

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4891638号
(P4891638)

(45) 発行日 平成24年3月7日(2012.3.7)

(24) 登録日 平成23年12月22日(2011.12.22)

(51) Int.Cl. F 1
G 0 6 F 17/30 (2006.01)
 G 0 6 F 17/30 2 1 0 D
 G 0 6 F 17/30 3 4 0 B
 G 0 6 F 17/30 4 1 9 B

請求項の数 8 (全 16 頁)

<p>(21) 出願番号 特願2006-71958 (P2006-71958) (22) 出願日 平成18年3月16日 (2006.3.16) (65) 公開番号 特開2007-249600 (P2007-249600A) (43) 公開日 平成19年9月27日 (2007.9.27) 審査請求日 平成21年1月16日 (2009.1.16)</p>	<p>(73) 特許権者 500257300 ヤフー株式会社 東京都港区赤坂9丁目7番1号 (74) 代理人 100106002 弁理士 正林 真之 (72) 発明者 藤田 澄男 東京都港区六本木6丁目10番1号 ヤフー株式会社内 審査官 岩田 淳</p>
--	---

最終頁に続く

(54) 【発明の名称】 目的データをカテゴリに分類する方法

(57) 【特許請求の範囲】

【請求項1】

端末装置と通信ネットワークを介して接続され、前記端末装置のユーザに閲覧させる目的データをカテゴリに分類するサーバであって、

前記目的データを複数の候補カテゴリデータと関連付けて商品分類テーブルに記憶する手段と、

(a) 前記端末装置から、前記ユーザによる前記複数の候補カテゴリデータのうちの選択を表すデータを受信したことに応じて、前記選択をした候補カテゴリデータに関連付けられた目的データを送信する手段と、

(b) 前記端末装置から前記ユーザによる前記目的データの選択を表すデータを受け付けたことに応じて、前記ユーザが前記選択をした候補カテゴリデータの選択回数データをカウントアップして、前記選択をした目的データを前記選択をした候補カテゴリデータに関連付けて記憶する手段と、

(c) 所定の期間における前記選択回数データに基づき、前記目的データを分類する前記カテゴリの決定をする手段と、

(d) 前記決定をしたカテゴリデータと関連付けて、前記目的データに含まれるキーワードデータを抽出して記憶する手段と、

前記目的データとは別の新たな目的データの登録を受け付けたことに応じて、前記新たな目的データと前記キーワードデータとの類似度の算出をする手段と、

前記算出をした類似度に基づくカテゴリの評価に従い、前記新たな目的データを分類す

10

20

る前記複数の候補カテゴリを決定して前記新たな目的データと関連付けて前記商品分類テーブルに記憶する手段と、

前記 (a) から (d) の処理を繰り返す手段と、
を備えるサーバ。

【請求項 2】

前記決定をするカテゴリは、前記候補カテゴリデータの選択回数が所定の回数以上のものであることを特徴とする請求項 1 に記載のサーバ。

【請求項 3】

前記決定をするカテゴリは、前記候補カテゴリデータの選択回数が所定の順位以上のものであることを特徴とする請求項 1 に記載のサーバ。

【請求項 4】

前記決定をするカテゴリは、前記候補カテゴリデータの選択回数の相対評価により決定するものであることを特徴とする請求項 1 に記載のサーバ。

【請求項 5】

前記目的データを複数の前記候補カテゴリデータと関連付けて記憶する手段が、前記目的データと前記キーワードデータとの類似度の算出をし、当該算出をした類似度に基づくカテゴリの評価に従い前記候補カテゴリデータを選択することを特徴とする請求項 1 から 4 のいずれかに記載のサーバ。

【請求項 6】

前記決定をしたカテゴリデータと関連付けて前記目的データに含まれるキーワードデータを抽出して記憶する手段が、前記カテゴリの選択回数に基づく選択確率データを更に関連付けて記憶し、

前記キーワードデータに関連付くカテゴリの評価が、前記類似度および前記選択確率データに基づくことを特徴とする請求項 1 から 5 のいずれかに記載のサーバ。

【請求項 7】

端末装置と通信ネットワークを介して接続されたサーバが、前記端末装置のユーザに閲覧させる目的データをカテゴリに分類する方法であって、

前記目的データを複数の候補カテゴリデータと関連付けて商品分類テーブルに記憶するステップと、

(a) 前記端末装置から、前記ユーザによる前記複数の候補カテゴリデータのうちの選択を表すデータを受信したことに応じて、前記選択をした候補カテゴリデータに関連付けられた目的データを送信するステップと、

(b) 前記端末装置から前記ユーザによる前記目的データの選択を表すデータを受け付けたことに応じて、前記ユーザが前記選択をした候補カテゴリデータの選択回数データをカウントアップして、前記選択をした目的データを前記選択をした候補カテゴリデータに関連付けて記憶するステップと、

(c) 所定の期間における前記選択回数データに基づき、前記目的データを分類する前記カテゴリの決定をするステップと、

(d) 前記決定をしたカテゴリデータと関連付けて、前記目的データに含まれるキーワードデータを抽出して記憶するステップと、

前記目的データとは別の新たな目的データの登録を受け付けたことに応じて、前記新たな目的データと前記キーワードデータとの類似度の算出をするステップと、

前記算出をした類似度に基づくカテゴリの評価に従い、前記新たな目的データを分類する前記複数の候補カテゴリを決定して前記新たな目的データと関連付けて前記商品分類テーブルに記憶するステップと、

前記 (a) から (d) の処理を繰り返すステップと、
を含む方法。

【請求項 8】

端末装置と通信ネットワークを介して接続されたサーバに、前記端末装置のユーザに閲覧させる目的データをカテゴリに分類させるプログラムであって、

10

20

30

40

50

前記目的データを複数の候補カテゴリデータと関連付けて商品分類テーブルに記憶させるステップと、

(a) 前記端末装置から、前記ユーザによる前記複数の候補カテゴリデータのうちの選択を表すデータを受信したことに応じて、前記選択をした候補カテゴリデータに関連付けられた目的データを送信させるステップと、

(b) 前記端末装置から前記ユーザによる前記目的データの選択を表すデータを受け付けたことに応じて、前記ユーザが前記選択をした候補カテゴリデータの選択回数データをカウントアップして、前記選択をした目的データを前記選択をした候補カテゴリデータに関連付けて記憶させるステップと、

(c) 所定の期間における前記選択回数データに基づき、前記目的データを分類する前記カテゴリの決定をさせるステップと、

(d) 前記決定をしたカテゴリデータと関連付けて、前記目的データに含まれるキーワードデータを抽出して記憶させるステップと、

前記目的データとは別の新たな目的データの登録を受け付けたことに応じて、前記新たな目的データと前記キーワードデータとの類似度の算出をさせるステップと、

前記算出をした類似度に基づくカテゴリの評価に従い、前記新たな目的データを分類する前記複数の候補カテゴリを決定して前記新たな目的データと関連付けて前記商品分類テーブルに記憶させるステップと、

前記(a)から(d)の処理を繰り返させるステップと、

を実行させるプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、端末装置のユーザに閲覧させる目的データをカテゴリに分類する方法、サーバ、およびプログラムに関する。

【背景技術】

【0002】

従来、インターネットにおいて情報(目的データ)を検索する方法としては、様々なものが提供されているが、その一つに、ツリー状に構成されたカテゴリをユーザが辿っていくことによって目的の情報(目的データ)に至る、ディレクトリ型と呼ばれる検索エンジンがある。この方法は、通常人手によりカテゴリ分けがなされた目的データを検索するもので、同種類の情報をまとめて閲覧することが可能であり、例えば商品情報の検索等に用いられている。

【0003】

この方法を実現するためには、予め商品等の情報提供項目に対してカテゴリ情報を記憶しておく必要がある。更に、ユーザによる検索が効果的に行われるためには、カテゴリに分類する際の精度の高さが求められる。こうした状況の中で、情報の分類に関しては、その精度を向上させる技術が知られている。

【0004】

例えば、特許文献1には、各クラス(カテゴリ)間の情報の重なりによる誤分類を少なくする為に、各クラスについて着目クラスには現れるが他のクラスでは現れにくい特徴、および他のクラスでは現れるが着目クラスでは現れにくい特徴を求め、そのデータを使って類似度の補正を行う方法が開示されている。

【特許文献1】特開2003-256441号公報

【発明の開示】

【発明が解決しようとする課題】

【0005】

しかしながら、上記方法によっても、分類の精度は学習事例(目的データを特徴付けるキーワード等とカテゴリとの組)の正確さ、あるいは分類の際の類似度評価の正確さに左右されるため、誤った分類がされる可能性は依然として残る。そのため、通常人手により

10

20

30

40

50

行っている各目的データに対するカテゴリ情報の付与を自動で行うと、自動で分類された学習事例に含まれる誤分類が蓄積されることによって、次第に分類精度が悪化していくことが懸念される。

【0006】

そこで本発明は、人手で行うのと近い精度で自動的に目的データをカテゴリに分類することができ、更には、そうした信頼度の高い分類サンプル（学習事例）を継続的に取得し、次回以降の新たな目的データの分類に利用できる方法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】

【0007】

具体的には、以下のようなものを提供する。

10

【0008】

(1) 端末装置と通信ネットワークを介して接続されたサーバが、前記端末装置のユーザに閲覧させる目的データをカテゴリに分類する方法であって、

前記目的データを複数の候補カテゴリデータと関連付けて記憶するステップと、

前記端末装置から、前記ユーザによる前記複数の候補カテゴリデータのうちの選択を表すデータを受信したことに応じて、前記選択をした候補カテゴリデータに関連付けられた目的データを送信するステップと、

前記端末装置から前記ユーザによる前記目的データの選択を表すデータを受け付けたことに応じて、前記ユーザが前記選択をした候補カテゴリデータの選択回数データをカウントアップして、前記選択をした目的データおよび前記選択をした候補カテゴリデータに関連付けて記憶するステップと、

20

所定の期間における前記選択回数データに基づき、前記目的データを分類する前記カテゴリの決定をするステップと、

を含む方法。

【0009】

(1)の発明によれば、サーバは、端末装置のユーザに閲覧させる目的データを複数の候補カテゴリデータと関連付けて記憶し、前記端末装置から、前記ユーザによる前記複数の候補カテゴリデータのうちの選択を表すデータを受信したことに応じて、前記選択をした候補カテゴリデータに関連付けられた目的データを送信し、前記端末装置から前記ユーザによる前記目的データの選択を表すデータを受け付けたことに応じて、前記ユーザが前記選択をした候補カテゴリデータの選択回数データをカウントアップして、前記選択をした目的データおよび前記選択をした候補カテゴリデータに関連付けて記憶し、所定の期間における前記選択回数データに基づき、前記目的データを分類する前記カテゴリの決定をする。

30

【0010】

このことにより、前記サーバは、複数の候補カテゴリについて、実際のユーザのアクセス履歴（どのカテゴリから当該目的データに辿り着いたか）を記憶できるので、当該アクセス履歴に基づいてカテゴリを選択することにより、仮に候補カテゴリの中に相応しくないものが混在していた場合にも、そのようなカテゴリは自動的に排除され、実際のユーザの思いに近い（精度の良い）カテゴリに対して自動的に分類できる可能性がある。

40

【0011】

(2) 前記決定をするカテゴリは、前記候補カテゴリデータの選択回数が所定の回数以上のものであることを特徴とする(1)に記載の方法。

【0012】

(2)の発明によれば、前記サーバは、前記目的データを選択回数が所定の回数以上のカテゴリに分類する。

【0013】

このことにより、前記サーバは、数多く（所定の回数以上）選択された実績のあるカテゴリに分類するので、実際のユーザが選択しやすいカテゴリに対して自動的に分類できる可能性がある。

50

(3) 前記決定をするカテゴリは、前記候補カテゴリデータの選択回数が所定の順位以上のものであることを特徴とする(1)に記載の方法。

【0014】

(3)の発明によれば、前記サーバは、前記目的データを前記選択回数が所定の順位以上のものに対応するカテゴリに分類する。

【0015】

このことにより、前記サーバは、選択される頻度が高いカテゴリを一または複数選択するので、多くのユーザが考える(思いつきやすい)カテゴリに対して自動的に分類できる可能性がある。

【0016】

(4) 前記決定をするカテゴリは、前記候補カテゴリデータの選択回数の相対評価により決定するものであることを特徴とする(1)に記載の方法。

【0017】

(4)の発明によれば、前記サーバは、前記カテゴリデータの選択回数の相対評価により前記目的データを分類するカテゴリを決定する。

【0018】

このことにより、前記サーバは、関連付けた複数のカテゴリの中から、例えば他と比べて選択される割合が突出するカテゴリを選択できるため、ユーザの志向により近いカテゴリに対して自動的に分類できる可能性がある。また、所定の期間において、どの候補カテゴリも所定の選択回数に達しなかった場合や、所定の順位以内の候補カテゴリ間で選択回数に大きな差がある場合にも適切にカテゴリを選択できる。

【0019】

(5) 前記決定をしたカテゴリデータと関連付けて、前記目的データに含まれるキーワードデータを抽出して記憶するステップを更に含む(1)から(4)のいずれかに記載の方法。

【0020】

(5)の発明によれば、前記サーバは、前記決定をしたカテゴリデータと関連付けて、前記目的データに含まれるキーワードデータを抽出して記憶する。

【0021】

このことにより、前記サーバは、前記カテゴリデータと前記キーワードデータとの対応付けを記憶するため、後に、分類の事例として利用することができる。

【0022】

(6) 前記目的データとは別の新たな目的データの登録を受け付けたことに応じて、前記新たな目的データと前記キーワードデータとの類似度の算出をするステップと、前記算出した類似度に基づくカテゴリの評価に従い、前記新たな目的データを分類するカテゴリを決定するステップと、を更に含む(5)に記載の方法。

【0023】

(6)の発明によれば、前記サーバは、前記目的データとは別の新たな目的データの登録を受け付けたことに応じて、前記新たな目的データと前記キーワードデータとの類似度の算出をし、前記算出した類似度に基づき、前記新たな目的データを分類するカテゴリを決定する。

【0024】

このことにより、前記サーバは、記憶した前記キーワードデータと前記カテゴリデータとの関連付けに基づき、キーワードが類似する前記目的データについて、対応するカテゴリに自動的に分類することができる。

【0025】

(7) 前記目的データを複数の前記カテゴリデータと関連付けて記憶するステップが、前記目的データと前記キーワードデータとの類似度の算出をし、当該算出した類似度に基づくカテゴリの評価に従い前記候補カテゴリデータを選択することを特徴とする(5)

10

20

30

40

50

)または(6)に記載の方法。

【0026】

(7)の発明によれば、前記サーバは、前記目的データと前記キーワードデータとの類似度の算出をし、当該算出をした類似度に基づき前記候補カテゴリデータを選択する。

【0027】

このことにより、前記サーバは、記憶したキーワードデータとカテゴリデータとの関連付けに基づき、学習を行う(分類するカテゴリをアクセス履歴により決定する)データについてキーワードが類似する複数のカテゴリを、分類する候補として前記目的データに関連付けることができる。よって、ユーザのアクセス履歴を取得する対象のカテゴリを自動的に抽出することができる。

10

【0028】

(8) 前記決定をしたカテゴリデータと関連付けて前記目的データに含まれるキーワードデータを抽出して記憶するステップが、前記カテゴリの選択回数に基づく選択確率データを更に関連付けて記憶し、

前記キーワードデータに関連付くカテゴリの評価が、前記類似度および前記選択確率データに基づくことを特徴とする(6)または(7)に記載の方法。

【0029】

(8)の発明によれば、前記サーバは、前記決定をしたカテゴリデータと関連付けて前記カテゴリの選択回数に基づく選択確率データを更に記憶し、前記キーワードデータに関連付くカテゴリの評価を、前記類似度および前記選択確率データに基づいて行う。

20

【0030】

このことにより、前記サーバは、例えば、類似度に基づいて抽出された分類候補となるカテゴリの中から、ユーザによる選択確率の高かったカテゴリに対して優先的に分類できる。よって、ユーザの志向に沿った、より自然な分類ができる可能性がある。

【0031】

(9) 端末装置と通信ネットワークを介して接続され、前記端末装置のユーザに閲覧させる目的データをカテゴリに分類するサーバであって、

前記目的データを複数の候補カテゴリデータと関連付けて記憶する手段と、

前記端末装置から、前記ユーザによる前記複数の候補カテゴリデータのうちの選択を表すデータを受信したことに応じて、前記選択をした候補カテゴリデータに関連付けられた目的データを送信する手段と、

30

前記端末装置から前記ユーザによる前記目的データの選択を表すデータを受け付けたことに応じて、前記ユーザが前記選択をした候補カテゴリデータの選択回数データをカウントアップして、前記選択をした目的データおよび前記選択をした候補カテゴリデータに関連付けて記憶する手段と、

所定の期間における前記選択回数データに基づき、前記目的データを分類する前記カテゴリの決定をする手段と、

を備えるサーバ。

【0032】

(9)の発明によれば、当該サーバを運用することにより、(1)と同様の効果が期待できる。

40

【0033】

(10) 前記決定をするカテゴリは、前記候補カテゴリデータの選択回数が所定の回数以上のものであることを特徴とする(9)に記載のサーバ。

【0034】

(10)の発明によれば、当該サーバを運用することにより、(2)と同様の効果が期待できる。

【0035】

(11) 前記決定をするカテゴリは、前記候補カテゴリデータの選択回数が所定の順位以上のものであることを特徴とする(9)に記載のサーバ。

50

【0036】

(11)の発明によれば、当該サーバを運用することにより、(3)と同様の効果が期待できる。

【0037】

(12) 前記決定をするカテゴリは、前記候補カテゴリデータの選択回数の相対評価により決定するものであることを特徴とする(9)に記載のサーバ。

【0038】

(12)の発明によれば、当該サーバを運用することにより、(4)と同様の効果が期待できる。

【0039】

(13) 前記決定をしたカテゴリデータと関連付けて、前記目的データに含まれるキーワードデータを抽出して記憶する手段を更に備える(9)から(12)のいずれかに記載のサーバ。

【0040】

(13)の発明によれば、当該サーバを運用することにより、(5)と同様の効果が期待できる。

【0041】

(14) 前記目的データとは別の新たな目的データの登録を受け付けたことに応じて、前記新たな目的データと前記キーワードデータとの類似度の算出をする手段と、前記算出した類似度に基づくカテゴリの評価に従い、前記新たな目的データを分類するカテゴリを決定する手段と、
を更に備える(13)に記載のサーバ。

【0042】

(14)の発明によれば、当該サーバを運用することにより、(6)と同様の効果が期待できる。

【0043】

(15) 前記目的データを複数の前記カテゴリデータと関連付けて記憶する手段が、前記目的データと前記キーワードデータとの類似度の算出をし、当該算出した類似度に基づくカテゴリの評価に従い前記候補カテゴリデータを選択することを特徴とする(13)または(14)に記載のサーバ。

【0044】

(15)の発明によれば、当該サーバを運用することにより、(7)と同様の効果が期待できる。

【0045】

(16) 前記決定をしたカテゴリデータと関連付けて前記目的データに含まれるキーワードデータを抽出して記憶する手段が、前記カテゴリの選択回数に基づく選択確率データを更に関連付けて記憶し、
前記キーワードデータに関連付くカテゴリの評価が、前記類似度および前記選択確率データに基づくことを特徴とする(14)または(15)に記載のサーバ。

【0046】

(16)の発明によれば、当該サーバを運用することにより、(8)と同様の効果が期待できる。

【0047】

(17) 端末装置と通信ネットワークを介して接続されたサーバに、前記端末装置のユーザに関連させる目的データをカテゴリに分類させるプログラムであって、
前記目的データを複数の候補カテゴリデータと関連付けて記憶させるステップと、
前記端末装置から、前記ユーザによる前記複数の候補カテゴリデータのうちの選択を表すデータを受信したことに応じて、前記選択をした候補カテゴリデータに関連付けられた目的データを送信させるステップと、
前記端末装置から前記ユーザによる前記目的データの選択を表すデータを受け付けたこ

10

20

30

40

50

とに応じて、前記ユーザが前記選択をした候補カテゴリデータの選択回数データをカウントアップして、前記選択をした目的データおよび前記選択をした候補カテゴリデータに関連付けて記憶させるステップと、

所定の期間における前記選択回数データに基づき、前記目的データを分類する前記カテゴリの決定をさせるステップと、

を実行させるプログラム。

【0048】

(17)の発明によれば、当該プログラムを当該サーバ上で実行することにより、(1)と同様の効果が期待できる。

【0049】

(18) 前記決定をするカテゴリは、前記候補カテゴリデータの選択回数が所定の回数以上のものであることを特徴とする(17)に記載のプログラム。

【0050】

(18)の発明によれば、当該プログラムを当該サーバ上で実行することにより、(2)と同様の効果が期待できる。

【0051】

(19) 前記決定をするカテゴリは、前記候補カテゴリデータの選択回数が所定の順位以上のものであることを特徴とする(17)に記載のプログラム。

【0052】

(19)の発明によれば、当該プログラムを当該サーバ上で実行することにより、(3)と同様の効果が期待できる。

【0053】

(20) 前記決定をするカテゴリは、前記候補カテゴリデータの選択回数の相対評価により決定するものであることを特徴とする(17)に記載のプログラム。

【0054】

(20)の発明によれば、当該プログラムを当該サーバ上で実行することにより、(4)と同様の効果が期待できる。

【0055】

(21) 前記決定をしたカテゴリデータと関連付けて、前記目的データに含まれるキーワードデータを抽出して記憶させるステップを更に含む(17)から(20)のいずれかに記載のプログラム。

【0056】

(21)の発明によれば、当該プログラムを当該サーバ上で実行することにより、(5)と同様の効果が期待できる。

【0057】

(22) 前記目的データとは別の新たな目的データの登録を受け付けたことに応じて、前記新たな目的データと前記キーワードデータとの類似度の算出をさせるステップと、前記算出をした類似度に基づくカテゴリの評価に従い、前記新たな目的データを分類するカテゴリを決定させるステップと、

を更に含む(21)に記載のプログラム。

【0058】

(22)の発明によれば、当該プログラムを当該サーバ上で実行することにより、(6)と同様の効果が期待できる。

【0059】

(23) 前記目的データを複数の前記カテゴリデータと関連付けて記憶させるステップが、前記目的データと前記キーワードデータとの類似度の算出をさせ、当該算出をした類似度に基づくカテゴリの評価に従い前記候補カテゴリデータを選択させることを特徴とする(21)または(22)に記載のプログラム。

【0060】

(23)の発明によれば、当該プログラムを当該サーバ上で実行することにより、(7)

10

20

30

40

50

)と同様の効果が期待できる。

【0061】

(24) 前記決定をしたカテゴリデータと関連付けて前記目的データに含まれるキーワードデータを抽出して記憶させるステップが、前記カテゴリの選択回数に基づく選択確率データを更に関連付けて記憶させ、

前記キーワードデータに関連付くカテゴリの評価が、前記類似度および前記選択確率データに基づくことを特徴とする(22)または(23)に記載のプログラム。

【0062】

(24)の発明によれば、当該プログラムを当該サーバ上で実行することにより、(8)と同様の効果が期待できる。

【発明の効果】

【0063】

本発明によれば、人手で行う場合と近い精度で自動的に目的データをカテゴリに分類できる可能性がある。更には、そうした信頼度の高い分類サンプル(学習事例)を継続的に取得し、次回以降の新たな目的データの分類に利用することで、分類の精度を高く維持できる可能性がある。

【発明を実施するための最良の形態】

【0064】

本発明に係る好適な実施形態の一例について、図面に基づいて以下に説明する。

【0065】

[全体構成]

図1は、本発明の好適な実施形態の一例に係るデータ分類方法の概念図である。

【0066】

カテゴリ分類を実施するサーバ10は、ユーザが目的データを閲覧するために使用する端末装置20と、通信ネットワーク30を介して接続され、端末装置20のユーザに対する情報提供サービスを運用する。

【0067】

端末装置20では、ユーザがサーバ10に記憶されている商品等の情報提供項目(目的データ)を検索する場合に、ツリー状にカテゴリ分けされたリンクを辿って、希望する種類の目的データ(商品データ等)群を閲覧する。目的データとカテゴリとの関連付けについてはサーバ10が記憶している。

【0068】

サーバ10は、分類するカテゴリを学習するデータ(商品A)を受け付けた際、分類の候補となるカテゴリ(分類1、分類2)を付与する。端末装置20のユーザは、付与されたカテゴリのうち、いずれかを選択して目的のデータ(商品A)に辿り着くが、このときに選択されたカテゴリをサーバ10は記憶する。

【0069】

このようなカテゴリ選択の履歴を所定の期間記憶することにより、サーバ10は、ユーザの考えに近いカテゴリ分類を学習することができる(処理の詳細は後述する)。その結果、サーバ10は、新たな目的データ(商品B)を受け付けた際に、学習したカテゴリ(例えば分類2)に分類する。

【0070】

[コンピュータの構成]

図2は、本発明の好適な実施形態の一例に係るコンピュータシステムを構成する各コンピュータ(サーバ10、または端末装置20)の構成を示すブロック図である。

【0071】

制御部110、記憶部120、入力部130、表示部140、通信制御部150は、バス160を介して接続されている。

【0072】

制御部110は、情報の演算、処理を行う情報演算処理装置(CPU)であり、当該コ

10

20

30

40

50

ンピュータ全体の制御を行う。制御部 110 は、記憶部 120 に記憶された各種プログラムを適宜読み出して実行することにより、上述のハードウェアと協働し、本発明に係る各種機能を実現している。

【0073】

記憶部 120 は、制御部 110 と組み合わせてプログラムの実行に使用するローカルメモリ、大容量のバルクメモリ、および当該バルクメモリの検索を効率的に行うために使用するキャッシュメモリを含んでよい。記憶部 120 を実現するコンピュータ可読媒体としては、電氣的、磁氣的、光学的、電磁的に実現するものを含んでよい。より具体的には、半導体記憶装置、磁気テープ、フロッピー（登録商標）ディスク、ランダム・アクセス・メモリ（RAM）、リードオンリー・メモリ（ROM）、CD-ROM と CD-R/W と DVD とを含む光ディスクが含まれる。

10

【0074】

入力部 130 は、ユーザによる入力の受付を行うものであり、キーボード、ポインティング・デバイス等を含んでよい。入力部 130 は、直接または介在 I/O コントローラを介して当該コンピュータと接続することができる。

【0075】

表示部 140 は、ユーザにデータの入力を受け付ける画面を表示したり、当該コンピュータによる演算処理結果の画面を表示したりするものであり、ブラウン管表示装置（CRT）、液晶表示装置（LCD）等のディスプレイ装置を含む。表示部 140 は、直接または介在 I/O コントローラを介して当該コンピュータと接続することができる。

20

【0076】

通信制御部 150 は、当該コンピュータを専用ネットワークまたは公共ネットワークを介して別の演算処理システムまたは記憶装置と接続できるようにするためのネットワーク・アダプタである。通信制御部 150 は、モデム、ケーブル・モデムおよびイーサネット（登録商標）・アダプタを含んでよい。

【0077】

[学習処理フロー]

図 3 は、本発明の好適な実施形態の一例に係るデータ分類の学習処理フローを示す図である。

【0078】

ステップ S105 では、サーバ 10 は、学習用の目的データを受け付ける。当該目的データは入力部 130 から受け付けることとしてもよいし、通信制御部 150 を介して別の演算処理システムあるいは記憶装置から受信することとしてもよい。目的データは、例えば商品に関するデータの場合には、商品名や商品説明等の情報を含んでおり、本学習処理を通じて、分類すべきカテゴリを付与する。

30

【0079】

ステップ S110 では、サーバ 10 は、受け付けた目的データに対して、分類の候補となる複数のカテゴリを付与する。このカテゴリの付与については、人手によってもよいが、既に分類済みのサンプルが存在すればサーバが自動で行う（詳細は図 4 にて後述する）ことが望ましい。

40

【0080】

ステップ S105 および S110 の結果、例えば商品データの場合には、図 5 に示すような商品分類テーブル 40 に、候補となるカテゴリ、商品名、商品説明を記憶する。これらのデータを基に、端末装置 20 においてカテゴリに紐付けて商品データを表示し、ユーザに対する情報提供サービスを運用する。

【0081】

ステップ S115 では、サーバ 10 は、端末装置 20 からカテゴリ選択を表すデータを受信する。具体的には、サーバ 10 は、端末装置 20 のユーザが辿ってきたカテゴリ（リンク）を示すデータを受信し、当該ユーザが目的データを選択（閲覧）した際に、ステップ S110 にて付与された候補カテゴリの中から選択されたカテゴリデータを判別する。

50

【 0 0 8 2 】

ステップ S 1 2 0 では、サーバ 1 0 は、候補カテゴリのそれぞれが選択された回数をカウントアップして記憶する。具体的には、例えば図 6 に示すアクセス履歴テーブル 5 0 を利用する。目的データ（商品名フィールド）に対してステップ S 1 1 0 にて付与した候補カテゴリ（カテゴリフィールド）のそれぞれについて、ステップ S 1 1 5 にて受信したカテゴリ選択を表すデータを受信する毎に選択回数フィールドの値をカウントアップして記憶する。

【 0 0 8 3 】

ステップ S 1 2 5 では、サーバ 1 0 は、目的データを分類するカテゴリを決定するか否かを判別する。具体的には、例えば、目的データに候補カテゴリを付与してユーザへの閲覧を開始してから所定の期間が経った場合に分類するカテゴリを決定することとしてもよいし、選択回数が所定の数に達した場合に決定することとしてもよい。

10

【 0 0 8 4 】

あるいは、サーバ 1 0 は、この時点で分類するカテゴリを決定せず、後に新たな目的データを受け付けたときの選択回数に基づいて分類するカテゴリを決定することとしてもよい。その場合には、サーバ 1 0 は、より長期間にわたってユーザの傾向を記憶し、カテゴリ分類に利用することができる。なお、このときには、ステップ S 1 2 5 および S 1 3 0 は不要となる。

【 0 0 8 5 】

ステップ S 1 3 0 では、サーバ 1 0 は、目的データを分類するカテゴリを決定する。具体的には、例えば以下のような方法で決定することができる。

20

【 0 0 8 6 】

(1) 選択回数の最も多いカテゴリにする。この場合、サーバ 1 0 は、目的データに付与した候補カテゴリの中から、最も選択回数の多いものを選択して目的データに関連付ける。

【 0 0 8 7 】

(2) 選択回数の上位複数のカテゴリにする。この場合、サーバ 1 0 は、目的データに付与した候補カテゴリのうち、選択回数の多いものから順に所定の数を選択して目的データに関連付ける。

【 0 0 8 8 】

(3) 選択回数が他と比べて多いカテゴリにする。この場合、サーバ 1 0 は、例えば各候補カテゴリについての選択回数の偏差値を算出し、当該偏差値が所定の値以上のものを選択して目的データに関連付ける。

30

【 0 0 8 9 】

分類するカテゴリの決定方法は、以上に限られないが、このような方法によれば、サーバ 1 0 は、多くのユーザが選択する傾向にあるカテゴリに対して目的データを効果的に分類することができる。このことにより、ユーザが辿りやすいカテゴリのみを残して、選択されにくい（精度の悪い）カテゴリを排除することもできる。

【 0 0 9 0 】

なお、所定の期間が経ってもなお、分類すべきカテゴリが決定できないと判断した場合には、サーバ 1 0 はその旨を管理者の端末に表示して候補カテゴリの再設定を促してもよい。

40

【 0 0 9 1 】

ステップ S 1 3 5 では、サーバ 1 0 は、ステップ S 1 3 0 にて決定したカテゴリ（ステップ S 1 2 5 およびステップ S 1 3 0 を行わない場合には全候補カテゴリ）と関連付けて、目的データに含まれるキーワードと、カテゴリの選択回数に基づいて計算する当該カテゴリの選択確率データを記憶する。具体的には、例えば、図 7 に示す分類学習テーブル 6 0 を利用する。

【 0 0 9 2 】

分類学習テーブル 6 0 には、目的データの名称（例えば商品名等）、目的データに含ま

50

れるキーワード、ステップS 1 3 0にて決定したカテゴリ、カテゴリの選択確率を関連付けて記憶する。ここで、キーワードは、目的データの特徴付ける単語であって、例えば商品データの場合には、商品分類テーブル4 0（図5）の商品説明を、形態素解析等によって語句に分解し、TF * IDFや確率言語モデルといった手法を用いることにより抽出することができる。

【0093】

なお、記憶手段はテーブルには限られず、転置索引ファイルとして記憶することもできる。更に、記憶するキーワードには、出現場所や出現回数、あるいは他の分類サンプル（学習事例）とあわせて中での出現頻度等を関連付けて記憶することとしてもよい。これらは後述する類似度の算出において利用することができる。

【0094】

なお、ステップS 1 3 5は、本処理フローにおいては、分類カテゴリの決定後に行うこととしたが、タイミングはこれに限られず、例えばキーワードについては、ステップS 1 0 5にて目的データを受け付けた際に記憶してもよいし、選択確率は、ステップS 1 2 0の選択回数とあわせて記憶することもできる。

【0095】

[分類処理フロー]

図4は、本発明の好適な実施形態の一例に係るカテゴリへの分類処理フローを示す図である。当該分類処理は、学習処理フロー（図3）において学習したカテゴリ分類に基づいて、新たな目的データを分類する処理であるが、当該学習処理フローのステップS 1 1 0における候補カテゴリの付与においても用いることができる。

【0096】

ステップS 2 0 5では、サーバ1 0は、カテゴリ分類を行う新たな目的データを受け付ける。当該新たな目的データは入力部1 3 0から受け付けることとしてもよいし、通信制御部1 5 0を介して別の演算処理システムあるいは記憶装置から受信することとしてもよい。

【0097】

ステップS 2 1 0では、サーバ1 0は、学習処理フロー（図3）のステップS 1 3 5において記憶した分類学習テーブル6 0または転置索引ファイル等に基づいて、当該新たな目的データとカテゴリ毎のキーワード群との類似度を算出する。具体的には、例えば、TF * IDF、あるいはk NN法やNaive Bayes法といった手法を用いることにより、類似度を算出することができる。このとき、キーワードと関連付けて出現頻度等のデータを前もって記憶していれば、これらを用いることができる。このことにより、サーバ1 0は、当該新たな目的データと類似度の高いキーワード群と、それに対応するカテゴリを抽出することができる。

【0098】

ステップS 2 1 5では、サーバ1 0は、ステップS 2 1 0にて算出した類似度に基づいて、分類するカテゴリを決定する。具体的には、サーバ1 0は当該類似度を用いて、対応するカテゴリそれぞれについて分類すべきか否かを評価する。このとき、サーバ1 0は、例えば以下のように分類するカテゴリを一つまたは複数決定する。

【0099】

（1）類似度が最も高いキーワード群に対応するカテゴリ、あるいは類似度の高い順に所定の個数のカテゴリに分類する（Naive Bayes法等）。この場合、例えば「バッグ」、「ヌメ革」を商品説明に含んだ目的データは、分類学習テーブル6 0（図7）の例によれば、1行目の学習事例のキーワードとの類似度が高くなり、サーバ1 0は、当該目的データを「ハンドバッグ」のカテゴリに分類する。

【0100】

このように、サーバ1 0は、当該新たな目的データを受け付けた時点で記憶している最新のキーワードとカテゴリの組み合わせを参照して、当該新たな目的データの分類に活用する。なお、この分類学習テーブル6 0は、前述のように学習処理フロー（図3）のステ

10

20

30

40

50

ップS130において実際のユーザによる選択実績に基づいて、不要と考えられる候補カテゴリは排除されているため、特許文献1のように誤分類が蓄積されることによる精度の悪化も抑止できる。

【0101】

(2)類似度の高い適切な個数のサンプル中で分類するカテゴリを多数決する(kNN法)。この場合、例えば、目的データとキーワードの類似する学習事例が3件あったとき、対応するカテゴリが「ハンドバッグ」2件、「トートバッグ」1件であれば、サーバ10は、当該目的データを多数決により「ハンドバッグ」に分類する。

【0102】

(3)類似度による重み付けをした上で、適切な個数のサンプル中で分類するカテゴリを多数決する(kNN法)。この場合、上記(2)の例では、「トートバッグ」よりも「ハンドバッグ」の方がキーワードの類似する学習事例の件数は多いが、「トートバッグ」の類似度が勝っていれば、サーバ10は、重み付けの結果「トートバッグ」をより高く評価して分類する可能性がある。

【0103】

(4)同一カテゴリに対応付けられたキーワード群との類似度の平均が最も高いカテゴリ、あるいは高い順に所定の個数のカテゴリに分類する。この場合は、サーバ10は、同じカテゴリに分類される学習事例の中で、目的データとの類似度にばらつきがあるものよりも、最高の類似度でなくても偏りが無いものに分類する可能性がある。

【0104】

(5)類似度に対して選択確率の重み付けをした上で、上述の(1)~(4)の方法により分類する。この場合は、選択確率が高ければ分類の精度が高いと仮定し、サーバ10は、例えば分類学習テーブル60(図7)の選択確率を類似度に掛け合わせて評価することにより、精度の高い学習事例に基づいたカテゴリに分類することができる。

【0105】

以上のように、本発明の実施形態によれば、情報提供サービスを運用することで目的データを分類するカテゴリの学習ができ、学習したカテゴリに対して新たな目的データを自動で分類することができる。

【0106】

ここで、学習の結果(学習事例)については、数多く、そして新しい方がより精度が高く、ユーザの意見を反映していることが期待できる。そこで、学習処理については定期的に行うことが好ましく、例えば、新たな目的データを自動で分類する際にも複数のカテゴリを付与し、学習を行うこととしてもよい。

【0107】

このようにして、サーバ10は、情報提供サービスを運用する中で学習事例を継続的に取得・利用することができ、カテゴリ分類の精度を高く維持することができる。

【0108】

以上、本発明の実施形態について説明したが、本発明は上述した実施形態に限るものではない。また、本発明の実施形態に記載された効果は、本発明から生じる最も好適な効果を列挙したに過ぎず、本発明による効果は、本発明の実施例に記載されたものに限定されるものではない。

【図面の簡単な説明】

【0109】

【図1】本発明の好適な実施形態の一例に係るデータ分類方法の概念図である。

【図2】本発明の好適な実施形態の一例に係るコンピュータシステムを構成する各コンピュータの構成を示すブロック図である。

【図3】本発明の好適な実施形態の一例に係るデータ分類の学習処理フローを示す図である。

【図4】本発明の好適な実施形態の一例に係るカテゴリへの分類処理フローを示す図である。

10

20

30

40

50

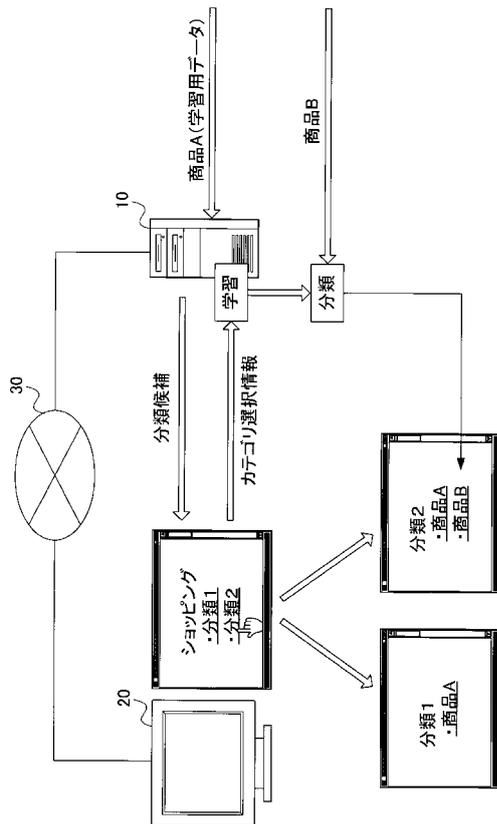
【図5】本発明の好適な実施形態の一例に係る商品分類テーブルを示す図である。
 【図6】本発明の好適な実施形態の一例に係るアクセス履歴テーブルを示す図である。
 【図7】本発明の好適な実施形態の一例に係る分類学習テーブルを示す図である。

【符号の説明】

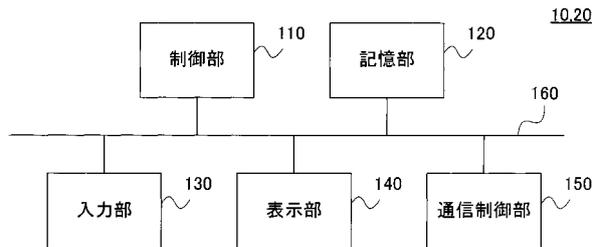
【0110】

- 10 サーバ
- 20 端末装置
- 30 通信ネットワーク
- 40 商品分類テーブル
- 50 アクセス履歴テーブル
- 60 分類学習テーブル
- 110 制御部
- 120 記憶部
- 130 入力部
- 140 表示部
- 150 通信制御部
- 160 バス

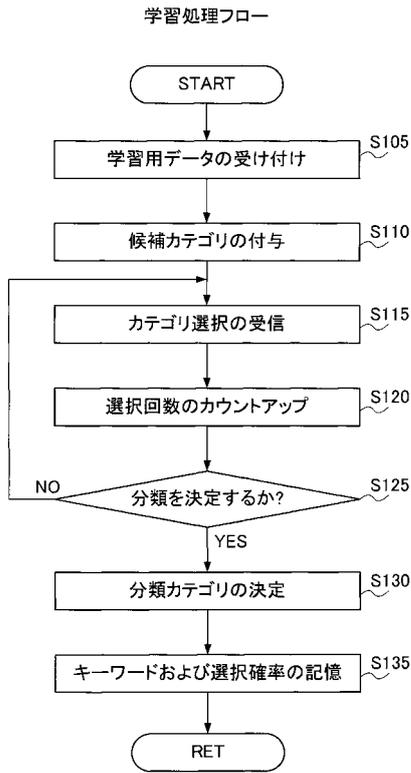
【図1】



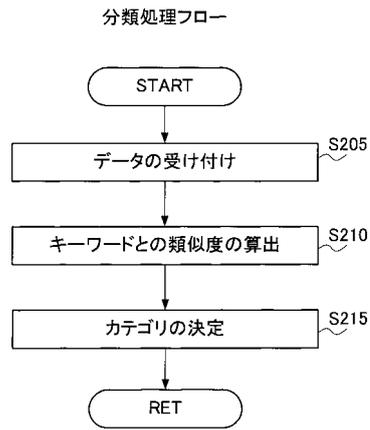
【図2】



【 図 3 】



【 図 4 】



【 図 5 】

商品分類テーブル 40

カテゴリ	商品名	商品説明
ハンドバッグ	AAA/バッグ モノグラム M12345	AAA/バッグの2006年新作 モデルです。シンプルかつ 機能的なデザインのバッ グです。側面のヌメ革のベ ルトがデザイン的にも…
トートバッグ		
セカンドバッグ、ポーチ		
⋮	⋮	⋮

【 図 6 】

アクセス履歴テーブル 50

商品名	カテゴリ	選択回数
AAA/バッグ	ハンドバッグ	51
モノグラム M12345	トートバッグ	29
	セカンドバッグ、ポーチ	3
BBBかばん	ハンドバッグ	36
P6789	トートバッグ	37
	セカンドバッグ、ポーチ	2
⋮	⋮	⋮

【 図 7 】

分類学習テーブル 60

商品名	キーワード	カテゴリ	選択確率
AAA/バッグ モノグラム M12345	バッグ ヌメ革 ⋮	ハンドバッグ	61%
BBBかばん P6789	かばん スポーティー ⋮	ハンドバッグ	48%
		トートバッグ	49%
⋮	⋮	⋮	⋮

フロントページの続き

(56)参考文献 特開2002-297642(JP,A)

徳永 健伸, 言語と計算5 情報検索と言語処理, 財団法人東京大学出版会, 1999年11月
25日, 第1版, P.193-199

(58)調査した分野(Int.Cl., DB名)

G06F 17/30