

利用者の特徴を考慮した メール分類機構の組み合わせ法

山本 泰隆 乃村 能成 谷口 秀夫

岡山大学 大学院自然科学研究科

発表構成

1. 組み合わせ機構の概要
2. 判定規則の動的な適応
3. システム設計
4. 発表のまとめ

はじめに

迷惑メールの増加が問題になっている

➡ メール分類機構が提案され、利用されている

- 正当メールと迷惑メールを分類
 - (1) 受信側メールサーバ上で動作
 - (2) 利用者宛のメールが迷惑メールか否かを判定
 - (3) 迷惑と判定したメールを利用者に配達しない
- 迷惑メール対策のひとつとして広く普及
メール分類機構の代表例
 - ISPの提供する迷惑メール対策サービス
(例) Yahooメール、Mail@nifty
 - メール処理ソフトウェア
(例) procmail、SpamAssassin

背景

迷惑メールの送信手口の巧妙化

➡ 単独のメール分類機構の利用

十分な精度でのメールの分類は困難になりつつある



(問題1) 正当メールの誤判定の発生

利用者に配達されるべき正当メールが配達されない

(問題2) 利用者の設定工数、メールサーバの処理負荷の増加



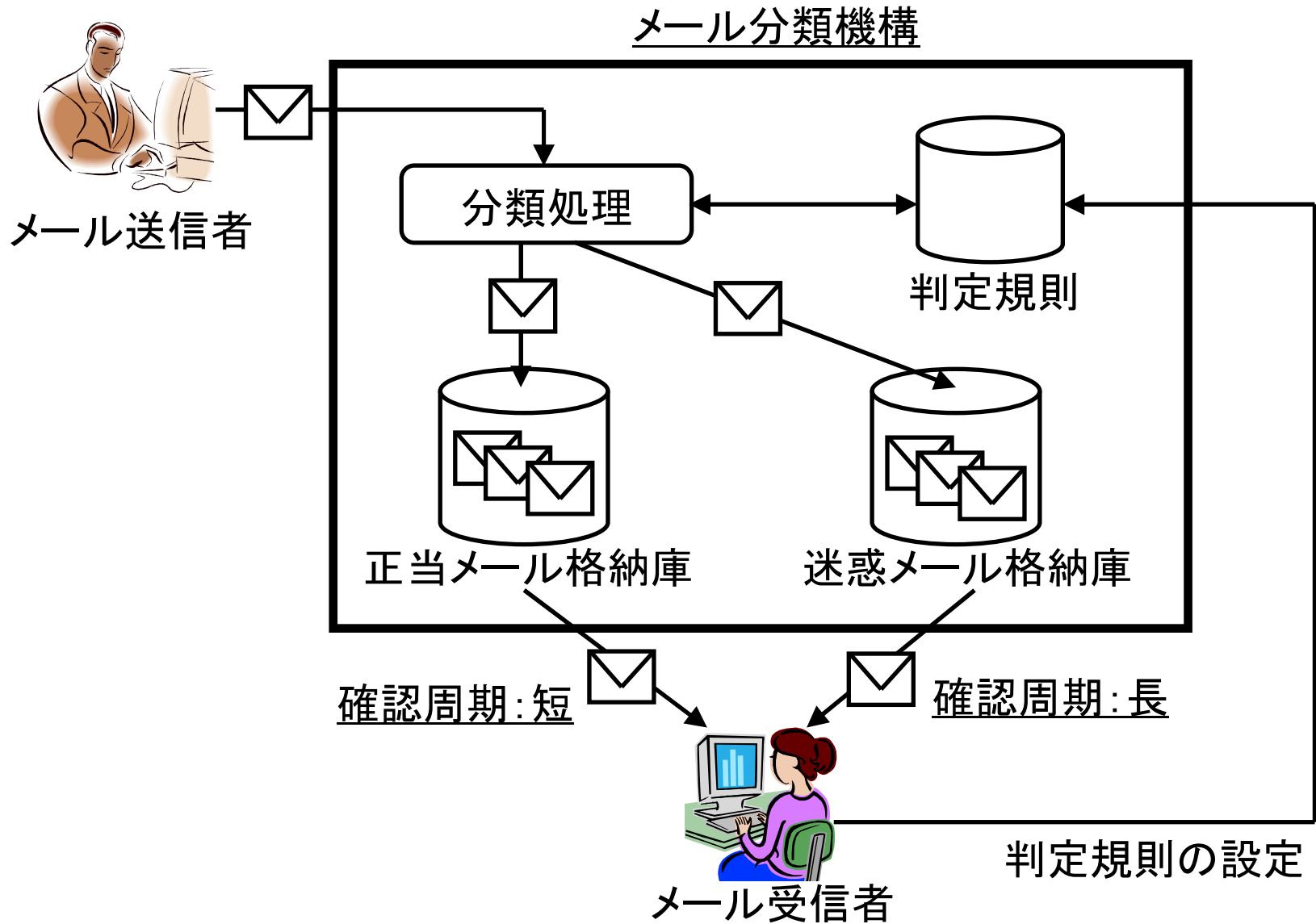
組み合わせ機構の提案

基本的な考え方: 組み合わせによる分類精度の向上

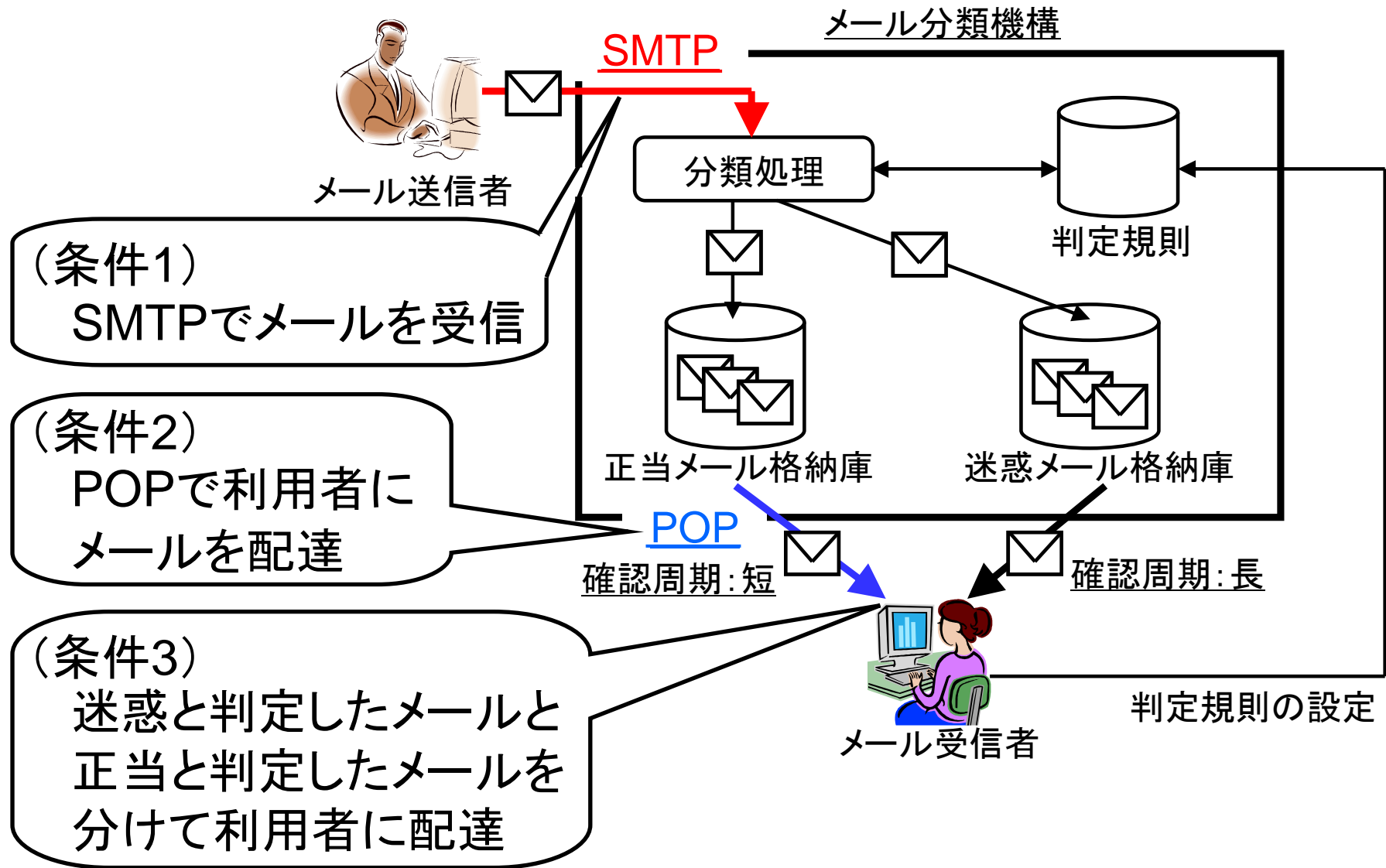
単独のメール分類機構の分類精度は不十分

- メール分類機構を複数利用
- 各メール分類機構の示す判定結果から、メールが迷惑メールか否かを判定

メール分類機構の構成

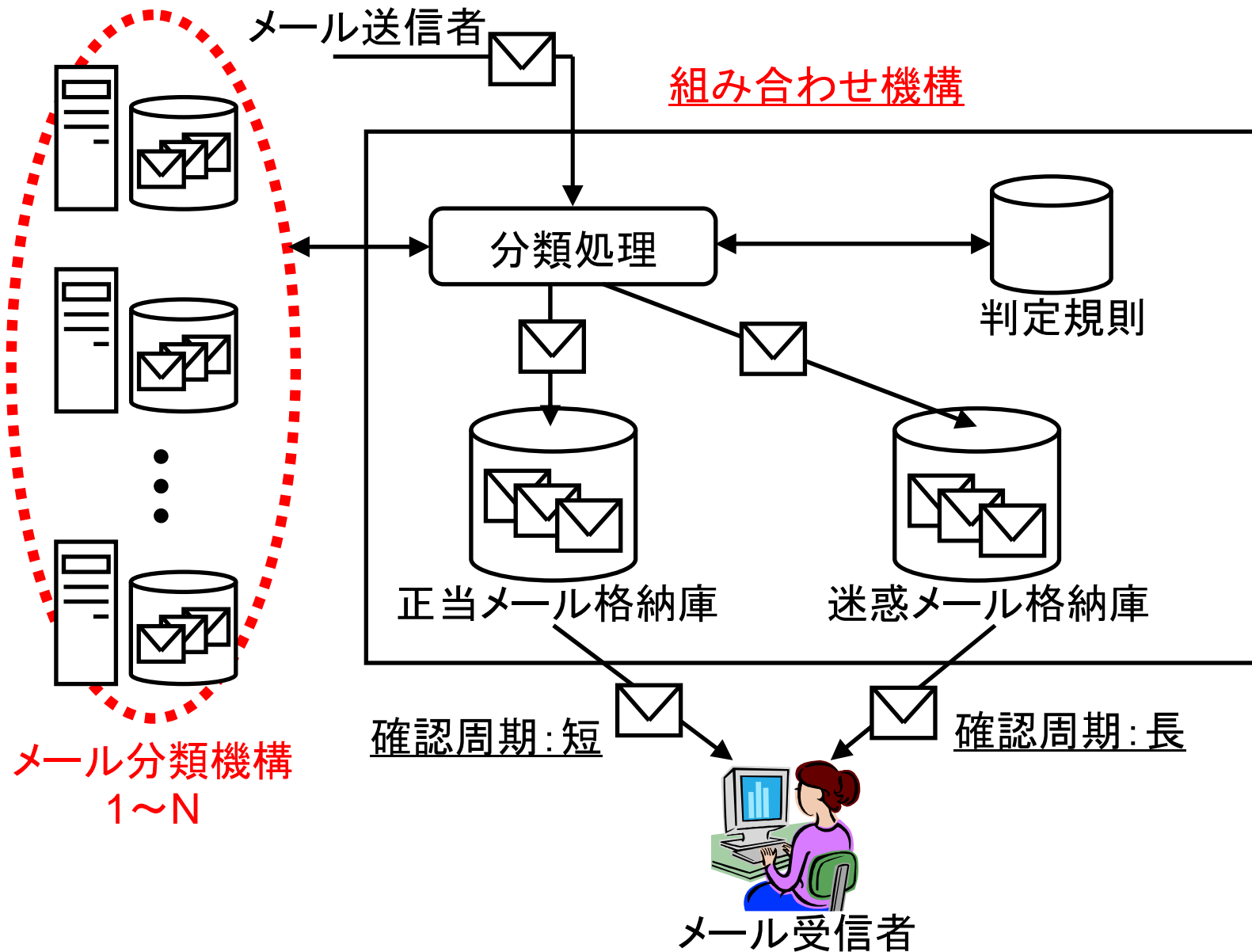


組み合わせ可能なメール分類機構の条件

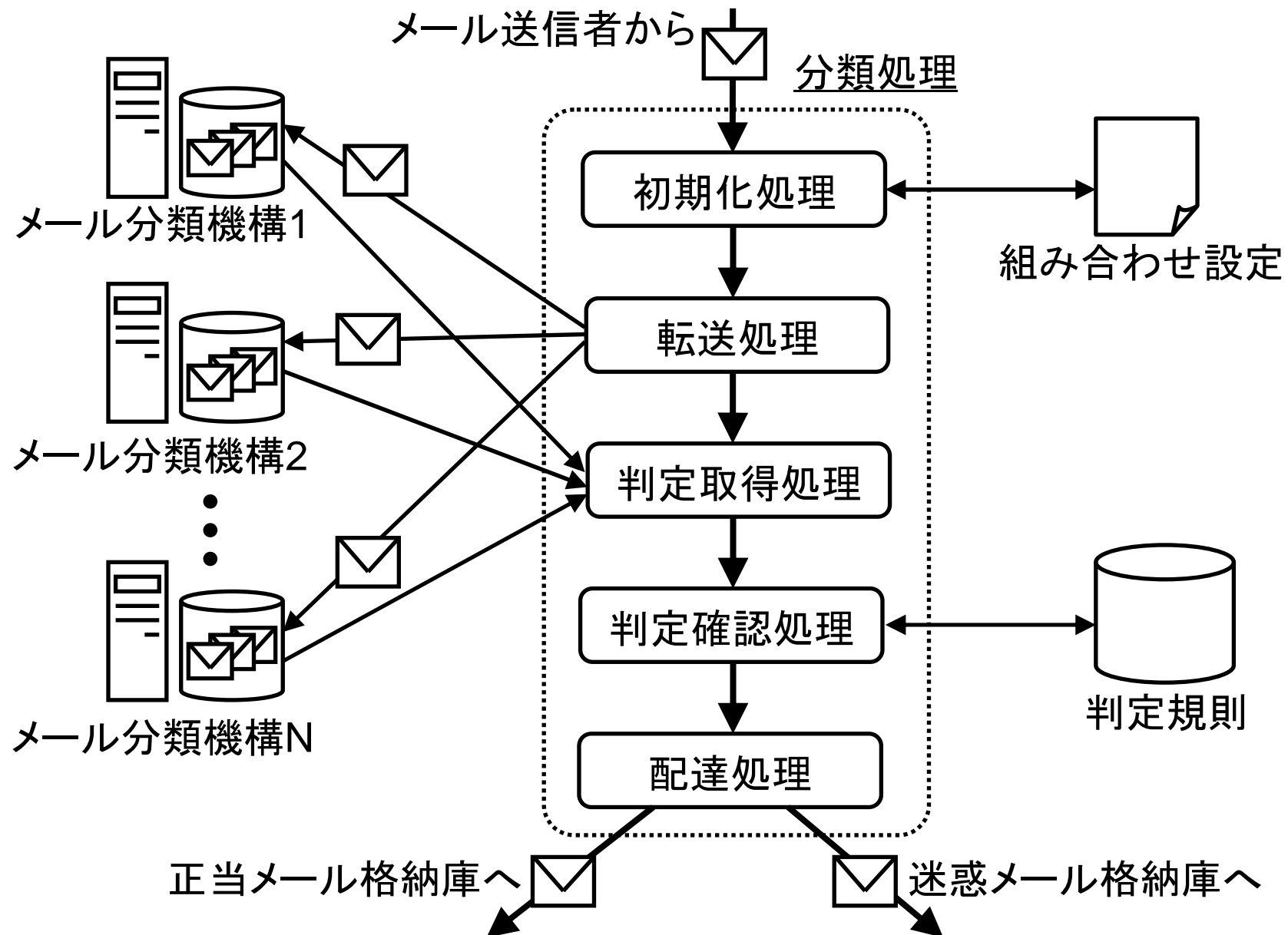


➡ 多くのメール分類機構がこれら条件を満たす

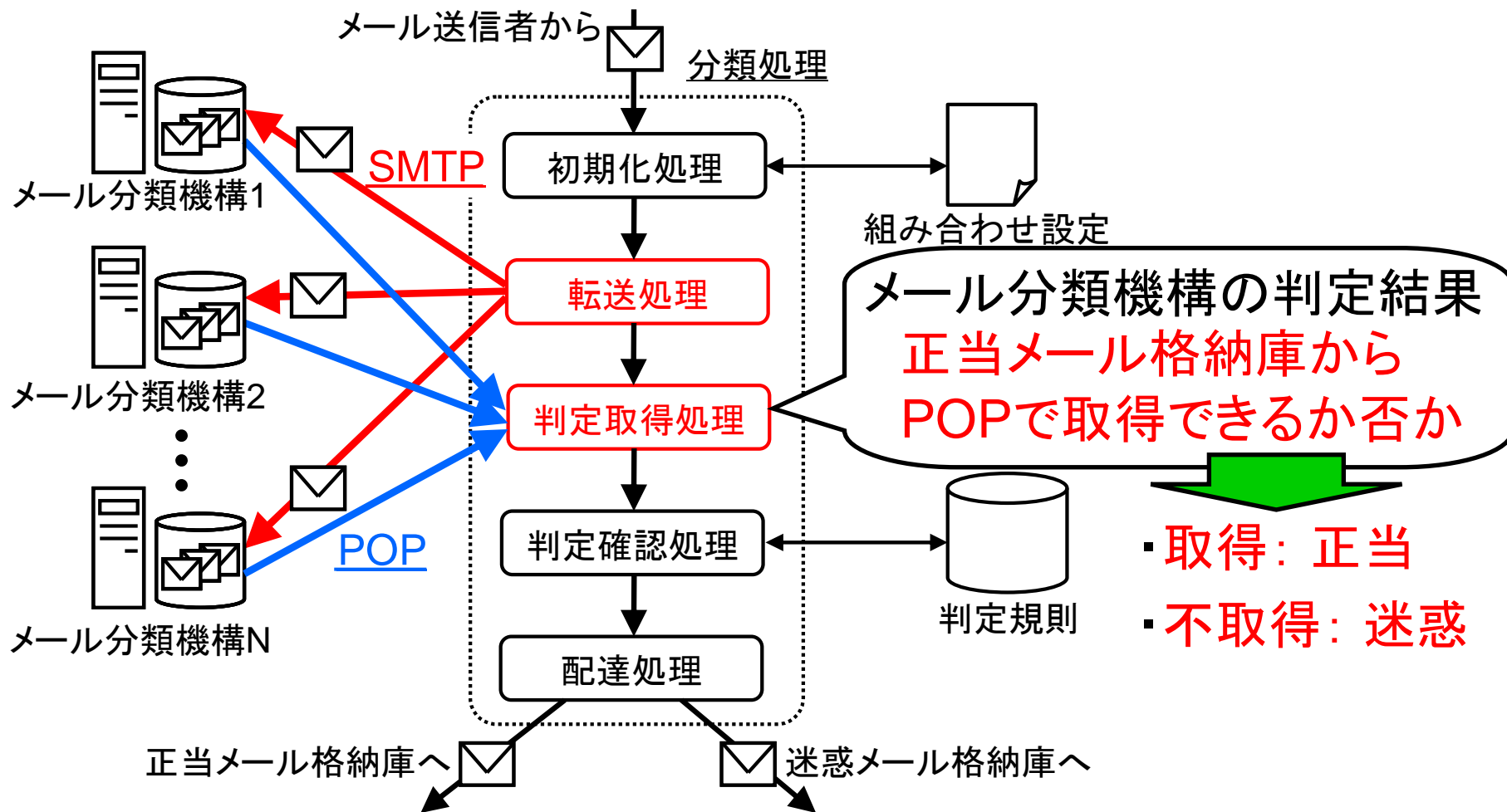
提案機構：組み合わせ機構



組み合わせ機構の分類処理



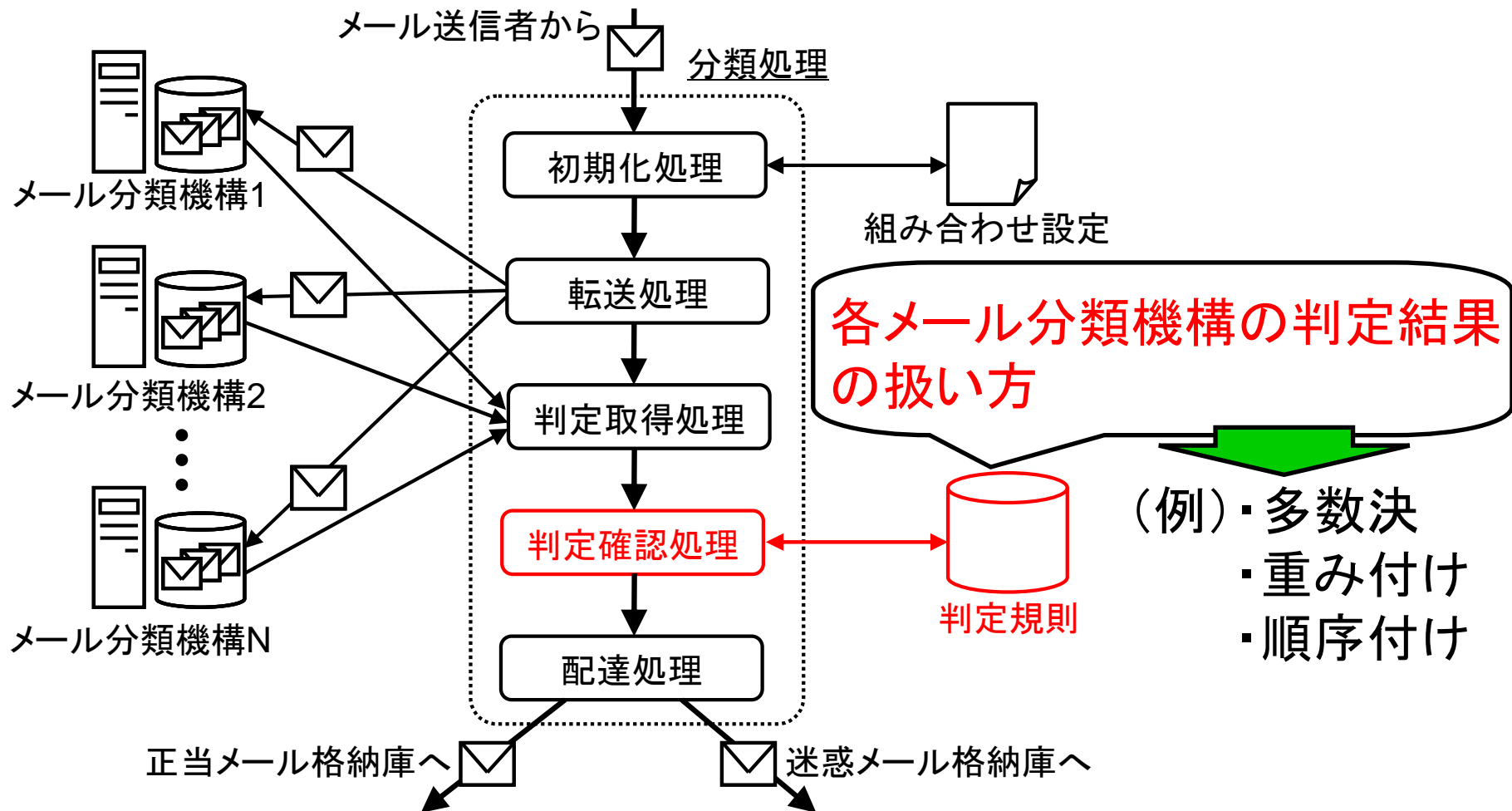
メール分類機構の判定結果の確認



判定結果の確認: SMTPとPOPを用いる

➡メール分類機構の内部構成を変更することなく利用可能

組み合わせ機構の判定規則



発表構成

1. 組み合わせ機構の概要
2. 判定規則の動的な適応
3. システム設計
4. 発表のまとめ

多数決を用いたメール判定の評価結果

組み合わせ機構の分類精度を評価

各メール分類機構の判定結果から、**多数決**でメールを判定

(結果1) 単独では効果が不十分なメール分類機構が含まれていても、高い分類精度を維持できる

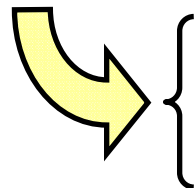
(結果2) 利用の初期から安定した分類精度が得られる



多数決の効果

個々のメール分類機構が持つ判定の偏りを平均化できる

➡ 相互の違いを埋めるための設定作業が不要



- 各メール分類機構の判定規則
- 利用者個々の嗜好にもとづく判断基準



利用者の初期導入の手間を軽減

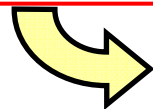
判定規則の動的な適応

動的な適応の必要性

- **利用者の嗜好は時間とともに変化する**
特定のメール分類機構の分類精度は極端に悪くなる場合がある
- **個々の利用者の嗜好は異なる**
効果的なメール分類機構の組み合わせは利用者ごとに異なる



判定規則に動的な適応性を持たせる



利用者の嗜好を反映した組み合わせの最適化

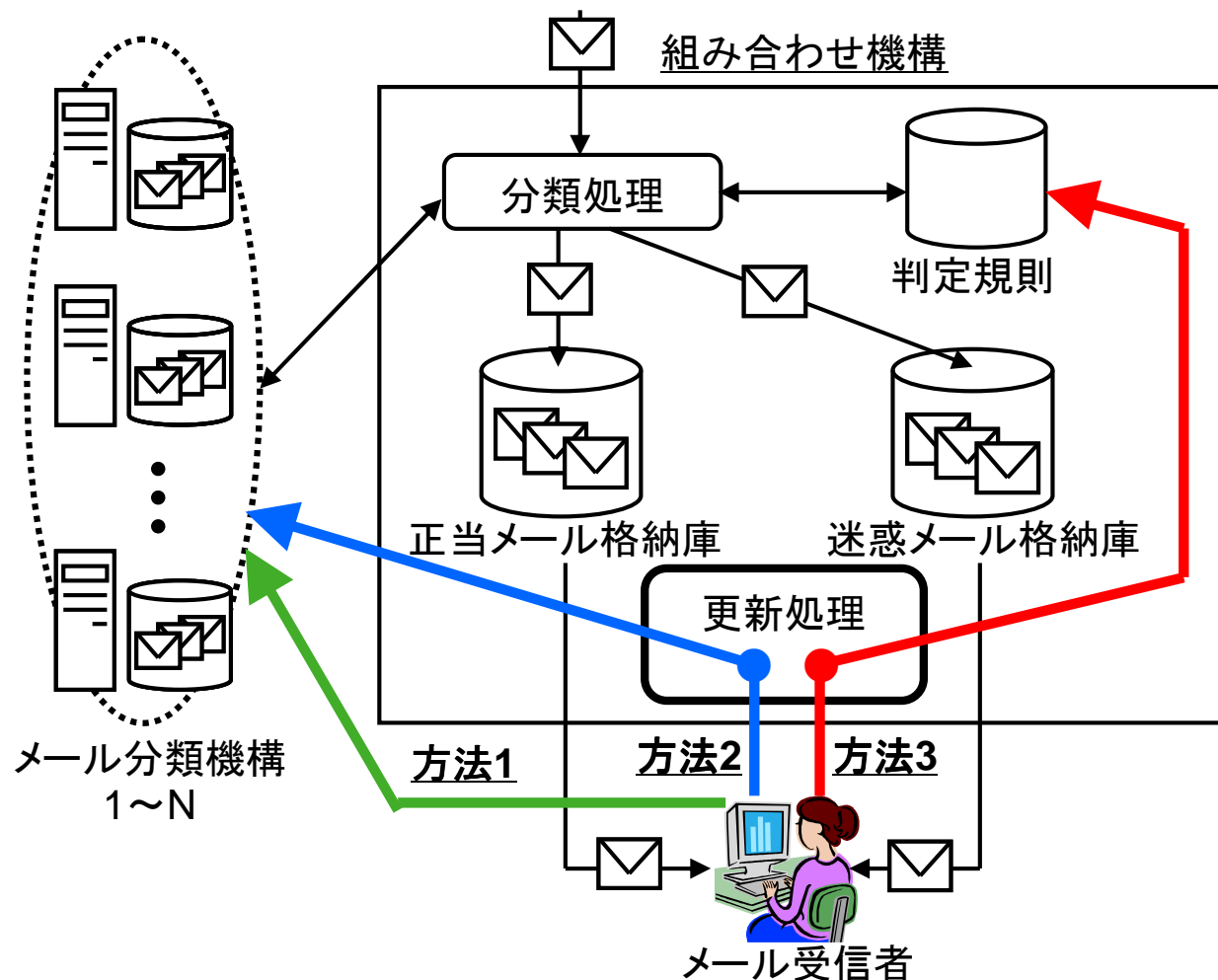
(例) 各メール分類機構の判定結果に対する重み付け
重みは各メール分類機構の分類精度で変化



利用者の嗜好を知らせる何らかの方法が必要

利用者の嗜好の反映方法

- (方法1) 判定規則を直接更新する方法
- (方法2) 組み合わせ機構を中継する方法
- (方法3) 組み合わせ機構で処理を完結する方法



各方法の比較

(方法1) 判定規則を直接更新する方法

(方法2) 組み合わせ機構を中継する方法

(方法3) 組み合わせ機構で処理を完結する方法

	利用者の手間の 少なさ	機構の処理の 簡潔さ	高度な判定規則 の実現
方法1	×	○	○
方法2	○	×	○
方法3	○	△	△

■ (方法1)、(方法2)について

- 高度な判定規則を実現可能
- それぞれ利用者の手間、処理の複雑さが欠点

■ (方法3)について

- 利用者の手間の少なさ、処理の簡潔さを満足
- 利用者の嗜好を反映したメールの判定は可能

(方法3)を採用

反映方法の事例

各メール分類機構の判定結果に対する重み付け

利用者ごとにメール分類機構への重み付けを変化

➡ 利用者の嗜好を組み合わせ機構のメール判定に反映

組み合わせ機構の判定結果の事例

利用者 j 宛のメールに対する組み合わせ機構の判定結果

$$D_j = \sum_i a_{ij} x_i \quad (D_j \text{が正ならば正当、負ならば迷惑と判定})$$

- x_i : メール分類機構 i の判定結果 (1/-1)
- a_{ij} : 利用者 j 宛のメールに対する分類機構 i の判定結果の重み

重みの計算方法

重み a_{ij} は各メール分類機構の**分類精度**をもとに決定

$$\text{重みの計算式: } a_{ij} = \frac{R_l + R_s}{2} \quad \left\{ \begin{array}{l} \bullet R_l: \text{正当メール判定率} \\ \bullet R_s: \text{迷惑メール判定率} \end{array} \right.$$

正当メール判定率 R_l 、迷惑メール判定率 R_s の計算方法:

• R_l の計算式:

$$R_l = \frac{L_1}{L_1 + L_2}$$

• R_s の計算式:

$$R_s = \frac{S_1}{S_1 + S_2}$$

利用者の判断	メール分類機構の判定	
正当	正当	L_1
正当	迷惑	L_2
迷惑	迷惑	S_1
迷惑	正当	S_2



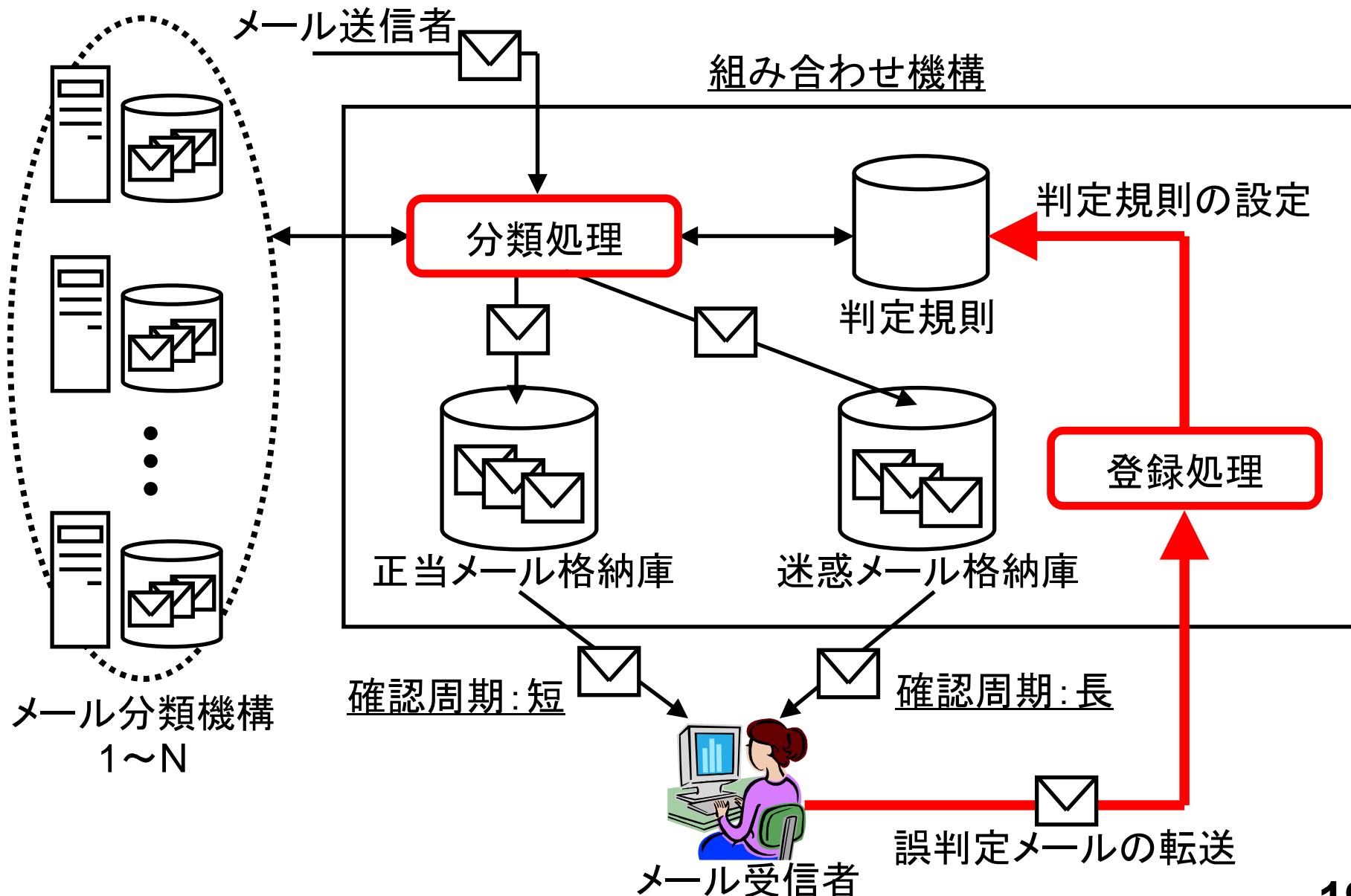
利用者からメールの正しい分類情報を受け取る必要がある

➡ 組み合わせ機構の**登録処理**により実現

発表構成

1. 組み合わせ機構の概要
2. 判定規則の動的な適応
3. システム設計
4. 発表のまとめ

各処理と利用者の連携



システム設計

登録処理の設計

利用者からの報告方法

■ メールを利用

- ・・・ {
 - 利用者の従来使用しているメーラを利用可能
 - メールを送受信のみで利用可能

■ 組み合わせ機構が誤判定したメールのみを登録処理に転送

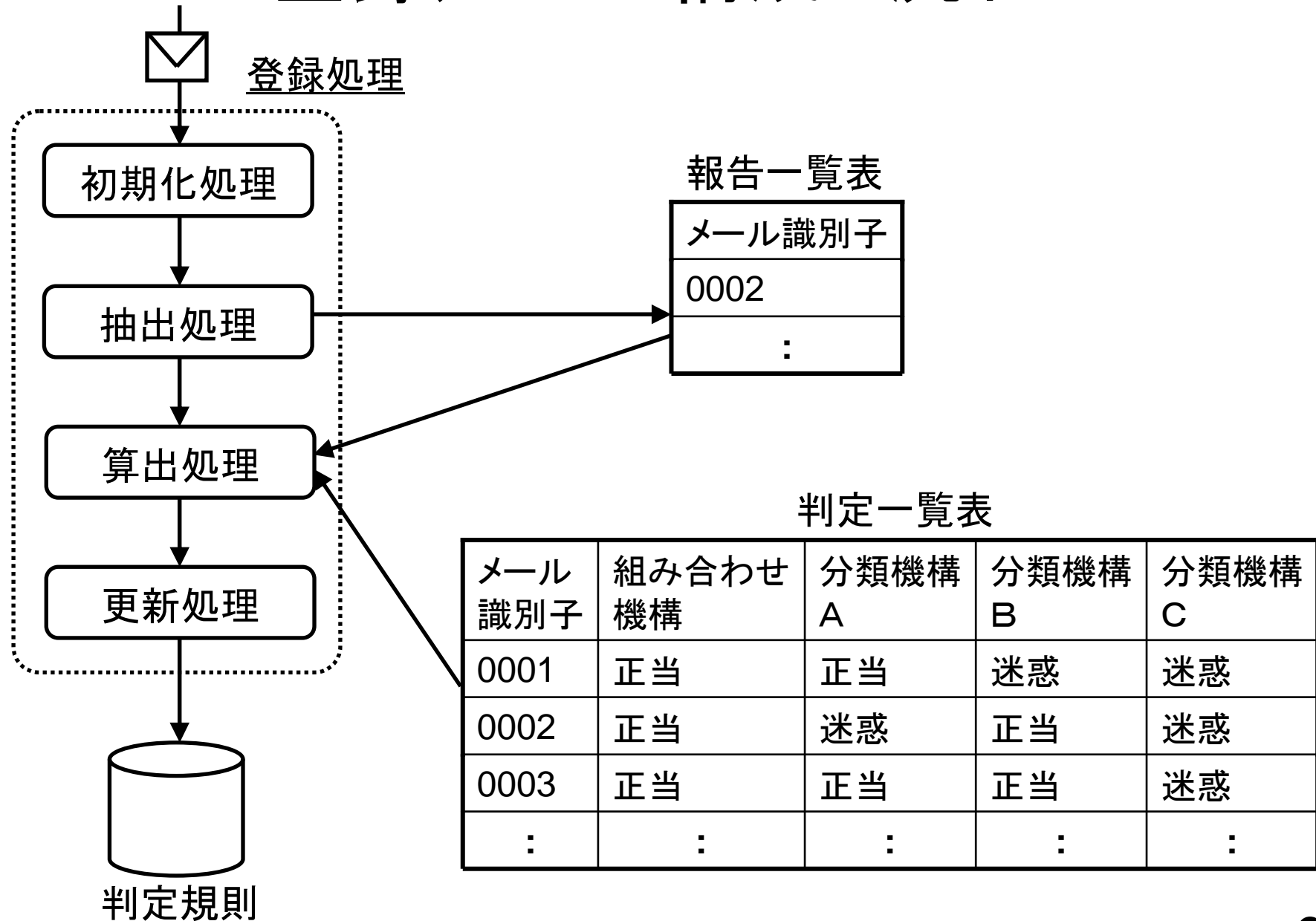
➡ 報告に要する利用者の手間を抑制

課題と対処

- 利用者から転送されたメール
 - 組み合わせ機構の判定したメール
- } 対応の一意な識別が必要

➡ 利用者へのメール配達時に識別子をメールに付加

登録処理の構成と流れ



おわりに

組み合わせ機構

- メール分類機構を複数利用
- 各メール分類機構の示す判定結果から、メールが迷惑メールか否かを判定

動的な判定規則の必要性

利用者の嗜好を反映した組み合わせの最適化

- 利用者の嗜好は時間とともに変化
- 利用者間で嗜好は異なる

利用者の嗜好の反映方法とシステム設計

各メール分類機構の判定結果に対しての重み付け

残された課題

登録処理の組み合わせ機構への実装、分類精度の評価