(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5990862号
(P5990862)
(45)発行日　平成28年9月14日(2016.9.14)　(24)登録日　平成28年8月26日(2016.8.26)

(51)Int.Cl.		F I		
C40B 30/02	(2006.01)	C40B	30/02	
G06F 17/18	(2006.01)	G06F	17/18	Z
G06F 17/30	(2006.01)	G06F	17/30	170F
G06F 19/16	(2011.01)	G06F	19/16	

請求項の数 8　(全 27 頁)

(21)出願番号　特願2012-219730(P2012-219730)
(22)出願日　平成24年10月1日(2012.10.1)
(65)公開番号　特開2014-71836(P2014-71836A)
(43)公開日　平成26年4月21日(2014.4.21)
審査請求日　平成27年9月30日(2015.9.30)

（出願人による申告）平成２０年度、独立行政法人科学技術振興機構戦略的創造研究事業「河岡感染宿主応答ネットワーク」委託研究、産業技術力強化法第１９条の適用を受ける特許出願

(73)特許権者　503360115
国立研究開発法人科学技術振興機構
埼玉県川口市本町四丁目１番８号
(74)代理人　100089118
弁理士　酒井　宏明
(72)発明者　ダ　シルヴァ　ロペス　ティアゴ　ジョゼ
東京都港区白金台４－６－１　独立行政法人科学技術振興機構内
(72)発明者　北野　宏明
東京都港区白金台４－６－１　独立行政法人科学技術振興機構内
(72)発明者　河岡　義裕
東京都港区白金台４－６－１　独立行政法人科学技術振興機構内

最終頁に続く

(54)【発明の名称】承認予測装置、承認予測方法、および、プログラム

(57)【特許請求の範囲】

【請求項１】

出力部と記憶部と制御部とを少なくとも備えた承認予測装置であって、

上記記憶部は、

タンパク質同士の類似性に基づき構成されたタンパク質類似性ネットワークに関する類似性ネットワーク情報を記憶する類似性ネットワーク情報記憶手段と、

薬剤の承認、または、非承認に関する承認属性を含む薬剤情報と、当該薬剤の標的となる上記タンパク質に関するタンパク質情報と、を対応付けて記憶する薬剤標的記憶手段と、

上記タンパク質間の相互作用に基づき構成されたタンパク質間相互作用ネットワークに関する相互作用ネットワーク情報を記憶する相互作用ネットワーク情報記憶手段と、

を備え、

上記制御部は、

上記類似性ネットワーク情報記憶手段に記憶された上記類似性ネットワーク情報に基づき、上記タンパク質類似性ネットワークを構成する上記タンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を含む中心性尺度である類似性中心性尺度を算出する類似性中心性尺度算出手段と、

上記相互作用ネットワーク情報記憶手段に記憶された上記相互作用ネットワーク情報に基づき、上記タンパク質間相互作用ネットワークを構成する上記タンパク質の上記次数中心性、上記媒介値中心性、上記距離の中心性、および、上記バートの制約値を含む中心性

尺度である相互作用中心性尺度を算出する相互作用中心性尺度算出手段と、
上記薬剤標的記憶手段に記憶された上記各薬剤の上記承認属性と、上記類似性中心性尺度算出手段により算出された上記類似性中心性尺度の上記各薬剤の上記標的毎の合計値および平均値と、上記相互作用中心性尺度算出手段により算出された上記相互作用中心性尺度の上記各薬剤の上記標的毎の合計値および平均値と、を訓練データとする分類器を用いて、検証対象の化合物が上記非承認の上記薬剤に分類される確率である拒否スコアを算出する拒否スコア算出手段と、
上記拒否スコア算出手段により算出された上記拒否スコアを上記出力部を介して出力させる拒否スコア出力手段と、
を備えたことを特徴とする承認予測装置。
【請求項２】
出力部と記憶部と制御部とを少なくとも備えた承認予測装置であって、
上記記憶部は、
類似性を有するタンパク質同士で構成されたタンパク質類似性ネットワークに関する類似性ネットワーク情報を記憶する類似性ネットワーク情報記憶手段と、
薬剤の承認、または、非承認に関する承認属性を含む薬剤情報と、当該薬剤の標的となる上記タンパク質に関するタンパク質情報と、を対応付けて記憶する薬剤標的記憶手段と、
を備え、
上記制御部は、
上記類似性ネットワーク情報記憶手段に記憶された上記類似性ネットワーク情報に基づき、上記タンパク質類似性ネットワークを構成する上記タンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を含む中心性尺度である類似性中心性尺度を算出する類似性中心性尺度算出手段と、
上記タンパク質類似性ネットワークを構成する、上記薬剤標的記憶手段に記憶された上記タンパク質情報に基づく上記タンパク質を上記標的とする上記薬剤の上記承認属性に基づき、当該タンパク質類似性ネットワークを構成する検証対象の上記タンパク質が、上記承認された上記薬剤、または、上記非承認の上記薬剤の上記標的の範囲に含まれるか否かの判定結果を、上記類似性中心性尺度算出手段により算出された当該検証対象のタンパク質の上記類似性中心性尺度を用いて取得する承認判定手段と、
上記承認判定手段により取得された上記判定結果を上記出力部を介して出力させる判定結果出力手段と、
を備えたことを特徴とする承認予測装置。
【請求項３】
請求項１または２に記載の承認予測装置において、
上記記憶部は、
上記タンパク質のアミノ酸配列に関する配列情報を記憶するタンパク質配列情報記憶手段、
を更に備え、
上記制御部は、
上記タンパク質配列情報記憶手段に記憶された上記配列情報に基づき、シグネチャベースアルゴリズムを用いて上記タンパク質同士の上記類似性が相互に検出された場合、当該相互に上記類似性が検出された上記タンパク質同士で構成される上記タンパク質類似性ネットワークを作成し、当該タンパク質類似性ネットワークに関する上記類似性ネットワーク情報を上記類似性ネットワーク情報記憶手段に格納する類似性ネットワーク情報格納手段、
を更に備えたことを特徴とする承認予測装置。
【請求項４】
請求項２に記載の承認予測装置において、
上記承認判定手段は、

上記タンパク質類似性ネットワークを構成する、上記薬剤標的記憶手段に記憶された上記タンパク質情報に基づく上記タンパク質を上記標的とする上記薬剤の上記承認属性に基づき、上記類似性中心性尺度算出手段により算出された上記検証対象のタンパク質の上記類似性中心性尺度に含まれる上記次数中心性が高く、上記距離の中心性が低く、上記バートの制約値が非常に低い場合、上記検証対象の上記タンパク質が、上記非承認の上記薬剤の上記標的の範囲に含まれるという判定結果を生成することを特徴とする承認予測装置。

【請求項５】

出力部と記憶部と制御部とを少なくとも備えた承認予測装置において実行される承認予測方法であって、

上記記憶部は、

タンパク質同士の類似性に基づき構成されたタンパク質類似性ネットワークに関する類似性ネットワーク情報を記憶する類似性ネットワーク情報記憶手段と、

薬剤の承認、または、非承認に関する承認属性を含む薬剤情報と、当該薬剤の標的となる上記タンパク質に関するタンパク質情報と、を対応付けて記憶する薬剤標的記憶手段と、

上記タンパク質間の相互作用に基づき構成されたタンパク質間相互作用ネットワークに関する相互作用ネットワーク情報を記憶する相互作用ネットワーク情報記憶手段と、

を備え、

上記制御部において実行される、

上記類似性ネットワーク情報記憶手段に記憶された上記類似性ネットワーク情報に基づき、上記タンパク質類似性ネットワークを構成する上記タンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を含む中心性尺度である類似性中心性尺度を算出する類似性中心性尺度算出ステップと、

上記相互作用ネットワーク情報記憶手段に記憶された上記相互作用ネットワーク情報に基づき、上記タンパク質間相互作用ネットワークを構成する上記タンパク質の上記次数中心性、上記媒介値中心性、上記距離の中心性、および、上記バートの制約値を含む中心性尺度である相互作用中心性尺度を算出する相互作用中心性尺度算出ステップと、

上記薬剤標的記憶手段に記憶された上記各薬剤の上記承認属性と、上記類似性中心性尺度算出ステップにて算出された上記類似性中心性尺度の上記各薬剤の上記標的毎の合計値および平均値と、上記相互作用中心性尺度算出ステップにて算出された上記相互作用中心性尺度の上記各薬剤の上記標的毎の合計値および平均値と、を訓練データとする分類器を用いて、検証対象の化合物が上記非承認の上記薬剤に分類される確率である拒否スコアを算出する拒否スコア算出ステップと、

上記拒否スコア算出ステップにて算出された上記拒否スコアを上記出力部を介して出力させる拒否スコア出力ステップと、

を含むことを特徴とする承認予測方法。

【請求項６】

出力部と記憶部と制御部とを少なくとも備えた承認予測装置において実行される承認予測方法であって、

上記記憶部は、

類似性を有するタンパク質同士で構成されたタンパク質類似性ネットワークに関する類似性ネットワーク情報を記憶する類似性ネットワーク情報記憶手段と、

薬剤の承認、または、非承認に関する承認属性を含む薬剤情報と、当該薬剤の標的となる上記タンパク質に関するタンパク質情報と、を対応付けて記憶する薬剤標的記憶手段と、

を備え、

上記制御部において実行される、

上記類似性ネットワーク情報記憶手段に記憶された上記類似性ネットワーク情報に基づき、上記タンパク質類似性ネットワークを構成する上記タンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を含む中心性尺度である類似性中心性尺

度を算出する類似性中心性尺度算出ステップと、

上記タンパク質類似性ネットワークを構成する、上記薬剤標的記憶手段に記憶された上記タンパク質情報に基づく上記タンパク質を上記標的とする上記薬剤の上記承認属性に基づき、当該タンパク質類似性ネットワークを構成する検証対象の上記タンパク質が、上記承認された上記薬剤、または、上記非承認の上記薬剤の上記標的の範囲に含まれるか否かの判定結果を、上記類似性中心性尺度算出ステップにて算出された当該検証対象のタンパク質の上記類似性中心性尺度を用いて取得する承認判定ステップと、

上記承認判定ステップにて取得された上記判定結果を上記出力部を介して出力させる判定結果出力ステップと、

を含むことを特徴とする承認予測方法。

【請求項７】

出力部と記憶部と制御部とを少なくとも備えた承認予測装置に実行させるためのプログラムであって、

上記記憶部は、

タンパク質同士の類似性に基づき構成されたタンパク質類似性ネットワークに関する類似性ネットワーク情報を記憶する類似性ネットワーク情報記憶手段と、

薬剤の承認、または、非承認に関する承認属性を含む薬剤情報と、当該薬剤の標的となる上記タンパク質に関するタンパク質情報と、を対応付けて記憶する薬剤標的記憶手段と、

上記タンパク質間の相互作用に基づき構成されたタンパク質間相互作用ネットワークに関する相互作用ネットワーク情報を記憶する相互作用ネットワーク情報記憶手段と、

を備え、

上記制御部において、

上記類似性ネットワーク情報記憶手段に記憶された上記類似性ネットワーク情報に基づき、上記タンパク質類似性ネットワークを構成する上記タンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を含む中心性尺度である類似性中心性尺度を算出する類似性中心性尺度算出ステップと、

上記相互作用ネットワーク情報記憶手段に記憶された上記相互作用ネットワーク情報に基づき、上記タンパク質間相互作用ネットワークを構成する上記タンパク質の上記次数中心性、上記媒介値中心性、上記距離の中心性、および、上記バートの制約値を含む中心性尺度である相互作用中心性尺度を算出する相互作用中心性尺度算出ステップと、

上記薬剤標的記憶手段に記憶された上記各薬剤の上記承認属性と、上記類似性中心性尺度算出ステップにて算出された上記類似性中心性尺度の上記各薬剤の上記標的毎の合計値および平均値と、上記相互作用中心性尺度算出ステップにて算出された上記相互作用中心性尺度の上記各薬剤の上記標的毎の合計値および平均値と、を訓練データとする分類器を用いて、検証対象の化合物が上記非承認の上記薬剤に分類される確率である拒否スコアを算出する拒否スコア算出ステップと、

上記拒否スコア算出ステップにて算出された上記拒否スコアを上記出力部を介して出力させる拒否スコア出力ステップと、

を実行させるためのプログラム。

【請求項８】

出力部と記憶部と制御部とを少なくとも備えた承認予測装置に実行させるためのプログラムであって、

上記記憶部は、

類似性を有するタンパク質同士で構成されたタンパク質類似性ネットワークに関する類似性ネットワーク情報を記憶する類似性ネットワーク情報記憶手段と、

薬剤の承認、または、非承認に関する承認属性を含む薬剤情報と、当該薬剤の標的となる上記タンパク質に関するタンパク質情報と、を対応付けて記憶する薬剤標的記憶手段と、

を備え、

上記制御部において、

上記類似性ネットワーク情報記憶手段に記憶された上記類似性ネットワーク情報に基づき、上記タンパク質類似性ネットワークを構成する上記タンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を含む中心性尺度である類似性中心性尺度を算出する類似性中心性尺度算出ステップと、

上記タンパク質類似性ネットワークを構成する、上記薬剤標的記憶手段に記憶された上記タンパク質情報に基づく上記タンパク質を上記標的とする上記薬剤の上記承認属性に基づき、当該タンパク質類似性ネットワークを構成する検証対象の上記タンパク質が、上記承認された上記薬剤、または、上記非承認の上記薬剤の上記標的の範囲に含まれるか否かの判定結果を、上記類似性中心性尺度算出ステップにて算出された当該検証対象のタンパク質の上記類似性中心性尺度を用いて取得する承認判定ステップと、

上記承認判定ステップにて取得された上記判定結果を上記出力部を介して出力させる判定結果出力ステップと、

を実行させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、承認予測装置、承認予測方法、および、プログラムに関する。

【背景技術】

【０００２】

従来から、既存の化合物のオフターゲットまたは副作用の予測を行う技術が開示されている。

【０００３】

非特許文献１に記載のタンパク質機能識別においては、リガンドの類似性に基づいてタンパク質のグループ化による薬物のオフターゲットを検出する技術が開示されており、以前に文献で報告されていない受容体と拮抗する点において、メタドン、エメチンおよびロペラミドのような薬剤の間に予期しない関係を発見している。

【０００４】

また、非特許文献２に記載の薬剤標的識別においては、出発点として市販医薬品によって引き起こされる副作用を用いて、オフターゲット効果を検討し、副作用に応じて薬をグループ化することにより適応や構造を持つ薬剤をグループ化し、以前に知られていなかった薬剤の追加標的タンパク質を決定する技術が開示されている。

【０００５】

また、非特許文献３に記載の既知薬剤の新規分子ターゲット予測においては、リガンドの類似性に基づいてタンパク質をグループ化し、オフターゲット効果を調べることで、既知の薬剤のうち、報告されている標的以外の標的を発見する技術が開示されている。

【０００６】

また、非特許文献４に記載の薬剤標的相互作用ネットワーク予測においては、タンパク質配列および薬剤標的に関する情報を結びつけて、薬理学的空間（ｐｈａｒｍａｃｏｌｏｇｉｃａｌ　ｓｐａｃｅ）と呼ぶリソースを新たに作成し、このリソースを使用して、既知の薬物の追加標的を明らかにし、薬剤標的を酵素、イオンチャネル、Ｇタンパク質結合受容体、および、核内受容体の４つのクラスに分類する技術が開示されている。

【０００７】

また、非特許文献５に記載の薬剤活性大規模予測においては、市販薬の副作用の予測および説明に使用される薬剤ターゲット副作用ネットワークを作成し、医薬品と特定のタンパク質との意図しない相互作用の様々な関連性から、以前には説明することができなかった副作用を見いだす技術が開示されている。

【０００８】

また、非特許文献６に記載の薬剤誘導肝臓損傷予測システムは、肝臓の損傷を引き起こす可能性が高い化合物を同定するための予測システムであり、予測対象を肝臓に限定し、

ある種の化合物が肝臓への損傷を起こしやすくしているという特性を科学文献の調査に基づいて予測する技術が開示されている。ここで、薬剤誘導肝臓損傷予測システムにおいては、肝臓への有害な影響を引き起こす可能性があるいくつかのタンパク質および経路を予測している。

【先行技術文献】

【非特許文献】

【０００９】

【非特許文献１】Ｋｅｉｓｅｒ ＭＪ, Ｒｏｔｈ ＢＬ, Ａｒｍｂｒｕｓｔｅｒ ＢＮ, Ｅｒｎｓｂｅｒｇｅｒ Ｐ, Ｉｒｗｉｎ ＪＪ, Ｓｈｏｉｃｈｅｔ ＢＫ. （２００７） Ｒｅｌａｔｉｎｇ ｐｒｏｔｅｉｎ ｐｈａｒｍａｃｏｌｏｇｙ ｂｙ ｌｉｇａｎｄ ｃｈｅｍｉｓｔｒｙ, Ｎａｔｕｒｅ Ｂｉｏｔｅｃｈｎｏｌｏｇｙ, ２５, １９７－２０６.

【非特許文献２】Ｃａｍｐｉｌｌｏｓ Ｍ, Ｋｕｈｎ Ｍ, Ｇａｖｉｎ ＡＣ, Ｊｅｎｓｅｎ ＬＪ, Ｂｏｒｋ Ｐ. （２００８） Ｄｒｕｇ Ｔａｒｇｅｔ Ｉｄｅｎｔｉｆｉｃａｔｉｏｎ Ｕｓｉｎｇ Ｓｉｄｅ－Ｅｆｆｅｃｔ Ｓｉｍｉｌａｒｉｔｙ, Ｓｃｉｅｎｃｅ, ３２１, ２６３－２６６.

【非特許文献３】Ｋｅｉｓｅｒ ＭＪ, Ｓｅｔｏｌａ Ｖ, Ｉｒｗｉｎ ＪＪ, Ｌａｇｇｎｅｒ Ｃ, Ａｂｂａｓ ＡＩ, Ｈｕｆｅｉｓｅｎ ＳＪ, Ｊｅｎｓｅｎ ＮＨ, Ｋｕｉｊｅｒ ＭＢ, Ｍａｔｏｓ ＲＣ, Ｔｒａｎ ＴＢ, Ｗｈａｌｅｙ Ｒ, Ｇｌｅｎｎｏｎ ＲＡ, Ｈｅｒｔ Ｊ, Ｔｈｏｍａｓ ＫＬ, Ｅｄｗａｒｄｓ ＤＤ, Ｓｈｏｉｃｈｅｔ ＢＫ, Ｒｏｔｈ ＢＬ. （２００９） Ｐｒｅｄｉｃｔｉｎｇ ｎｅｗ ｍｏｌｅｃｕｌａｒ ｔａｒｇｅｔｓ ｆｏｒ ｋｎｏｗｎ ｄｒｕｇｓ, Ｎａｔｕｒｅ, ４６２, １７５－１８１.

【非特許文献４】Ｙａｍａｎｉｓｈｉ Ｙ, Ａｒａｋｉ Ｍ, Ｇｕｔｔｅｒｉｄｇｅ Ａ, Ｈｏｎｄａ Ｗ, Ｋａｎｅｈｉｓａ Ｍ. （２００８） Ｐｒｅｄｉｃｔｉｏｎ ｏｆ ｄｒｕｇ－ｔａｒｇｅｔ ｉｎｔｅｒａｃｔｉｏｎ ｎｅｔｗｏｒｋｓ ｆｒｏｍ ｔｈｅ ｉｎｔｅｇｒａｔｉｏｎ ｏｆ ｃｈｅｍｉｃａｌ ａｎｄ ｇｅｎｏｍｉｃ ｓｐａｃｅｓ, Ｂｉｏｉｎｆｏｒｍａｔｉｃｓ, ２４, ｉ２３２－ｉ２４０.

【非特許文献５】Ｌｏｕｎｋｉｎｅ Ｅ, Ｋｅｉｓｅｒ ＭＪ, Ｗｈｉｔｅｂｒｅａｄ Ｓ, Ｍｉｋｈａｉｌｏｖ Ｄ, Ｈａｍｏｎ Ｊ, Ｊｅｎｋｉｎｓ ＪＬ, Ｌａｖａｎ Ｐ, Ｗｅｂｅｒ Ｅ, Ｄｏａｋ ＡＫ, Ｃｏｔｅ Ｓ, Ｓｈｏｉｃｈｅｔ ＢＫ, Ｕｒｂａｎ Ｌ. （２０１２） Ｌａｒｇｅ－ｓｃａｌｅ ｐｒｅｄｉｃｔｉｏｎ ａｎｄ ｔｅｓｔｉｎｇ ｏｆ ｄｒｕｇ ａｃｔｉｖｉｔｙ ｏｎ ｓｉｄｅ－ｅｆｆｅｃｔ ｔａｒｇｅｔｓ, Ｎａｔｕｒｅ, ４８６, ３６１－３６７.

【非特許文献６】Ｌｉｕ Ｚ, Ｓｈｉ Ｑ, Ｄｉｎｇ Ｄ, Ｋｅｌｌｙ Ｒ, Ｆａｎｇ Ｈ, ｅｔ ａｌ. （２０１１） Ｔｒａｎｓｌａｔｉｎｇ Ｃｌｉｎｉｃａｌ Ｆｉｎｄｉｎｇｓ ｉｎｔｏ Ｋｎｏｗｌｅｄｇｅ ｉｎ Ｄｒｕｇ Ｓａｆｅｔｙ Ｅｖａｌｕａｔｉｏｎ － Ｄｒｕｇ Ｉｎｄｕｃｅｄ Ｌｉｖｅｒ Ｉｎｊｕｒｙ Ｐｒｅｄｉｃｔｉｏｎ Ｓｙｓｔｅｍ （ＤＩＬＩｐｓ）. ＰＬｏＳ Ｃｏｍｐｕｔ Ｂｉｏｌ ７（１２）： ｅ１００２３１０.

【発明の概要】

【発明が解決しようとする課題】

【００１０】

しかしながら、非特許文献１乃至６に記載の従来の薬剤標的予測技術においては、標的タンパク質の特性に基づいて医薬品の承認の可能性を定量化することができないという問題点を有していた。

【００１１】

本発明は、上記問題点に鑑みてなされたもので、医薬品の承認または拒否の可能性を定量化することができる承認予測装置、承認予測方法、および、プログラムを提供すること

を目的とする。

【課題を解決するための手段】

【００１２】

このような目的を達成するため、本発明の承認予測装置は、出力部と記憶部と制御部とを少なくとも備えた承認予測装置であって、上記記憶部は、タンパク質同士の類似性に基づき構成されたタンパク質類似性ネットワークに関する類似性ネットワーク情報を記憶する類似性ネットワーク情報記憶手段と、薬剤の承認、または、非承認に関する承認属性を含む薬剤情報と、当該薬剤の標的となる上記タンパク質に関するタンパク質情報と、を対応付けて記憶する薬剤標的記憶手段と、上記タンパク質間の相互作用に基づき構成されたタンパク質間相互作用ネットワークに関する相互作用ネットワーク情報を記憶する相互作用ネットワーク情報記憶手段と、を備え、上記制御部は、上記類似性ネットワーク情報記憶手段に記憶された上記類似性ネットワーク情報に基づき、上記タンパク質類似性ネットワークを構成する上記タンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を含む中心性尺度である類似性中心性尺度を算出する類似性中心性尺度算出手段と、上記相互作用ネットワーク情報記憶手段に記憶された上記相互作用ネットワーク情報に基づき、上記タンパク質間相互作用ネットワークを構成する上記タンパク質の上記次数中心性、上記媒介値中心性、上記距離の中心性、および、上記バートの制約値を含む中心性尺度である相互作用中心性尺度を算出する相互作用中心性尺度算出手段と、上記薬剤標的記憶手段に記憶された上記各薬剤の上記承認属性と、上記類似性中心性尺度算出手段により算出された上記類似性中心性尺度の上記各薬剤の上記標的毎の合計値および平均値と、上記相互作用中心性尺度算出手段により算出された上記相互作用中心性尺度の上記各薬剤の上記標的毎の合計値および平均値と、を訓練データとする分類器を用いて、検証対象の化合物が上記非承認の上記薬剤に分類される確率である拒否スコアを算出する拒否スコア算出手段と、上記拒否スコア算出手段により算出された上記拒否スコアを上記出力部を介して出力させる拒否スコア出力手段と、を備えたことを特徴とする。

【００１３】

また、本発明の承認予測装置は、出力部と記憶部と制御部とを少なくとも備えた承認予測装置であって、上記記憶部は、類似性を有するタンパク質同士で構成されたタンパク質類似性ネットワークに関する類似性ネットワーク情報を記憶する類似性ネットワーク情報記憶手段と、薬剤の承認、または、非承認に関する承認属性を含む薬剤情報と、当該薬剤の標的となる上記タンパク質に関するタンパク質情報と、を対応付けて記憶する薬剤標的記憶手段と、を備え、上記制御部は、上記類似性ネットワーク情報記憶手段に記憶された上記類似性ネットワーク情報に基づき、上記タンパク質類似性ネットワークを構成する上記タンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を含む中心性尺度である類似性中心性尺度を算出する類似性中心性尺度算出手段と、上記タンパク質類似性ネットワークを構成する、上記薬剤標的記憶手段に記憶された上記タンパク質情報に基づく上記タンパク質を上記標的とする上記薬剤の上記承認属性に基づき、当該タンパク質類似性ネットワークを構成する検証対象の上記タンパク質が、上記承認された上記薬剤、または、上記非承認の上記薬剤の上記標的の範囲に含まれるか否かの判定結果を、上記類似性中心性尺度算出手段により算出された当該検証対象のタンパク質の上記類似性中心性尺度を用いて取得する承認判定手段と、上記承認判定手段により取得された上記判定結果を上記出力部を介して出力させる判定結果出力手段と、を備えたことを特徴とする。

【００１４】

また、本発明の承認予測装置は、上記記載の承認予測装置において、上記記憶部は、上記タンパク質のアミノ酸配列に関する配列情報を記憶するタンパク質配列情報記憶手段、を更に備え、上記制御部は、上記タンパク質配列情報記憶手段に記憶された上記配列情報に基づき、シグネチャベースアルゴリズムを用いて上記タンパク質同士の上記類似性が相互に検出された場合、当該相互に上記類似性が検出された上記タンパク質同士で構成される上記タンパク質類似性ネットワークを作成し、当該タンパク質類似性ネットワークに関

する上記類似性ネットワーク情報を上記類似性ネットワーク情報記憶手段に格納する類似性ネットワーク情報格納手段、を更に備えたことを特徴とする。

【００１５】

また、本発明の承認予測装置は、上記記載の承認予測装置において、上記承認判定手段は、上記タンパク質類似性ネットワークを構成する、上記薬剤標的記憶手段に記憶された上記タンパク質情報に基づく上記タンパク質を上記標的とする上記薬剤の上記承認属性に基づき、上記類似性中心性尺度算出手段により算出された上記検証対象のタンパク質の上記類似性中心性尺度に含まれる上記次数中心性が高く、上記距離の中心性が低く、上記バートの制約値が非常に低い場合、上記検証対象の上記タンパク質が、上記非承認の上記薬剤の上記標的の範囲に含まれるという判定結果を生成することを特徴とする。

【００１６】

また、本発明の承認予測方法は、出力部と記憶部と制御部とを少なくとも備えた承認予測装置において実行される承認予測方法であって、上記記憶部は、タンパク質同士の類似性に基づき構成されたタンパク質類似性ネットワークに関する類似性ネットワーク情報を記憶する類似性ネットワーク情報記憶手段と、薬剤の承認、または、非承認に関する承認属性を含む薬剤情報と、当該薬剤の標的となる上記タンパク質に関するタンパク質情報と、を対応付けて記憶する薬剤標的記憶手段と、上記タンパク質間の相互作用に基づき構成されたタンパク質間相互作用ネットワークに関する相互作用ネットワーク情報を記憶する相互作用ネットワーク情報記憶手段と、を備え、上記制御部において実行される、上記類似性ネットワーク情報記憶手段に記憶された上記類似性ネットワーク情報に基づき、上記タンパク質類似性ネットワークを構成する上記タンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を含む中心性尺度である類似性中心性尺度を算出する類似性中心性尺度算出ステップと、上記相互作用ネットワーク情報記憶手段に記憶された上記相互作用ネットワーク情報に基づき、上記タンパク質間相互作用ネットワークを構成する上記タンパク質の上記次数中心性、上記媒介値中心性、上記距離の中心性、および、上記バートの制約値を含む中心性尺度である相互作用中心性尺度を算出する相互作用中心性尺度算出ステップと、上記薬剤標的記憶手段に記憶された上記各薬剤の上記承認属性と、上記類似性中心性尺度算出ステップにて算出された上記類似性中心性尺度の上記各薬剤の上記標的毎の合計値および平均値と、上記相互作用中心性尺度算出ステップにて算出された上記相互作用中心性尺度の上記各薬剤の上記標的毎の合計値および平均値と、を訓練データとする分類器を用いて、検証対象の化合物が上記非承認の上記薬剤に分類される確率である拒否スコアを算出する拒否スコア算出ステップと、上記拒否スコア算出ステップにて算出された上記拒否スコアを上記出力部を介して出力させる拒否スコア出力ステップと、を含むことを特徴とする。

【００１７】

また、本発明の承認予測方法は、出力部と記憶部と制御部とを少なくとも備えた承認予測装置において実行される承認予測方法であって、上記記憶部は、類似性を有するタンパク質同士で構成されたタンパク質類似性ネットワークに関する類似性ネットワーク情報を記憶する類似性ネットワーク情報記憶手段と、薬剤の承認、または、非承認に関する承認属性を含む薬剤情報と、当該薬剤の標的となる上記タンパク質に関するタンパク質情報と、を対応付けて記憶する薬剤標的記憶手段と、を備え、上記制御部において実行される、上記類似性ネットワーク情報記憶手段に記憶された上記類似性ネットワーク情報に基づき、上記タンパク質類似性ネットワークを構成する上記タンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を含む中心性尺度である類似性中心性尺度を算出する類似性中心性尺度算出ステップと、上記タンパク質類似性ネットワークを構成する、上記薬剤標的記憶手段に記憶された上記タンパク質情報に基づく上記タンパク質を上記標的とする上記薬剤の上記承認属性に基づき、当該タンパク質類似性ネットワークを構成する検証対象の上記タンパク質が、上記承認された上記薬剤、または、上記非承認の上記薬剤の上記標的の範囲に含まれるか否かの判定結果を、上記類似性中心性尺度算出ステップにて算出された当該検証対象のタンパク質の上記類似性中心性尺度を用いて取得す

る承認判定ステップと、上記承認判定ステップにて取得された上記判定結果を上記出力部を介して出力させる判定結果出力ステップと、を含むことを特徴とする。

【００１８】

また、本発明のプログラムは、出力部と記憶部と制御部とを少なくとも備えた承認予測装置に実行させるためのプログラムであって、上記記憶部は、タンパク質同士の類似性に基づき構成されたタンパク質類似性ネットワークに関する類似性ネットワーク情報を記憶する類似性ネットワーク情報記憶手段と、薬剤の承認、または、非承認に関する承認属性を含む薬剤情報と、当該薬剤の標的となる上記タンパク質に関するタンパク質情報と、を対応付けて記憶する薬剤標的記憶手段と、上記タンパク質間の相互作用に基づき構成されたタンパク質間相互作用ネットワークに関する相互作用ネットワーク情報を記憶する相互作用ネットワーク情報記憶手段と、を備え、上記制御部において、上記類似性ネットワーク情報記憶手段に記憶された上記類似性ネットワーク情報に基づき、上記タンパク質類似性ネットワークを構成する上記タンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を含む中心性尺度である類似性中心性尺度を算出する類似性中心性尺度算出ステップと、上記相互作用ネットワーク情報記憶手段に記憶された上記相互作用ネットワーク情報に基づき、上記タンパク質間相互作用ネットワークを構成する上記タンパク質の上記次数中心性、上記媒介値中心性、上記距離の中心性、および、上記バートの制約値を含む中心性尺度である相互作用中心性尺度を算出する相互作用中心性尺度算出ステップと、上記薬剤標的記憶手段に記憶された上記各薬剤の上記承認属性と、上記類似性中心性尺度算出ステップにて算出された上記類似性中心性尺度の上記各薬剤の上記標的毎の合計値および平均値と、上記相互作用中心性尺度算出ステップにて算出された上記相互作用中心性尺度の上記各薬剤の上記標的毎の合計値および平均値と、を訓練データとする分類器を用いて、検証対象の化合物が上記非承認の上記薬剤に分類される確率である拒否スコアを算出する拒否スコア算出ステップと、上記拒否スコア算出ステップにて算出された上記拒否スコアを上記出力部を介して出力させる拒否スコア出力ステップと、を実行させることを特徴とする。

【００１９】

また、本発明のプログラムは、出力部と記憶部と制御部とを少なくとも備えた承認予測装置に実行させるためのプログラムであって、上記記憶部は、類似性を有するタンパク質同士で構成されたタンパク質類似性ネットワークに関する類似性ネットワーク情報を記憶する類似性ネットワーク情報記憶手段と、薬剤の承認、または、非承認に関する承認属性を含む薬剤情報と、当該薬剤の標的となる上記タンパク質に関するタンパク質情報と、を対応付けて記憶する薬剤標的記憶手段と、を備え、上記制御部において、上記類似性ネットワーク情報記憶手段に記憶された上記類似性ネットワーク情報に基づき、上記タンパク質類似性ネットワークを構成する上記タンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を含む中心性尺度である類似性中心性尺度を算出する類似性中心性尺度算出ステップと、上記タンパク質類似性ネットワークを構成する、上記薬剤標的記憶手段に記憶された上記タンパク質情報に基づく上記タンパク質を上記標的とする上記薬剤の上記承認属性に基づき、当該タンパク質類似性ネットワークを構成する検証対象の上記タンパク質が、上記承認された上記薬剤、または、上記非承認の上記薬剤の上記標的の範囲に含まれるか否かの判定結果を、上記類似性中心性尺度算出ステップにて算出された当該検証対象のタンパク質の上記類似性中心性尺度を用いて取得する承認判定ステップと、上記承認判定ステップにて取得された上記判定結果を上記出力部を介して出力させる判定結果出力ステップと、を実行させることを特徴とする。

【発明の効果】

【００２０】

また、この発明によれば、タンパク質類似性ネットワークを構成するタンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を含む中心性尺度である類似性中心性尺度を算出し、タンパク質間相互作用ネットワークを構成するタンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を含む中心性尺度で

ある相互作用中心性尺度を算出し、各薬剤の承認属性と、算出された類似性中心性尺度の各薬剤の標的毎の合計値および平均値と、算出された相互作用中心性尺度の各薬剤の標的毎の合計値および平均値と、を訓練データとする分類器を用いて、検証対象の化合物が非承認の薬剤に分類される確率である拒否スコアを算出し、算出された拒否スコアを出力部を介して出力させるので、１つの化合物の標的として全タンパク質の特性を考慮することで、複数標的化合物の承認・非承認の予測に利用することができるという効果を奏する。また、この発明によれば、機械学習の分類を用いた、候補化合物が望ましくない副作用を引き起こす可能性を示すスコアリングを行うことで、医薬品開発の初期段階で使用され、承認される高い可能性を有する化合物の優先順位を決定する手助けになるという効果を奏する。

【００２１】

この発明によれば、タンパク質類似性ネットワークを構成するタンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を含む中心性尺度である類似性中心性尺度を算出し、タンパク質類似性ネットワークを構成する、タンパク質を標的とする薬剤の承認属性に基づき、当該タンパク質類似性ネットワークを構成する検証対象のタンパク質が、承認された薬剤、または、非承認の薬剤の標的の範囲に含まれるか否かの判定結果を、算出された当該検証対象のタンパク質の類似性中心性尺度を用いて取得し、取得された判定結果を出力部を介して出力させるので、個々のタンパク質の特性を特定し、有害な影響を与える可能性がないかを判断することができるという効果を奏する。また、この発明によれば、ｓｉＲＮＡ治療、または、シングル標的化合物（ａｋａ　ｍａｇｉｃ　ｂｕｌｌｅｔｓ）等の単独標的の評価、または、単一の特定のタンパク質の活性を調節する技術に利用することができるという効果を奏する。

【００２２】

また、この発明によれば、シグネチャベースアルゴリズムを用いてタンパク質同士の類似性が相互に検出された場合、当該相互に類似性が検出されたタンパク質同士で構成されるタンパク質類似性ネットワークを作成し、当該タンパク質類似性ネットワークに関する類似性ネットワーク情報を格納するので、従来公表されているネットワークデータよりも高い類似性を有するネットワークデータを提供することができるという効果を奏する。

【００２３】

また、この発明によれば、タンパク質類似性ネットワークを構成する、タンパク質を標的とする薬剤の承認属性に基づき、算出された検証対象のタンパク質の類似性中心性尺度に含まれる次数中心性が高く、距離の中心性が低く、バートの制約値が非常に低い場合、検証対象のタンパク質が、非承認の薬剤の標的の範囲に含まれるという判定結果を生成するので、非特異的な結合をしやすく、副作用を起こしやすいタンパク質を的確に特定することができるという効果を奏する。

【図面の簡単な説明】

【００２４】

【図１】図１は、本実施の形態の基本原理を示すフローチャートである。

【図２】図２は、本実施の形態の基本原理を示すフローチャートである。

【図３】図３は、本実施の形態における承認予測装置の構成の一例を示すブロック図である。

【図４】図４は、本実施の形態の承認予測装置の処理の一例を示すフローチャートである。

【図５】図５は、本実施の形態における配列情報の一例を示す図である。

【図６】図６は、本実施の形態における類似性ネットワーク情報の一例を示す図である。

【図７】図７は、本実施の形態におけるバートの制約の一例を示す図である。

【図８】図８は、本実施の形態におけるタンパク質の中心性尺度の一例を示す図である。

【図９】図９は、本実施の形態における薬剤標的データベースに記憶される情報の一例を示す図である。

【図１０】図１０は、本実施の形態における承認または非承認の薬剤の標的の中心性尺度

の一例を示す図である。
【図１１】図１１は、本実施の形態における相互作用ネットワーク情報の一例を示す図である。
【図１２】図１２は、本実施の形態における分類器の性能改善の一例を示す図である。
【図１３】図１３は、本実施の形態における分類器の分類精度の一例を示す図である。
【図１４】図１４は、本実施の形態における分類器の一例を示す図である。
【図１５】図１５は、本実施形態における出力情報の一例を示す図である。
【発明を実施するための形態】
【００２５】
以下に、本発明にかかる承認予測装置、承認予測方法、および、プログラムの実施の形態を図面に基づいて詳細に説明する。なお、この実施の形態によりこの発明が限定されるものではない。
【００２６】
［本発明の実施の形態の概要］
以下、本発明の実施の形態の概要について図１および図２を参照して説明し、その後、本実施の形態の構成および処理等について詳細に説明する。
【００２７】
［概要（その１）］
まず、図１を参照して、本発明の実施の形態の概要の一例について説明する。図１は、本実施の形態の基本原理を示すフローチャートである。本実施の形態は、概略的に、以下の基本的特徴を有する。
【００２８】
すなわち、本実施の形態の承認予測装置の制御部は、図１に示すように、タンパク質類似性ネットワークを構成するタンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を含む中心性尺度である類似性中心性尺度を算出する（ステップＳＡ－１）。
【００２９】
そして、承認予測装置の制御部は、タンパク質類似性ネットワークを構成する、タンパク質を標的とする薬剤の承認属性に基づき、当該タンパク質類似性ネットワークを構成する検証対象のタンパク質が、承認された薬剤、または、非承認の薬剤の標的の範囲に含まれるか否かの判定結果を、ステップＳＡ－１にて算出された当該検証対象のタンパク質の類似性中心性尺度を用いて取得する（ステップＳＡ－２）。
【００３０】
そして、承認予測装置の制御部は、ステップＳＡ－２にて取得された判定結果を出力部を介して出力させ（ステップＳＡ－３）、処理を終了する。
【００３１】
以上で、概要（その１）の説明を終える。
【００３２】
［概要（その２）］
また、図２を参照して、本発明の実施の形態の概要の一例について説明する。図２は、本実施の形態の基本原理を示すフローチャートである。
【００３３】
本実施の形態の承認予測装置の制御部は、図２に示すように、タンパク質類似性ネットワークを構成するタンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を含む中心性尺度である類似性中心性尺度を算出する（ステップＳＢ－１）。
【００３４】
そして、承認予測装置の制御部は、タンパク質間相互作用ネットワークを構成するタンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を含む中心性尺度である相互作用中心性尺度を算出する（ステップＳＢ－２）。
【００３５】

そして、承認予測装置の制御部は、各薬剤の承認属性と、ステップＳＢ－１にて算出された類似性中心性尺度の各薬剤の標的毎の合計値および平均値と、ステップＳＢ－２にて算出された相互作用中心性尺度の各薬剤の標的毎の合計値および平均値と、を訓練データとする分類器を用いて、検証対象の化合物が非承認の薬剤に分類される確率である拒否スコアを算出する（ステップＳＢ－３）。

【００３６】

そして、承認予測装置の制御部は、ステップＳＢ－３にて算出された拒否スコアを出力部を介して出力させ（ステップＳＢ－４）、処理を終了する。

【００３７】

以上で、本実施の形態の概要の説明を終える。

【００３８】

［承認予測装置１００の構成］

次に、本実施の形態における承認予測装置１００の構成の詳細について、図３を参照して以下に説明する。図３は、本実施の形態における承認予測装置１００の構成の一例を示すブロック図であり、該構成のうち本発明に関係する部分のみを概念的に示している。ここで、本実施の形態における承認予測装置１００においては、各構成が一筐体内に全て備えられ、単独で処理を行うもの（スタンドアローン型）を、承認予測装置１００として説明するが、当該実施例に限らず、各構成が分離した筐体内に備えられ、ネットワーク３００等を介して接続されて一つの概念としての装置を構成するもの（例えば、クラウドコンピューティング等）であってもよい。

【００３９】

図３において、外部システム２００は、ネットワーク３００を介して、承認予測装置１００と相互に接続され、タンパク質配列情報、薬剤情報、薬剤標的情報、および／もしくは、タンパク質間相互作用情報等に関する外部データベース、ならびに／または、ユーザインターフェース等を実行するウェブサイトを提供する機能等を有していてもよい。

【００４０】

ここで、外部システム２００は、ＷＥＢサーバやＡＳＰサーバ等として構成していてもよい。また、外部システム２００のハードウェア構成は、一般に市販されるワークステーション、パーソナルコンピュータ等の情報処理装置およびその付属装置により構成していてもよい。また、外部システム２００の各機能は、外部システム２００のハードウェア構成中のＣＰＵ、ディスク装置、メモリ装置、入力装置、出力装置、通信制御装置等およびそれらを制御するプログラム等により実現されてもよい。

【００４１】

また、ネットワーク３００は、承認予測装置１００と外部システム２００とを相互に接続する機能を有し、例えば、インターネット等である。

【００４２】

また、承認予測装置１００は、概略的に、制御部１０２と通信制御インターフェース部１０４と記憶部１０６と入出力制御インターフェース部１０８とを備える。ここで、承認予測装置１００は、更に、表示部１１２を少なくとも含む出力部、および、入力部１１４を備えていてもよい。また、出力部は、更に、音声出力部、および、印刷出力部等を含んでいてもよい。ここで、制御部１０２は、承認予測装置１００の全体を統括的に制御するＣＰＵ等である。また、通信制御インターフェース部１０４は、通信回線等に接続されるルータ等の通信装置（図示せず）に接続されるインターフェースであり、入出力制御インターフェース部１０８は、出力部、および、入力部１１４に接続されるインターフェースである。また、記憶部１０６は、各種のデータベースやテーブルなどを格納する装置である。これら承認予測装置１００の各部は任意の通信路を介して通信可能に接続されている。更に、この承認予測装置１００は、ルータ等の通信装置および専用線等の有線または無線の通信回線を介して、ネットワーク３００に通信可能に接続されている。

【００４３】

記憶部１０６に格納される各種のデータベースやテーブル（タンパク質配列情報データ

ベース１０６ａ、類似性ネットワーク情報データベース１０６ｂ、薬剤標的データベース１０６ｃ、および、相互作用ネットワーク情報データベース１０６ｄ）は、固定ディスク装置等のストレージ手段である。例えば、記憶部１０６は、各種処理に用いる各種のプログラム、テーブル、ファイル、データベース、および、ウェブページ等を格納する。

【００４４】

これら記憶部１０６の各構成要素のうち、タンパク質配列情報データベース１０６ａは、タンパク質のアミノ酸配列に関する配列情報を記憶するタンパク質配列情報記憶手段である。ここで、アミノ酸配列は、ヒトタンパク質のアミノ酸配列であってもよい。また、配列情報は、ＦＡＳＴＡ形式であってもよい。これら配列情報は、タンパク質配列情報データベース１０６ａに予め記憶されており、承認予測装置１００の制御部１０２は、定期的に、および／または、制御部１０２による処理に応じてネットワーク３００を介して最新のデータを外部システム２００（例えば、ＮＣＢＩ、または、ＵＮＩＰＲＯＴ等）からダウンロードしてタンパク質配列情報データベース１０６ａに記憶された配列情報をアップデートしてもよい。

【００４５】

また、類似性ネットワーク情報データベース１０６ｂは、類似性を有するタンパク質同士で構成されたタンパク質類似性ネットワーク（ＰＳＩＮ）に関する類似性ネットワーク情報を記憶する類似性ネットワーク情報記憶手段である。

【００４６】

また、薬剤標的データベース１０６ｃは、薬剤の承認、または、非承認に関する承認属性を含む薬剤情報と、当該薬剤の標的となるタンパク質に関するタンパク質情報と、を対応付けて記憶する薬剤標的記憶手段である。ここで、非承認の薬剤とは、本実施の形態において、薬剤承認において撤回（Ｗｉｔｈｄｒａｗ）、または、却下（Ｉｌｌｉｃｉｔ）された薬剤を、問題がある薬剤として１つのグループと見なした薬剤であってもよい。すなわち、問題がある薬剤とは、有害な影響を与えるため市場から撤廃しなければならない薬剤、または、社会的に許されない違法な薬剤（例えば、覚せい剤または幻覚剤等）であって、承認される薬剤と区別する必要のある薬剤であってもよい。これら薬剤承認に関する薬剤情報およびタンパク質情報は、薬剤標的データベース１０６ｃに予め記憶されており、承認予測装置１００の制御部１０２は、定期的に、および／または、制御部１０２による処理に応じてネットワーク３００を介して最新のデータを外部システム２００（例えば、Ｄｒｕｇｂａｎｋ（ｈｔｔｐ：／／ｗｗｗ．ｄｒｕｇｂａｎｋ．ｃａ／）等）からダウンロードして薬剤標的データベース１０６ｃに記憶された薬剤承認に関する薬剤情報およびタンパク質情報をアップデートしてもよい。

【００４７】

また、相互作用ネットワーク情報データベース１０６ｄは、タンパク質間の相互作用に基づき構成されたタンパク質間相互作用ネットワーク（ＰＰＩ）に関する相互作用ネットワーク情報を記憶する相互作用ネットワーク情報記憶手段である。これら相互作用ネットワーク情報は、相互作用ネットワーク情報データベース１０６ｄに予め記憶されており、承認予測装置１００の制御部１０２は、定期的に、および／または、制御部１０２による処理に応じてネットワーク３００を介して最新のデータを外部システム２００（例えば、ＨＩＰＰＩＥ（ｈｔｔｐ：／／ｃｂｄｍ．ｍｄｃ－ｂｅｒｌｉｎ．ｄｅ／ｔｏｏｌｓ／ｈｉｐｐｉｅ／）等）からダウンロードして相互作用ネットワーク情報データベース１０６ｄに記憶された相互作用ネットワーク情報をアップデートしてもよい。

【００４８】

また、通信制御インターフェース部１０４は、承認予測装置１００とネットワーク３００（またはルータ等の通信装置）との間における通信制御を行う。すなわち、通信制御インターフェース部１０４は、外部システム２００、および、他の端末等と通信回線を介してデータを通信する機能を有する。

【００４９】

また、入出力制御インターフェース部１０８は、出力部（表示部１１２）、および、入

力部１１４の制御を行う。

【００５０】

ここで、表示部１１２としては、アプリケーション等の表示画面を表示する表示手段（例えば、液晶または有機ＥＬ等から構成されるディスプレイ、モニタ、または、タッチパネル等）であってもよい。また、入力部１１４は、例えば、キー入力部、タッチパネル、コントロールパッド（例えば、タッチパッド、および、ゲームパッド等）、マウス、キーボード、または、マイク等であってもよい。また、音声出力部としては、例えば、スピーカ等であってもよい。また、印刷出力部としては、例えば、プリンタ等であってもよい。

【００５１】

また、図３において、制御部１０２は、ＯＳ（Ｏｐｅｒａｔｉｎｇ　Ｓｙｓｔｅｍ）等の制御プログラムや、各種の処理手順等を規定したプログラム、および、所要データを格納するための内部メモリを有する。そして、制御部１０２は、これらのプログラム等により、種々の処理を実行するための情報処理を行う。制御部１０２は、機能概念的に、類似性ネットワーク情報格納部１０２ａ、類似性中心性尺度算出部１０２ｂ、承認判定部１０２ｃ、判定結果出力部１０２ｄ、相互作用中心性尺度算出部１０２ｅ、拒否スコア算出部１０２ｆ、および、拒否スコア出力部１０２ｇを備える。

【００５２】

このうち、類似性ネットワーク情報格納部１０２ａは、タンパク質配列情報データベース１０６ａに記憶された配列情報に基づき、シグネチャベースアルゴリズムを用いてタンパク質同士の類似性が相互に検出された場合、当該相互に類似性が検出されたタンパク質同士で構成されるタンパク質類似性ネットワーク（ＰＳＩＮ）を作成し、当該タンパク質類似性ネットワークに関する類似性ネットワーク情報を類似性ネットワーク情報データベース１０６ｂに格納する類似性ネットワーク情報格納手段である。

【００５３】

また、類似性中心性尺度算出部１０２ｂは、類似性ネットワーク情報データベース１０６ｂに記憶された類似性ネットワーク情報に基づき、タンパク質類似性ネットワークを構成するタンパク質の次数中心性（Ｄｅｇｒｅｅ）、媒介値中心性（Ｂｅｔｗｅｅｎｎｅｓｓ）、距離の中心性（Ｃｌｏｓｅｎｅｓｓ　Ｃｅｎｔｒａｌｉｔｙ）、および、バートの制約値（Ｂｕｒｔ’ｓ　ｃｏｎｓｔｒａｉｎｔ）を含む中心性尺度である類似性中心性尺度を算出する類似性中心性尺度算出手段である。ここで、次数中心性は、ネットワーク内で他のノードとどの程度（いくつ）直接つながっているのかを示す指標である。また、媒介値中心性は、ネットワーク内の他のノードと接続するために通過しなければならない最短経路の数を数えることによって、タンパク質のネットワークの中心性を測るものである。また、距離の中心性は、ネットワーク内の他の全てのノードに到達するために必要なステップ数を測定したものである。また、バートの制約値は、グループ内の個人の位置関係と利点を勉強する社会学的文脈で提案された指標である。

【００５４】

また、承認判定部１０２ｃは、タンパク質類似性ネットワークを構成する、薬剤標的データベース１０６ｃに記憶されたタンパク質情報に基づくタンパク質を標的とする薬剤の承認属性に基づき、当該タンパク質類似性ネットワークを構成する検証対象のタンパク質が、承認された薬剤、または、非承認の薬剤の標的の範囲に含まれるか否かの判定結果を、類似性中心性尺度算出部１０２ｂにより算出された当該検証対象のタンパク質の類似性中心性尺度を用いて取得する承認判定手段である。ここで、承認判定部１０２ｃは、タンパク質類似性ネットワークを構成する、薬剤標的データベース１０６ｃに記憶されたタンパク質情報に基づくタンパク質を標的とする薬剤の承認属性に基づき、類似性中心性尺度算出部１０２ｂにより算出された検証対象のタンパク質の類似性中心性尺度に含まれる次数中心性が高く、距離の中心性が低く、バートの制約値が非常に低い場合、検証対象のタンパク質が、非承認の薬剤の標的の範囲に含まれるという判定結果を生成してもよい。また、検証対象のタンパク質は、ユーザにより入力部１１４を介して入力されたタンパク質情報に基づくものであってもよい。

【００５５】
また、判定結果出力部１０２ｄは、承認判定部１０２ｃにより取得された判定結果を出力部を介して出力させる判定結果出力手段である。ここで、判定結果出力部１０２ｄは、判定結果を表示部１１２に表示させてもよい。また、判定結果出力部１０２ｄは、判定結果を印刷出力部を介して出力させてもよい。
【００５６】
また、相互作用中心性尺度算出部１０２ｅは、相互作用ネットワーク情報データベース１０６ｄに記憶された相互作用ネットワーク情報に基づき、タンパク質間相互作用ネットワークを構成するタンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を含む中心性尺度である相互作用中心性尺度を算出する相互作用中心性尺度算出手段である。
【００５７】
また、拒否スコア算出部１０２ｆは、薬剤標的データベース１０６ｃに記憶された各薬剤の承認属性と、類似性中心性尺度算出部１０２ｂにより算出された類似性中心性尺度の各薬剤の標的毎の合計値および平均値と、相互作用中心性尺度算出部１０２ｅにより算出された相互作用中心性尺度の各薬剤の標的毎の合計値および平均値と、を訓練データとする分類器を用いて、検証対象の化合物が非承認の薬剤に分類される確率である拒否スコアを算出する拒否スコア算出手段である。ここで、検証対象の化合物（薬剤）は、ユーザにより入力部１１４を介して入力された化合物情報に基づくものであってもよい。
【００５８】
また、拒否スコア出力部１０２ｇは、拒否スコア算出部１０２ｆにより算出された拒否スコアを出力部を介して出力させる拒否スコア出力手段である。ここで、拒否スコア出力部１０２ｇは、拒否スコアを表示部１１２に表示させてもよい。また、拒否スコア出力部１０２ｇは、拒否スコアを印刷出力部を介して出力させてもよい。
【００５９】
以上で、本実施の形態における承認予測装置１００の構成の一例の説明を終える。
【００６０】
［承認予測装置１００の処理］
次に、このように構成された本実施の形態における承認予測装置１００の処理の詳細について、以下に図４乃至図１５を参照して詳細に説明する。図４は、本実施の形態における承認予測装置１００の処理の一例を示すフローチャートである。
【００６１】
図４に示すように、類似性ネットワーク情報格納部１０２ａは、ヒトタンパク質データベース（タンパク質配列情報データベース）１０６ａに記憶された配列情報に基づき、タンパク質のホモログ間の類似点を探索するタンパク質シグネチャベースアルゴリズムを用いてタンパク質同士の類似性が相互に検出された場合、当該相互に類似性が検出されたタンパク質同士で構成されるタンパク質類似性ネットワーク（ＰＳＩＮ）を作成し、当該タンパク質類似性ネットワークに関する類似性ネットワーク情報を類似性ネットワーク情報データベース１０６ｂに格納する（ステップＳＣ－１）。ここで、類似したタンパク質を見つけるために、ＮＣＢＩヒトタンパク質データベースから２２０００のヒトタンパク質の各々を照会し、比較するＰＳＩ－ＢＬＡＳＴツール（Ｓｃｈａｆｆｅｒ，ｅｔ ａｌ．，２００１）を使用した場合、従来の研究（Ａｔｋｉｎｓｏｎ，ｅｔ ａｌ．，２００９；Ｃａｍｏｇｌｕ，ｅｔ ａｌ．，２００６；Ｒａｔｔｅｉ，ｅｔ ａｌ．，２０１０；Ｖａｌａｖａｎｉｓ，ｅｔ ａｌ．，２０１０；Ｗｅｓｔｏｎ，ｅｔ ａｌ．，２００４；Ｚｈａｎｇ ａｎｄ Ｇｒｉｇｏｒｏｖ，２００６）とは異なり、相互関連性（すなわち、タンパク質Ａを照会してタンパク質Ｂが類似していることが同定される場合に、その逆のタンパク質Ｂを照会してタンパク質Ａを類似として同定されるということ）を示さないという結果を得た。そこで、この結果を受けて、類似性ネットワーク情報格納部１０２ａは、グラフ理論表現を用いた新しいタンパク質類似性ネットワーク（ＰＳＩＮ）を作成した。ここで、当該タンパク質類似性ネットワー

ク（ＰＳＩＮ）では、ノードがタンパク質を表しており、タンパク質の配列が高い類似性を有し、且つ、双方向のヒット（すなわち、タンパク質Ａがタンパク質Ｂに類似することが同定される場合、その逆のタンパク質Ｂがタンパク質Ａに類似することが同定されること）が確認されたときにのみ、２つのノードがエッジで接続されている。これにより、類似性ネットワーク情報格納部１０２ａは、１９，７２１ノードおよび７７６，５９８エッジを含むタンパク質類似性ネットワーク（ＰＳＩＮ）を作成した。

【００６２】

ここで、図５を参照して、本実施の形態における配列情報の一例について説明する。図５は、本実施の形態における配列情報の一例を示す図である。

【００６３】

図５に示すように、タンパク質配列情報データベース１０６ａに記憶された配列情報は、Ｐ６３２６１、および、Ｐ４９２８１等のヒトタンパク質のＦＡＳＴＡ形式のタンパク質配列情報であってもよい。

【００６４】

また、図６を参照して、本実施の形態における類似性ネットワーク情報の一例について説明する。図６は、本実施の形態における類似性ネットワーク情報の一例を示す図である。

【００６５】

図６に示すように、本実施の形態における類似性ネットワーク情報には、タンパク質の名称、当該タンパク質と類似するタンパク質（Ｎｅｉｇｈｂｏｕｒｓ）の名称、配列スコア、および、２つのタンパク質が類似する領域の配列情報を含んでいてもよい。図６においては、Ｑ３ＭＩ９４とＱ９Ｙ４７３との類似性に関する類似性ネットワーク情報、および、Ｑ９Ｐ２Ｖ４とＱ８Ｎ０Ｖ４との類似性に関する類似性ネットワーク情報を一例として示している。

【００６６】

図４に戻り、類似性中心性尺度算出部１０２ｂは、類似性ネットワーク情報データベース１０６ｂに記憶された類似性ネットワーク情報に基づき、中心性基準を算出するアルゴリズムを用いて、タンパク質類似性ネットワーク（ＰＳＩＮ）を構成するタンパク質の次数中心性（Ｄｅｇｒｅｅ）、媒介値中心性（Ｂｅｔｗｅｅｎｎｅｓｓ）、距離の中心性（Ｃｌｏｓｅｎｅｓｓ　Ｃｅｎｔｒａｌｉｔｙ）、および、バートの制約値（Ｂｕｒｔ’ｓ　ｃｏｎｓｔｒａｉｎｔ）を算出する（ステップＳＣ－２）。

【００６７】

ここで、本実施形態におけるＰＳＩＮを構成するタンパク質の中心性尺度について説明する。まず、類似性中心性尺度算出部１０２ｂは、ＰＳＩＮ内のノードとどの程度直接つながっているのかを示す指標である次数中心性として、１（最もつながりの少ないもの）から４４１（最もつながりの多いもの）の範囲で算出している。

【００６８】

また、類似性中心性尺度算出部１０２ｂは、ノードｉとノードｊとの間の最短経路の数であるs_{ij}、および、ノードｖを通過する最短経路の割合である$s_{ij}(v)$からなる以下の数式（１）を用いて、媒介値中心性Ｂ（ｖ）を算出している。

【数１】

$$B(v)=\sum\frac{s_{ij}(v)}{s_{ij}},\qquad \text{with } i\neq j, v\neq i \text{ and } v\neq j \qquad \cdots(1)$$

【００６９】

また、類似性中心性尺度算出部１０２ｂは、ノードｖとノードｉとの間のステップで表す距離であるｄ（ｖ，ｉ）からなる以下の数式（２）を用いて、距離の中心性Ｃ（ｖ）を算出している。

【数２】

$$C(v)=\frac{1}{\sum d(v,i)},\qquad \text{with } i\neq v \qquad \cdots(2)$$

【００７０】

また、類似性中心性尺度算出部１０２ｂは、ノードｊとノードｉとの関係の比例強度、および、ノードｊとノードｑとの関係の比例強度の積である$p_{iq}p_{qj}$からなる以下の数式（３）を用いて、バートの制約値Ｃ（ｉ）を算出している。

【数３】

$$C(i)=\sum_{j}(p_{ij}+\sum_{q}p_{iq}p_{qj})^2,\qquad \text{with } q\neq i,j,\text{and } j\neq i \qquad \cdots(3)$$

【００７１】

ここで、図７を参照して、本実施の形態におけるバートの制約値について説明する。図７は、本実施の形態におけるバートの制約の一例を示す図である。

【００７２】

バートの制約値は、グループ内の個人の位置関係と利点とを学習する社会学的文脈で提案された手法であり、図７においてノードが個人であるとすると、図７左図において、すべてのノードには代替接続があるので、他人との交渉および取引をすることができる。一方、図７右図のように、構造的な穴（Ｓｔｒｕｃｔｕｒａｌ　ｈｏｌｅ）が存在する場合、ノード２とノード３との間では互いの存在を認識できないため、ノード１が交渉に有利な位置にあることになる。これを、本実施の形態においては、タンパク質であるノードの類似のコンテキストに適用することで、小さなバートの制約値を持つタンパク質（ノード）は、一般的に異なるタンパク質ファミリーの間に位置する複数のドメインを有するものとなり、大きなバートの制約値を持つタンパク質（ノード）は、わずかな隣接タンパク質と配列類似性と示すものとなる。

【００７３】

また、図８を参照して、本実施の形態におけるタンパク質の中心性尺度の一例を説明する。図８は、本実施の形態におけるタンパク質の中心性尺度の一例を示す図である。

【００７４】

図８に示すように、類似性中心性尺度算出部１０２ｂは、ＰＳＩＮを構成するタンパク質（Ｐ１４７８４、Ｐ１４８５４、Ｐ１４８５９、Ｐ１４８６７、Ｐ１４８６８、Ｐ１４９０２、および、Ｐ１４９２０）の中心性尺度として、次数中心性（Ｄｅｇｒｅｅ）、媒介値中心性（Ｂｅｔｗｅｅｎｎｅｓｓ）、距離の中心性（Ｃｌｏｓｅｎｅｓｓ　Ｃｅｎｔｒａｌｉｔｙ）、および、バートの制約値（Ｂｕｒｔ’ｓ　ｃｏｎｓｔｒａｉｎｔ）を算出し、当該中心性尺度のリストを出力してもよい。

【００７５】

図４に戻り、承認判定部１０２ｃは、タンパク質類似性ネットワークを構成する、薬剤標的データベース１０６ｃに記憶されたタンパク質情報に基づくタンパク質を標的とする薬剤の承認属性に基づき、当該タンパク質類似性ネットワークを構成する検証対象のタンパク質が、承認された薬剤、または、非承認の薬剤の標的の範囲に含まれるか否かの判定結果（標的タンパク質の安全性）を、ステップＳＣ－２にて類似性中心性尺度算出部１０２ｂにより算出された当該検証対象のタンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を用いて取得する（ステップＳＣ－３）。すなわち、承認判定部１０２ｃは、タンパク質類似性ネットワークを構成するタンパク質の中心性尺度、および、薬剤標的データベース１０６ｃに記憶されたリストを必要とし、薬剤標的の承認、または、非承認（撤回（Ｗｉｔｈｄｒａｗ）、および、却下（Ｉｌｌｉｃｉｔ））と仮定できる値の範囲を決定してもよい。ここで、当該ステップにおいては、個々のタンパク質だけを考慮し、化合物の標的可能なタンパク質全セットは考慮しない。また、個々の薬剤ターゲットの特性を先に決定するのは、単一標的の化合物（魔法の弾丸（ｍａｇｉｃ

ｂｕｌｌｅｔ））、または、ｓｉＲＮＡベースの治療法等が単一標的のみを阻害するように設計されているため、標的の治療抑制が安全であるという前提で、標的の選択がなされる必要があるからである。

【００７６】

また、承認判定部１０２ｃは、タンパク質類似性ネットワークを構成する、薬剤標的データベース１０６ｃに記憶されたタンパク質情報に基づくタンパク質を標的とする薬剤の承認属性に基づき、ステップＳＣ－２にて類似性中心性尺度算出部１０２ｂにより算出された検証対象のタンパク質の類似性中心性尺度に含まれる次数中心性が高く、距離の中心性が低く、バートの制約値が非常に低い場合、検証対象のタンパク質が、非承認の薬剤の標的の範囲に含まれるという判定結果を生成してもよい。

【００７７】

ここで、図９を参照して、本実施の形態における薬剤標的データベース１０６ｃに記憶される情報の一例について説明する。図９は、本実施の形態における薬剤標的データベース１０６ｃに記憶される情報の一例を示す図である。

【００７８】

図９に示すように、本実施の形態における薬剤標的データベース１０６ｃに記憶される情報には、薬剤名称（Ｄｒｕｇ）、当該薬剤の標的となるタンパク質の名称（Ｔａｒｇｅｔｓ）、および、（日本の厚生労働省、または、米国のＦＤＡ等による）当該薬剤の承認、または、非承認に関する承認属性（Ｓｔａｔｕｓ）が含まれていてもよい。

【００７９】

ここで、図１０を参照して、本実施の形態における承認または非承認の薬剤の標的の中心性尺度の一例について説明する。図１０は、本実施の形態における承認または非承認の薬剤の標的の中心性尺度の一例を示す図である。

【００８０】

図１０に示すように、非承認（問題のある）薬剤の標的タンパク質は、承認薬剤の標的タンパク質よりも、高い次数中心性（Ｄｅｇｒｅｅ）、非常に低いバートの制約値（Ｂｕｒｔ’ｓ ｃｏｎｓｔｒａｉｎｔ）、および、負の対数スケールにおいて低い距離の中心性（Ｃｌｏｓｅｎｅｓｓ Ｃｅｎｔｒａｌｉｔｙ）を示していてもよい。図１０に示すように、承認薬剤の標的では、他の多くのタンパク質と構造共有度が低いが（ｌｏｗ－ｄｅｇｒｅｅ）、非承認の標的では、複数のタンパク質間での構造共有が高いことから、非承認薬物の標的は、非特異的な結合および副作用を起こしやすいという特徴を持っていることがわかる。

【００８１】

図４に戻り、判定結果出力部１０２ｄは、承認判定部１０２ｃにより取得された標的タンパク質の安全性を表示部１１２に表示させる（ステップＳＣ－４）。ここで、判定結果出力部１０２ｄは、判定結果を印刷出力部を介して出力させてもよい。また、判定結果出力部１０２ｄは、ユーザが興味のあるタンパク質が、安全な薬剤標的の範囲内にあるか危険な薬剤標的の範囲内にあるかどうかを確認のために照会可能なリストを出力してもよい。

【００８２】

一方、相互作用中心性尺度算出部１０２ｅは、相互作用ネットワーク情報データベース１０６ｄに記憶された相互作用ネットワーク情報に基づき、タンパク質間相互作用ネットワーク（ＰＰＩ）を構成するタンパク質の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値を算出する（ステップＳＣ－５）。

【００８３】

ここで、図１１を参照して、本実施の形態における相互作用ネットワーク情報の一例について説明する。図１１は、本実施の形態における相互作用ネットワーク情報の一例を示す図である。

【００８４】

図１１に示すように、本実施の形態における相互作用ネットワーク情報は、物理的に相

互作用するタンパク質同士を組としたタンパク質リストを含んでいてもよい。

【００８５】

図４に戻り、拒否スコア算出部１０２ｆは、薬剤標的データベース１０６ｃに記憶された各薬剤の承認属性と、ステップＳＣ－２にて類似性中心性尺度算出部１０２ｂにより算出された次数中心性、媒介値中心性、距離の中心性、および、バートの制約値の各薬剤の標的毎の合計値および平均値と、ステップＳＣ－５にて相互作用中心性尺度算出部１０２ｅにより算出された次数中心性、媒介値中心性、距離の中心性、および、バートの制約値の各薬剤の標的毎の合計値および平均値と、を訓練データとする機械学習の分類器を用いて、検証対象の化合物が非承認の薬剤に分類される確率である拒否スコアを算出する（ステップＳＣ－６）。ここで、薬剤標的データベース１０６ｃでは、ほとんどの既存の薬剤（化合物）は、一度に複数のタンパク質に結合してその活性を阻害する、すなわち、複数の薬剤標的が報告されているため、各化合物について標的となるすべてのタンパク質の中心性尺度を考慮する必要がある。そこで、拒否スコア算出部１０２ｆは、タンパク質類似性ネットワーク（ＰＳＩＮ）、および、タンパク質相互作用ネットワーク（ＰＰＩ）を使用して、各薬剤について標的毎の次数中心性、媒介値中心性、距離の中心性、および、バートの制約値の合計および平均を算出し、ＰＳＩＮから８属性、ＰＰＩから８属性、および、化合物（承認または非承認）のクラスを示す１属性を分類器に入力する最終データセットとして用いている。ここで、機械学習の分類器は、ＷＥＫＡ等の既存のパッケージ（Ｗｉｓｈａｒｔ，２００６）などの機械学習の分類器のセットであってもよい。

【００８６】

なお、本実施の形態においては、トレーニング、および、予測ステップのガイドとして、機械学習の分類と薬剤分類（承認、または、非承認）を用いて、最終データセットを１０倍クロス検証法を使用して処理した。さらに、本実施の形態においては、いくつかの異なる分類アルゴリズムを使用してこの手順を実行し、前処理技術の使用、および、同じデータセットに対しタンパク質類似性ネットワーク（ＰＳＩＮ）からの中心性尺度と、タンパク質相互作用ネットワーク（ＰＰＩ）からの中心性尺度の使用の２つのケースで予測のパフォーマンスが強化されていることを確認した。

【００８７】

ここで、本実施の形態における前処理は、下記の三つのステップで行ってもよい。最初に、合成したデータセットの他のインスタンスの手段およびモードで欠損値を充填し、第二に、より小さいクラスのインスタンスの数を拡大し、最後に、データセットをサンプリングしてもよい。本実施の形態におけるデータセットにおいては、承認された薬剤標的クラスと最大３００例の非承認（問題のある）薬剤標的クラスの複数のインスタンスから構成されているため、より小さいクラスに対する標本からサンプルを多く集める必要がある。そのため、新規化合物の開発コストを鑑みれば、承認された薬剤が問題のある薬剤と誤分類されることで生じる不都合は、問題のある薬剤を承認された薬剤として誤分類することで引き起こされる不都合よりも小さいため、本実施の形態においては、小さなクラスを多く集め、大規模なクラスを少なくサンプリングするためにＳＭＯＴＥアルゴリズムを使用してもよい。このような戦略は、様々なサイズを含むデータセットでの分類器の性能を向上させることになる。また、２番目のステップであるリサンプリングを実行するには、データセットからランダムにインスタンスを選択、すなわち、同じインスタンスが二回選択することができてもよい。更に、新しいデータセットには、オリジナルのデータセットと同じインスタンス数および属性の数があってもよく、５０から６０の一意のインスタンスであってもよい。

【００８８】

ここで、図１２を参照して、本実施の形態における分類器の性能改善の一例について説明する。図１２は、本実施の形態における分類器の性能改善の一例を示す図である。

【００８９】

図１２に示すように、本実施の形態における分類器においては、前処理技術の使用、および、同じデータセットに対しＰＳＩＮからの中心性尺度と、ＰＰＩからの中心性尺度の

使用をすることで、問題のある薬剤クラスに対する分類器の感度をかなり向上させることができることを示している。

【００９０】

更に、本実施の形態においては、三つの異なる戦略を使用した１５機械学習の分類器の予測力を比較した。最初の方法では、１０倍交差検証を用いて比較した。また、第二の方法では、元のデータセットを、インスタンスの７０／１００と３５／１００とでそれぞれ設定したトレーニングセットとテストセットとに分割して比較した。本実施の形態においては、元のデータセットから、５００回以上ランダムに薬剤を選択して、偏りがないように調整した。また、トレーニングセットとテストセットとにデータセットを分割する際、トレーニングセットのみを事前処理した。

【００９１】

ここで、図１３を参照して、本実施の形態における分類器の分類精度の一例について説明する。図１３は、本実施の形態における分類器の分類精度の一例を示す図である。

【００９２】

図１３に示すように、本実施の形態における分類器の現実的な精度測定には、薬剤の承認クラス、または、問題のあるクラスに対する、真陽性率の調和平均（ｈａｒｍｏｎｉｃ　ｍｅａｎ　ｏｆ　ｔｈｅ　ｔｒｕｅ　ｐｏｓｉｔｉｖｅ　ｒａｔｅｓ）を用いた。ここで、図１３に示すように、（パラメータの最適化、および、前処理技術の使用により）ほとんどすべての分類子（分類器）が同様の性能を持っていたため、本実施の形態においては、更なる医薬品の安全性予測のために、すべてのアルゴリズムが必然的に持っているバイアスを補正することを目的として、異なる原理を使用して構築された７つの最高の性能を発揮するアルゴリズム（ＫＳＴＡＲ、ＩＢＫ、Ｄｅｃｏｒａｔｅ、ＥＮＤ　ＣｌａｓｓＢａｌａｎｃｅｄＮＤ、ＪＲｉｐ、または、ＲｏｔａｔｉｏｎＦｏｒｅｓｔ等）を使用した。

【００９３】

また、図１４を参照して、本実施の形態における分類器の一例について説明する。図１４は、本実施の形態における分類器の一例を示す図である。

【００９４】

図１４に示すように、元のデータセットの調整の有無にかかわらず、ＫＳｔａｒ、Ｄｅｃｏｒａｔｅ、Ｒｏｔａｔｉｏｎ　Ｆｏｒｅｓｔ、および、Ｒａｎｄｏｍ　Ｆｏｒｅｓｔが最高のパフォーマンスを示すことを確認したので、本実施の形態においては、これらベスト４のアルゴリズムを更なる分析に使用した。テストフェーズでは、分類器がこれまで検出されていないインスタンスを分類する際、これらの７つの最適なアルゴリズムは、各薬剤が問題のあるクラスに属する確率を算出し、算出した確率を使用して、拒否スコア（ＲＳ）という名前のインデックスを作成した。ここで、本実施の形態においては、反対調和平均（ｃｏｎｔｒａ　ｈａｒｍｏｎｉｃ　ｍｅａｎ）を用いて、これらの確率を平均した値をＲＳとしてもよい。ＲＳの値は、化合物が安全（ＲＳが０．０に近い）、または、有害（ＲＳが１．０に近い）であるかを予測しているかを示唆していてもよい。

【００９５】

図４に戻り、拒否スコア出力部１０２ｇは、拒否スコア算出部１０２ｆにより算出された化合物の拒否スコアを表示部１１２に表示させ（ステップＳＣ－７）、処理を終了する。ここで、拒否スコア出力部１０２ｇは、化合物の拒否スコアを印刷出力部を介して出力させてもよい。

【００９６】

ここで、図１５を参照して、本実施形態における出力情報の一例について説明する。図１５は、本実施形態における出力情報の一例を示す図である。

【００９７】

図１５に示すように、拒否スコア出力部１０２ｇは、薬剤とそれぞれの拒否スコア（０．００－１．００の間の値）とのリストを出力してもよい。ここで、問題のある薬剤は１．００に近いスコア値を持っているが、承認薬剤は０．００に近いスコアを持っている。

また、図１５は、Ｄｒｕｇｂａｎｋデータベースから得られた既存の薬剤を入力して得られた一例を示している。このように、ユーザが興味のある薬剤候補となる化合物を入力することで、標的タンパク質と化合物との拒否スコアを確認することができる。なお、本実施の形態による手法を用いることで、既存の１０００の承認薬と非承認薬との間で正確に区別することに成功し、本手法の有効性を確認した。

【００９８】

以上で、本実施の形態における承認予測装置１００の処理の一例の説明を終える。

【００９９】

［他の実施の形態］

さて、これまで本発明の実施の形態について説明したが、本発明は、上述した実施の形態以外にも、特許請求の範囲に記載した技術的思想の範囲内において種々の異なる実施の形態にて実施されてよいものである。

【０１００】

例えば、承認予測装置１００がスタンドアローンの形態で処理を行う場合を一例に説明したが、承認予測装置１００は、クライアント端末（承認予測装置１００とは別筐体である）からの要求に応じて処理を行い、その処理結果を当該クライアント端末に返却するようにしてもよい。

【０１０１】

また、実施の形態において説明した各処理のうち、自動的に行われるものとして説明した処理の全部または一部を手動的に行うこともでき、あるいは、手動的に行われるものとして説明した処理の全部または一部を公知の方法で自動的に行うこともできる。

【０１０２】

このほか、上記文献中や図面中で示した処理手順、制御手順、具体的名称、各処理の登録データや検索条件等のパラメータを含む情報、画面例、データベース構成については、特記する場合を除いて任意に変更することができる。

【０１０３】

また、承認予測装置１００に関して、図示の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。

【０１０４】

例えば、承認予測装置１００の各装置が備える処理機能、特に制御部１０２にて行われる各処理機能については、その全部または任意の一部を、ＣＰＵ（Ｃｅｎｔｒａｌ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ）および当該ＣＰＵにて解釈実行されるプログラムにて実現してもよく、また、ワイヤードロジックによるハードウェアとして実現してもよい。尚、プログラムは、後述する、コンピュータに本発明に係る方法を実行させるためのプログラム化された命令を含む、一時的でないコンピュータ読み取り可能な記録媒体に記録されており、必要に応じて承認予測装置１００に機械的に読み取られる。すなわち、ＲＯＭまたはＨＤＤ（Ｈａｒｄ　Ｄｉｓｋ　Ｄｒｉｖｅ）などの記憶部１０６などには、ＯＳ（Ｏｐｅｒａｔｉｎｇ　Ｓｙｓｔｅｍ）と協働してＣＰＵに命令を与え、各種処理を行うためのコンピュータプログラムが記録されている。このコンピュータプログラムは、ＲＡＭにロードされることによって実行され、ＣＰＵと協働して制御部を構成する。

【０１０５】

また、このコンピュータプログラムは、承認予測装置１００に対して任意のネットワーク３００を介して接続されたアプリケーションプログラムサーバに記憶されていてもよく、必要に応じてその全部または一部をダウンロードすることも可能である。

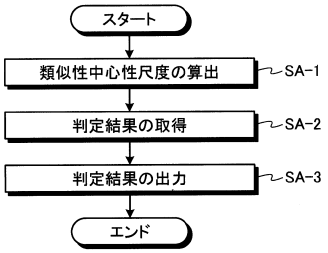
【０１０６】

また、本発明に係るプログラムを、コンピュータ読み取り可能な記録媒体に格納してもよく、また、プログラム製品として構成することもできる。ここで、この「記録媒体」とは、メモリーカード、ＵＳＢメモリ、ＳＤカード、フレキシブルディスク、光磁気ディスク、ＲＯＭ、ＥＰＲＯＭ、ＥＥＰＲＯＭ、ＣＤ－ＲＯＭ、ＭＯ、ＤＶＤ、および、Ｂｌｕ－ｒａｙ　Ｄｉｓｃ等の任意の「可搬用の物理媒体」を含むものとする。

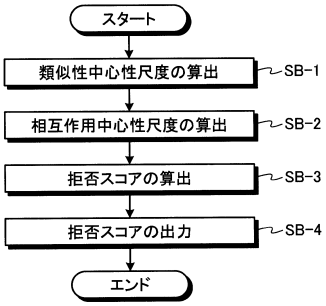
【０１０７】
また、「プログラム」とは、任意の言語や記述方法にて記述されたデータ処理方法であり、ソースコードやバイナリコード等の形式を問わない。なお、「プログラム」は必ずしも単一的に構成されるものに限られず、複数のモジュールやライブラリとして分散構成されるものや、ＯＳ（Ｏｐｅｒａｔｉｎｇ　Ｓｙｓｔｅｍ）に代表される別個のプログラムと協働してその機能を達成するものをも含む。なお、実施の形態に示した各装置において記録媒体を読み取るための具体的な構成、読み取り手順、あるいは、読み取り後のインストール手順等については、周知の構成や手順を用いることができる。
【０１０８】
記憶部１０６に格納される各種のデータベース等（タンパク質配列情報データベース１０６ａ、類似性ネットワーク情報データベース１０６ｂ、薬剤標的データベース１０６ｃ、および、相互作用ネットワーク情報データベース１０６ｄ）は、ＲＡＭ、ＲＯＭ等のメモリ装置、ハードディスク等の固定ディスク装置、フレキシブルディスク、および、光ディスク等のストレージ手段であり、各種処理やウェブサイト提供に用いる各種のプログラム、テーブル、データベース、および、ウェブページ用ファイル等を格納する。
【０１０９】
また、承認予測装置１００は、既知のデスクトップ型またはノート型のパーソナルコンピュータ、携帯電話、スマートフォン、ＰＨＳ、およびＰＤＡ等の携帯端末装置、ならびに、ワークステーション等の情報処理装置として構成してもよく、また、該情報処理装置に任意の周辺装置を接続して構成してもよい。また、承認予測装置１００は、該情報処理装置に本発明の方法を実現させるソフトウェア（プログラム、データ等を含む）を実装することにより実現してもよい。
【０１１０】
更に、装置の分散・統合の具体的形態は図示するものに限られず、その全部または一部を、各種の付加等に応じて、または、機能負荷に応じて、任意の単位で機能的または物理的に分散・統合して構成することができる。すなわち、上述した実施の形態を任意に組み合わせて実施してもよく、実施の形態を選択的に実施してもよい。
【産業上の利用可能性】
【０１１１】
以上詳述に説明したように、本発明によれば、医薬品の承認または拒否の可能性を定量化することができる承認予測装置、承認予測方法、および、プログラムを提供することができるので、特に医療、製薬、創薬、および、生物学研究などの様々な分野において極めて有用である。
【符号の説明】
【０１１２】
１００　承認予測装置
　　１０２　制御部
　　１０２ａ　類似性ネットワーク情報格納部
　　１０２ｂ　類似性中心性尺度算出部
　　１０２ｃ　承認判定部
　　１０２ｄ　判定結果出力部
　　１０２ｅ　相互作用中心性尺度算出部
　　１０２ｆ　拒否スコア算出部
　　１０２ｇ　拒否スコア出力部
　　１０４　通信制御インターフェース部
　　１０６　記憶部
　　１０６ａ　タンパク質配列情報データベース
　　１０６ｂ　類似性ネットワーク情報データベース
　　１０６ｃ　薬剤標的データベース
　　１０６ｄ　相互作用ネットワーク情報データベース

１０８　入出力制御インターフェース部
１１２　表示部
１１４　入力部
２００　外部システム
３００　ネットワーク

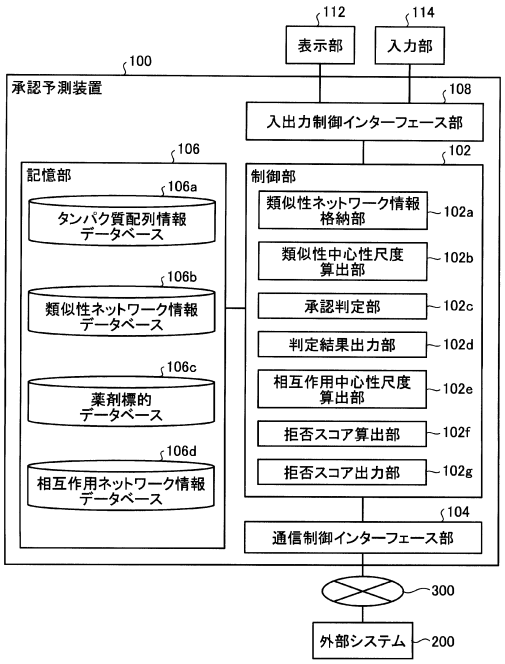
【図１】



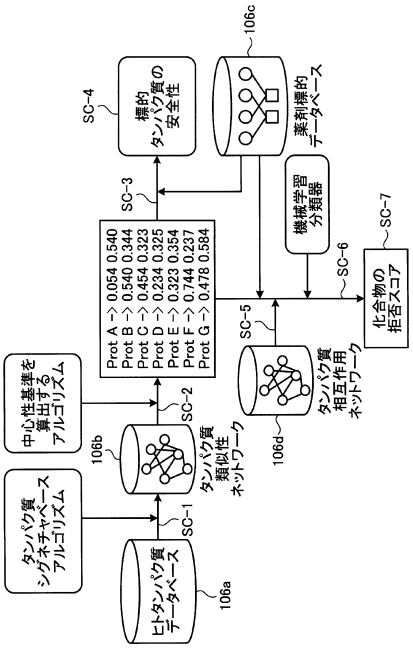
【図２】



【図３】



【図４】



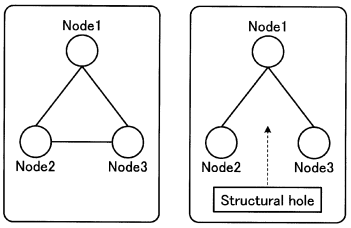
【図５】

```
>P63261

MEEEIAALVIDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEAQS
KRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKMT
QIMFETFNTPAMYVAIQAVLSLYASGRTTGIVMDSGDGVTHTVPIYEGYALPHAILRLDL
AGRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKSY
ELPDGQVITIGNERFRCPEALFQPSFLGMESCGIHETTFNSIMKCDVDIRKDLYANTVLS
GGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWISKQ
EYDESGPSIVHRKCF

>P49281

MVLGPEQKMSDDSVSGDHGESASLGNINPAYSNPSLSQSPGDSEEYFATYFNEKISIPEE
EYSCFSFRKLWAFTGPGFLMSIAYLDPGNIESDLQSGAVAGFKLLWILLLATLVGLLLQR
LAARLGVVTGLHLAEVCHRQYPKVPRVILWLMVELAIIGSDMQEVIGSAIAINLLSVGRI
PLWGGVLITIADTFVFLFLDKYGLRKLEAFFGFLITIMALTFGYEYVTVKPSQSQVLKGM
FVPSCSGCRTPQIEQAVGIVGAVIMPHNMYLHSALVKSRQVNRNNKQEVREANKYFFIES
CIALFVSFIINVFVVSVFAEAFFGKTNEQVVEVCTNTSSPHAGLFPKDNSTLAVDIYKGG
VVLGCYFGPAALYIWAVGILAAGQSSTMTGTYSGQFVMEGFLNLKWSRFARVVLTRSIAI
IPTLLVAVFQDVEHLTGMNDFLNVLQSLQLPFALIPILTFTSLRPVMSDFANGLGWRIAG
GILVLIICSINMYFVVVYVRDLGHVALYVVAAVVSVAYLGFVFYLGWQCLIALGMSFLDC
GHTCHLGLTAQPELYLLNTMDADSLVSR
```

【図６】

```
Q3MI94   Q9Y473   210.901 536      1.670035e-54      82          119         1           214         377
GKMFGQNSTLVIHKAIHTGEKPYKCNECGKAFNQQSHLSRHHRLHTGEKPYKCNDCGKAFIHQSSLARHHRLHTGEKSYKCEECDRVFS
QKSNLERHKIIHTGEKPYKCNECHKTFSHRSSLPCHRRLHSGEKPYKCNECGKTFNVQSHLSRHHRLHTGEKPYKCKVCDKAFMCHSYLA
N H T R I H S G E K P Y K C N E C G K A H N H L I D S S I K P C M S S      5 9 0                2 3 0
GQLFSHSSSDACSKMIHTGETFCKGNQCRKVCGHKQSLKQHQ-IHTQKKPDGCSECGGSFTQKSHLFAQQRIHSVGNLHECGKCGKAFM
PQLKLSVYLTDHTGDIPCICKECGKVFIQRSELLTHQKTHTRKKPYKCHDCGKAFFQMLSLFRHQRTHSREKLYECSECGKGFSQNSTLIIH
QKIHTGERQYACSECGKAFTQKSTLSLHQRIHS  442     G++F  +S+    K IHTGE    K N+C K    +  L +HI  +HT +KP   C++CG +F
+S L    R+H+    ++C +C + F  +  L  +   HTG+ P   C EC K F  RS L   H++ H+ +KPYKC++CGK F     L RH R H+ EK Y+C   C K F
+S L  H +IH+GE+ Y C+ECGKA        S+  + S

Q9P2V4   Q8N0V4   94.6147 232      1.58178e-19       47          82          23          291         23
CPSQCSCSLHIMGDGSKARTVVCNDPDMTLPPASIPPDTSRLRLERTAIRRVPGEAFRPLGRLEQLWLPYNALSELNALMLRGLRRLRELR
LPGNRLAAFPWAALRDAPKLRLLDLQANRLSAVPAEAARFLENLTFLDLSSNQLMRLPQELIVSWAHLETGIFPPGHHPRRVLGLQDNPWA
CDCRLYDLVHLLDGWAPNLAFIETELRCASPRSLAGVAFSQLELRKCQGPELHPGVASIRSLLGQTALLRCGATGVPGPEMSWRRANGRPL
N G T V H Q E V S S D G T S W T L L                                   3 1 3                    3 6
CPATCSCTKESIICVGS————————SWVPRVPGDISSLSLVNGTFSEIKDRMFSHLPSLQLLLLNSNSFTIIRDDAFAGLFHLEYLFIEGNKIETISR
NAFRGLRDLTHLSLANNHIKALPRDVFSDLDSLIELDLRGNKFECDCKAKWLYLWLKMTNS————————TVSDVLCIGPPEYQEKKLNDVTSFD
YECTTTDFVVHQTLPYQSVSVDTFNSKNDVYVAIAQPSMENCMVLEWDHIEMNFRSYDNITGQSIVGCKAILIDDQVFVVVAQLFGGSHIY
303      CP+ CSC+   +                  + P  +P D S L L          +   F  L  L+ L L  N+ + +      GL  L   L + GN++
A R    L L L  N + A+P +    L++L  LDL  N+    +    +      T                +      Y   L D  + +     T+        V+      +
+      +L   +    +    S        ++ V  V+    +
```

【図７】



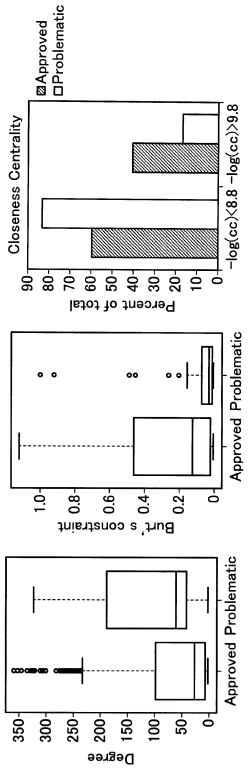
【図８】

Protein ID	Degree	Betweenness	Closeness Centrality	Burts Constraint
P14784	55	12671.535	0.000	0.043
P14854	1	0.000	5.0709939148073e-05	1.000
P14859	224	544.657	0.000	0.016
P14867	59	0.000	0.000	0.066
P14868	4	0.000	5.07176548156413e-05	0.766
P14902	1	0.000	5.0709939148073e-05	1.000
P14920	1	0.000	5.0709939148073e-05	1.000

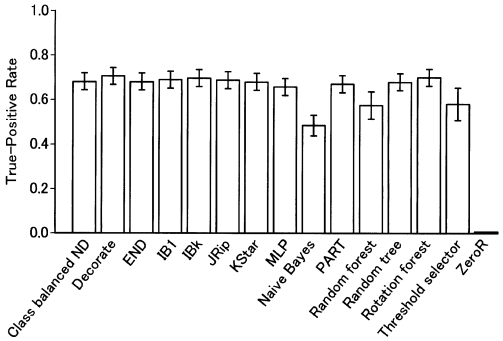
【図９】

Drug	Targets	Status
Acetylsalicylic acid	P23219	Approved
Glycodiazine	P48048, Q09428	Approved
Phenacetin	P23219	Withdraw

【図１０】



【図１３】



【図１４】

Classifier	Reference
ZeroR	(Witten, et al., 2011)
Naive Bayes	(John and Langley, 1995)
Threshold Selector	(Witten, et al., 2011)
MLP	(Cybenko, 1992)
JRip	(Cohen, 1995)
IB1	(Aha, et al., 1991)
PART	(Frank and Witten, 1998)
END	(Dong, et al., 2005)
Class Balanced ND	(Dong, et al., 2005)
IBk	(Aha, et al., 1991)
Random Tree	(Witten, et al., 2011)
Rotation Forest	(Rodriguez and Kuncheva, 2006)
Random Forest	(Breiman, 2001)
Decorate	(Melville and Mooney, 2003)
KStar	(Cleary and Trigg, 1995)

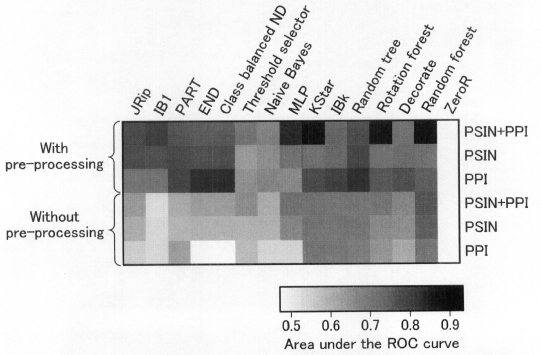
【図１５】

Drug	True Class	Algorithm 1	Algorithm 2	Algorithm 3	Algorithm 4	Rejection Score
AZD8075	Problematic	1.00	1.00	1.00	1.00	0.98
Zomepirac	Problematic	1.00	1.00	1.00	1.00	0.98
BMS 275291	Problematic	1.00	1.00	0.99	1.00	0.98
Yohimbine	Approved	0.00	0.00	0.00	0.01	0.01
Doxepin	Approved	0.00	0.00	0.00	0.01	0.00
Rifabutin	Approved	0.00	0.00	0.00	0.00	0.00

【図１１】

Protein 1	Protein 2
AL1A1_HUMAN	AL1A1_HUMAN
ITA7_HUMAN	ACHA_HUMAN
NEB1_HUMAN	ACTG_HUMAN
SRGN_HUMAN	CD44_HUMAN
GRB7_HUMAN	ERBB2_HUMAN
PAK1_HUMAN	ERBB2_HUMAN
DLG4_HUMAN	ERBB2_HUMAN
P85B_HUMAN	ERBB2_HUMAN
PTN18_HUMAN	ERBB2_HUMAN

【図１２】



フロントページの続き

審査官　渡邉　加寿磨

(56)参考文献　特開２００８－１７６３８９（ＪＰ，Ａ）
特開２０１０－１６５２３０（ＪＰ，Ａ）
特開２００６－１４６３８０（ＪＰ，Ａ）
国際公開第２００５／０６９１８８（WO，Ａ１）
Monica Campillos et al, Drug Target Identification Using Side-Effect Similarity, Science, ２００８年　7月, Vol.321
Michael J. Keiser et al, Prediciting new molecular targets for known drugs, Nature, ２００９年１１月, Vol.462
Eugen Lounkine et al, Large-scale prediction and testing of drug activity on side-effect targets, Nature, ２０１２年　6月２１日, Vol.486

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　１９／１０　－　１９／２８
Ｃ４０Ｂ　３０／０２
Ｇ０６Ｆ　１７／１８
Ｇ０６Ｆ　１７／３０
Ｇ０６Ｑ　１０／００　－　９９／００