

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6822182号
(P6822182)

(45) 発行日 令和3年1月27日 (2021.1.27)

(24) 登録日 令和3年1月12日 (2021.1.12)

(51) Int.Cl.

F I

G O 6 F 16/906 (2019.01)

G O 6 F 16/906

請求項の数 3 (全 13 頁)

(21) 出願番号	特願2017-18450 (P2017-18450)	(73) 特許権者	000005496
(22) 出願日	平成29年2月3日 (2017.2.3)		富士ゼロックス株式会社
(65) 公開番号	特開2018-124898 (P2018-124898A)		東京都港区赤坂九丁目7番3号
(43) 公開日	平成30年8月9日 (2018.8.9)	(74) 代理人	110001210
審査請求日	令和1年12月20日 (2019.12.20)		特許業務法人 Y K I 国際特許事務所
		(72) 発明者	岡本 洋
			神奈川県横浜市西区みなとみらい六丁目1番 富士ゼロックス株式会社内
		(72) 発明者	邱 シュウレ
			神奈川県横浜市西区みなとみらい六丁目1番 富士ゼロックス株式会社内
		審査官	三橋 電太郎

最終頁に続く

(54) 【発明の名称】 プログラム及び情報処理装置

(57) 【特許請求の範囲】

【請求項1】

コンピュータを、

ノード群とそれらノード群を結ぶ実リンク群からなるネットワークの情報を取得する取得手段、

前記ネットワークに対し、前記実リンクで結ばれていないノード同士を結ぶ少なくとも1つの仮想リンクを追加することにより拡張ネットワークを生成する追加手段、

前記拡張ネットワークにおける前記実リンク及び前記仮想リンクを介するノード間の遷移の確率過程を定常状態とみなせるまで繰り返し計算することにより、前記ノード群のクラスタリングを行うクラスタリング手段、

として機能させるためのプログラム。

【請求項2】

前記クラスタリング手段は、

前記拡張ネットワークにおける前記ノード間の確率の遷移の定常状態に基づいて前記実リンク及び前記仮想リンクのリンク確率を計算し、求めたリンク確率のうち実リンクに係るもののみを抽出して規格化し、規格化後の実リンクのリンク確率を用いて前記確率過程の計算を実行する、

ことを特徴とする請求項1に記載のプログラム。

【請求項3】

ノード群とそれらノード群を結ぶ実リンク群からなるネットワークの情報を取得する取

得手段と、

前記ネットワークに対し、前記実リンクで結ばれていないノード同士を結ぶ少なくとも1つの仮想リンクを追加することにより拡張ネットワークを生成する追加手段と、

前記拡張ネットワークにおける前記実リンク及び前記仮想リンクを介するノード間の遷移の確率過程を定常状態とみなせるまで繰り返し計算することにより、前記ノード群のクラスタリングを行うクラスタリング手段と、

を有する情報処理装置。

10

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、プログラム及び情報処理装置に関する。

【背景技術】

【0002】

ネットワーク内のノード群をクラスタリング（コミュニティ抽出とも呼ばれる）する方法はすでに数多く提案されている（非特許文献1及び2参照）。しかし、それらの多くは、互いに重なりを持つクラスタ（コミュニティ）群や、階層構造を持つクラスタ群を抽出することができない。また、リンクに方向があるネットワークに適用することも難しい。

20

【0003】

これに対し、発明者らは、特許文献1、2、3にて「マルコフ連鎖のモジュール分解」に基づいて、ネットワークから重なりと階層を持つクラスタ構造を検出する方法を提案した。これらの方法はリンクに方向があるネットワークにも適用することができる。

【0004】

マルコフ連鎖のモジュール分解に基づくクラスタリングの計算では、クラスタ（コミュニティ） k におけるノードの確率（分類割合） $p(n|k)$ 及びクラスタ k の事前確率（重要度） π_k について繰り返し計算（EMアルゴリズムのMステップ、Eステップ）を行い、定常状態に達したときのこれらの量（ $p(n|k)$ および π_k ）から各クラスタの情報を得る。

30

【先行技術文献】

【特許文献】

【0005】

【特許文献1】特開2013-168127号公報

【特許文献2】特開2016-029526号公報

【特許文献3】特開2016-218531号公報

【非特許文献】

【0006】

【非特許文献1】Newman, M.E.J. Communities, modules and large-scale structure in networks. Nature Physics 8, 25-31 (2011)

40

【非特許文献2】Fortunato, S. Community detection in graphs. Physics Report 486, 75-174 (2010)

【発明の概要】

【発明が解決しようとする課題】

【0007】

マルコフ連鎖等の確率過程の計算を繰り返し行うことでネットワークのクラスタリングを行う方法では、繰り返し計算が定常状態に収束するまでに非常に長い時間がかかる（あるいは非常に大きな計算コストを要する）ことがある。特に、クラスタが非クリーク的な構造を持つ場合に、この問題が顕著に起こる。ここで、複数のノードが相互に完全に結合しているとき、これらのノードは「クリーク」を形成するという。クリークにおいては、

50

任意のノードは任意のノードに対し1リンクでつながっている。一方、クラスタがこれとは反対の構造を持つ（あるノードから別のノードに達するまでに経由するリンクの数が非常に多い）とき、このクラスタは「非クリークの」であるという。

【0008】

確率過程の繰り返し計算は、ネットワーク中の確率の伝播として行われる。物理学とのアナロジーに基づけば、これはネットワークにおけるリンクに沿った拡散現象（ブラウン運動・ランダムウォーク）に他ならない。確率分布の定常状態への収束は、各ノードからリンクをたどって流れ出た確率がネットワーク内に十分拡散して混じりあった後に達成される。しかしながら、コミュニティが非クリーク的な構造を持つときには、あるノードから流れ出た確率が別のあるノードにまで達するのに、多くのリンクを経なければならぬことがある。そのため、各ノードからリンクをたどって流れ出た確率がネットワーク内に十分拡散して混じり合うまでに、非常に時間がかかる。さらに、ネットワークが非クリーク的な構造を持つ場合に、確率が十分混ざり合う前に偽の定常状態に陥ってしまうことが、経験上、しばしば起こる。

10

【0009】

本発明は、単にネットワークのリンクを介したノード間の確率の遷移の確率過程を繰り返し計算する場合よりも、その繰り返し演算が定常状態に収束するまでの繰り返し計算の回数を低減することを目的とする。

【課題を解決するための手段】

【0010】

20

請求項1に係る発明は、コンピュータを、ノード群とそれらノード群を結ぶ実リンク群からなるネットワークの情報を取得する取得手段、前記ネットワークに対し、前記実リンクで結ばれていないノード同士を結ぶ少なくとも1つの仮想リンクを追加することにより拡張ネットワークを生成する追加手段、前記拡張ネットワークにおける前記実リンク及び前記仮想リンクを介するノード間の遷移の確率過程を定常状態とみなせるまで繰り返し計算することにより、前記ノード群のクラスタリングを行うクラスタリング手段、として機能させるためのプログラムである。

【0011】

請求項2に係る発明は、前記クラスタリング手段は、前記拡張ネットワークにおける前記ノード間の確率の遷移の定常状態に基づいて前記実リンク及び前記仮想リンクのリンク確率を計算し、求めたリンク確率のうち実リンクに係るもののみを抽出して規格化し、規格化後の実リンクのリンク確率を用いて前記確率過程の計算を実行する、ことを特徴とする請求項1に記載のプログラムである。

30

【0012】

請求項3に係る発明は、ノード群とそれらノード群を結ぶ実リンク群からなるネットワークの情報を取得する取得手段と、前記ネットワークに対し、前記実リンクで結ばれていないノード同士を結ぶ少なくとも1つの仮想リンクを追加することにより拡張ネットワークを生成する追加手段と、前記拡張ネットワークにおける前記実リンク及び前記仮想リンクを介するノード間の遷移の確率過程を定常状態とみなせるまで繰り返し計算することにより、前記ノード群のクラスタリングを行うクラスタリング手段と、を有する情報処理装置である。

40

【発明の効果】

【0013】

請求項1又は3に係る発明によれば、単にネットワークのリンクを介したノード間の確率の遷移の確率過程を繰り返し計算する場合よりも、その繰り返し演算が定常状態に収束するまでの繰り返し計算の回数を低減することができる。

【0014】

請求項2に係る発明によれば、確率過程の繰り返し計算において、仮想リンクについてのリンク確率も考慮して計算する場合よりも、計算コストを低減することができる。

【図面の簡単な説明】

50

【0015】

【図1】実施形態の装置構成の例を示す図である。

【図2】仮想リンク（テレポーテーション）の説明のための図である。

【図3】実施形態の装置の処理手順の例を示す図である。

【発明を実施するための形態】

【0016】

以下、本発明の実施の形態について、図面を参照しながら説明する。

【0017】

図1は、本発明の実施形態に係る情報処理装置1の構成図である。情報処理装置1は、記憶部10、入力部12、計算部14、表示部16を含む。

10

【0018】

記憶部10は、例えばRAM（Random Access Memory）やROM（Read Only Memory）を含む。記憶部10は、計算部14が実行するプログラムを格納するとともに、計算部14のワークメモリとしても機能する。なお、記憶部10に格納される計算部14が実行するプログラムは、電気通信回線を介して提供されるものであってもよいし、半導体記憶素子等のコンピュータで読み取り可能な情報記憶媒体に格納されて提供されるものであってもよい。

【0019】

本実施形態に係る情報処理装置1の記憶部10には、ネットワーク情報が記憶される。ネットワーク情報は、複数のノードとそれら複数のノードを結ぶ複数のリンクとを含むネットワークを表す情報である。ネットワーク情報が表すネットワークは、例えばハイパーテキストが構成するネットワークや、友人関係のネットワーク等、どのようなものであってもよい。ネットワーク情報は、少なくともノード間の結び付きの関係（ノードとリンクの関係）を示すものであればよく、ノードが含む具体的な内容（HTMLデータの内容等）を示す情報は含まなくてもよい。

20

【0020】

入力部12は、例えばキーボードやマウス等であり、ユーザの指示を計算部14に伝達する。

【0021】

計算部14は、例えばCPU（Central Processing Unit）を含んでおり、記憶部10に格納されるプログラムを実行することにより、ネットワークのクラスタリング（コミュニティ抽出）のための計算処理を実行する。計算部14は、機能的に、取得部140、仮想リンク追加部142、定常リンク確率計算部144、リンク確率修正部146及びクラスタリング計算部148を含む。計算部14の行うクラスタリングのための計算処理については、後に詳細に説明する。

30

【0022】

表示部16は、計算部14により処理された情報をユーザに表示するものであり、例えば液晶ディスプレイである。

【0023】

計算部14について更に詳しく説明する。

40

【0024】

計算部14の取得部140は、記憶部10からネットワーク情報等の情報を取得する。

【0025】

仮想リンク追加部142は、取得されたネットワーク情報が表すネットワークに対して仮想リンクを追加する。仮想リンクとは、そのネットワークを構成するリンク（仮想リンクとの区別のために実リンクと呼ぶ）とは別に設けられる仮想的なリンクである。仮想リンク追加部142は、実リンクが存在しないノード同士の間を結ぶ仮想リンクを追加してもよいし、これに加えて実リンクが存在するノード同士の間を結ぶか沿うリンクを追加してもよい。一つの例では、個々のノードについて、そのノード（自ノード）と、ネットワーク内の各ノード（自ノードも含む）と、を結ぶ仮想ノードを追加する。この例では、ネ

50

ネットワークのノード数がN個の場合、1つのノードについてN個の仮想リンクを追加することとなる。ネットワーク情報が表すネットワークに対して、仮想リンク追加部142により仮想リンクを追加したのもネットワークであり、この後者のネットワークを拡張ネットワークと呼ぶ。

【0026】

図2にネットワークと仮想リンクを模式的に示す。この例では、楕円で示すノード200同士が直線の実リンク202で結ばれて構成されたネットワークが、ネットワーク情報が表す元々のネットワークである。このネットワークはクリークではない（すなわち、ネットワーク内のノードのペアの中に、そのペアのノード間に実リンク202が存在しないものがある）。一方、仮想リンク204は、図中に矢印付き曲線として示す。煩雑さを避けるため、ノード「n」を起点とする仮想リンク204のみを図示している。ノード「n」から自分自身へとつながる仮想リンクは図示を省略している。この例は、個々のノードについて、そのノード（自ノード）と、ネットワーク内の各ノード（自ノードも含む）と、を結ぶ仮想ノードを追加した場合の例である。ノード「n」以外のノードについても、そのノードから全ノードへと延びる仮想リンクが追加される。ネットワーク情報が表すノード200群及び実リンク群202群と、仮想リンク追加部142により追加された仮想リンク204群から構成されるネットワークが拡張ネットワークである。この例では、仮想リンクは有効リンクであり、仮想リンクの起点のノードの確率の一部が終点のノードへと遷移する。

【0027】

前述の通り、非クリーク的なクラスタ（すなわち所属するノード同士の間を結ぶリンクが少ないクラスタ。例えば各ノードはそれぞれ両隣の2つのノードを結ぶリンクしか持たない等）では、あるノードから別のノードに確率が伝播（移動）するのに多数のリンクを経由する必要がある、これが確率過程の繰り返し計算を収束しにくくする重要な原因の一つであった。解析対象である実際のネットワークには、非クリーク的なクラスタを含んだものが少なくない。

【0028】

これに対し、本実施形態のように仮想リンクを追加すれば、繰り返し計算の際に、ノードが持つ確率が実リンクでつながっていない他のノードにも1つの仮想リンクを介して伝播可能となる。今まで2つのノード間を複数の実リンクを介して伝播していた確率が1つの仮想リンクで伝播するので、ノードの持つ確率が他のノードに移動しやすくなる。これにより、繰り返し計算の収束が早くなる。

【0029】

仮想リンクは、実際のネットワーク上の2つのノード間に介在する複数の実リンクに対するショートカット（近道）であり、実空間（実リンク群）ではすぐに行けない別の場所（ノード）に瞬間的（すなわち1リンクで）に移動するテレポーテーションと捉えてもよい。

【0030】

なお、以上では、ノードからネットワーク内のすべてのノードに対して仮想リンクを追加する例を示したが、これは一例に過ぎない。仮想リンクの追加は、ネットワーク内の実リンク202が存在しないノードのペアのうち少なくとも1つに対して行えば、元のネットワークよりも仮想リンクの分だけ確率が混ざりやすくなり、繰り返し計算の収束が早まる。

【0031】

定常リンク確率計算部144は、仮想リンク群が追加済みのネットワーク（すなわち拡張ネットワーク）について、定常状態のリンク確率（詳細は後述）を計算する。

【0032】

リンク確率修正部146は、仮想リンクの追加による計算量の増大を抑えるための、学習データモデルの設定方法の修正の一環として、定常リンク確率計算部144が求めた定常リンク確率を修正する。

【0033】

すなわち、後述するクラスタリング計算部148での計算のために、ランダムウォーカーがどのリンクを移動中であるかを観測すると仮想して、どのような観測結果（学習データ。通過情報とも呼ぶ）が得られるかを数学的にモデル化する。しかし、仮想リンクを含んだ拡張ネットワークはリンク数が膨大な数となる（特に全ノードについて全ノードへの仮想リンクを追加した場合）ので、仮想的な観測により用意すべき学習データの数もリンク数に応じて膨大なものとなり、確率過程の繰り返し計算においてそれら膨大な学習データを参照するための計算コストも膨大になってしまう。一方、翻って考えると、クラスタリングしたいネットワークは、仮想リンクが追加された拡張ネットワークではなく、実リンクのみで仮想リンクを含まない元のネットワークである。そこで、本実施形態では、仮想的な観測は実リンクに対してのみ行われるものとして学習データのモデルを修正する。リンク確率修正部146が行うリンク確率の修正も、このモデル修正に応じたものである。リンク確率修正部146の修正では、定常リンク確率計算部144が計算した仮想リンクも含む各リンクの定常リンク確率のうち、実リンクについての定常リンク確率を抽出する。ただし、定常リンク確率は、仮想リンクも含んだ拡張ネットワークについて計算されているので、その一部である実リンクについての定常リンク確率の合計は1とはならない。このため、抽出した実リンクについての定常リンク確率はそのままでは確率の要件を満たさず、確率過程の計算を適用できない。そこで、リンク確率修正部146は、抽出した実リンクについての定常リンク確率の合計が1となるように規格化し、規格化した値を実リンクについての定常リンク確率の修正値とする。なお、仮想リンクについての定常リンク確率は0とする。

10

20

【0034】

クラスタリング計算部148は、拡張ネットワークの情報、修正済みの定常リンク確率等を用いて、クラスタリングのための確率過程の繰り返し計算を実行する。この繰り返し計算としては、例えば、特許文献1～3に示されたマルコフ連鎖のモジュール分解の考え方に沿った計算を用いればよい。

【0035】

次に図3を参照して、計算部14が実行する処理の例を説明する。

【0036】

この処理では、まず計算部14内の取得部140が、記憶部10からネットワーク情報を取得する（S10）。

30

【0037】

次に、そのネットワーク情報が示すネットワークに対して、仮想リンク追加部142が、仮想リンクの追加を行う（S12）。仮想リンクの追加は、例えば、ネットワーク情報から生成される遷移確率行列 T_{nm} を、仮想リンクを追加した形に修正することで行う。

【0038】

ここで、遷移確率行列 T_{nm} は、ネットワーク内のノード m からノード n へリンクを辿ってエージェント（言い換えれば、ノード m が持つ確率値）が遷移（ランダムウォーク）する確率を行列として表現したものである。例えばエージェントがノードから出る1以上のリンクを等確率で選択するとみなした場合、遷移確率行列は、ネットワーク情報が示すネットワークの構造、すなわちノード間がどのようにリンクされているか、のみに依存する。もちろん、ネットワークの構造以外の情報（例えば各ノードが持つ重要度）を考慮に入れて遷移確率行列 T_{nm} を求めることもできるが、この場合も遷移確率行列 T_{nm} はネットワーク情報に基づく点は同様である。遷移確率行列については、特許文献1～3も参照されたい。

40

【0039】

1つの具体例では、S12では、仮想リンク追加部142は、元のネットワークについての遷移確率行列 T_{nm} から拡張ネットワーク（仮想リンク追加済み）についての遷移確率行列 T_{nm} （ただし T の上にチルダ（波線）付き）を、例えば次の式1により求める。

【数 1】

$$\tilde{T}_{nm} = (1 - \rho)T_{nm} + \rho/N \quad \cdots(\text{式1})$$

【0040】

ここで、Nは拡張ネットワークのノード数であり、これはネットワーク情報が示す元のネットワークのノード数と等しい。またρは、0より大きく1より小さい実数であり、拡張ネットワークにおけるノード間の確率の遷移のうち、仮想リンクによる遷移の割合を示す。なお、仮想リンクの重み（割合ρ）を重くすると、本来クラスタリングしたい実リンクにより構成されるネットワークの構造がマルコフ連鎖の計算に（仮想リンクにくらべて）反映されにくくなり、本末転倒となる。そこで、ρの値はある程度低い値（例えば0.5未満）する。式1の右辺第2項は、仮想リンクによる確率の遷移の成分を示し、この例ではノードmが持つ確率（エージェント）のうちの割合ρが、ノードmから全ノード（合計N個）に延びるN本の仮想リンクを介して均等に、それら全ノードに遷移することを表している。一方、式1の右辺第1項は、実リンクを介して遷移する成分であり、これはノードmの持つ確率のうち、仮想リンクで遷移する割合ρを引いた残りの割合（1-ρ）が、元のネットワークの遷移確率行列 T_{nm} に従って遷移することを表す。

【0041】

式1は、ノードの持つ確率が、そのノードから出る全仮想リンクに対して均等に分配されて遷移するものであるが、これは一例にすぎない。別の例として、次の式2を用いてもよい。

【数 2】

$$\tilde{T}_{nm} = (1 - \rho)T_{nm} + \rho \sum_m W_{nm} / \sum_{n,m} W_{nm} \quad \cdots(\text{式2})$$

【0042】

ここで W_{nm} は、ネットワークの隣接行列であり、ノードmからノードnに実リンクがあれば $W_{nm} = 1$ 、なければ $W_{nm} = 0$ となるN行N列の行列である。式1では、ノードmの持つ確率のうち仮想リンク経由で遷移する割合ρの部分は、起点のノードmから出る全仮想リンクに対して均等に分配された。これに対して、式2では、他のノードとの実リンクの数が多いいノードn（他のノードからのいわば「人気」が高いノード）への仮想リンクほど高い割合で分配される。

更に別の例として、拡張ネットワークについての遷移確率行列を、次の式に従って求めてもよい。

【数 3】

$$\tilde{T} = (1 - \rho)T_{nm} + \rho \frac{f(p_n)}{\sum_{m=1}^N f(p_m)}$$

10

20

30

40

50

【0043】

この式の左辺が、求めるべき拡張ネットワークの遷移確率行列 T_{nm} （ただし T の上にチルダ（波線）付き）を示す。右辺の関数 $f(x)$ は変数 x についての単調増加関数であればどのような関数であってもよい。この式では、ノード m の持つ確率のうち仮想リンク経由で遷移する割合 ρ の部分は、ノード n の確率が高いほどそのノード n に大きい割合で分配される。

【0044】

次に、定常リンク確率計算部 144 が、定常リンク確率を計算する（S14）。

【0045】

この計算では、まず S12 で求められた拡張ネットワークの遷移確率行列 T_{nm} （チルダ付き）を用いて、拡張ネットワークにおける確率遷移（ランダムウォーク）の定常状態において各ノードが持つ確率（定常状態のノード確率）を計算する。この計算では、例えば次の式 3 の計算を定常状態となるまで繰り返す。

【数 4】

$$p_t(n) = \sum_{m=1}^N \tilde{T}_{nm} p_{t-1}(m) \quad \dots(\text{式3})$$

10

20

【0046】

式 3 において、 $p_t(n)$ は、離散的な時刻 t においてノード n が持つ確率（そのノードにランダムウォーカーが存在する確率）である。

【0047】

式 3 を繰り返し計算して定常状態となったときの $p_t(n)$ が、ノード n の定常状態でのノード確率 $p^{\text{stead}}(n)$ である。

【0048】

次に定常リンク確率計算部 144 は、各ノード n の定常状態でのノード確率 $p^{\text{stead}}(n)$ から、定常状態でのリンク確率を次の式 4 に従って計算する。

【数 5】

$$\tilde{p}_l^{\text{stead}} = T(\text{terminal end of link } l | \text{initial end of link } l) p^{\text{stead}}(\text{initial end of link } l) \quad \dots(\text{式4})$$

30

【0049】

リンク確率とは、ノード確率 $p_t(n)$ に対してそのノードから出るリンク l （エル）の遷移確率を乗じたものである。リンク l についての定常状態のリンク確率（式 4 の左辺）は、そのリンク l の起点のノード（initial end of link l ）の定常状態のノード確率に対して、遷移確率行列 T に含まれる、そのリンク l の起点ノードから終点ノード（terminal end of link l ）への遷移確率を乗じたものである。ここで用いる遷移確率行列 T は、S10 での仮想リンクの追加により拡張したものではなく、仮想リンク追加前の元の遷移確率行列 T である。このように、各ノードのノード確率は仮想リンクを含む拡張ネットワークを用いて計算するが、そのノード確率から求めるリンク確率は実リンク l についてのみ計算する。このようにして求めた実リンク l についてのリンク確率の総和は 1 にはならず、確率の要件を満たさない。

40

【0050】

50

そこで、次にリンク確率修正部 146 が、S12 で求められた各実リンク 1 の定常リンク確率を、学習データモデルの修正に応じて修正する (S16)。この修正は、次の式 5 の計算により行われる。

【数 6】

$$\tilde{p}_l^{\text{stead}} = p_l^{\text{stead}} / \sum_l p_l^{\text{stead}} \quad \dots(\text{式5})$$

10

【0051】

この修正の後の計算では、仮想リンクについての定常リンク確率は 0 として取り扱う。

【0052】

この後、これまで生成したデータを用いて、クラスタリング計算部 148 が、クラスタリングのための繰り返し計算を行う。この例では、特許文献 2 及び 3 に開示されたものと同様の、マルコフ連鎖のモジュール分解の計算手法を用いる。以下で説明するこの例での計算手法は、学習データの形式や計算の順序等が特許文献 2 及び 3 に開示されたものと一部異なるが、計算の基本的な考え方自体は特許文献 2 及び 3 のものと同じである。

【0053】

特許文献 2 及び 3 では、D 回の仮想的な観測で得られる観測データである通過情報 $\tau^{(d)}(k)$ (d は 1 から D までの整数。 k はクラスタ(コミュニティ)の番号) を学習データとして用いた。これに対して以下に説明する例では、観測回数 D が十分大きい(ノード数 N よりもはるかに多い)という妥当な想定の下、 $\tau^{(d)}(k)$ の代わりに実リンク 1 に関する通過情報

20

【数 7】

$$\tilde{\tau}_n^{(l)} = \delta_{n, \text{terminal end of link } l} + \delta_{n, \text{initial end of link } l} \quad (l=1, \dots, L). \quad \dots(\text{式6})$$

30

を用いる。ここで n はノードの識別番号である。また δ はクロネッカーの δ である。すなわち、式 6 が定義するノード n の実リンク l に関する通過情報(学習データ)は、そのノード n がその実リンク l の終点ノード又は起点ノードに一致する場合に値が 1 となり、それ以外の場合は値が 0 となる。

【0054】

また、この例では、特許文献 2 等での仮想的な観測の各回 d における複数のクラスタ(成分)全体に対するクラスタ k が占める割合 $y^{(d)}(k)$ の代わりに、実リンク 1 に関して後述する式 (I I I) で定義される割合 y_{lk} (チルダ付き)を用いる。

【0055】

また、このような観測回数 d から実リンクの番号 l への置換えにより、関数の総和の表現は以下のように置き換えられる。

40

【数 8】

$$\sum_{d=1}^D f(d) \rightarrow D \sum_{l=1}^L \tilde{p}_l^{\text{stead}} \tilde{f}(l) \quad \dots(\text{式7})$$

【0056】

後述する式 (I) の右辺第 2 項は、特許文献 2 等に説明した同様の式に対してこのよう

50

な置き換えを行ったものである。

【0057】

図3の手順の説明に戻ると、次にクラスタリング計算部148は、分類割合 $p_t(n|k)$ 及び重要度 π_k^{new} 、及び割合 γ_{lk} (チルダ付き) の初期値を仮決めし、繰返し回数のカウンタ g の値を0に初期化する (S18)。分類割合 $p_t(n|k)$ は、ノード n がクラスタ k に分類される割合 (度合い) である。また、重要度 π_k^{new} は、クラスタ k の重要度である。

【0058】

そしてクラスタリング計算部148は、次に示す式 (I)、(II)、(III) を用いて EM (Expectation and Maximization) の繰返し演算を行う。

【数9】

$$(I) \quad p_t(n|k) = \frac{\tilde{\alpha}}{\tilde{\alpha} + \pi_k} \sum_{m=1}^N \tilde{T}_{mn} p_{t-1}(m|k) + \frac{1}{\tilde{\alpha} + \pi_k} \frac{1}{2} \sum_{l=1}^L \tilde{p}_l \tilde{\gamma}_{lk} \tilde{\tau}_n^{(l)}$$

$$(II) \quad \pi_k = \sum_{l=1}^L \tilde{p}_l \tilde{\gamma}_{lk}$$

where

$$\tilde{\alpha} = \frac{\alpha}{2D}.$$

$$(III) \quad \tilde{\gamma}_{lk} = \frac{\pi_k \prod_{n=1}^N [p_t(n|k)]^{\tilde{\tau}_n^{(l)}}}{\sum_{k=1}^K \pi_k \prod_{n=1}^N [p_t(n|k)]^{\tilde{\tau}_n^{(l)}}}$$

【0059】

すなわちまずクラスタリング計算部148は、割合 γ_{lk} (チルダ付き) を、式 (III) を用いて計算する (S20)。この計算の最初の繰返しでは、S18で仮決めした初期値を用いる。

【0060】

次にクラスタリング計算部148は、現在の分類割合 $p_t(n|k)$ 及び重要度 π_k^{new} を一時刻前の値 $p_{t-1}(n|k)$ 及び重要度 π_k^{old} とする置き換えを行う (S22)。そして、式 (I) 及び (II) に従って、分類割合 $p_t(n|k)$ 及び重要度 π_k^{new} を計算する (S24)。ここで α は、正の実数であって、クラスタの大きさを定めるパラメータであり、予め定めた値を用いればよい。

【0061】

そして、クラスタリング計算部148は、繰返し計算の回数のカウンタ g を1インクリメントし (S26)、そのカウンタ g が予め定めた値 G に達したかどうかを判定し (S28)、達していなければ S20～S26 の処理を繰返す。値 G は、本実施形態の計算手法において S20～S24 の計算が収束するのに必要な繰返し回数であり、実験や経験的知識等により予め定めておく。仮想リンク (テレポーテーション) を追加する本実施形態の手法では、仮想リンクの追加を行わない従来手法 (例えば特許文献1～3) よりも繰返し計算の収束が早いので、 G の値として従来より小さい値を用いてよい。

【0062】

S28で、カウンタ g が値 G に達したと判定した場合は、繰返し計算が収束したものとして、処理を終了する。このときに得られた分類割合 $p_t(n|k)$ 及び重要度 π_k が、

10

20

30

40

50

特許文献 2 及び 3 に開示された処理の場合と同様、元のネットワークのソフトクラスタリングの結果を示す。

【0063】

なお、S 28 における収束の判定では、図 3 に例示した方法の代わりに、特許文献 1 ～ 3 で説明したの同様の、繰り返し毎の評価値 Q_t の変化量が微小な値（閾値未満）となったときに、繰り返し計算が収束したと判定してもよい。

【0064】

図 3 の手順は、特許文献 2 及び 3 の計算手法と同様の手法を用いたものであったが、仮想リンク（テレポーテーション）の追加という本実施形態の処理は、特許文献 1 の計算手法を用いる場合にも適用可能である。また、特許文献 1 ～ 3 の手法はいずれもネットワーク上でのノード間の確率の遷移を離散的なマルコフ連鎖として取り扱うものであったが、離散的なマルコフ連鎖の代わりに、連続時間マスター方程式等のような連続的なマルコフ過程をベースとした計算手法を用いてもよく、このような場合にも、仮想リンク（テレポーテーション）の追加という本実施形態の手法は利用可能である。

【0065】

以上、本発明の実施形態を説明した。以上に例示した情報処理装置 1 は、コンピュータに上述の各機能を表すプログラムを実行させることにより実現される。ここで、コンピュータは、例えば、ハードウェアとして、CPU 等のマイクロプロセッサ、ランダムアクセスメモリ（RAM）およびリードオンリメモリ（ROM）等のメモリ（一次記憶）、フラッシュメモリや SSD（ソリッドステートドライブ）、HDD（ハードディスクドライブ）や等の固定記憶装置を制御するコントローラ、各種 I/O（入出力）インタフェース、ローカルエリアネットワークなどのネットワークとの接続のための制御を行うネットワークインタフェース等が、たとえばバス等を介して接続された回路構成を有する。それら各機能の処理内容が記述されたプログラムがネットワーク等の経路でフラッシュメモリ等の固定記憶装置に保存され、コンピュータにインストールされる。固定記憶装置に記憶されたプログラムが RAM に読み出され CPU 等のマイクロプロセッサにより実行されることにより、上に例示した機能モジュール群が実現される。

【符号の説明】

【0066】

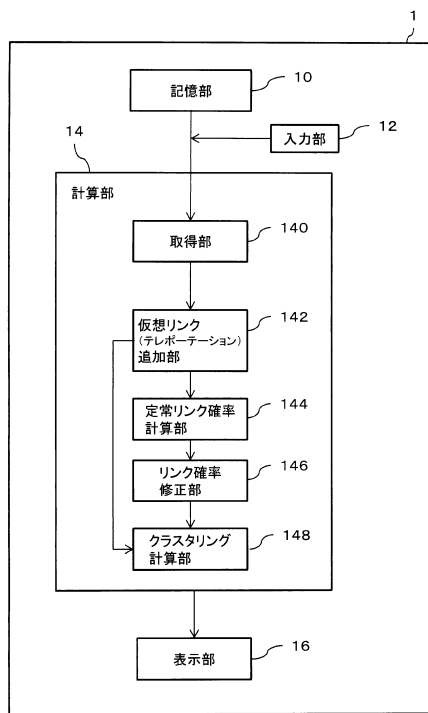
10 記憶部、12 入力部、14 計算部、16 表示部、140 取得部、142 仮想リンク追加部、144 定常リンク確率計算部、146 リンク確率修正部、148 クラスタリング計算部。

10

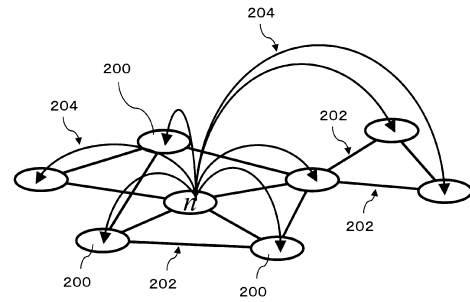
20

30

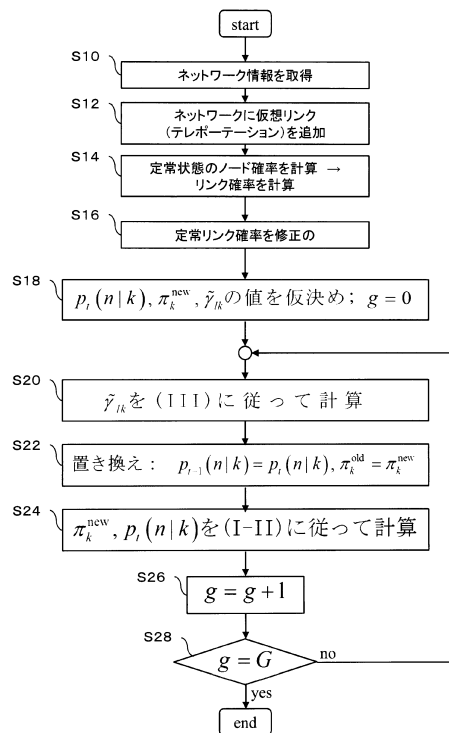
【図 1】



【図 2】



【図 3】



フロントページの続き

(56)参考文献 特開2013-003626 (JP, A)
特開2003-271669 (JP, A)
特開2016-029526 (JP, A)
米国特許出願公開第2003/0035429 (US, A1)

(58)調査した分野(Int.Cl., DB名)
G06F 16/00-16/958