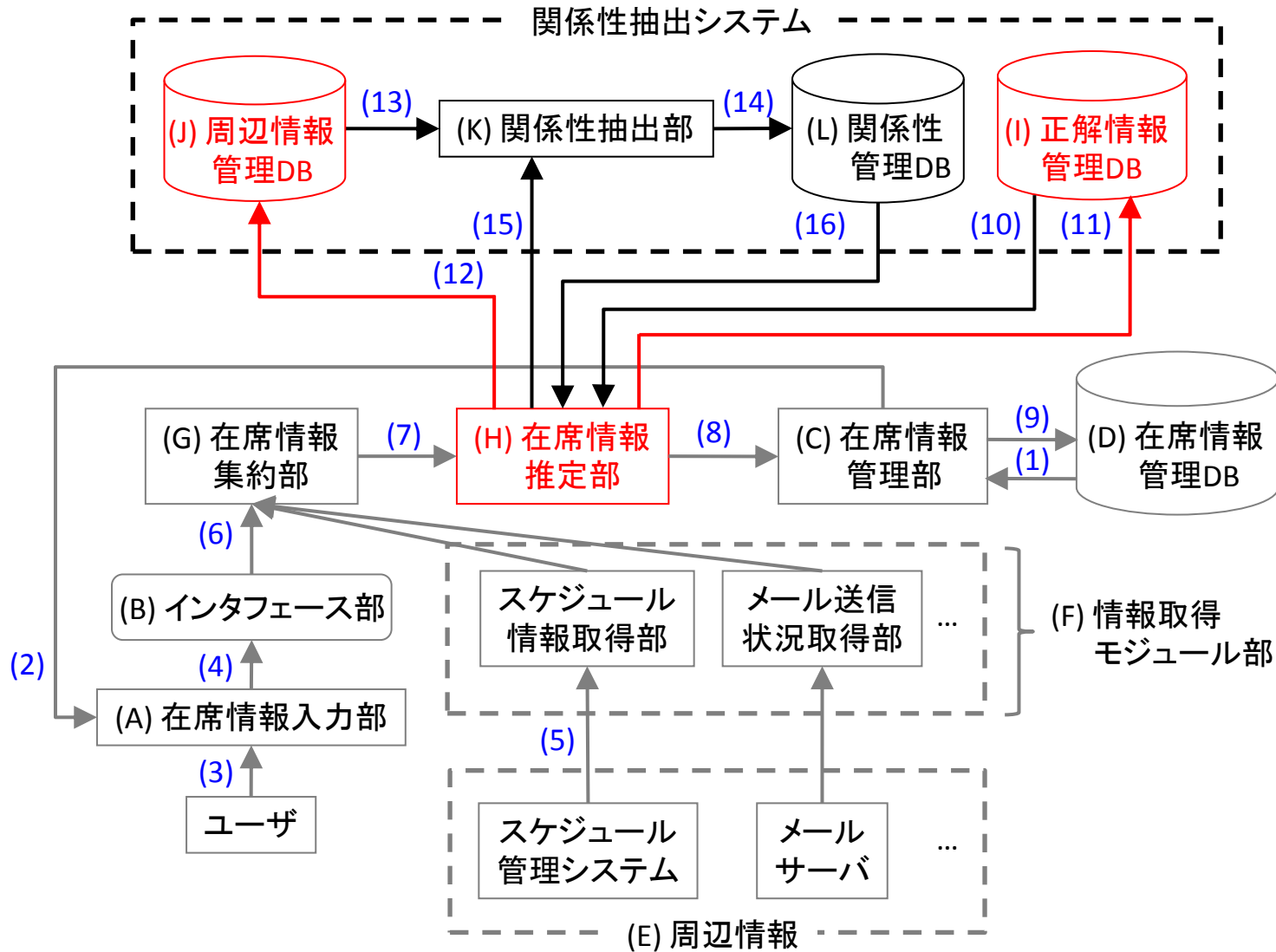
(d) 推定結果を通知する

学習データ取得の流れ



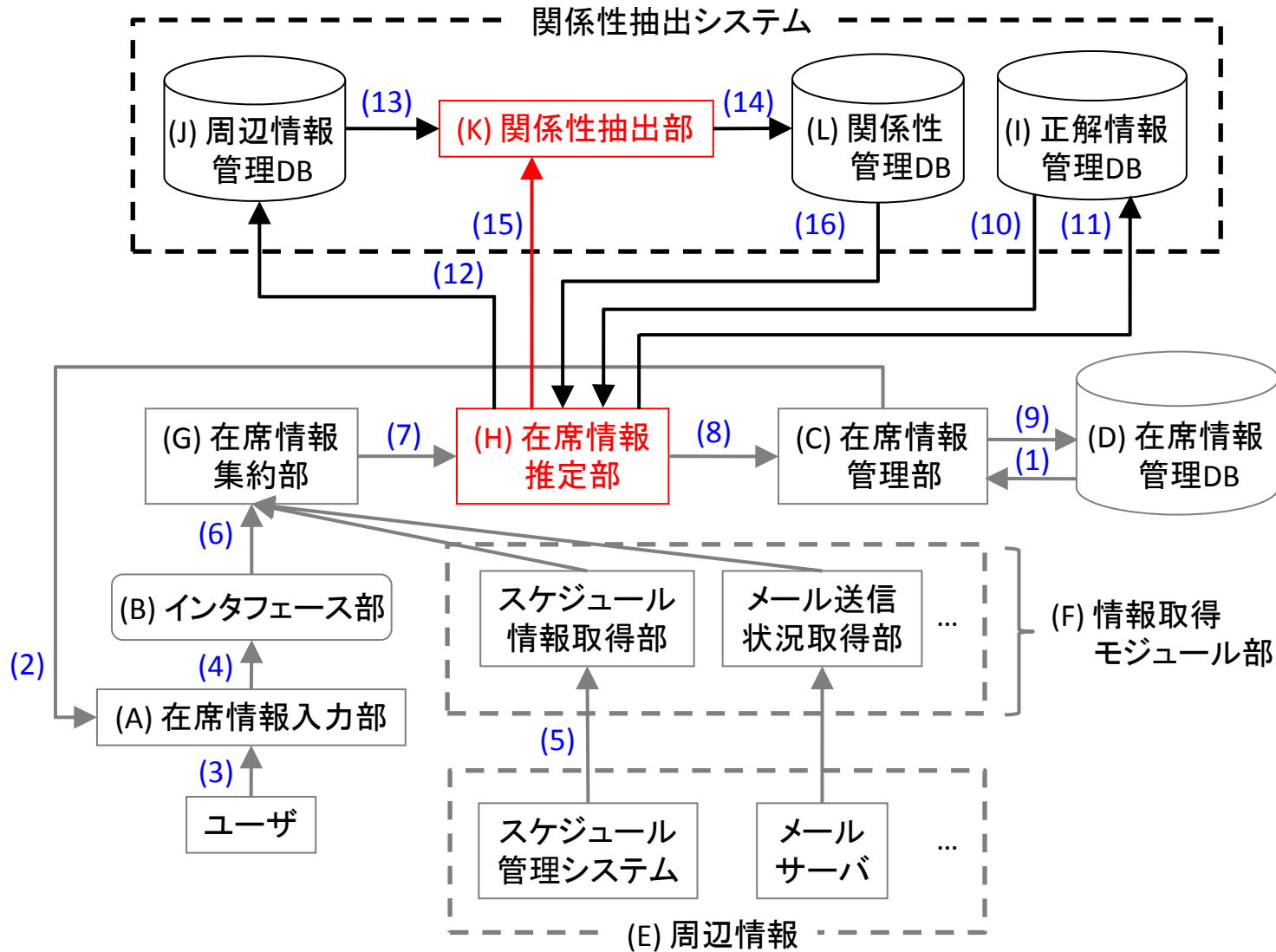
(a) 在席情報をユーザに評価してもらい、
評価結果を教師値として通知する

学習データ取得の流れ



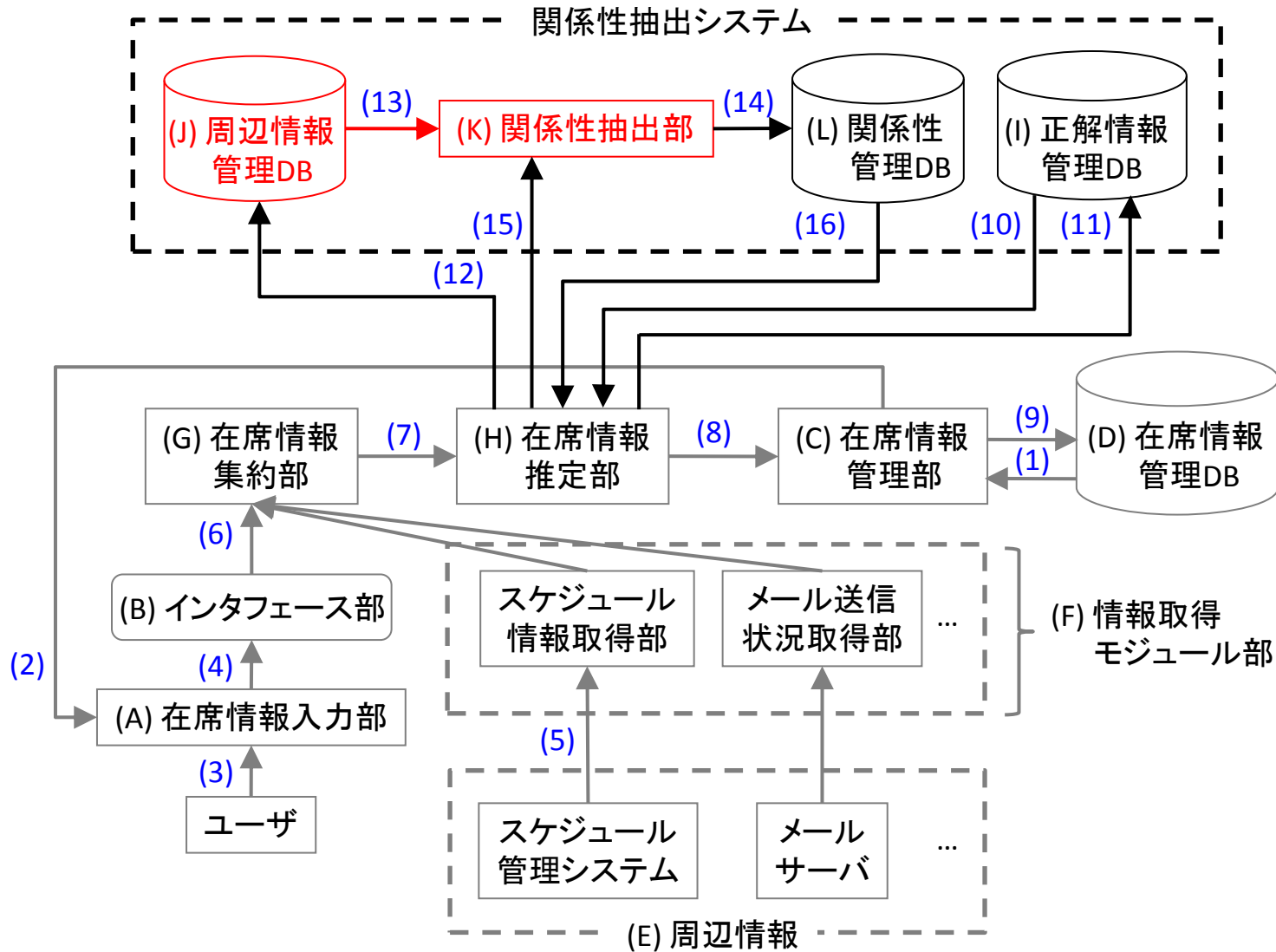
(b) 学習データとして記録する
また、推定結果の正誤を判断し記録する

ルール抽出処理の流れ



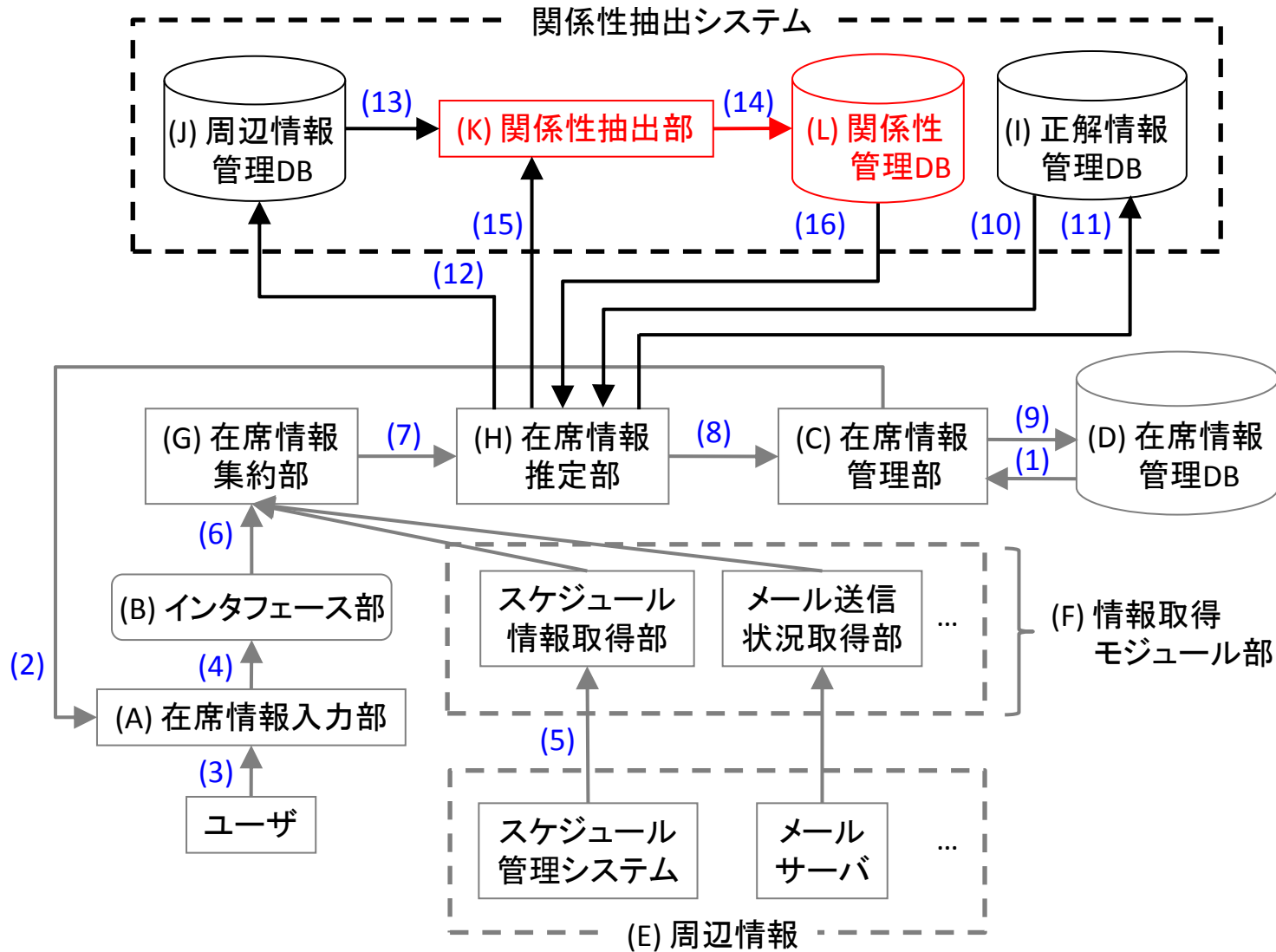
(a) ルール抽出を依頼する

ルール抽出処理の流れ



(b) データを取得する

ルール抽出処理の流れ



評価実験

<実験内容>

- (1) 同じ研究室に所属する10人を対象に実験
- (2) 2週間分の在席情報を取得
- (3) 前半1週間を学習データとして使用し、後半1週間の情報に対し
推定実験を実行
- (4) 判定する在席情報は「在席」と「不在」の2種類

<使用した周辺情報と値の区分>

- (1) 曜日 ... 曜日毎(7種類)
- (2) メール ... 10分以内にメールが送信されたかどうか(2種類)
- (3) スケジュール... 現在予定中かどうか(2種類)
- (4) 気温 ... 5°C区切り(11種類)
- (5) 天気 ... 晴れ, 曇り, 雨, 雪, その他(5種類)
- (6) DHCP ... 15分以内にリースされたかどうか(2種類)
- (7) 時刻 ... 時間毎(24種類)

実験条件と結果

<実験結果>

被験者	正答数(回)	正答率(%)
A	2343	23.2
B	3458	34.3
C	1412	14.0
D	2224	22.1
E	2497	24.8
F	2407	23.9
G	1215	12.1
H	2476	24.6
I	1469	14.6
J	1685	16.7
平均	2118.6	21.0

<閾値>

- (1) エントロピー
...0.99ビット
- (2) 相互情報量
...0.5ビット
- (3) 組み合わせ回数
...5回
- (4) ルール出現確率
...30パーセント

<試行回数>

1人あたり10080回
...1分に1回推定

周辺情報の組のみを使用した場合

＜推定結果が悪い原因＞

単体でも信頼できる周辺情報が多く抽出されていた

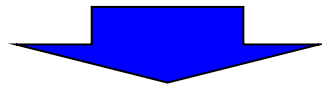
限られた期間では周辺情報が出力する値の範囲は**限定的信頼できない周辺情報も有用と判定していた**と考えられる

＜実際のデータ＞

- (1) 在席ルールと不在ルールのほとんどが共通
- (2) 在席ルールは不在ルールより多い
半分以上の推定において在席と推定したと思われる
- (3) ほとんどの被験者は不在時間が在席時間より長かった

＜周辺情報の組のみを使用した場合＞

平均正答率は48.1%に向上(+27.1%)

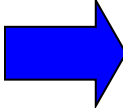


学習データ量が少ない場合は**周辺情報の組のみを使用すべき**である

主観的な周辺情報のみを使用した場合

<情報抽出の問題点>

情報抽出は主観的である

 主観的な周辺情報のみを使用しても影響を排除できるか実験

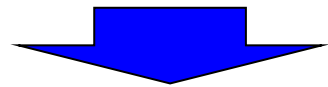
<使用した要素>

- (1) 曜日
- (2) メール
- (3) スケジュール
- (4) DHCP

<使用しなかった要素>

- (1) 気温
- (2) 天気
- (3) 時刻

実験結果...変化なし



在席情報に無関係な要素群を排除できたといえる

まとめ

<複数情報源から在席情報を推定する在席管理手法>

- (1) 複数情報源の関係性抽出に相互情報量を用いる手法を提案
- (2) 関係性抽出システム実装に必要な用語を定義
- (3) 関係性抽出システムを使用した在席管理システムを実装
- (4) 関係性抽出システムを使用する際の諸問題を検討
- (5) 関係性抽出システムを用いた在席推定実験の実行
- (6) 各種検討事項の評価

<残された課題>

- (1) 長期的な情報収集を必要とする要素群の対処
- (2) 関係性抽出システムの性能改善
- (3) 在席管理に関連すると思われる周辺情報の発見
- (4) 主観的な要素の導入
- (5) 在席管理システムの運用実験
- (6) 関係性抽出に使用する手法の変更