

【公報種別】特許法第 17 条の 2 の規定による補正の掲載
 【部門区分】第 6 部門第 3 区分
 【発行日】平成28年11月4日 (2016.11.4)

【公表番号】特表2016-502701(P2016-502701A)
 【公表日】平成28年1月28日 (2016.1.28)
 【年通号数】公開・登録公報2016-006
 【出願番号】特願2015-537832(P2015-537832)
 【国際特許分類】

G 0 6 F 17/22 (2006.01)

G 0 6 N 99/00 (2010.01)

G 0 6 F 11/00 (2006.01)

【 F I 】

G 0 6 F 17/22 6 8 2

G 0 6 N 99/00 1 5 0

G 0 6 F 9/06 6 3 0 A

【手続補正書】

【提出日】平成28年9月14日 (2016.9.14)

【手続補正 1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項 1】

各々が部分表現を含むプログラム表現を含む候補変換プログラムを順位付けして、ユーザーによって入力される入力文字列から、1つまたは複数のユーザー供給される入力 - 出力例の各々と調和する、各々がユーザー所望の形式で出力文字列を生成する1つまたは複数の変換プログラムの順位付けされたグループを確立するために、コンピューターにより実装されるプロセスであって、

コンピューターを使用して、以下のプロセス動作、すなわち、

前記1つまたは複数のユーザー供給される入力 - 出力例から帰納的に合成された、各々が各ユーザー供給される入力例から各ユーザー供給される出力例によって提示される形式で出力文字列を生成する候補変換プログラムのセットを入力することと、

各候補変換プログラムについて、

より小さい方からより大きい方への順に、順位付け体系が確立された、前記候補変換プログラムの各部分表現について、前記部分表現についての前記順位付け体系を使用して尤度スコアを確立することと、

前記候補変換プログラムについて確立された前記部分表現の尤度スコアから、前記候補変換プログラムについての全体的な順位付けスコアを算出することとを実行させることを含む、コンピューターにより実装されるプロセス。

【請求項 2】

部分表現が、正規表現または位置表現またはアトミック表現または連結表現のいずれかであり、前記順位付け体系が、前記正規表現、位置表現、アトミック表現、および連結表現の各々をこの順序で順位付けするために確立される、請求項 1 に記載のプロセス。

【請求項 3】

アトミック表現が、定数文字列表現または部分文字列表現のいずれかであり、アトミック表現についての順位付け体系を確立するプロセス動作が、

前記アトミック表現が定数文字列表現であるのか、または部分文字列表現であるのかを

識別する動作と、

各タスクが複数の入力 - 出力例から成るタスクのトレーニングセットを入力する動作と

、

前記トレーニングセットからトレーニングデータを生成する動作と、

各部分文字列表現および各定数文字列表現に尤度スコアを割り当てるために、機械学習分類子をトレーニングする動作と含む、請求項 2 に記載のプロセス。

【請求項 4】

各部分文字列表現に尤度スコアを割り当てるプロセス動作が、0 から 1 までの範囲における値を有する尤度スコアを割り当てることを含み、各定数文字列表現に尤度スコアを割り当てるプロセス動作が、所定の尤度スコア値を割り当てることを含み、請求項 3 に記載のプロセス。

【請求項 5】

前記タスクのトレーニングセットからトレーニングデータを生成するプロセス動作が、前記トレーニングセット内の各タスクについて、

前記タスク内の入力 - 出力例ごとに、

前記入力 - 出力例から帰納的に合成され、各々が前記例内の入力文字列のタプルから前記例内の前記出力文字列を生成する、変換プログラムのセットを確立する動作と、

前記変換プログラムのセットから正のトレーニング部分表現のセットを確立する動作と、

前記変換プログラムのセットから負のトレーニング部分表現のセットを確立する動作と、

前記正および負のトレーニング部分表現のセット内の各部分文字列表現から前記トレーニングデータを生成する動作とを含む、請求項 3 に記載のプロセス。

【請求項 6】

前記正および負のトレーニング部分表現のセット内の各部分文字列表現からトレーニングデータを生成する前記プロセス動作が、

特徴の各々が、前記出力文字列内の部分文字列が前記正および負のトレーニング部分表現のセット内の部分文字列表現によって生成されるかを示す、1 つまたは複数の特徴のセットを入力する動作と、

前記正および負のトレーニング部分表現のセット内の各部分文字列表現について、各要素が、クラスラベルのために予約される要素を除いて、前記特徴のうちの 1 つにそれぞれ関連付けられる特徴値に対応する、要素の所定の順序を含む特徴ベクトルを生成する動作とを含み、前記特徴ベクトルの生成は、

前記要素の所定の順序に対応する順序における前記特徴の各々について、前記特徴が考慮中の前記部分文字列表現によって生成される前記出力文字列内の部分文字列において提示される場合、第 1 のバイナリ値を有する特徴値を、対応する特徴ベクトル要素に割り当て、前記特徴が考慮中の前記部分文字列表現によって生成される前記出力文字列中の部分文字列において提示されない場合、第 2 のバイナリ値を有する特徴値を、前記対応する特徴ベクトル要素に割り当てることと、

前記クラスラベルのために予約される前記特徴ベクトル要素について、前記部分文字列表現が前記正のトレーニング部分表現のセットに存在する場合、1 つのバイナリ値を割り当て、前記部分文字列表現が前記負のトレーニング部分表現のセットに存在する場合、異なるバイナリ値を割り当てることとを含む、請求項 5 に記載のプロセス。

【請求項 7】

前記 1 つまたは複数の特徴のセットは、

前記出力部分文字列の左端位置がトークンによって認識されること、

前記出力部分文字列の前記左端位置の文字が定数であること、

前記出力部分文字列の右端位置がトークンによって認識されること、

前記出力部分文字列の前記右端位置の文字が定数であること、

出力部分文字列はトークンを示すこと、

前記入力部分文字列の左端位置がトークンによって認識されること、
前記入力部分文字列の右端位置がトークンによって認識されること、
入力部分文字列はトークンであること、
前記出力部分文字列を取得するために実行されるケーシング、
前記部分文字列の長さ、
前記入力文字列の前記長さに対する前記部分文字列の相対的な長さ、
前記出力文字列の前記長さに対する前記部分文字列の相対的な長さ、または
構成要素となる位置表現の尤度スコア、
の少なくとも1つを含む、請求項6に記載のプロセス。

【請求項8】

前記機械学習分類子は、サポートベクターマシン（SVM）バイナリ分類子である、請求項6に記載のプロセス。

【請求項9】

タスク内の入力 - 出力例から確立される前記変換プログラムのセットから正のトレーニング部分表現のセットを確立する前記プロセス動作が、前記タスク内の前記入力 - 出力例から合成される変換プログラムの全てのセットの共通部分において見出される各部分表現を、正のトレーニング部分表現として指定することを含む、請求項5に記載のプロセス。

【請求項10】

前記タスク内の前記入力 - 出力例から確立される前記変換プログラムのセットから負のトレーニング部分表現のセットを確立する前記プロセス動作が、正のトレーニング部分表現として識別されない前記変換プログラムのセット内の各部分表現を、負のトレーニング部分表現として指定することを含む、請求項9に記載のプロセス。

【請求項11】

正規表現および位置表現について順位付け体系を確立する前記プロセス動作が、有限な数の予め定義されたトークンを有するセットから値を取る出力部分文字列を生成する部分表現を、前記正のトレーニング部分表現のセット内のそれらの部分表現の頻度に基づいて、頻度に基づくスコアにマッピングする辞書を算出する動作を含む、請求項5に記載のプロセス。

【請求項12】

連結表現について順位付け体系を確立する前記プロセス動作が、連結表現において見出される個々のアトミック表現についてのスコアを結合することを含む、請求項2に記載のプロセス。

【請求項13】

連結表現において見出だされる個々のアトミック表現についてスコアを連結する前記プロセス動作が、前記スコアを乗算すること、または前記スコアを加算することの1つを含む、請求項12に記載のプロセス。

【請求項14】

1つまたは複数の変換プログラムの前記順位付けされたグループとして、所定の数の上位の候補変換プログラムを選択するプロセス動作をさらに含む、請求項1に記載のプロセス。

【請求項15】

1つまたは複数の変換プログラムの前記順位付けされたグループとして、所定の数の上位の候補変換プログラムを選択する前記プロセス動作が、前記最高位の候補変換プログラムを選択することを含む、請求項14に記載のプロセス。

【請求項16】

文字列変換を実行するためにコンピューターにより実装されるプロセスであって、コンピューターを使用して、以下のプロセス動作、すなわち、
1つまたは複数のユーザー供給される入力 - 出力例から帰納的に合成された、各々が各ユーザー供給される入力例から各ユーザー供給される出力例によって提示される形式で出力文字列を生成する、各々が部分表現を含むプログラム表現を含む候補変換プログラムの

セットを入力することと、

最高位の変換プログラムを識別するために前記候補変換プログラムのセットを順位づけることと、

出力文字列を生成するためにユーザー供給された入力文字列に前記最高位の変換プログラムを適用することとを実行させることを含む、コンピュータにより実装されるプロセス。

【請求項 17】

各々のプログラムが部分表現を含むプログラム表現を含む候補変換プログラムを順位付けして、ユーザーによって入力される入力文字列から、1つまたは複数のユーザー供給される入力 - 出力例の各々と調和する、各々がユーザー所望の形式で出力文字列を生成する1つまたは複数の変換プログラムの順位付けされたグループを確立するためのシステムであって、

1つまたは複数のコンピューティング装置であって、前記コンピューティング装置は複数のコンピューティング装置がある場合はいつでもコンピュータネットワークを介して相互に通信する、1つまたは複数のコンピューティング装置と、

前記1つまたは複数のコンピューティング装置によって実行されるプログラムモジュールを有するコンピュータプログラムとを含み、前記1つまたは複数のコンピューティング装置は、

前記1つまたは複数のユーザー供給される入力 - 出力例から帰納的に合成された、各々が各ユーザー供給される入力例から各ユーザー供給される出力例によって提示される形式で出力文字列を生成する前記候補変換プログラムのセットを入力し、

各候補変換プログラムについて、

より小さい方からより大きい方への順に、順位付け体系が確立された、前記候補変換プログラムの各部分表現について、前記部分表現について確立された前記順位付け体系を使用して尤度スコアを確立し、

前記候補変換プログラムについて確立された前記部分表現の尤度スコアから、前記候補変換プログラムについての全体的な順位付けスコアを算出するように、前記コンピュータプログラムの前記プログラムモジュールにより誘導される、システム。

【請求項 18】

候補変換プログラムの順位付けの前に前記順位付け体系を確立するためのプログラムモジュールをさらに含み、各順位付け体系は前記部分表現が、ユーザーによって入力される入力文字列から、前記1つまたは複数のユーザー供給される入力 - 出力例の各々と調和する、ユーザー所望の形式で出力文字列を生成することができるプログラムの一部である尤度を示す尤度スコアを生成し、前記順位付け体系確立プログラムモジュールが、正規表現、位置表現、アトミック表現、および連結表現の各々をこの順序で順位付けする前記順位付け体系を確立するためのサブモジュールを含み、部分表現は、正規表現または位置表現またはアトミック表現または連結表現のいずれかである、請求項 17 に記載のシステム。

【請求項 19】

順位付け体系が確立された前記候補変換プログラムの各部分表現について、前記部分表現のために確立された前記順位付け体系を使用して尤度スコアを確立する前記プログラムモジュールは、

正規表現および位置表現を頻度に基づくスコアにマッピングする辞書を採用することを含む、正規表現および位置表現のための頻度に基づく順位付けを使用することと、

アトミック表現のための特徴に基づく順位付けを使用することであって、アトミック表現は定数文字列表現または部分文字列表現のどちらかであり、前記特徴に基づく順位付けは、

前記アトミック表現が定数文字列表現または部分文字列表現のどちらかあるか特定することと、

前記アトミック表現が定数文字列表現である場合はいつでも、所定の尤度スコア値を割り当てることと、

前記アトミック表現が部分文字列表現である場合にはいつでも、尤度スコアを各部分文字列表現に割り当てるようにトレーニングされた機械学習分類子を使用して尤度スコア値を割り当てることと、を含む、アトミック表現のための特徴に基づく順位付けを使用することと、

連結表現において見出される個々のアトミック表現についてのスコアを結合することを含む、連結表現を順位付けるためのパスに基づく順位付けを使用すること、のためのサブモジュールを含む、請求項 18 に記載のシステム。

【請求項 20】

1 つまたは複数の変換プログラムの前記順位付けされたグループとして、所定の数の前記上位の候補変換プログラムを選択するためのプログラムモジュールをさらに含む、請求項 17 に記載のシステム。