

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7181439号
(P7181439)

(45)発行日 令和4年11月30日(2022.11.30)

(24)登録日 令和4年11月21日(2022.11.21)

(51)国際特許分類 F I
G 1 6 H 10/40 (2018.01) G 1 6 H 10/40

請求項の数 18 (全30頁)

(21)出願番号	特願2022-531991(P2022-531991)	(73)特許権者	000001926
(86)(22)出願日	令和3年6月22日(2021.6.22)		塩野義製薬株式会社
(86)国際出願番号	PCT/JP2021/023524		大阪府大阪市中央区道修町3丁目1番8号
(87)国際公開番号	WO2021/261468	(74)代理人	100145403
(87)国際公開日	令和3年12月30日(2021.12.30)		弁理士 山尾 憲人
審査請求日	令和4年7月7日(2022.7.7)	(74)代理人	100156144
(31)優先権主張番号	特願2020-108248(P2020-108248)		弁理士 落合 康
(32)優先日	令和2年6月23日(2020.6.23)	(74)代理人	100132241
(33)優先権主張国・地域又は機関	日本国(JP)		弁理士 岡部 博史
早期審査対象出願		(74)代理人	100135703
			弁理士 岡部 英隆
		(74)代理人	100189544
			弁理士 柏原 啓伸
		(72)発明者	木口 亮

最終頁に続く

(54)【発明の名称】 臨床試験の解析プログラムの生成を補助するためのプログラム生成補助システム

(57)【特許請求の範囲】

【請求項1】

臨床試験の解析プログラムの生成を補助するためのプログラム生成補助システムであって、

前記臨床試験の解析方法および解析結果の出力形式を指定する解析図表計画書から生成されたテキストデータおよび画像データを取得するインタフェース装置と、

前記テキストデータおよび前記画像データを記憶する記憶装置と、

前記テキストデータを利用する第1の候補予測方法および前記画像データを利用する第2の候補予測方法をそれぞれ実行することにより、前記解析図表計画書を、予め定められた複数のパターンのうちの少なくとも1つのパターンに分類し、結果を分類候補として出力する処理回路と

を備え、

前記複数のパターンは、前記臨床試験の解析方法および解析結果の出力形式に応じて予め分類されており、

前記プログラム生成補助システムには、

複数のテキスト列の各々と、各テキスト列が分類されるパターンとの対応関係を規定した判定ルールが予め用意されており、かつ、

過去の複数の臨床試験の各解析図表計画書の画像データと前記各解析図表計画書の画像データが分類された正解パターンとを教師データとして学習が行われた学習済みの人工ニューラルネットワークが構築されており、

前記処理回路は、

前記第1の候補予測方法を実行することにより、前記解析図表計画書の前記テキストデータに含まれるテキスト列と前記判定ルールとを利用して、前記複数のパターンのうちの第1パターンを示すデータを出力し、

前記第2の候補予測方法を実行することにより、前記解析図表計画書の画像データを前記人工ニューラルネットワークへ入力し、前記人工ニューラルネットワークから出力された第2パターンを示すデータを取得し、

前記第1パターンおよび前記第2パターンを、解析資料の前記分類候補として出力する、プログラム生成補助システム。

【請求項2】

前記判定ルールは、臨床試験の中止理由、解析対象集団、患者背景、服薬、有害事象、有効性を示す各テキスト列のうちの2種類以上のテキスト列の各々と、各テキスト列が分類されるパターンとの対応関係を規定する、請求項1に記載のプログラム生成補助システム。

【請求項3】

前記人工ニューラルネットワークは、畳み込みニューラルネットワークである、請求項1または請求項2に記載のプログラム生成補助システム。

【請求項4】

前記畳み込みニューラルネットワークは、前記解析図表計画書の画像データが前記複数のパターンの各々に該当する確率値を算出する、請求項3に記載のプログラム生成補助システム。

【請求項5】

前記畳み込みニューラルネットワークは、各パターンに該当する確率値のうち、最大の確率値をとるパターンを示すデータを出力する、請求項4に記載のプログラム生成補助システム。

【請求項6】

臨床試験の解析プログラムの生成を補助するためのプログラム生成補助システムであって、

データベースを記憶した記憶装置と、

前記臨床試験の解析方法および解析結果の出力形式を指定する解析図表計画書のデータを取得するインタフェース装置と、

前記データベースおよび前記解析図表計画書のデータを参照して、前記臨床試験の解析に使用される少なくとも1つのモデルデータの名称を候補として出力する処理回路とを備え、

前記解析図表計画書は1以上の帳票を含み、各帳票は前記臨床試験の解析の種類ごとの解析の内容を指定し、

前記データベースは、過去の臨床試験の解析資料の各帳票と、前記各帳票で使用されたモデルデータの名称であって、前記解析のためのデータセットおよびメタデータの集合であるモデルデータの名称とを記憶しており、

前記処理回路は、

取得された前記解析図表計画書の帳票ごとに、前記帳票の記載と前記過去の臨床試験の解析資料の各帳票の記載との類似度を所定の類似度評価関数を利用して算出し、

算出した前記類似度に応じて、前記帳票の記載に対応する前記過去の臨床試験の解析資料の帳票で使用された少なくとも1つのモデルデータの名称を抽出し、

抽出した前記モデルデータの名称を、前記候補として出力する、プログラム生成補助システム。

【請求項7】

前記データベースは、前記過去の臨床試験の解析資料の帳票で使用されたモデルデータと、前記モデルデータを記述する項目であって前記帳票で使用された変数とをリンクさせたリンク情報をさらに記憶しており、

10

20

30

40

50

前記処理回路は、

前記解析図表計画書の各帳票のテキストデータの第1項目と、前記リンク情報の変数の第2項目とを比較し、

一致する場合には、前記解析図表計画書の各帳票のテキストデータの第1項目と前記リンク情報の変数の第2項目とを媒介に、結合した結合帳票を生成し、

抽出した前記モデルデータの名称と一致するモデルデータが前記結合帳票内に存在する場合には、名称が一致する前記結合帳票内のモデルデータを記述する項目である変数を、抽出した前記モデルデータの変数として採用し、

抽出した前記モデルデータの名称に加えて、採用した前記変数を、前記候補として出力する、請求項6に記載のプログラム生成補助システム。

10

【請求項8】

前記インタフェース装置は、前記臨床試験で作成するモデルデータおよび前記モデルデータを記述する変数の定義が対応付けられて記述されたモデル仕様書のデータをさらに取得し、

前記処理回路は、

所定のアルゴリズムを用いて、前記解析図表計画書の各帳票のテキストデータの第1項目と、前記モデル仕様書に記述された変数を示す第3項目との一致度を算出し、

算出した前記一致度が予め定められた閾値よりも大きい変数に対応付けられたモデルデータを前記モデル仕様書から選択し、

選択した前記モデルデータが、抽出した前記モデルデータと一致する場合には、前記閾値よりも大きい変数を、抽出した前記モデルデータの変数として採用し、

20

抽出した前記モデルデータの名称に加えて、採用した前記変数を、前記候補として出力する、請求項6または請求項7に記載のプログラム生成補助システム。

【請求項9】

前記解析図表計画書の各帳票の第1項目は、前記変数のセル情報を含み、

前記第3項目は、前記変数のラベル情報を含む、請求項8に記載のプログラム生成補助システム。

【請求項10】

前記解析図表計画書の各帳票の第1項目は、前記変数のセル情報を含み、

前記第3項目は、前記変数を表すフォーマット情報を含む、請求項8に記載のプログラム生成補助システム。

30

【請求項11】

前記処理回路は、

Nグラムにより前記第1項目の文字列群および前記第3項目の文字列群を抽出し、

前記第1項目の文字列群および前記第3項目の文字列群に共通する文字列の数に基づいて前記一致度を算出する、請求項8から請求項10のいずれかに記載のプログラム生成補助システム。

【請求項12】

前記処理回路は、前記第1項目の文字列群と前記第3項目の文字列群とのすべての組み合わせの数に対する、共通する文字列の数の割合を利用して前記一致度を算出する、請求項11に記載のプログラム生成補助システム。

40

【請求項13】

前記処理回路は、取得された前記解析図表計画書の帳票ごとに、前記帳票のタイトルの記載と前記過去の臨床試験の解析図表計画書の各帳票のタイトルの記載との類似度を所定の類似度評価関数を利用して算出する、請求項6から請求項12のいずれかに記載のプログラム生成補助システム。

【請求項14】

前記処理回路は、

取得された前記解析図表計画書の帳票ごとに、前記帳票の記載をベクトル化して第1ベクトルを生成し、

50

前記過去の臨床試験の解析図表計画書に含まれる複数の帳票の各々をベクトル化して第2ベクトル群を生成し、

前記第1ベクトルと前記第2ベクトル群の各々との距離を前記類似度として出力する関数を前記類似度評価関数として用いて、前記類似度を算出する、請求項6から請求項13のいずれかに記載のプログラム生成補助システム。

【請求項15】

前記類似度はTanimoto係数またはJaccard係数である、請求項14に記載のプログラム生成補助システム。

【請求項16】

インタフェース装置と、
記憶装置と、
処理回路と

を備えた、臨床試験の解析プログラムを生成する解析プログラム生成システムであって、前記インタフェース装置は、

請求項1から請求項5のいずれかに記載のプログラム生成補助システムから出力された前記第1パターンおよび前記第2パターンを取得し、

請求項7から請求項15のいずれかに記載のプログラム生成補助システムから出力された前記モデルデータの前記名称および前記変数を取得し、

前記第1パターン、前記第2パターン、前記名称および／または前記変数を修正するための修正データを取得可能であり、

前記記憶装置は、前記複数のパターンの各々に対応するテンプレートプログラムであって、前記臨床試験の解析方法および前記解析結果の出力形式を指定可能なテンプレートプログラムを記憶しており、

前記処理回路は、

前記修正データが取得された場合には、修正された前記第1パターン、前記第2パターン、前記名称および／または前記変数と、前記テキストデータとに基づいて、

前記修正データが取得されていない場合には、取得された前記第1パターン、前記第2パターン、前記名称および／または前記変数と、前記テキストデータとに基づいて、

前記複数のパターンのいずれかを指定するパターン指定データを含むメタデータを生成し、

前記パターン指定データによって指定されるパターンに対応するテンプレートプログラムを読み出し、

前記メタデータを利用して前記テンプレートプログラムから前記解析プログラムを生成する、解析プログラム生成システム。

【請求項17】

前記テキストデータは、前記解析結果の出力形式の指定を含み、

前記処理回路は、前記テキストデータに基づいて前記解析結果の出力形式を前記テンプレートプログラムに指定し、指定された前記出力形式に従うレイアウトで前記解析結果を出力することを可能とする前記解析プログラムを生成する、請求項16に記載の解析プログラム生成システム。

【請求項18】

請求項1から請求項15のいずれかに記載のプログラム生成補助システム、又は請求項16若しくは請求項17に記載の解析プログラム生成システムにおいて、前記処理回路に処理を実行させるコンピュータプログラムを記憶するコンピュータ可読記憶媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、医薬品開発における臨床試験の解析プログラムの生成を補助するためのプログラム生成補助システムに関する。

【背景技術】

10

20

30

40

50

【0002】

医薬品開発においては、臨床試験における解析業務により幾つかのデータセットや資料が作成される。それらデータセットや資料は、医薬品の当局（PMDA／FDA／EMA）への申請時に提出されるものであり、病院等の医療機関から収集される臨床データを解析することで作成される。

【0003】

データセットや資料の作成のため、製薬会社は種々のプログラミングを行っている。例えば、臨床データを解析して所定のデータセットを作成するプログラミング、当該所定のデータセットから他のデータセットを作成するプログラミングである。また製薬会社は、当該他のデータセットから解析資料を作成するプログラミングも行っている。

10

【0004】

データセットのフォーマットは、一部または全部について製薬業界内にて標準化が図られている。よって標準化されている範囲内ではプログラミングは概ね自動化することが可能である。一方、解析資料のフォーマットは標準化されておらず、解析資料のプログラミングを自動化することは行われていなかった。解析資料には解析結果が正しく反映されていればよい程度の制約しかなく、各製薬会社、ひいては各製薬会社の各プログラマに与えられた解析資料のフォーマットに関する裁量は大きかった。

【先行技術文献】

【特許文献】

【0005】

20

【文献】特開2002-58650号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

1つの治験で得られた臨床データを利用して、多数の解析、例えば100以上の解析、が行われる。解析ごとに個々のプログラマが自由裁量で解析資料を作成するためのプログラミングを行っている現状には、業務効率を改善し得る余地がある。

本開示は、医薬品開発の臨床試験における解析業務により作成される解析資料の自動生成のためのプログラミングの自動化を補助することを目的とする。

【課題を解決するための手段】

30

【0007】

本発明の例示的な実施形態に係る、臨床試験の解析プログラムの生成を補助するための第1のプログラム生成補助システムは、臨床試験の解析方法および解析結果の出力形式を指定する解析図表計画書から生成されたテキストデータおよび画像データを取得するインタフェース装置と、テキストデータおよび画像データを記憶する記憶装置と、テキストデータを利用する第1の候補予測方法および画像データを利用する第2の候補予測方法をそれぞれ実行することにより、解析図表計画書を、予め定められた複数のパターンのうちの少なくとも1つのパターンに分類し、結果を分類候補として出力する処理回路とを備える。

このような第1のプログラム生成補助システムにおいて、複数のパターンは、臨床試験の解析方法および解析結果の出力形式に応じて予め分類されている。第1のプログラム生成補助システムには、複数のテキスト列の各々と、各テキスト列が分類されるパターンとの対応関係を規定した判定ルールが予め用意されており、かつ、過去の複数の臨床試験の各解析図表計画書の画像データと各解析図表計画書の画像データが分類された正解パターンとを教師データとして学習が行われた学習済みの人工ニューラルネットワークが構築されている。処理回路は、第1の候補予測方法を実行することにより、解析図表計画書のテキストデータに含まれるテキスト列と判定ルールとを利用して、複数のパターンのうちの第1パターンを示すデータを出力する。同様に、処理回路は、第2の候補予測方法を実行することにより、解析図表計画書の画像データを人工ニューラルネットワークへ入力し、人工ニューラルネットワークから出力された第2パターンを示すデータを取得する。更に、処理回路は、第1パターンおよび第2パターンを、解析資料の分類候補として出力する。

40

50

【0008】

本発明の別の例示的な実施形態に係る、臨床試験の解析プログラムの生成を補助するための第2のプログラム生成補助システムは、データベースを記憶した記憶装置と、臨床試験の解析方法および解析結果の出力形式を指定する解析図表計画書のデータを取得するインタフェース装置と、データベースおよび解析図表計画書のデータを参照して、臨床試験の解析に使用される少なくとも1つのモデルデータの名称を候補として出力する処理回路とを備える。

解析図表計画書は1以上の帳票を含み、各帳票は臨床試験の解析の種類ごとの解析の内容を指定する。データベースは、過去の臨床試験の解析資料の各帳票と、各帳票で使用されたモデルデータの名称であって、解析のためのデータセットおよびメタデータの集合であるモデルデータの名称とを記憶している。処理回路は、取得された解析図表計画書の帳票ごとに、帳票の記載と過去の臨床試験の解析資料の各帳票の記載との類似度を所定の類似度評価関数を利用して算出する。更に、算出した類似度に応じて、帳票の記載に対応する過去の臨床試験の解析資料の帳票で使用された少なくとも1つのモデルデータの名称を抽出し、抽出したモデルデータの名称を、候補として出力する。

【0009】

本発明の更に別の例示的な実施形態に係る、臨床試験の解析プログラムを生成する解析プログラム生成システムは、インタフェース装置と、記憶装置と、処理回路とを備える。

インタフェース装置は、第1のプログラム生成補助システムから出力された第1パターンおよび第2パターンを取得し、第2のプログラム生成補助システムから出力されたモデルデータの名称および変数を取得する。第1パターン、第2パターン、名称および／または変数を修正するための修正データを取得可能である。記憶装置は、複数のパターンの各々に対応するテンプレートプログラムであって、臨床試験の解析方法および解析結果の出力形式を指定可能なテンプレートプログラムを記憶している。処理回路は、修正データが取得された場合には、修正された第1パターン、第2パターン、名称および／または変数と、テキストデータとに基づいて、複数のパターンのいずれかを指定するパターン指定データを含むメタデータを生成する。また、処理回路は、修正データが取得されていない場合には、取得された第1パターン、第2パターン、名称および／または変数と、テキストデータとに基づいて、複数のパターンのいずれかを指定するパターン指定データを含むメタデータを生成する。更に、処理回路は、パターン指定データによって指定されるパターンに対応するテンプレートプログラムを読み出し、メタデータを利用してテンプレートプログラムから解析プログラムを生成する。

【発明の効果】

【0010】

本発明に係る実施形態によれば、医薬品開発の臨床試験における解析業務により作成される解析資料の自動生成のためのプログラミングの自動化を補助することができ、更に、解析資料作成用プログラムを自動生成することができる。

【図面の簡単な説明】

【0011】

【図1】実施の形態に係る解析資料作成用プログラム自動生成システムのシステム構成図である。

【図2】実施の形態に係る解析資料作成用プログラム自動生成システムの概略の全体機能を示すブロック図である。

【図3】ADaMデータセットの具体例である。

【図4】図表計画書である解析図表計画書の一部の例である。

【図5】認識機能部におけるデータの流れを示す図である。

【図6】学習機能部におけるデータの流れを示す図である。

【図7】「学習」後、自動生成用メタデータ作成までのデータの流れを示す図である。

【図8】解析資料におけるパターン例である。

【図9A】解析図表計画書から抽出されるテキストデータであるフォーマット情報とレイ

10

20

30

40

50

アウト情報のうち、フォーマット情報の例である。

【図9B】解析図表計画書から抽出されるテキストデータであるフォーマット情報とレイアウト情報のうち、レイアウト情報の例である。

【図10】ルールベースに基づく予測の、判定ルールの例である。

【図11】CNNのインプットデータの例(1)(2)である。

【図12】CNNモデルの例である。

【図13】モデル仕様書であるADaMデータセット定義書(ADaM specification)の例である。

【図14】帳票作成の関する定義書(TFL specification)の例である。

【図15】帳票作成に関する定義書における、修正入力前の例である。

【図16】帳票作成に関する定義書における、修正入力後の例である。

【図17】ADaM予測のアルゴリズムの例を表すフローチャートである。

【図18】変数予測のアルゴリズム(1)を表すフローチャートである。

【図19】ADaM／変数の紐づけ情報の例である。

【図20】変数予測のアルゴリズム(2)を表すフローチャートである。

【図21】変数予測のアルゴリズム(3)を表すフローチャートである。

【図22】自動生成用メタデータの例である。

【図23】解析用テンプレートプログラムの例である。

【発明を実施するための形態】

【0012】

以下、適宜図面を参照しながら、実施の形態を詳細に説明する。但し、必要以上に詳細な説明は省略する場合がある。例えば、既によく知られた事項の詳細説明や実質的に同一の構成に対する重複説明を省略する場合がある。これは、以下の説明が不必要に冗長になるのを避け、当業者の理解を容易にするためである。

【0013】

なお、発明者らは、当業者が本開示を十分に理解するために添付図面および以下の説明を提供するのであって、これらによって特許請求の範囲に記載の主題を限定することを意図するものではない。

【0014】

1. 本開示に至る経緯

医薬品開発においては、臨床試験における解析業務により幾つかのデータセットや資料が作成される。それらデータセットや資料は、医薬品の当局(PMDA/FDA/EMA)への申請時に提出されるものであり、病院等の医療機関から収集される臨床データを解析することで作成される。

【0015】

これらデータセットや資料には、

(1) EDC(Electric Data Capture; 電子的臨床検査情報収集)のRawデータ、

(2) SDTM(Study Data Tabulation Model)に関するデータセット、

(3) ADaM(Analysis Data Model)に関するデータセット、

(4) 解析資料

などが含まれる。

【0016】

これらのうち、SDTMとは、当局(PMDA/FDA/EMA)への申請時に提出する臨床試験データセットの標準モデルである。ADaMとは、当局(PMDA/FDA/EMA)への申請時に提出する臨床試験データセットの解析データモデルである。図3は、ADaMデータセットの具体例(一部)を示す図である。

【0017】

これらのデータセットや資料を作成するため、製薬会社は種々のプログラミングを行っ

10

20

30

40

50

ている。プログラミングの一例は、EDCのRawデータからSDTMを作成するためのプログラミング、及び、SDTMからADaMを作成するためのプログラミングである。さらに、ADaMから解析資料を作成するためのプログラミングも行われる。解析資料は、医薬品の承認審査を申請する際に提出される添付文書等の作成に利用される。

【0018】

SDTMに関するデータセットは、世界的な標準開発機関により規格化されており、更に、作成のためのガイダンス文書が存在している。そのため、データセット作成のためのプログラミングも自動生成に馴染む。ADaMに関するデータセットは、世界的な標準開発機関により一部は規格化されており、また、作成のためのガイダンス文書が存在している。ADaMに関するデータセットについても、データセット作成のためのプログラミングは部分的ではあるが自動生成に馴染むと言える。

10

【0019】

しかしながら、解析資料の解析結果の正しさが担保されているという前提の下で解析者が比較的自由に作成できるものであることから、有効性や安全性といった当局が求める解析資料の内容自体に関するガイダンス文書はあるが、その形式に関するガイダンス文書はなく、規格化もされていない。このため、解析資料作成のためのプログラミングは、自動化には馴染まないとされており、実際にも自動化の先例は無い。

【0020】

解析資料の自動生成のためのプログラミングの自動化が困難であった理由は、以下のような事情が複合的に存在していたからである。

20

(1) 解析資料のフォーマットは規格化されていないため、SDTMやADaMに関するデータセットを作成するプログラムと比較すると、解析資料のプログラム作成の自由度が高く、画一的なアルゴリズム化が困難である。

(2) 解析資料の作成に当たっては、それに先立って1または複数の解析図表計画書を作成する必要がある。しかし、それら解析図表計画書からプログラム作成に必要な情報を抽出することは困難であった。その理由は、解析図表計画書の作成に使用するアプリケーションソフトウェアに応じてファイルフォーマットが異なり得るからである。加えて、ファイル内の各種データの記述位置、記述内容、表現、改行コードの位置、数等のデータフォーマットの構成の自由度が大きく、データを自動的に抽出するために、読み出すべきデータの位置、サイズ等を事前に指定することは困難だった。なお、解析図表計画書は、解析を行う者が臨床試験の解析方法及び解析結果の出力形式を指定する仕様書である。

30

(3) 解析に使用するADaMに関するデータセットや変数の予測に使用するデータは殆どがテキストデータである。コンピュータを用いてテキストデータの内容を認識し、解析資料を生成するためには、高度なテキスト解析技術が必要である。更に単語等で構成される各テキスト情報間の関連性が踏まえらる必要がある。しかし、臨床試験に特化したテキスト情報間の関連性を示す電子的な辞書が存在せず、テキスト解析技術、及び予測アルゴリズムの獲得、実施が困難である。

【0021】

本開示は、これらの問題点を克服しつつ、医薬品開発の臨床試験における解析業務により作成される解析資料の自動生成のためのプログラミングの自動化を補助する、プログラム生成補助システムを提供するものである。更に本開示は、解析資料作成用プログラムを自動生成する解析プログラム生成システムを提供するものである。

40

【0022】

2. [実施の形態]

以下、添付の図面を参照して、本開示の好ましい実施の形態を説明する。

【0023】

2. 1. [システム構成]

実施の形態に係る解析資料作成用プログラム自動生成システムは、医薬品開発の臨床試験における解析業務により作成される解析資料の自動生成のためのプログラミングの自動化を補助し、更に、解析資料作成用プログラムを自動生成するための、システムである。

50

図1は、本実施の形態に係る解析資料作成用プログラム自動生成システム2のシステム構成図である。以下では「解析資料作成用プログラム自動生成システム2」を、「システム2」と略記する。

【0024】

システム2は、コンピュータ装置4及び記憶装置12を有する。コンピュータ装置4及び記憶装置12は有線または無線の通信回線で接続されており、互いにデータを送受信可能である。システム2はさらに外部ネットワーク14に接続されていてもよく、外部ネットワーク14に接続された他のコンピュータシステムとの間でデータを授受してもよい。

【0025】

コンピュータ装置4は、一つ以上のプロセッサを搭載するPC、タブレットコンピュータ、若しくはワークステーションコンピュータ等である。

【0026】

記憶装置12は、ディスクドライブやフラッシュメモリ等の、コンピュータ装置4の外部に設けられる記憶装置であり、コンピュータ装置4で用いられる各種データベース16、各種データセット、及び、各種コンピュータプログラムを記憶する。各種データベース16には、例えば、後で説明する、コンピュータ装置4のインタフェース装置6を介してコンピュータ装置4が取得するデータや、システム2が生成する各種データが、格納される。

【0027】

外部ネットワーク14は、例えば、インターネットであり、ネットワーク端子等のインタフェース装置6を介して、コンピュータ装置4と接続する。

【0028】

更にコンピュータ装置4は、インタフェース装置6、処理回路8、及びメモリ10を含む。

【0029】

インタフェース装置6は、ネットワーク端子、映像入力端子、USB端子、キーボード、マウス等を含む、外部からデータを取得可能なインタフェースユニットである。外部から、インタフェース装置6を介して、種々のデータが取得される。データは、例えば、後で説明する、対象の臨床試験の解析図表計画書データ、及び過去の臨床試験の解析図表計画書データである。データの他の例は、蓄積された過去の臨床試験における、帳票作成に関する定義書(TFL specification)、及びモデル仕様書であるADaMデータセット定義書(ADaM specification)である。取得後、これらのデータは、記憶装置12に格納され得る。記憶装置12に格納されるデータは適宜、インタフェース装置6を介して、コンピュータ装置4内に取得され得る。

【0030】

更に、システム2により生成される各種データは、記憶装置12に適宜格納される。各種データは、例えば、後で説明する、解析図表計画書から作成されるテキストデータ(フォーマット情報とレイアウト情報)である。各種データの他の例は、解析図表計画書の画像データ、帳票及び帳票の記載とモデルデータ(ADaMデータ)との対応テーブル、抽出される「ADaM specificationの情報」、モデルデータ/変数紐づけ情報(ADaM/変数紐づけ情報)等である。このようなシステム2により生成され記憶装置12に適宜格納される各種データは、インタフェース装置6を介して、コンピュータ装置4内に再び取得され得る。

【0031】

処理回路8は、プロセッサにより構成される。ここでのプロセッサは、CPU(Central Processing Unit;中央処理ユニット)やGPU(Graphics Processing Unit;画像処理ユニット)を包括するものである。本実施の形態に係るシステム2の諸機能は、各種プログラムを処理回路8が実行することによって実現される。なお、当該諸機能は、ASIC(Application Specific Integrated Circuit)などによって実現されてもよいし、これらの組合せによって実現されてもよい。

10

20

30

40

50

【0032】

本開示における処理回路8は、複数台の信号処理回路から構成されてもよい。各信号処理回路は、例えばCPU（Central Processing Unit）またはGPU（Graphics Processing Unit）であり、「プロセッサ」と呼ばれ得る。本実施の形態に係るシステム2における様々な処理の一部をあるプロセッサが実行し、他の一部の処理を他のプロセッサが実行してもよい。例えば、後で説明する「マイクロ・ルールベースに基づく予測」と「マクロ・CNNに基づく予測」とを別のプロセッサが行ってもよい。つまり、例えば、「マイクロ・ルールベースに基づく予測」を或るCPUが実行し、「マクロ・CNNに基づく予測」を或るGPUが実行する、というものであってもよい。なお「CNN」という語は、畳み込みニューラルネットワーク（Convolutional Neural Network）を意味する。本明細書では「CNN」と略記する。

10

【0033】

メモリ10は、コンピュータ装置4内部のデータ書き換えが可能な記憶部であり、例えば、多数の半導体記憶素子を含むRAM（Random Access Memory）により構成される。メモリ10は、処理回路8が様々な処理を実行する際の、具体的なコンピュータプログラムや、変数値や、パラメータ値等を一時的に格納する。なおメモリ10は、いわゆるROM（Read Only Memory）を含んでもよい。ROMには、以下で説明するシステム2の処理を実現するコンピュータプログラムが予め格納されている。処理回路8がROMからコンピュータプログラムを読み出し、RAMに展開することにより、処理回路8が当該コンピュータプログラムを実行可能になる。

20

【0034】

本開示に係るシステム2は、米国SAS institute社のSAS言語や、Python等のコンピュータ言語を用いて構築されている。本開示に係る解析資料作成用プログラム自動生成システムの構築のために用いられ得るコンピュータ言語はこれらに限定されるものではなく、勿論、他のコンピュータ言語が用いられてもよい。

【0035】

2.2. [システム動作]

図2は、本開示の実施の形態に係るシステム2の処理の流れに沿った概略の全体機能を示すブロック図である。本実施の形態に係るシステム2は、概ね以下の三つの機能に分けることができる。機能ごとのまとまりを便宜的に「機能部」と呼ぶ。

30

- (1) 認識機能部
- (2) 学習機能部
- (3) 行動機能部

【0036】

(1) 認識機能部では、対象の臨床試験の解析図表計画書（計画文書データ）から、解析資料を作成するためのプログラム生成に必要な情報が取得される。それと共に、認識機能部では、過去の臨床試験データや外部標準メタデータから、教師データや実績データが収集されて格納される。外部標準データには、例えば、CDISC標準に関するデータが含まれる。

図4にて、図表計画書である解析資料の解析図表計画書（一部）の例を示している。

40

【0037】

(2) 学習機能部では、教師データや実績データに基づき、対象の臨床試験の解析資料作成用プログラムの生成に必要な情報が予測され検索される。

【0038】

(3) 行動機能部では、学習機能部により予測され検索された情報に対して、必要に応じて、人であるシステム利用者により修正データが入力され得る。なお、修正データが入力されないこともある。続いて、行動機能部では、修正データが入力された予測情報及び検索情報、並びに、対象の臨床試験の解析図表計画書から取得された情報に基づいて、自動生成用メタデータが生成される。行動機能部では、更に、生成された自動生成用メタデータを利用してテンプレートプログラムから解析プログラムが生成される。

50

【0039】

一方で、実施の形態に係るシステム2は、概略以下の三つのサブシステムにより、構築されている。

〔第1サブシステム〕 帳票のパターンを予測する第1のプログラム生成補助システム。

〔第2サブシステム〕 モデルデータ及び変数を予測する第2のプログラム生成補助システム。

〔第3サブシステム〕 解析プログラム生成システム。

【0040】

機能部との関係では、第1及び第2サブシステムは、認識機能部の処理結果を用いる学習機能部に対応し、第3サブシステムは、学習機能部の処理結果を用いる行動機能部に対応する。

10

【0041】

なお、解析資料における「帳票」は個々の解析結果を示すものであり、テーブル、リスティング、フィギュア等により表される。モデルデータは、例えば、ADaMデータであり、解析のためのデータ及びメタデータの集合である。変数は、モデルデータ（ADaMデータ）における変数である。

【0042】

なお、上述の認識機能部、学習機能部及び行動機能部は、1つの、つまり共通の、ハードウェアによって実現する必要はない。認識機能部、学習機能部及び行動機能部の各々は、別個独立のハードウェアを用いて実現され得る。

20

【0043】

また、システム2は、上述した第1～第3サブシステムの全てを常に含む必要はない。例えば、システム2は第1及び第2サブシステムを有するが、第3サブシステムを有していなくてもよい。このとき第3サブシステムは、別個のコンピュータシステムによって実現され得る。ただし、第3サブシステムのみを含むコンピュータシステムもまた、本実施形態にかかるシステム2の範疇である。さらに、第1及び第2サブシステムもまた、別個のコンピュータシステムによって実現されてもよい。

【0044】

第1サブシステムは、解析図表計画書の「フォーマット情報とレイアウト情報」及び「画像データ」を用いて、帳票のパターン予測を、第1の候補予測方法である「ミクロ・ルールベースに基づく予測」と第2の候補予測方法である「マクロ・CNNに基づく予測」との二つのアプローチで行い、人による正解（即ち、修正）入力支援に繋げるシステムである。

30

【0045】

第2サブシステムは、解析図表計画書の「テキストデータ（フォーマット情報とレイアウト情報）」、モデル仕様書であるADaMデータセット定義書の「ADaM specificationの情報」、及び、過去の臨床試験の帳票におけるモデルデータ（ADaMデータ）と変数とのリンク情報の「モデルデータ／変数紐づけ情報（ADaM／変数の紐づけ情報）」を用いて、モデルデータ／変数予測を行い、モデルデータ／変数予測とパターン予測とを組み合わせ提示し、人による正解（即ち、修正）入力支援に繋げるシステムである。

40

【0046】

第3サブシステムでは、第1サブシステムと第2サブシステムとにより予測され検索された情報に対して、人であるシステム利用者により修正データが入力される。続いて、第3サブシステムは、修正データが入力された予測情報及び検索情報、並びに、対象の臨床試験の解析図表計画書から取得された情報に基づいて、自動生成用メタデータを生成し、更に、生成された自動生成用メタデータを利用してテンプレートプログラムから解析プログラムを生成する。

【0047】

なお本開示におけるADaMデータセット定義書（ADaM specification

50

on)とは、データモデルのモデル仕様書である。ADaMデータセット定義書(ADaM specification)は、ADaMデータをプログラミングで作成する際に不可欠な定義書であり、変数の定義が含まれる。図13は、ADaMデータセット定義書(ADaM specification)の一部である。図13(1)はデータセットの定義部分であり、図13(2)は変数の定義部分であり、図13(3)はコードリストの定義部分である。

【0048】

2.2.1. [帳票のパターンを予測する第1のプログラム生成補助システムの動作]

帳票のパターンを予測する第1のプログラム生成補助システム(第1サブシステム)(図2)の動作を説明する。

【0049】

図8は、解析資料の帳票におけるパターンの分類例を示す図である。パターンは、基本的には資料の表す内容(性質、属性)に着目して分類される。例えば、以下のような観点(a)~(e)で分類される。図8に示す分類例では、階層的な分類がなされており、21のパターン(パターン1~パターン16)に分類されることが示されている。

(a) 帳票の種類: テーブル、リスティング、フィギュア

(b) 帳票の内容: 背景情報, 有効性情報, 安全性情報

(c) 解析の方法: 要約統計量(平均、標準偏差など)、カテゴリ集計(件数、割合)

(d) 使用するデータベース構造

(e) その他(単位の有無、解析項目の種類など)

要するに、解析資料の帳票におけるパターンの分類は、臨床試験の解析方法及び解析結果の出力形式に応じて予め分類されるものである。

【0050】

第1のプログラム生成補助システムは、帳票のパターンが、21のパターンのどれであるか、分類候補を予測する。

【0051】

図5は、認識機能部の全体におけるデータの流れを示す図である。このうち、認識機能部は、準備された対象の臨床試験の解析図表計画書から、解析資料作成用プログラム作成に必要な情報を抽出して、テキストデータ(フォーマット情報とレイアウト情報)52を生成する(図5・処理(1)参照)。第1のプログラム生成補助システムは、テキストデータ(フォーマット情報とレイアウト情報)52を取得して、その後の処理を実行する。なお、第1のプログラム生成補助システムがテキストデータ(フォーマット情報とレイアウト情報)52を生成する処理を行ってもよい。

【0052】

また、認識機能部は、準備された対象の臨床試験の解析図表計画書、及び過去の臨床試験の解析図表計画書から、CNNに使用するための画像データ58を作成する(図5・処理(2)参照)。第1のプログラム生成補助システムは、画像データ58を取得して、その後の処理を実行する。

【0053】

(処理(1))において生成されるテキストデータ52は、「フォーマット情報」と「レイアウト情報」を含む。図9Aと図9Bにて、テキストデータの例の一部を示す。図9Aは、「フォーマット情報」である「cell_infoシート」を示し、図9Bは、「レイアウト情報」である「layout_infoシート」を示している。

【0054】

「フォーマット情報」である「cell_infoシート」は、例えば、以下の表1のようなデータを一部として記録している。

10

20

30

40

50

【表 1】

PRODUCT：開発品番号
PROTOCOL：試験名
TYPE_TLF_CONTENTS：出力形式（テーブル、リスティング、フィギュア）
TLF_DISPLAY_ID：帳票番号
COLUMN_NO：列番号（0から始まる）
ROW_NO：行番号（0から始まる）
CELL_DATA：該当の例・行番号のセルに含まれる情報
IS_GROUP：薬剤群情報に該当するか（該当1，非該当0）
IS_CATEGORY：変数予測対象に該当するか（該当1，非該当0）
ALL_SINGLE_ROW_LAST_IDX：最後に実線がある行番号

10

【0055】

「レイアウト情報」である「layout_infoシート」は、例えば、以下の表2のようなデータを一部として記録している。

【表 2】

PRODUCT：開発品番号
PROTOCOL：試験名
TYPE_TLF_CONTENTS：出力形式（テーブル、リスティング、フィギュア）
TLF_DISPLAY_ID：帳票番号
TITLE：帳票のタイトル
FOOTNOTE：帳票の脚注（全ての帳票に脚注がある訳ではない）
PROGRAMMING：プログラミング時の注意事項
POPULATION：解析対象集団（各帳票の2行目に記載した情報）
COLUMNS_WIDTH：表の列の幅（単位：cm）
COLUMNS_ALIGN：表の列の文字等の出力位置（L：左寄せ，C：中央寄せ）
SECTION_ORIENT：出力時の用紙の向き（landscape：横，portrait：縦）

20

【0056】

（処理（2））において生成される画像データの例を、図11（1）（2）に示す。ここでの画像データのフォーマットは、例えばpngであればよい。図11（1）（2）の画像データでは、テキスト文字列も含めて画像化されている。

【0057】

続いて、図6は、学習機能部の全体におけるデータの流れを示す図である。このうち、第1のプログラム生成補助システムは、取得したテキストデータ（フォーマット情報とレイアウト情報）52を用いて、ミクロ・ルールベースに基づく予測の処理を行い、帳票のパターンを分類候補として予測する。

【0058】

図10は、ミクロ・ルールベースに基づく予測の判定ルール（一部）の例を纏めた表である。判定のための条件が、a-1、a-2・・・sまで25個用意されている。各々の帳票、特に、表番号（表1及び表2における「帳票番号」）とタイトル（表2における「帳票のタイトル」）が、個々の条件を満たすならば、判別される可能性のあるパターンの列には「Y」が配されている。第1のプログラム生成補助システムでは、処理回路8（図1）が、取得したテキストデータ（フォーマット情報とレイアウト情報）52に含まれるテキスト文字列が上述の判定ルールに含まれる文字列に該当するか否かを照合する。この表により表される判定ルールにより、処理回路8は、パターンの予測結果（以下、『パターン（Main）の予測結果』という）を分類候補として出力する。

30

40

【0059】

50

判定ルールは、複数のテキスト列の各々と、各テキスト列の、分類されるパターン候補との対応関係を規定している。即ち、判定ルールは、臨床試験の中止理由、解析対象集団、患者背景、服薬、有害事象、及び、有効性を示す、各テキスト列のうちの２種類以上のテキスト列の各々と、各テキスト列の、分類されるパターンとの対応関係を規定している（図１０参照）。

【００６０】

なお、図１０の表におけるミクロ・ルールベースに基づく予測の判定ルールは、例示のものであり、これらは変動するものであってもよい。

【００６１】

更に、図６に示すように、第１のプログラム生成補助システムは、取得した解析図表計画書の画像データ５８（図１１参照）を用いて、マクロ・CNNに基づく予測の処理を行い、帳票のパターンを分類候補として予測する。

【００６２】

マクロ・CNNに基づく予測の処理では、以下の手順で予測が行われる。

（１）過去に実施された臨床試験の解析図表計画書から、学習のためのインプットデータを作成する。対象の臨床試験は、CDISC（製薬業界標準のデータ作成の規準）に準拠するものである。具体的には、過去の臨床試験の解析図表計画書から各帳票の画像データを作成する。同時に、各画像に関する正解パターンを示す正解データを作成する。

（２）次に、学習のためのインプットデータから、学習データを作成する。具体的には、例えば、トリミングを行い、一つの画像データから複数の画像データを作成して、学習データとする。トリミング処理を行うことにより、学習データの絶対数を増加させることが可能である。学習データの絶対数が十分存する場合には、学習のためのインプットデータを、そのまま学習データとしてもよい。

（３）次に、（２）の学習データ、及び（１）の正解データを用いて、人工ニューラルネットワーク、例えば、CNN（畳み込みニューラルネットワーク）、を学習させる。図１２は、用いられるCNNの例の概念図である。入力層に入力されるデータのサイズ、中間層の数、各中間層に入力されるデータのサイズ等は一例である。図１２に示すCNNとは別のCNNを採用してもよい。勿論、CNN以外の人工ニューラルネットワークや学習モデルが用いられてもよい。

（４）マクロ・CNNに基づく予測の処理では、図１２に示すように、解析図表計画書の画像データ（図１１参照）について、予定された２１パターンの各々に該当する確率値（０～１で表現され、予定された２１パターンの合計確率値が１になるもの）が、アウトプットされる。更に、最大確率値であるパターンが、予測データとしてアウトプットされる。例えば、ある帳票Ａについて、パターン１である確率値が０．２、パターン２である確率値が０．７８、パターン３である確率値が０．０２、それ以外のパターンである確率値が夫々０である場合、帳票Ａのアウトプットデータは「パターン２」となる。

【００６３】

上述の手順（１）～（３）は、過去の複数の臨床試験の各解析図表計画書の画像データと前記各解析図表計画書の画像データが分類された正解パターンとを教師データとしてCNNを学習させる処理である。手順（４）は、学習が行われた結果、システム上に構築された学習済みの人工ニューラルネットワークに、分類を行いたい解析図表計画書の画像データ（図１１参照）を入力として用いて、どの帳票のパターンに該当するかを分類して分類候補を推論させる処理である。

【００６４】

第１のプログラム生成補助システムは、少なくとも推論処理を行う。すなわち処理回路８（図１）は、解析図表計画書の画像データをCNNへ入力し、CNNから出力された帳票のパターンを取得し、取得した帳票のパターンを分類候補として出力する。ただし、第１のプログラム生成補助システムが上述の学習処理を行ってもよい。ここで処理回路８は、パターンの予測結果（以下、『パターン（Vari）の予測結果』という）を分類候補として出力する。

【0065】

ミクロ・ルールベースに基づく予測の処理とマクロ・CNNに基づく予測の処理とにおける、予測結果の例を図15に示している。図15は、帳票作成に関する定義書（TFL specification）における、修正入力前の例である。この表にて、解析図表計画書における帳票の帳票番号（「Display_identifier」）に対する、ミクロ・ルールベースに基づく予測の処理の予測結果＝「Pattern_m」と、マクロ・CNNに基づく予測の処理の予測結果＝「Pattern_v」とが、分類候補として示されている。

【0066】

帳票作成に関する定義書（TFL specification）については、後で説明する。

【0067】

2.2.2. [モデルデータ及び変数を予測する第2のプログラム生成補助システムの動作]

モデルデータ及び変数を予測する第2のプログラム生成補助システム（第2サブシステム）（図2）の動作を説明する。

【0068】

第2のプログラム生成補助システムは、解析資料の各帳票が、どのモデルデータのどの変数を用いるのか、を予測する。

【0069】

再び図5を参照して、図5は、認識機能部の全体におけるデータの流れを示す図である。認識機能部は、蓄積された過去の臨床試験における、帳票作成に関する定義書（TFL specification）から、帳票及び帳票の記載と、モデルデータ（ADaMデータ）との対応テーブル60を作成する（図5・処理（3）参照）。

【0070】

ここで、本開示における、帳票作成に関する定義書（TFL specification）とは、帳票の作成で使用するモデルデータ、変数、データ抽出条件等を記載する定義書であり、帳票作成のプログラミングにおいて通常用いられるものである。図14は、帳票作成の関する定義書（TFL specification）の一部である。図14（1）はspecificationの部分であり、図14（2）はcontentsの部分である。

【0071】

また、認識機能部は、準備されたモデル仕様書であるADaMデータセット定義書（ADaM specification）から、解析資料作成用プログラム作成に必要な情報を抽出して「ADaM specificationの情報」54を生成する（図5・処理（4）参照）。準備されるモデル仕様書であるADaMデータセット定義書（ADaM specification）には、対象の臨床試験のみならず、過去の臨床試験のものも含まれる。

【0072】

この「ADaM specificationの情報」54は、例えば、ADaMデータセット定義書（ADaM specification）のデータ容量が少なければ、特段作成されなくてもよい。

【0073】

更に、認識機能部は、過去の臨床試験における帳票作成に関する定義書（TFL specification）と、過去の臨床試験における、モデル仕様書であるADaMデータセット定義書（ADaM specification）とから、モデルデータ／変数紐づけ情報（ADaM／変数紐づけ情報）56を作成する（図5・処理（5）参照）。モデルデータ／変数紐づけ情報（ADaM／変数紐づけ情報）56は、過去の臨床試験の解析資料の帳票で使用されたモデルデータ（ADaMデータ）と、当該モデルデータ（ADaMデータ）を記述する項目である変数であって当該帳票で使用された変数とを、リン

10

20

30

40

50

くさせたリンク情報である。モデルデータ／変数紐づけ情報（A D a M／変数紐づけ情報）56では、使用されたモデルデータ（A D a Mデータ）と変数、及びラベル（L A B E L）の情報がリンクされている。つまり、ラベル（L A B E L）の情報の解析のために使用されたA D a Mデータと変数とがリンクされている。例えば、ラベル（L A B E L）情報がA g eであった場合、「過去の臨床試験では、L A B E L = A g eを解析するためにA D a M = A D S Lの変数 = A G Eを使用していた」という情報が格納されている。

【0074】

図19は、モデルデータ／変数紐づけ情報（「A D a M／変数紐づけ情報」）56の例である。データセット（D a t a s e t）は、モデルデータ（A D a Mデータ）の名称を示し、変数（V a r i a b l e）は、変数名を示し、更に、ラベル（L A B E L）は、各セルに含まれる情報を示している。

10

【0075】

続いて、再び図6を参照して、図6は、学習機能部の全体におけるデータの流れを示す図である。このうち、第2のプログラム生成補助システムは、まずモデルデータ（A D a Mデータ）の予測を出力する。

【0076】

つまり、第2のプログラム生成補助システムは、対象の臨床試験の解析図表計画書における各帳票の記載と、過去の臨床試験の解析資料における各帳票の記載との、類似度を算出する。更に、第2のプログラム生成補助システムは、算出した類似度に応じて、対象の臨床試験の解析図表計画書における各帳票に対応する、過去の臨床試験の解析資料における各帳票で使用された一つ以上のモデルデータ（A D a Mデータ）の名称を、抽出してレコメンドする（図6・「レコメンドによる予測」参照）。

20

【0077】

ここで、「対象の臨床試験の解析図表計画書における各帳票の記載」は、図5、図6、及び図9Aに示す、テキストデータ52の「フォーマット情報」に含まれるデータである。「過去の臨床試験の解析資料における各帳票の記載」は、図5及び図6に示す「帳票及び帳票の記載とA D a Mとの対応テーブル」60に含まれるものである。更に、過去の臨床試験の解析資料における各帳票で使用された一つ以上のモデルデータ（A D a Mデータ）の名称も、図5及び図6に示す「帳票及び帳票の記載とA D a Mとの対応テーブル」60に含まれるものである。

30

【0078】

上述の、対象の臨床試験の解析図表計画書における各帳票の記載と、過去の臨床試験の解析資料における各帳票の記載との類似度は、所定の類似度評価関数を利用して算出されればよい。所定の類似度評価関数は、例えば、T a n i m o t o係数やJ a c c a r d係数である。また、各帳票の記載とは、例えば、各帳票のタイトルであればよい。

【0079】

図17は、A D a Mデータについての、レコメンドによる予測の、アルゴリズムの例を表すフローチャートである。まず、処理回路8（図1）は、対象の臨床試験の帳票（例えば、テーブル）のタイトルを分解する（ステップS1702）。タイトルは通常英語で記載されているため、本実施形態では、処理回路8はブランク（空白スペース）の位置でタイトルを区切り、単語単位に分解する。

40

【0080】

例えば、帳票「t a b l e X」のタイトルが「S u m m a r y o f a d v e r s e e v e n t s」であるならば、そのタイトルを「S u m m a r y」、「o f」、「a d v e r s e」、「e v e n t s」に分解する。

【0081】

次に、処理回路8は、分解結果をベクトル化する（ステップS1704）。上述の例では、処理回路8は以下の数1のようなベクトルを作成する。

【数1】

50

*Summary
of
adverse
events*

【0082】

次に、処理回路8は、分解結果と、ベクトル化された、過去の他の臨床試験の、他の帳票のタイトルとの、Tanimoto係数を求める（ステップS1706）。

【0083】

過去の他の臨床試験の、他の帳票に対するTanimoto係数について、他の帳票で用いられているADaMデータ毎に、和、及び、平均値を求める（ステップS1708）。最終的には、平均値がゼロではないADaMデータが、その平均値と共に、抽出されてレコメンドされる。なお、平均値が大きいほど、そのADaMデータのレコメンドの度合いは高い、ということになる。

【0084】

なお、帳票「table A」と帳票「table B」の、タイトルに関するTanimoto係数は、一般に以下のように算出されている。

【表3】

Table AとTable BのTanimoto係数 (T)	
n_{11}	Table AとTable Bのタイトルに共に含まれる単語数
n_{10}	Table Aのタイトルにのみ含まれる単語数
n_{01}	Table Bのタイトルにのみ含まれる単語数
$T = \frac{n_{11}}{n_{11} + n_{10} + n_{01}}$	

【0085】

続いて、図6に示すように、第2のプログラム生成補助システムは、変数の予測を出力する。変数の予測は、以下の三つのアルゴリズムに拠って行われる。三つのアルゴリズムの各々で予測された変数の全てが、予測結果の候補として出力される。

(1) アルゴリズム(1)による予測

(2) アルゴリズム(2)による予測

(3) アルゴリズム(3)による予測

【0086】

(1) アルゴリズム(1)による予測

図18は、変数予測のアルゴリズム(1)を表すフローチャートである。まず、処理回路8(図1)は、テキストデータ52の「フォーマット情報」における「帳票のセル情報の変数CELL」と、過去の臨床試験から作成したモデルデータ／変数紐づけ情報(ADaM／変数紐づけ情報)56(図5・処理(5)参照)における変数LABELが一致しているものを抽出する。そして処理回路8(図1)は、一致している二者を媒介にして二つのデータセットを合わせて結合帳票を生成する(ステップS1802)。これにより、対象の臨床試験の帳票のセル情報と、モデルデータ(ADaMデータ)と、及び、変数とが、結び付けられる。

【0087】

次に、処理回路8は、帳票(帳票番号)毎に、前ステップ(ステップS1802)で作成した結合帳票と、モデルデータ(ADaMデータ)の予測結果(図6参照)を比較する

。そして処理回路8は、比較結果からモデルデータ（ADaMデータ）の予測結果と一致するもののみを抽出し、変数予測結果を最終化する（ステップS1804）。即ち、処理回路8（図1）は、モデルデータ（ADaMデータ）の予測結果と、変数の予測結果とを比較して、変数を絞ってレコメンドしている。このとき、抽出されたモデルデータ（ADaMデータ）の名称に加えて、採用された変数が、候補として出力される。

【0088】

（2）アルゴリズム（2）による予測

図20は、変数予測のアルゴリズム（2）を表すフローチャートである。まず、処理回路8（図1）は、テキストデータ52の「フォーマット情報」における「帳票のセル情報の変数CELL」と、「ADaM specificationの情報」54に記述された変数を示す所定項目との一致度を測り、その一致度が閾値を超えている変数を抽出する（ステップS2002）。閾値は、例えば、0.7である。

10

【0089】

所定項目とは、変数を示すものであればよく、ここでは、所定項目は、変数のラベル情報である。ラベルとは、変数の名前を指すものであり、例えば、変数SITEIDの場合、ラベルは“Study Site Identifier”となる。対象の臨床試験の帳票のセル情報に“Site”などの文言が含まれていれば一致度が高く、当該帳票では変数SITEIDを使用する確率が高いと予測される。ADaMデータセット定義書（ADaM specification）の一部を示す図13において、図13（2）は変数の定義部分「variableシート」の一部であるが、4列目が変数名、5列目がラベルである。

20

【0090】

一致度は、二つの言葉の一致具合を示すものであればよく、例えば、Nグラム（N-gram）を用いて算出され得る。Nグラム（N-gram）は、一般的に、対象となる文字列をN文字毎に（例えば、N=2文字毎に）分割してインデクス化する手法である。以下では、例として、N=2との条件下で、「SAFFL」と「COMPFL」との一致度を算出する。

【0091】

まず、処理回路8は、2（=N）文字毎に「SAFFL」と「COMPFL」とを分割して、夫々の文字列群を作成する。以下の【数2】が「SAFFL」の文字列群、【数3】が「COMPFL」の文字列群である。

30

【数2】

SA
AF
FF
FL

【数3】

CO
OM
MP
PF
FL

40

【0092】

処理回路8は、「SAFFL」からの文字列群と、「COMPFL」からの文字列群とに共通する文字列の数に基づいて、一致度を算出する。例えば、具体的には共通する文字

50

列の数を全ての組み合わせ数（上記例では、 $4 \times 5 = 20$ の、組み合わせ数）で割って、一致度とする。

【0093】

次に、処理回路8は、帳票（帳票番号）毎に、前ステップ（ステップS2002）で抽出したデータセットにて、モデルデータ（ADaMデータ）の予測結果と一致するもののみを抽出し、変数予測結果を最終化する（ステップS2004）。即ち、処理回路8は、モデルデータ（ADaMデータ）の予測結果と、変数の予測結果とを比較して、変数を絞ってレコメンドしている。このとき、抽出されたモデルデータ（ADaMデータ）の名称に加えて、採用された変数が、候補として出力される。

【0094】

（3）アルゴリズム（3）による予測

図21は、変数予測のアルゴリズム（3）を表すフローチャートである。まず、処理回路8（図1）は、テキストデータ52の「フォーマット情報」における「帳票のセル情報の変数CELL」と、「ADaM specificationの情報」54に記述された変数を示す第2の所定項目との一致度を測り、その一致度が閾値を超えている変数を抽出する（ステップS2102）。閾値は、例えば、0.7である。

【0095】

第2の所定項目についても、変数を示すものであればよく、ここでは、第2の所定項目は、変数のフォーマット情報である。フォーマットとは、カテゴリ変数の各カテゴリを指すものであり、例えば、変数SEXであればフォーマットは“Male”と“Female”となる。対象の臨床試験の帳票のセル情報にMaleやFemaleと一致度が高い情報が含まれていれば、当該帳票では変数SEXを使用する確率が高いと予測される。ADaMデータセット定義書（ADaM specification）の一部である図13において、図13（3）はコードリストの定義部分「Codelistシート」の一部であるが、1列目が変数名、7列目がフォーマットである。

【0096】

ここでの一致度も、二つの言葉の一致具合を示すものであればよく、例えば、前に説明した、Nグラム（N-gram）を用いて算出され得る。

【0097】

次に、処理回路8は、帳票（帳票番号）毎に、前ステップ（ステップS2102）で抽出したデータセットにて、モデルデータ（ADaMデータ）の予測結果と一致するもののみを抽出し、変数予測結果を最終化する（ステップS2104）。即ち、モデルデータ（ADaMデータ）の予測結果と、変数の予測結果とを比較して、変数を絞ってレコメンドしている。このとき、抽出されたモデルデータ（ADaMデータ）の名称に加えて、採用された変数が、候補として出力される。

【0098】

以上、第2のプログラム生成補助システムの変数の予測について、（1）アルゴリズム（1）による予測、（2）アルゴリズム（2）による予測、及び、（3）アルゴリズム（3）による予測を、それぞれ説明した。第2のプログラム生成補助システムの変数の予測においては、必ずしも、全てのアルゴリズムによる予測が行われなくてもよい。三つの内の、少なくとも一つのアルゴリズムによる予測が行われればよい。

【0099】

2.2.3. [解析プログラム生成システムの動作]

解析プログラム生成システム（第3サブシステム）の動作を説明する。

【0100】

解析プログラム生成システムでは、第1のプログラム生成補助システムと第2のプログラム生成補助システムとにより予測され検索される情報である、パターン（Main）の予測結果、パターン（Vari）の予測結果、モデルデータであるADaMの予測結果、及び、変数の予測結果（いずれも、図6参照）を、図15に示すような、修正入力前の、帳票作成に関する定義書（TFL specification）という形で、システム

10

20

30

40

50

利用者に提示する。

【0101】

図7は、「学習」後、自動生成用メタデータ作成までのデータの流れを示す図である。

【0102】

図15は、修正入力前の、帳票作成に関する定義書(TFL specification)の例である。この表にて、解析図表計画書における帳票の帳票番号(2列目:「Display__identifier」)に対する、ミクロ・ルールベースに基づく予測の処理の予測結果=「Pattern__m」(7列目)と、マクロ・CNNに基づく予測の処理の予測結果=「Pattern__v」(8列目)と、モデルデータであるADaMデータの予測結果=「P__Dataset」(9列目)と、変数の予測結果=「P__Analysis variable」(10列目)とが示されている。

10

【0103】

これに対して、システム利用者により修正データが入力される。図16は、修正入力後の、帳票作成に関する定義書(TFL specification)の例である。なお、修正すべき箇所が無ければ、修正データは入力されず、修正入力前の帳票作成に関する定義書(TFL specification)のデータがそのまま用いられる。

【0104】

続いて、解析プログラム生成システムは、修正データが反映された帳票作成に関する定義書(TFL specification)、並びに、対象の臨床試験の帳票についてのテキストデータ(フォーマット情報とレイアウト情報)52における、解析結果の出力形式を指定する「レイアウト情報」に基づいて、自動生成用メタデータ70を生成する(図7参照)。図22は、自動生成用メタデータ70の例である。自動生成用メタデータ70は、解析資料の帳票のパターンを指定する、パターン指定データを含んでいる。

20

【0105】

続いて、解析プログラム生成システムは、記憶装置12に記憶された、解析資料の帳票の複数パターンの各々に対応するテンプレートプログラムであって、臨床試験の解析方法及び解析結果の出力形式を指定可能であるテンプレートプログラムを、利用する。即ち、解析プログラム生成システムは、自動生成用メタデータ70におけるパターン指定データによって指定されるパターンに対応するテンプレートプログラムを読み出す。更に、解析プログラム生成システムは、自動生成用メタデータ70を利用して、読み出されたテンプレートプログラムから、解析プログラムを自動生成する。図23は、テンプレートプログラムの一つの例である。

30

【0106】

ここで、対象の臨床試験の帳票についてのテキストデータ(フォーマット情報とレイアウト情報)52は、上述のように、解析結果の出力形式の指定を含んでいる。従って、解析プログラム生成システムは、当該テキストデータ52に基づいて、解析結果の出力形式を対応するテンプレートプログラムに指定し、指定された出力形式に従うレイアウトで解析結果を出力し得る解析プログラムを自動生成することとなる。

【0107】

2.3. [実施の形態のまとめ]

40

以上のように、本実施の形態に係る、臨床試験の解析プログラムの生成を補助するための第1のプログラム生成補助システムは、臨床試験の解析方法および解析結果の出力形式を指定する解析図表計画書から生成されたテキストデータおよび画像データを取得するインタフェース装置6と、テキストデータおよび画像データを記憶する記憶装置12と、テキストデータを利用する第1の候補予測方法および画像データを利用する第2の候補予測方法をそれぞれ実行することにより、解析図表計画書を、予め定められた複数のパターンのうちの少なくとも1つのパターンに分類し、結果を分類候補として出力する処理回路8とを備える。複数のパターンは、臨床試験の解析方法および解析結果の出力形式に応じて予め分類されている。第1のプログラム生成補助システムには、複数のテキスト列の各々と、各テキスト列が分類されるパターンとの対応関係を規定した判定ルールが予め用意さ

50

れており、かつ、過去の複数の臨床試験の各解析図表計画書の画像データと各解析図表計画書の画像データが分類された正解パターンとを教師データとして学習が行われた学習済みの人工ニューラルネットワークが構築されている。処理回路8は、第1の候補予測方法を実行することにより、解析図表計画書の前記テキストデータに含まれるテキスト列と判定ルールとを利用して、複数のパターンのうちの第1パターンを示すデータを出力し、第2の候補予測方法を実行することにより、解析図表計画書の画像データを人工ニューラルネットワークへ入力し、人工ニューラルネットワークから出力された第2パターンを示すデータを取得し、第1パターンおよび第2パターンを、解析資料の分類候補として出力する。

【0108】

また、本実施の形態に係る、臨床試験の解析プログラムの生成を補助するための第2のプログラム生成補助システムは、データベース16を記憶した記憶装置12と、臨床試験の解析方法および解析結果の出力形式を指定する解析図表計画書のデータを取得するインタフェース装置6と、データベース16および解析図表計画書のデータを参照して、臨床試験の解析に使用される少なくとも1つのモデルデータの名称を候補として出力する処理回路8とを備える。解析図表計画書は1以上の帳票を含み、各帳票は臨床試験の解析の種類ごとの解析の内容を指定する。データベース16は、過去の臨床試験の解析資料の各帳票と、各帳票で使用されたモデルデータの名称であって、解析のためのデータセットおよびメタデータの集合であるモデルデータの名称とを記憶している。処理回路8は、取得された解析図表計画書の帳票ごとに、帳票の記載と過去の臨床試験の解析資料の各帳票の記載との類似度を所定の類似度評価関数を利用して算出し、算出した類似度に応じて、帳票の記載に対応する過去の臨床試験の解析資料の帳票で使用された少なくとも1つのモデルデータの名称を抽出し、抽出した前記モデルデータの名称を、候補として出力する。

【0109】

更に、本実施の形態に係る、臨床試験の解析プログラムの生成を補助するための第2のプログラム生成補助システムにおいては、データベース16は、過去の臨床試験の解析資料の帳票で使用されたモデルデータと、モデルデータを記述する項目であって帳票で使用された変数とをリンクさせたリンク情報を記憶している。処理回路8は、解析図表計画書の各帳票のテキストデータの第1項目と、リンク情報の変数の第2項目とを比較し、一致する場合には、解析図表計画書の各帳票のテキストデータの第1項目とリンク情報の変数の第2項目とを媒介に、結合した結合帳票を生成する。処理回路8は、抽出したモデルデータの名称と一致するモデルデータが結合帳票内に存在する場合には、名称が一致する結合帳票内のモデルデータを記述する項目である変数を、抽出した前記モデルデータの変数として採用し、抽出した前記モデルデータの名称に加えて、採用した変数を、候補として出力する。

【0110】

更に、本実施の形態に係る、臨床試験の解析プログラムの生成を補助するための第2のプログラム生成補助システムにおいては、インタフェース装置6は、臨床試験で作成するモデルデータおよびモデルデータを記述する変数の定義が対応付けられて記述されたモデル仕様書のデータをさらに取得する。処理回路8は、所定のアルゴリズムを用いて、解析図表計画書の各帳票のテキストデータの第1項目と、モデル仕様書に記述された変数を示す第3項目との一致度を算出し、算出した一致度が予め定められた閾値よりも大きい変数に対応付けられたモデルデータをモデル仕様書から選択する。処理回路8は、選択したモデルデータが、抽出したモデルデータと一致する場合には、閾値よりも大きい変数を、抽出したモデルデータの変数として採用し、抽出した前記モデルデータの名称に加えて、採用した変数を、候補として出力する。

【0111】

また、本実施の形態に係る、臨床試験の解析プログラムを生成する解析プログラム生成システムは、インタフェース装置6と、記憶装置12と、処理回路8とを備える。インタフェース装置6は、第1のプログラム生成補助システムから出力された第1パターンおよ

10

20

30

40

50

び第2パターンを取得し、第2のプログラム生成補助システムから出力されたモデルデータの名称および変数を取得する。インタフェース装置6は、第1パターン、第2パターン、名称および／または変数を修正するための修正データを取得可能である。記憶装置12は、複数のパターンの各々に対応するテンプレートプログラムであって、臨床試験の解析方法および前記解析結果の出力形式を指定可能なテンプレートプログラムを記憶している。処理回路8は、修正データが取得された場合には、修正された第1パターン、第2パターン、名称および／または変数と、テキストデータとに基づいて、修正データが取得されていない場合には、取得された第1パターン、第2パターン、名称および／または変数と、テキストデータとに基づいて、前記複数のパターンのいずれかを指定するパターン指定データを含むメタデータを生成する。処理回路8は、パターン指定データによって指定されるパターンに対応するテンプレートプログラムを読み出し、メタデータを利用してテンプレートプログラムから解析プログラムを生成する。

10

【0112】

以上の、本実施の形態に係る、第1のプログラム生成補助システムと、第2のプログラム生成補助システムと、解析プログラム生成システムとを備える解析資料作成用プログラム自動生成システムは、医薬品開発の臨床試験における解析業務により作成される解析資料の自動生成のためのプログラミングの自動化を補助し、更に、解析資料作成用プログラムを自動生成することができる。

【0113】

3. [他の実施の形態]

20

以上のように、本出願において開示する技術の例示として、実施の形態を説明した。しかしながら、本開示における技術は、これに限定されず、適宜、変更、置き換え、付加、省略などを行った実施の形態にも適用可能である。

【0114】

上述の実施の形態に係る解析プログラム生成システムでは、第1のプログラム生成補助システムと第2のプログラム生成補助システムとにより予測され検索される情報である、パターン(Main)の予測結果、パターン(Vari)の予測結果、モデルデータであるADaMの予測結果、及び、変数の予測結果(いずれも、図6参照)を、解析プログラムの自動生成にて利用する自動生成用メタデータ70の生成に繋げている。ここで、例えば、CDISC標準を正確に遵守するADaMを使用した臨床試験に関連するデータが、学習データとして十分に蓄積され準備されるようになれば、モデルデータであるADaM、及び、変数の予測は不要になり、第1のプログラム生成補助システムの出力データとADaM、変数の検索結果とを用いて、自動生成用メタデータ70の生成に繋げることも可能になる。つまり、パターン(Main)の予測結果、パターン(Vari)の予測結果、及び、ADaM、変数の検索結果から、自動生成用メタデータ70を生成することも可能になる。

30

【0115】

また、実施の形態を説明するために、添付図面および詳細な説明を提供した。したがって、添付図面および詳細な説明に記載された構成要素の中には、課題解決のために必須な構成要素だけでなく、上記技術を例示するために、課題解決のためには必須でない構成要素も含まれ得る。そのため、それらの必須ではない構成要素が添付図面や詳細な説明に記載されていることをもって、直ちに、それらの必須ではない構成要素が必須であるとの認定をするべきではない。

40

【0116】

また、上述の実施の形態は、本開示における技術を例示するためのものであるから、特許請求の範囲またはその均等の範囲において種々の変更、置き換え、付加、省略などを行うことができる。

【符号の説明】

【0117】

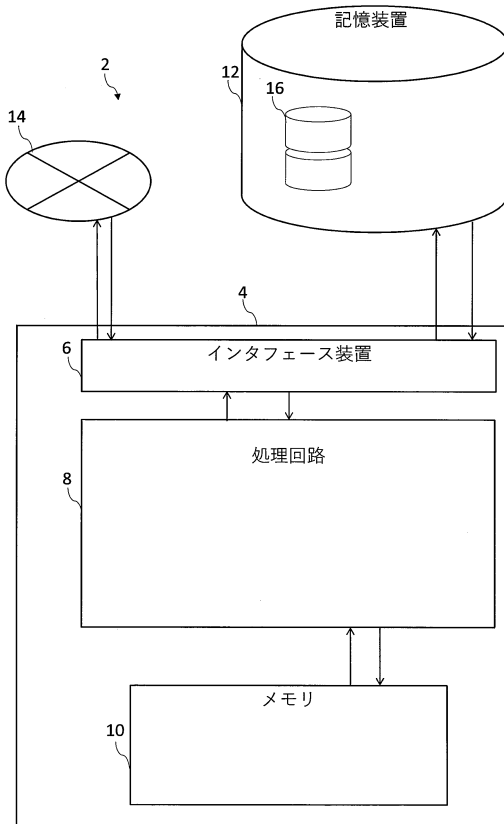
2・・・解析資料作成用プログラム自動生成システム、4・・・コンピュータ装置、6・

50

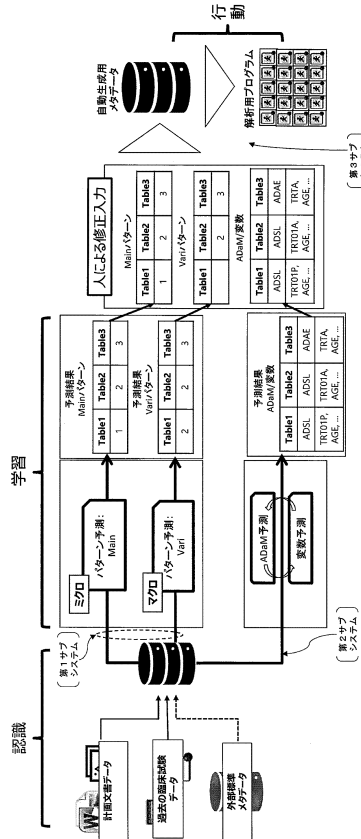
・・・インタフェース装置、８・・・処理回路、１０・・・メモリ、１２・・・記憶装置、１４・・・外部ネットワーク、１６・・・データベース、５２・・・テキストデータ（フォーマット情報とレイアウト情報）、５４・・・ＡＤａＭ specification の情報、５６・・・モデルデータ／変数紐づけ情報（ＡＤａＭ／変数紐づけ情報）、５８・・・解析図表計画書の画像データ、６０・・・帳票及び帳票の記載とＡＤａＭとの対応テーブル、７０・・・自動生成用メタデータ。

【図面】

【図１】



【図２】



10

20

30

40

50

【図 3】

ADaMデータセットの具体例：ADEG					変数の情報	
患者ID	投薬情報	解析時点 (week)	検査項目名	臨床検査値	コメント	
SUBJID	TRTDP	AVSIT	PARAM	AVAL	EGABCOM	
S123456-A7890-6AB001	New drug	Baseline	QT interval	451		
S123456-A7890-6AB001	New drug	5	QT interval	432		
S123456-A7890-6AB001	New drug	10	QT interval	438		
S123456-A7890-6AB001	New drug	12	QT interval	426		
S123456-A7890-6AB002	Placebo	Baseline	QT interval	425		
S123456-A7890-6AB003	Placebo	Baseline	QT interval	413		
S123456-A7890-6AB003	Placebo	5	QT interval	400		
S123456-A7890-6AB003	Placebo	10	QT interval	384		
S123456-A7890-6AB004	Placebo	Baseline	QT interval	400		
S123456-A7890-6AB004	Placebo	5	QT interval	403		
S123456-A7890-6AB004	Placebo	10	QT interval	510		
S123456-A7890-6AB004	Placebo	12	QT interval	512		
S123456-A7890-6AB001	Placebo	Baseline	QT interval	454		
S123456-A7890-6AB001	Placebo	5	QT interval	421		
S123456-A7890-6AB001	Placebo	10	QT interval	430		

(データ一部抽出)

【図 4】

Table 14.1.1.1 Patient Disposition All Randomized Patients			Placebo N=** n (%)	S-123456 N=** n (%)
Treatment period	Completed	Withdrawn	** (***)	** (***)
	Reason for withdrawal	Adverse event	** (***)	** (***)
		Protocol violation	** (***)	** (***)
		Lack of efficacy	** (***)	** (***)
		Withdrawal by subject	** (***)	** (***)
		Lost to follow-up	** (***)	** (***)
		Other	** (***)	** (***)
Proceeded to taper period	Completed	Adverse event	** (***)	** (***)
Taper period	Withdrawn	Protocol violation	** (***)	** (***)
	Reason for withdrawal	Lack of efficacy	** (***)	** (***)
		Withdrawal by subject	** (***)	** (***)
		Lost to follow-up	** (***)	** (***)
		Other	** (***)	** (***)

Enrolled in A3133

For the summary in treatment period, the row of "Proceeded to taper period" or the row of "Enrolled in A3133"; N is used as the denominator of percentage.

For the summary in taper period, the number of patients for the row of "Proceeded to taper period" is used as the denominator of percentage.

Programmer's Note

Reference SAP Section 6.1.1

Output FILE: SAR-A7890-TAB14.1.1.1

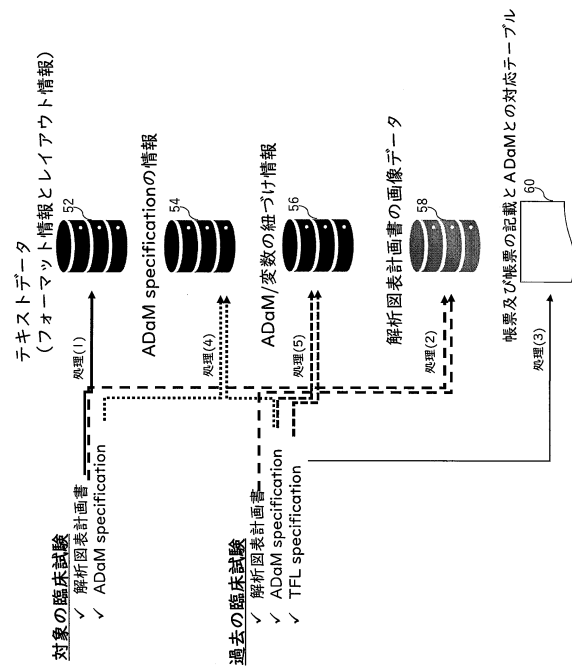
(N=*)は、全群の期待値列を算出する。治療期における全群の期待値はNとする。

Proceeded to taper period: 治療期終了後の期待値を算出する。治療期における全群の期待値はNとする。

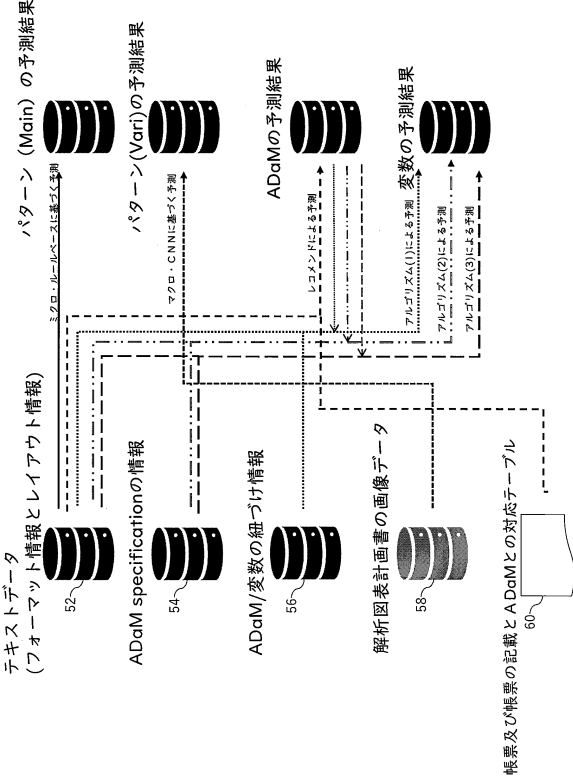
Enrolled in A3133: 中止理由の表記。表示欄は、データベースの情報を参照する。

該当列が 0 列の場合、0 を表示する。

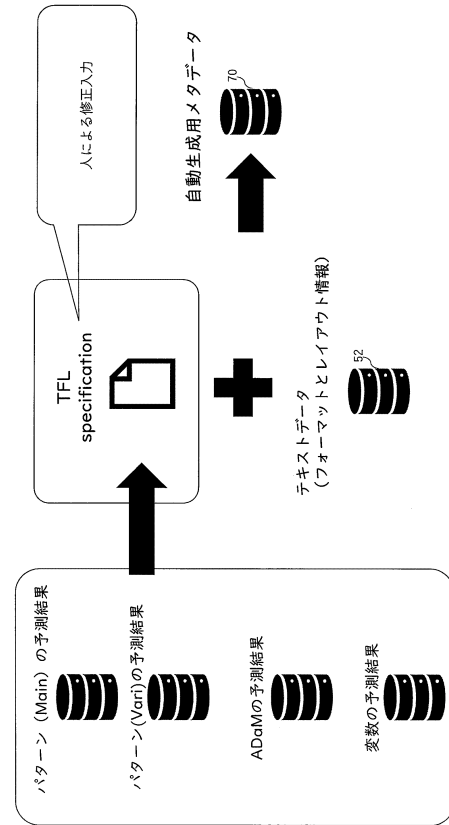
【図 5】



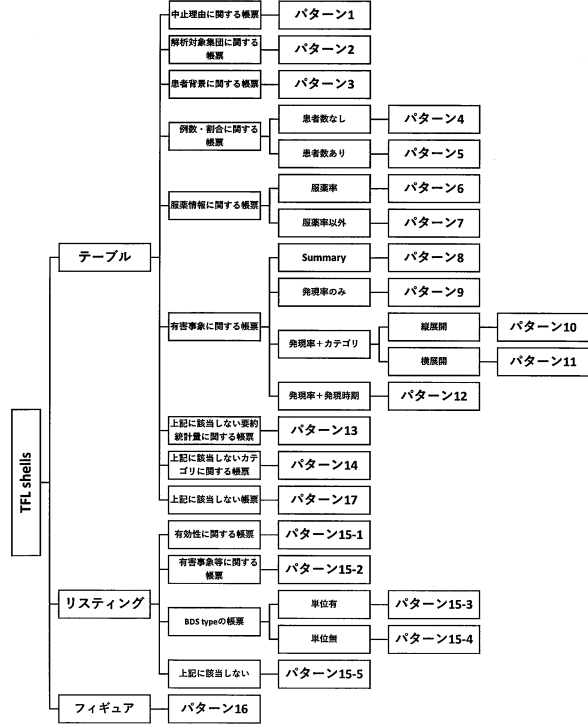
【図 6】



【圖 7】



【図 8】



10

20

【図 9 A】

PRODUCT	PROTOCOL	TYPE	TUF_CONTENT	TUF_DISPLAY_ID	COLUMN_NO	ROW_NO	CELL_DATA	IS_GROUP	IS_CATEGORY	ALL_SINGLE_ROW_LAST_IDX
S-123456	A7890	Table	Table 14.1.1.1	1	0	0	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	1	0	0	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	2	0	0	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	3	0	Placebo(0%~**0s (3)	1	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	4	0	S-123456(0%~**0s (3)	1	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	0	1	Treatment period	0	1	0	2
S-123456	A7890	Table	Table 14.1.1.1	1	1	Completed	0	1	0	2
S-123456	A7890	Table	Table 14.1.1.1	2	1	0	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	3	1	** (***)	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	4	1	** (***)	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	0	2	0	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	1	2	Withdrawn	0	1	0	2
S-123456	A7890	Table	Table 14.1.1.1	2	2	0	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	3	2	** (***)	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	4	2	** (***)	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	0	3	Reason for withdrawal	0	1	0	2
S-123456	A7890	Table	Table 14.1.1.1	1	3	Adverse event	0	1	0	2
S-123456	A7890	Table	Table 14.1.1.1	2	3	** (***)	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	4	3	** (***)	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	0	4	0	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	1	4	0	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	2	4	Protocol violation	0	1	0	2
S-123456	A7890	Table	Table 14.1.1.1	3	4	** (***)	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	4	4	** (***)	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	0	5	0	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	1	5	0	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	2	5	Lack of efficacy	0	1	0	2
S-123456	A7890	Table	Table 14.1.1.1	3	5	** (***)	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	4	5	** (***)	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	0	6	0	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	1	6	0	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	2	6	Withdrawal by subject	0	1	0	2
S-123456	A7890	Table	Table 14.1.1.1	3	6	** (***)	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	4	6	** (***)	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	0	7	0	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	1	7	0	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	2	7	Lost to follow-up	0	1	0	2
S-123456	A7890	Table	Table 14.1.1.1	3	7	** (***)	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	4	7	** (***)	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	0	8	0	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	1	8	0	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	2	8	Other	0	1	0	2
S-123456	A7890	Table	Table 14.1.1.1	3	8	** (***)	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	4	8	** (***)	0	0	0	2
S-123456	A7890	Table	Table 14.1.1.1	0	9	0	0	0	0	2

【図 9 B】

[illegible]

30

40

【図 10】

		TLF shells 由(9)→分層																	
要件	詳細	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
a-1	表層中に table を含む	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
a-2	表層中に listing を含む																		
b-1	表層中に table を含む																		
b-2	表層中に listing を含む	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
c-1	Day 1 表層																		
c-2	Day 1 表層でない																		
d-1	{severity}, {outcome}, {causality}, {onset} を含む																		
d-2	{severity}, {outcome}, {causality}, {onset} を含む																		
e-1	Overall 表層																		
e-2	Incidence 表層																		
f-1	classification_TLFs-ass の 90% 要素 ae_type が 1 である (重症度等の値に 100% のスコアを算出する)																		
f-2	classification_TLFs-ass の 90% 要素 ae_type が 2 である (重症度等の値に 100% のスコアを算出する)																		
g	Summary 表層																		
h	{categorized}, {analysis}, {electrocardiographic assessment} 表層																		
i	Exposure 表層																		
j	Analysis population 表層																		
k	Demographic, {demographic} 表層																		
l	Concomitant 表層																		
m	{concurrent medical conditions}, {prior therapies}, {prior drugs}, {previous medical conditions}, {previous disease}, {concurrent disease} 表層																		
n	Treatment compliance 表層																		
o	Exposure 表層																		
p	Adverse event, {death} 表層																		
q	表層中に listing 16.2.8 表層を含む。スコア 100% に達する																		
r	表層中に listing 16.2.8 表層を含む。スコア 100% に達する																		
s	16.2.8 表層																		

【図 11】

(1)

		Phase 1	Phase 2	Phase 3	Phase 4
		100%	100%	100%	100%
Treatment Period	Completed	100%	100%	100%	100%
	Withdrawn	100%	100%	100%	100%
	Reason for withdrawal	100%	100%	100%	100%
	Lost to follow-up	100%	100%	100%	100%
Type Period	Completed	100%	100%	100%	100%
	Withdrawn	100%	100%	100%	100%
	Reason for withdrawal	100%	100%	100%	100%
	Lost to follow-up	100%	100%	100%	100%

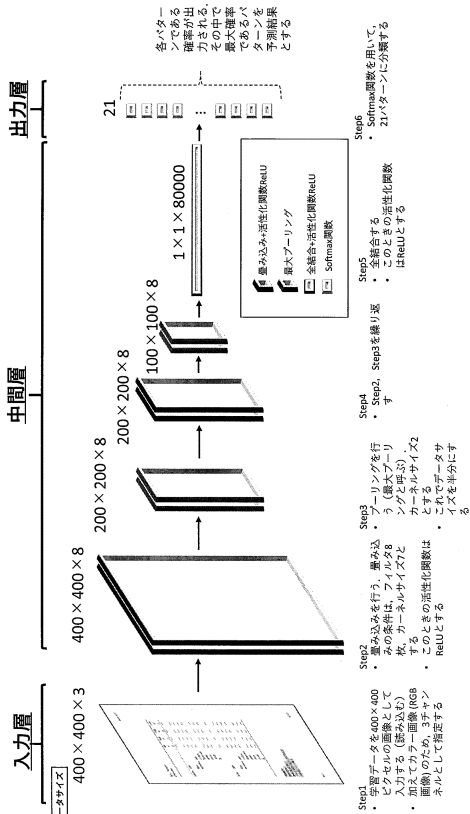
(2)

		Phase 1	Phase 2	Phase 3	Phase 4
		100%	100%	100%	100%
Period (week) [a]	Reason for Withdrawal	100%	100%	100%	100%
	Completed	100%	100%	100%	100%
	Withdrawn	100%	100%	100%	100%
	Lost to follow-up	100%	100%	100%	100%
>=1 to <2	Completed	100%	100%	100%	100%
	Withdrawn	100%	100%	100%	100%
	Reason for withdrawal	100%	100%	100%	100%
	Lost to follow-up	100%	100%	100%	100%
>=2 to <3	Completed	100%	100%	100%	100%
	Withdrawn	100%	100%	100%	100%
	Reason for withdrawal	100%	100%	100%	100%
	Lost to follow-up	100%	100%	100%	100%
>=3 to <4	Completed	100%	100%	100%	100%
	Withdrawn	100%	100%	100%	100%
	Reason for withdrawal	100%	100%	100%	100%
	Lost to follow-up	100%	100%	100%	100%
>=4 to <5	Completed	100%	100%	100%	100%
	Withdrawn	100%	100%	100%	100%
	Reason for withdrawal	100%	100%	100%	100%
	Lost to follow-up	100%	100%	100%	100%
>=5 to <6	Completed	100%	100%	100%	100%
	Withdrawn	100%	100%	100%	100%
	Reason for withdrawal	100%	100%	100%	100%
	Lost to follow-up	100%	100%	100%	100%

Version 1.0

Page 8 of 101

【図 12】



【図 13】

		Company	Terminology	Code	Order	Term	NCI Term	Recorded Value
AC001	Age Unit							Years
	Description of Planned Arm							
	Description of Planned Arm							
	Description of Planned Arm							
AC002	Race							American Indian or Alaska Native
	Race							Asian
	Race							Black or African American
	Race							Hispanic or Latino

10

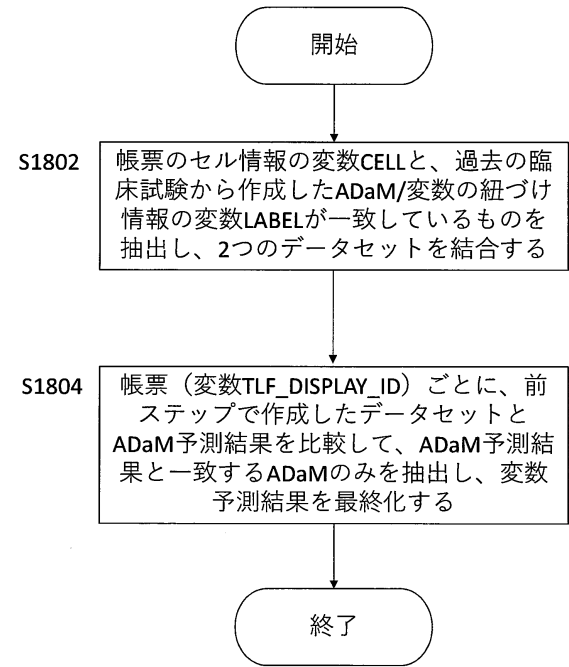
20

30

40

50

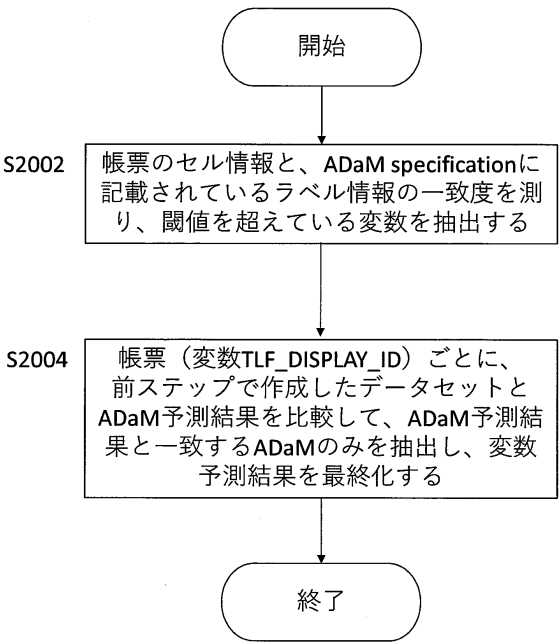
【図 18】



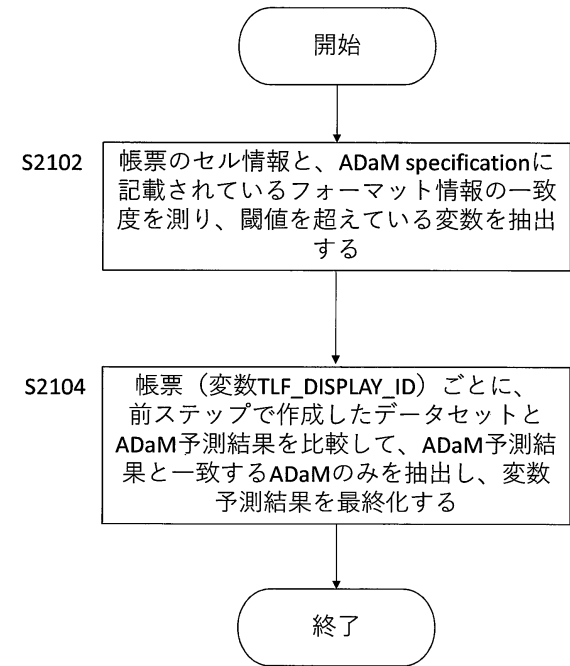
【図 19】

Dataset	Variable	LABEL
ADSL	COMPLFL	Completed
ADSL	COMPLFL	Withdrawn
ADDS	DSDECODC	Reason for withdrawn
ADSL	TCOMPLFL	Completed
ADSL	TCOMPLFL	Withdrawn
ADDS	DSDECODC	Reason for withdrawn
ADSL	TRT01PC	S-877503 N=** n (%)
ADSL	TRT01PC	Placebo N=** n (%)
ADSL	SAFEL	Subjects included in safety population
ADSL	SAFEL	Subjects excluded from safety population
ADSL	SAFECLAC	Reason for exclusion
ADSL	TRT01PC	S-877503 N=** n (%)
ADSL	TRT01PC	Placebo N=** n (%)
ADSL	SEXC	Sex
ADSL	AGE	Age (years)
ADSL	HEIGHT	Height (cm)
ADSL	WEIGHT	Weight (kg)
ADSL	BMI	BMI (kg/m2)
ADSL	RACEC	Race
ADSL	ETHICC	Ethnicity
ADSL	BLDHYN	Blood donation history
ADSL	PRTHYN	Prior therapy
ADSL	PRDRYN	Prior drug
ADSL	HAODYN	Drinking habit
ADSL	HAOSYN	Smoking habit

【図 20】



【図 21】



10

20

30

40

50

【図 2 2】

Period	PRODUCT	PROTOCOL ID	Table	Table	TITLE	T_OURS	T_FUSER	T_LENGTH
s-1254561000	s125456	s7800	Table 14.1.1	Table 14.1.1	US Market SS: Changes from Baseline to ADOSS (SS) via Adverse Events: Full Screen by Time Point	001 002 003 004 005 006 007 008		1 sec 2 sec 3 sec 4 sec 5 sec
s-1254561000	s125456	s7800	Table 14.1.1	Table 14.1.1	Withdrawn in Screening Period	001 002 003 004 005 006 007 008		1 sec 2 sec 3 sec 4 sec 5 sec
s-1254561000	s125456	s7800	Table 14.1.1	Table 14.1.1	Withdrawn in Treatment Period	001 002 003 004 005 006 007 008		1 sec 2 sec 3 sec 4 sec 5 sec
s-1254561000	s125456	s7800	Table 14.1.1	Table 14.1.1	Withdrawn in Post-Treatment Period	001 002 003 004 005 006 007 008		1 sec 2 sec 3 sec 4 sec 5 sec
s-1254561000	s125456	s7800	Table 14.1.1	Table 14.1.1	Patients Excluded from the Analysis Population	001 002 003 004 005 006 007 008		1 sec 2 sec 3 sec 4 sec 5 sec
s-1254561000	s125456	s7800	Table 14.1.1	Table 14.1.1	Demographic and Baseline Characteristics	001 002 003 004 005 006 007 008		1 sec 2 sec 3 sec 4 sec 5 sec
s-1254561000	s125456	s7800	Table 14.1.1	Table 14.1.1	Pharmacokinetics	001 002 003 004 005 006 007 008		1 sec 2 sec 3 sec 4 sec 5 sec
s-1254561000	s125456	s7800	Table 14.1.1	Table 14.1.1	Pharmacodynamics	001 002 003 004 005 006 007 008		1 sec 2 sec 3 sec 4 sec 5 sec
s-1254561000	s125456	s7800	Table 14.1.1	Table 14.1.1	Adverse Events	001 002 003 004 005 006 007 008		1 sec 2 sec 3 sec 4 sec 5 sec
s-1254561000	s125456	s7800	Table 14.1.1	Table 14.1.1	Medical History	001 002 003 004 005 006 007 008		1 sec 2 sec 3 sec 4 sec 5 sec
s-1254561000	s125456	s7800	Table 14.1.1	Table 14.1.1	Treatment Compliance	001 002 003 004 005 006 007 008		1 sec 2 sec 3 sec 4 sec 5 sec
s-1254561000	s125456	s7800	Table 14.1.1	Table 14.1.1	Physical Examination	001 002 003		1 sec 2 sec 3 sec 4 sec 5 sec
s-1254561000	s125456	s7800	Table 14.1.1	Table 14.1.1	Physical Examination in Treatment Period	001 002 003		1 sec 2 sec 3 sec 4 sec 5 sec
s-1254561000	s125456	s7800	Table 14.1.1	Table 14.1.1	Physical Examination in Post-Treatment Period	001 002 003		1 sec 2 sec 3 sec 4 sec 5 sec
s-1254561000	s125456	s7800	Table 14.1.1	Table 14.1.1	Analysis Populations	001 002 003		1 sec 2 sec 3 sec 4 sec 5 sec
s-1254561000	s125456	s7800	Table 14.1.1	Table 14.1.1	Demographic and Baseline Characteristics	001 002 003		1 sec 2 sec 3 sec 4 sec 5 sec

【図 2 3】

```
/*-----* ;  
PRODUCT ID      : @PRODUCT@  
PROTOCOL ID     : @PROTOCOLID@  
TITLE           : @Title@  
PROGRAM NAME    : @PROTOCOLID@_@program_name@_@ID_num@.sas  
-----  
AUTHOR          : @AUTHOR@  
COMPLETION DATE : &date.  
-----  
MODIFIED 01    :  
CHANGES       :  
-----  
SAS VERSION    : 9.4(DBCS)  
-----  
DATA           :  
-----  
NOTE           :  
*-----*/  
;  
%let studypath= Q:\$SD\%project%\@PRODUCT%\@PROTOCOLID@;  
%inc    &studypath.\%production%\analysis\%programs%\pc_macros\%init_def.sas ;  
%init_def(SHQ);  
%include    &trpg_v.\%fmt.sas ;  
  
options orientation = landscape;  
  
*===== ADaM 読み込み ;  
proc sort data = adam.ADDS out = _adds ;  
by USUBJID ;  
where @SC_adds@ ;  
run ;  
  
*===== Work Format ;  
proc format;  
value _wk_fmt 1 = Completed  
2 = Withdrawn;
```

10

20

30

40

50

フロントページの続き

- 大阪府大阪市中央区道修町3丁目1番8号 塩野義製薬株式会社内
(72)発明者 藤原 正和
大阪府大阪市中央区道修町3丁目1番8号 塩野義製薬株式会社内
(72)発明者 北西 由武
大阪府大阪市中央区道修町3丁目1番8号 塩野義製薬株式会社内
審査官 牧 裕子
(56)参考文献 特開2016-192085 (JP, A)
特開2020-35036 (JP, A)
特開2011-159176 (JP, A)
特開2019-114152 (JP, A)
特開2002-58650 (JP, A)
(58)調査した分野 (Int.Cl., DB名)
G06Q 10/00 - 99/00
G16H 10/00 - 80/00