

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2008-135926
(P2008-135926A)

(43) 公開日 平成20年6月12日(2008.6.12)

(51) Int.Cl.	F I	テーマコード (参考)
HO4L 12/58 (2006.01)	HO4L 12/58 100F	5K030
GO6F 13/00 (2006.01)	GO6F 13/00 610Q	

審査請求 未請求 請求項の数 8 O L (全 10 頁)

(21) 出願番号	特願2006-320004 (P2006-320004)	(71) 出願人	304020177
(22) 出願日	平成18年11月28日 (2006.11.28)		国立大学法人山口大学
			山口県山口市吉田1677-1
		(72) 発明者	杉井 学
			山口県山口市吉田1677-1
		(72) 発明者	松野 浩嗣
			山口県山口市吉田1677-1
		Fターム(参考)	5K030 GA15 HA06 LC15 LD17

(54) 【発明の名称】 迷惑メールのフィルタ機能を有する電子メールシステム

(57) 【要約】

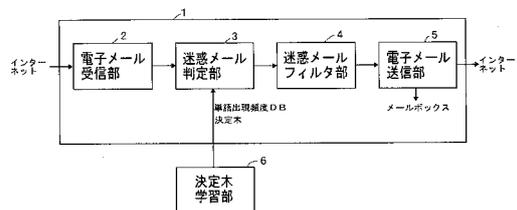
【課題】

学習型の決定木アルゴリズムを用いて、迷惑メールを正確に効率良くフィルタリングする。

【解決手段】

電子メール受信部2と、迷惑メール判定部3と、迷惑メールフィルタ部4と、電子メール送信部5とを有する電子メールシステム1であって、迷惑メール判定部は、決定木学習部6によって予め生成された単語出現頻度データベースにより電子メール中の全ての単語を出現頻度に応じた符号に変換する単語符号化部と、単語符号化部により符号化された電子メール符号化データに決定木学習部によって予め生成された決定木を適用することにより迷惑メールか否かを判定する判定部とを有し、決定木学習部は単語出現頻度データベース及び迷惑メールと通常メールとを振り分ける最適な決定木を生成するものであり、迷惑メール判定部及び決定木学習部は、電子メールにおけるヘッダ部分及び本文を分けずに同一アルゴリズムにより処理する。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

外部からの電子メールを受信する電子メール受信部と、
前記電子メール受信部によって受信された電子メールが迷惑メールか否かを判定する迷惑メール判定部と、

前記迷惑メール判定部の判定結果に応じて前記電子メールをフィルタリングする迷惑メールフィルタ部と、

前記迷惑メールフィルタ部によってフィルタリングされた前記電子メールをローカルメールボックス又は外部に送信する電子メール送信部と、

を有する電子メールシステムであって、

前記迷惑メール判定部は、

決定木学習部によって予め生成された単語出現頻度データベースにより、前記電子メール中の全ての単語を出現頻度に応じた符号に変換する単語符号化部と、

前記単語符号化部により符号化された電子メール符号化データに、決定木学習部によって予め生成された決定木を適用することにより迷惑メールか否かを判定する判定部と、

を有し、

前記決定木学習部は、

前記迷惑メール判定部と同一サーバ内又は異なるサーバ内にあり、

迷惑メールを保存した迷惑メールデータベースと、通常メールを保存した通常メールデータベースと、

前記迷惑メールデータベース及び前記通常メールデータベース内の電子メール中の単語の出願頻度を求めて前記単語出現頻度データベースを生成する単語出現頻度データベース生成部と、

前記単語出現頻度データベースにより、前記迷惑メールデータベース及び前記通常メールデータベース内の電子メール中の全ての単語を出現頻度に応じた符号に変換する単語符号化部と、

前記単語符号化部により符号化された電子メール符号化データのパターンに基づいて、迷惑メールと通常メールとを振り分ける最適な決定木を生成する学習部と、

を有し、

前記迷惑メール判定部及び前記決定木学習部における電子メールはヘッダ部分及び本文の両方を含むものであり、前記迷惑メール判定部及び前記決定木学習部は、前記電子メールにおけるヘッダ部分及び本文を分けずに同一アルゴリズムにより処理することを特徴とする電子メールシステム。

【請求項 2】

前記単語出現頻度データベース生成部は、単語の出現頻度とともに、前記単語が迷惑メールと通常メールのどちらに多く含まれるかを示す出現偏りも求めて前記単語出現頻度データベースを生成し、

前記単語符号化部は、前記電子メール中の全ての単語を前記出現頻度及び前記出現偏りに応じた符号に変換する

ことを特徴とする請求項 1 記載の電子メールシステム。

【請求項 3】

前記学習部は、前記電子メール符号化データ内の符号を、最適な決定木を求められるグループに分け、前記グループ分けの結果により前記符号をさらに第 2 の符号に変換する機能を有することを特徴とする請求項 1 又は 2 記載の電子メールシステム。

【請求項 4】

前記学習部に、BONSAI プログラムを用いることを特徴とする請求項 3 記載の電子メールシステム。

【請求項 5】

外部からの電子メールを受信する電子メール受信部と、

前記電子メール受信部によって受信された電子メールが迷惑メールか否かを判定する迷

10

20

30

40

50

惑メール判定部と、

前記迷惑メール判定部の判定結果に応じて前記電子メールをフィルタリングする迷惑メールフィルタ部と、

前記迷惑メールフィルタ部によってフィルタリングされた前記電子メールをローカルメールボックス又は外部に送信する電子メール送信部と、

を有する電子メールプログラムであって、

前記迷惑メール判定部は、

決定木学習部によって予め生成された単語出現頻度データベースにより、前記電子メール中の全ての単語を出現頻度に応じた符号に変換する単語符号化部と、

前記単語符号化部により符号化された電子メール符号化データに、決定木学習部によって予め生成された決定木を適用することにより迷惑メールか否かを判定する判定部と、

を有し、

前記決定木学習部は、

前記迷惑メール判定部と同一サーバ内又は異なるサーバ内にあり、

迷惑メールを保存した迷惑メールデータベースと、通常メールを保存した通常メールデータベースと、

前記迷惑メールデータベース及び前記通常メールデータベース内の電子メール中の単語の出願頻度を求めて前記単語出現頻度データベースを生成する単語出現頻度データベース生成部と、

前記単語出現頻度データベースにより、前記迷惑メールデータベース及び前記通常メールデータベース内の電子メール中の全ての単語を出現頻度に応じた符号に変換する単語符号化部と、

前記単語符号化部により符号化された電子メール符号化データのパターンに基づいて、迷惑メールと通常メールとを振り分ける最適な決定木を生成する学習部と、

を有し、
前記迷惑メール判定部及び前記決定木学習部における電子メールはヘッダ部分及び本文の両方を含むものであり、前記迷惑メール判定部及び前記決定木学習部は、前記電子メールにおけるヘッダ部分及び本文を分けずに同一アルゴリズムにより処理することを特徴とする電子メールプログラム。

【請求項 6】

前記単語出現頻度データベース生成部は、単語の出現頻度とともに、前記単語が迷惑メールと通常メールのどちらに多く含まれるかを示す出現偏りも求めて前記単語出現頻度データベースを生成し、

前記単語符号化部は、前記電子メール中の全ての単語を前記出現頻度及び前記出現偏りに応じた符号に変換することを特徴とする請求項 5 記載の電子メールプログラム。

【請求項 7】

前記学習部は、前記電子メール符号化データ内の符号を、最適な決定木を求められるグループに分け、前記グループ分けの結果により前記符号をさらに第 2 の符号に変換する機能を有することを特徴とする請求項 5 又は 6 記載の電子メールプログラム。

【請求項 8】

前記学習部に、BONSAI プログラムを用いることを特徴とする請求項 7 記載の電子メールプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、学習型の決定木アルゴリズムを用いた、迷惑メールのフィルタ機能を有する電子メールシステム及びプログラムに関する。

【背景技術】

【0002】

10

20

30

40

50

インターネット上を流れる迷惑メールの割合は、全電子メール中の60%を越えると言われており、現在では、迷惑メール対策として、さまざまな自動分類方法が用いられている。開発初期のもっとも単純な方法に、メールヘッダに記述される特定の迷惑メール送信サーバやメールのFrom行記載のメールアドレスを、管理者やユーザがひとつひとつ登録し、合致するメールを排除する方法がある。しかし迷惑メール送信者は、このような対策をかいくぐる新しい方法で次々に迷惑メールを送信してくるため、手作業で分類やアドレスの登録などを行うには作業コストが大きすぎ、現実的ではなくなっている。また、これまでの方法では、通常電子メールを迷惑メールと間違えて判断するケースおよびその逆のケースも増えている。近年、メール本文などの単語の出現頻度による特徴を分類に役立てるベイズ理論を応用した方法が注目されているが、未だ利用者および管理者の作業コストは大きく、分類精度もそれほど高くない。

10

【0003】

従来技術として、特許文献1乃至3が挙げられる。

特許文献1には、文字列の一部をわざと間違えたり文字間に無意味な記号を挿入した電子メールであっても、迷惑メール等の電子メールを効果的に分類できる電子メール処理装置が記載されている。電子メールに含まれる単語について単語情報データベース内の迷惑メール対象文字列と相同性検索をすることによって迷惑メールの判定を行っている。

特許文献2には、電子メールのヘッダ情報に含まれるメール中継装置によって、該当電子メールが迷惑メールか否かをベイズ確率モデルを用いて判定する電子メールフィルタリングシステムが記載されている。

20

特許文献3には、ユーザが通常メールと迷惑メールとを分類し、その分類された内容を分析してフィルタルールを追加する電子メールフィルタリングシステムが記載されている。

特許文献1乃至3のいずれにも、迷惑メールの判定に学習型の決定木アルゴリズムを用いることについて記載されていない。

【特許文献1】特開2006-293573号公報

【特許文献2】特開2006-260515号公報

【特許文献3】特開2006-245813号公報

【発明の開示】

【発明が解決しようとする課題】

30

【0004】

本発明は、学習型の決定木アルゴリズムを用いて、迷惑メールを正確に効率良くフィルタリングする電子メールシステムを提供することを目的とする。

【課題を解決するための手段】

【0005】

前記目的を達成するため、本発明は以下の構成を有する。

外部からの電子メールを受信する電子メール受信部と、前記電子メール受信部によって受信された電子メールが迷惑メールか否かを判定する迷惑メール判定部と、前記迷惑メール判定部の判定結果に応じて前記電子メールをフィルタリングする迷惑メールフィルタ部と、前記迷惑メールフィルタ部によってフィルタリングされた前記電子メールをローカルメールボックス又は外部に送信する電子メール送信部と、を有する電子メールシステム及びプログラムであって、前記迷惑メール判定部は、決定木学習部によって予め生成された単語出現頻度データベースにより、前記電子メール中の全ての単語を出現頻度に応じた符号に変換する単語符号化部と、前記単語符号化部により符号化された電子メール符号化データに、決定木学習部によって予め生成された決定木を適用することにより迷惑メールか否かを判定する判定部と、を有し、前記決定木学習部は、前記迷惑メール判定部と同一サーバ内又は異なるサーバ内にあり、迷惑メールを保存した迷惑メールデータベースと、通常メールを保存した通常メールデータベースと、前記迷惑メールデータベース及び前記通常メールデータベース内の電子メール中の単語の出願頻度を求めて前記単語出現頻度データベースを生成する単語出現頻度データベース生成部と、前記単語出現頻度データベース

40

50

により、前記迷惑メールデータベース及び前記通常メールデータベース内の電子メール中の全ての単語を出現頻度に応じた符号に変換する単語符号化部と、前記単語符号化部により符号化された電子メール符号化データのパターンに基づいて、迷惑メールと通常メールとを振り分ける最適な決定木を生成する学習部と、を有し、前記迷惑メール判定部及び前記決定木学習部における電子メールはヘッダ部分及び本文の両方を含むものであり、前記迷惑メール判定部及び前記決定木学習部は、前記電子メールにおけるヘッダ部分及び本文を分けずに同一アルゴリズムにより処理することを特徴とする電子メールシステム及びプログラム。

【0006】

また、以下の実施態様を有する。

前記単語出現頻度データベース生成部は、単語の出現頻度とともに、前記単語が迷惑メールと通常メールのどちらに多く含まれるかを示す出現偏りも求めて前記単語出現頻度データベースを生成し、前記単語符号化部は、前記電子メール中の全ての単語を前記出現頻度及び前記出現偏りに応じた符号に変換する。

前記学習部は、前記電子メール符号化データ内の符号を、最適な決定木を求められるグループに分け、前記グループ分けの結果により前記符号をさらに第2の符号に変換する機能を有する。

前記学習部に、BONSAIプログラムを用いる。

【発明の効果】

【0007】

学習型の決定木アルゴリズムを用いることで、従来のシステムに比べて、迷惑メールを正確に効率よくフィルタリングできる。また、決定木の学習及び適用の前に、電子メールを単語の出現頻度及び出現偏りに応じて符号化しておくことで、効果的に決定木の学習及び適用ができる。本発明のアルゴリズムは電子メールのヘッダ情報及び本文の両方に分け隔てなく適用でき、両方の情報を用いることでより簡単に電子メールのフィルタリングが可能である。

決定木の学習には時間が掛かるが、予め生成された決定木に基づいて電子メールを分類するのは短時間でできる。本発明の決定木学習部と迷惑メール判定部とは独立して実行可能であるので、決定木を事前に学習しておいたり、決定木の学習を別サーバで実行することが可能である。迷惑メール判定部は、既に生成された決定木に基づいて電子メールを分類すればよいので、リアルタイムで電子メールのフィルタリングが可能である。

【発明を実施するための最良の形態】

【0008】

図面を用いて本発明の実施形態について説明する。図1は、本電子メールシステムのブロック図である。電子メールシステム1は、インターネットから電子メールを受信する電子メール受信部2と、電子メール受信部2で受信された電子メールが迷惑メールか否かを判定する迷惑メール判定部3と、迷惑メール判定部3の判定結果に応じて電子メールをフィルタリングする迷惑メールフィルタ部4と、迷惑メールフィルタ部4によってフィルタリングされた電子メールをローカルメールボックス又は外部に送信する電子メール送信部5とからなる。迷惑メールフィルタ部4は、迷惑メールの削除、迷惑メールにフラグを付与、迷惑メールを別フォルダに移動などの動作を行う。電子メール送信部5は、本電子メールシステムの使用形態に応じて、フィルタリングされた電子メールを同一サーバ内のローカルメールボックスに振り分けて送信しても良いし、外部のメールサーバに転送しても良い。この電子メールシステムは、インターネットに接続されたサーバ上で動作させても良いし、電子メールを受信する端末上で動作させても良い。

【0009】

図2は、迷惑メール判定部3のブロック図である。迷惑メール判定部3は、決定木学習部6によって予め生成された単語出現頻度データベース33及び決定木34と、単語出現頻度データベース33により電子メール中の全ての単語を出現頻度に応じた符号に変換する単語符号化部31と、単語符号化部31により符号化された電子メール符号化データに

10

20

30

40

50

決定木 3 4 を適用することにより迷惑メールか否かを判定する判定部 3 2 とからなる。単語出現頻度データベース 3 3 及び決定木 3 4 は決定木学習部 6 により予め生成しておき、必要に応じて転送しておくなどして迷惑メール判定部 3 で利用可能にしておく。迷惑メール判定部 3 と決定木学習部 6 とは同一サーバ上で実行させても良いし、それぞれ異なるサーバ上で実行させても良い。決定木学習部 6 で生成された単語出現頻度データベース 3 3 及び決定木 3 4 を、複数のサーバ上の迷惑メール判定部 3 で利用しても良い。

【 0 0 1 0 】

図 3 は、決定木学習部 6 のブロック図である。決定木学習部 6 は、迷惑メールを保存した迷惑メールデータベース 6 1 と、通常メールを保存した通常メールデータベース 6 2 と、迷惑メールデータベース 6 1 及び通常メールデータベース 6 2 内の電子メール中の単語の出願頻度を求めて単語出現頻度データベース 3 3 を生成する単語出現頻度データベース生成部 6 3 と、単語出現頻度データベース 3 3 により迷惑メールデータベース 6 1 及び前記通常メールデータベース 6 2 内の電子メール中の全ての単語を出現頻度に応じた符号に変換する単語符号化部 6 4 と、単語符号化部 6 4 により符号化された電子メール符号化データのパターンに基づいて迷惑メールと通常メールとを分類する最適な決定木 3 4 を生成する学習部 6 5 とからなる。迷惑メールデータベース 6 1 及び通常メールデータベース 6 2 には、予めユーザによって分類された迷惑メール及び通常メールが蓄積されている。単語出現頻度データベース生成部 6 3 は、単語の出現頻度及び出現偏り（単語が迷惑メールと通常メールのどちらに多く含まれるか）を求めて単語出現頻度データベース 3 3 を生成する。単語符号化部 6 4 は、単語出現頻度データベース 3 3 に含まれる単語の出現頻度及び偏りの情報から、各単語を A、B、C などの符号に変換する。迷惑メール判定部 3 内の単語符号化部 3 1 も、単語符号化部 6 4 と同様な動作を行う。学習部 6 5 は、後述の B O N S A I プログラムを用いて決定木 3 4 の生成を行う。迷惑メール判定部 3 及び決定木学習部 6 は、電子メールにおけるヘッダ部分及び本文を分けずに同一アルゴリズムにより処理する。

【 0 0 1 1 】

以下、B O N S A I について簡単に説明する（B O N S A I の詳細については、Shimozono, S., Shinohara, A., Miyano, S., Kuhara, S., Arikawa, S. “Knowledge Acquisition from Amino Acid Sequence by Machine Learning System B O N S A I”, Trans. Inform. Process. Soc. Japan, 35(10): 2009-2018, 1994 参照)。B O N S A I は、確率的近似学習と呼ばれる学習パラダイムに基づいて開発された機会学習プログラムで、正の学習グループと負の学習グループを与えると決定木を作成する。決定木の作成については、J. R. Quinlan の決定木学習アルゴリズム ID 3 の枝狩り規準を改良した “C 4 . 5 ” というアルゴリズムに基づいている。さらに、B O N S A I は indexing というグルーピングの機能を持っている。もともと B O N S A I は生物ゲノム情報から重要な遺伝子配列などを抽出する目的で開発された機械学習システムであるが、本発明者の工夫によって迷惑メールの分類に利用可能であることが見出された。B O N S A I は、正の例と負の例として二つのデータ集団を入力すると、正の例には存在するが負の例には存在しないパターンを見つけ出すことができるので、この機能を利用して迷惑メールの分類を行う。

【 0 0 1 2 】

図 4 は、本システムの決定木学習の流れ図である。図 4 に示すように正の学習グループとして迷惑メール群、負の学習グループとして通常メール群を作成し、迷惑メール群に存在する特徴的なパターンの抽出を試みる。まず、両群の電子メールの文字列を単語に分解し、両群に存在するすべての単語について正の学習グループでの出現頻度を算出し、出現頻度の高いものから A ~ E までのグルーピングを行う。出現頻度を表す A ~ E の文字で電子メール内のすべての文字列を置換してから、機械学習システム B O N S A I に投入する。B O N S A I (東京大学医科学研究所ヒトゲノム解析センター宮野研究室開発) は、正の

10

20

30

40

50

学習グループと負の学習グループとして二つのデータ集団を入力すると、正の学習グループには存在するが、負の学習グループには存在しないといったパターンを見つけ出し、二つの学習例を正しく分けることができる決定木 (Decision Tree) を作成する。また同時に、正の学習グループと負の学習グループを最も効率よく分類できる条件で、それぞれのグループ例を構成する要素もグループ分けする機能を持っている (Indexing)。単語の出現頻度を反映させた学習グループ例を BONS AI に投入することで、単語の出現頻度とその語順を考慮したパターン抽出が可能になる。つまり BONS AI は、出現頻度を反映した A ~ E の文字で置換された電子メール内の文字列を、図 4 のように indexing によってさらにグルーピングし、例えば 0 ~ 2 のような文字で置き換えながら、電子メール内の文字列中に存在するパターンを抽出する。また同時に正および負の学習グループ例を最も正しく分ける規則を提示する。

10

【0013】

図 5 は、決定木の例である。例えば図 4 のケースでは、電子メールを単語分解及び 2 段階グルーピング (単語出現頻度、Indexing) して 0 ~ 2 の符号に変換されたデータが、パターン「20」を含んでいたなら「迷惑メール」(正の学習グループ) と判定する。パターン「20」を含んでいない場合は、さらにパターン「021」の検索を行い、パターン「021」を含んでいたなら「迷惑メール」、含んでいなければ「通常メール」(負の学習グループ) と判定する。図 4 及び 5 は説明のための簡単な事例であるが、実際に利用する場合はパターン長はもっと長く、枝分岐ももっと複雑である。決定木学習及び決定木の適用には単語を符号化したものを利用するので、単語分解できるデータであれば何でも利用可能であり、電子メールのヘッダ部分及び本文について同じアルゴリズムを適用できる。

20

【実施例】

【0014】

以下、実施例について説明する。

決定木の学習手順は、以下の通りである。

1. サンプル電子メール (迷惑メール [正の例] : 500 通、通常メール [負の例] : 500 通) の準備。
2. サンプル電子メール (ヘッダ及び本文) を単語に分解。
3. 単語の出現率と出現偏りの計算
 - ・ 出現率 = \log (出現数の総和)
 - (出現率が小さいものは除外)
 - ・ 出現偏り = 正の例での出現数 / 正の例及び負の例での出現数の総和
4. 出現頻度に応じた符号化。
 - X : 0.8 < (出現偏り)
 - Y : 0.6 (出現偏り) 0.8
 - Z : (出現偏り) < 0.6
 - O : その他 [出現数少]
5. BONS AI により最適な決定木の生成。

30

40

図 6 に、BONS AI により生成された決定木の例を示す。この例では、BONS AI のグルーピング機能 (indexing) により、X 0、Y 0、Z 1、O 1 のさらなる符号化が行われている。

【0015】

生成された決定木に基づいて、712 通の一般の受信メールを振り分けてみた結果は以下の通りである。

通常メール分類の正解率 : 94.4% (238 / 252 通)

迷惑メール分類の正解率 : 97.8% (450 / 460 通)

この結果から、高い正解率で迷惑メールと通常メールの振り分けが可能であることがわかる。

50

【 0 0 1 6 】

別の実施例について説明する。前述の実施例では、単語出現頻度による符号化の符号数は4個(X, Y, Z, O)、BONS AIのグルーピング機能(indexing)による符号化の符号数は2個(0, 1)であったが、単語出現頻度による符号化の符号数を6個(X, Y, Z, O, A, B)、BONS AIのグルーピング機能(indexing)による符号化の符号数を3個(0, 1, 2)にした場合の決定木の例を図7に示す。

この決定木に基づいて、806通の一般の受信メールを振り分けてみた結果は以下の通りである。

通常メール分類の正解率：96.1% (273 / 284 通)

迷惑メール分類の正解率：98.6% (515 / 522 通)

10

前述の実施例よりもさらに高い正解率であることがわかる。

「単語出現頻度による符号化の符号数」及び「BONS AIのグルーピング機能(indexing)による符号化の符号数」はこの他の組み合わせも可能であり、演算速度、サンプル電子メール数、学習に掛けられる時間等に応じて任意に設定できる。

【 0 0 1 7 】

以上、本発明の実施形態の一例を説明したが、本発明はこれに限定されるものではなく、特許請求の範囲に記載された技術的思想の範疇において各種の変更が可能であることは言うまでもない。

【 図面の簡単な説明 】

【 0 0 1 8 】

20

【 図 1 】 本システムのブロック図

【 図 2 】 迷惑メール判定部のブロック図

【 図 3 】 決定木学習部のブロック図

【 図 4 】 決定木学習の流れ図

【 図 5 】 決定木の例

【 図 6 】 実施例における決定木

【 図 7 】 別の実施例における決定木

【 符号の説明 】

【 0 0 1 9 】

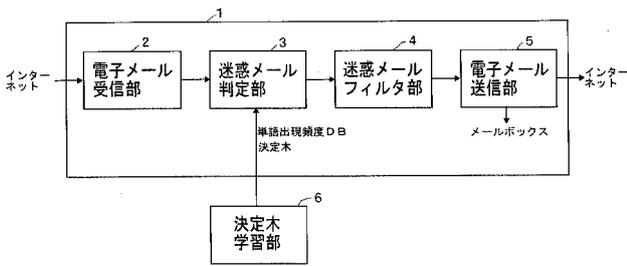
1：電子メールシステム、 2：電子メール受信部、 3：迷惑メール判定部、 4：迷惑メールフィルタ部、 5：電子メール送信部、 6：決定木学習部、

30

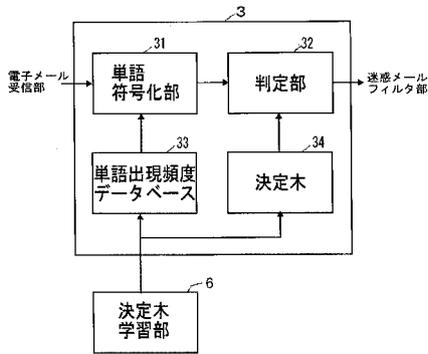
31：単語符号化部、 32：判定部、 33：単語出現頻度データベース、 34：決定木、

61：迷惑メールデータベース、 62：通常メールデータベース、 63：単語出現頻度データベース生成部、 64：単語符号化部、 65：学習部(BONS AI)

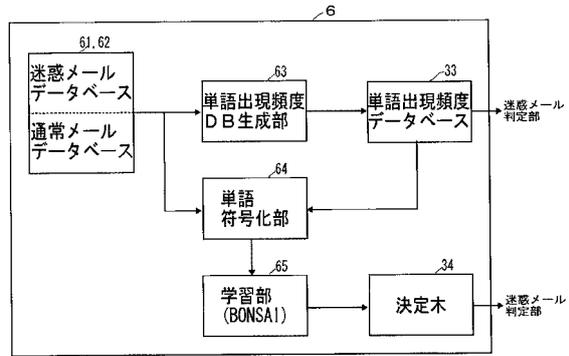
【 図 1 】



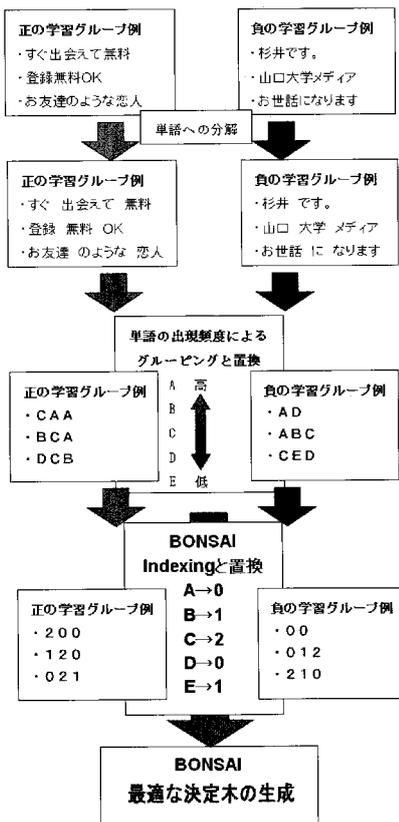
【 図 2 】



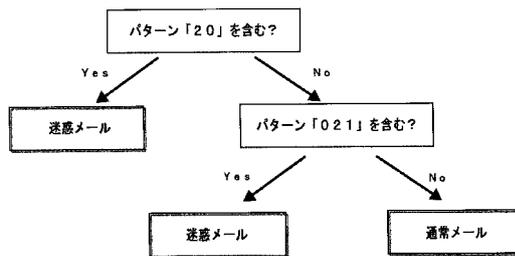
【 図 3 】



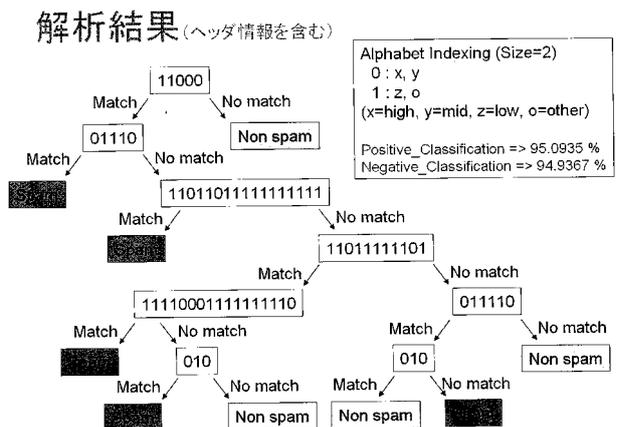
【 図 4 】



【 図 5 】



【 図 6 】



【 図 7 】

解析結果 (ヘッダ情報を含む)

Alphabet Indexing (Size=3)	(x=high, y=mid, z=low, o=other, a=neg_high, b=neg_mid)
0 : a	
1 : x, y,	Positive_Classification => 96.3387 %
2 : z, o, b	Negative_Classification => 96.888 %

