氏名	すずき もとゆき 鈴木基之
授与学位	博士（工学）
学位授与年月日	平成16年3月　日
学位授与の根拠法規	学位規則第4条第2項
最終学歴	平成7年3月 東北大学大学院工学研究科電気及通信工学専攻 博士課程前期課程修了
学位論文題目	隠れマルコフ網を用いた時系列データのモデル化に関する研究
論文審査委員	主査　東北大学教授　牧野正三　東北大学教授　阿曽弘具 東北大学教授　鈴木陽一（情報科学研究科） 東北大学助教授　伊藤彰則

論文内容要旨

第1章　序論

古くから自然現象や社会現象を解明する手段のひとつとして，観測データのモデル化やクラスタリングが行われてきた。特に観測データが時系列(時間的な系列に限らず，1次元的に並んだもの)である場合，一般に系列長が不定であるために高精度なモデル化やクラスタリングが困難となる。そこで本論文では，時系列データのモデル化とクラスタリングのそれぞれについて問題点を指摘し，それらを統一的に解決する方法として隠れマルコフ網(Hidden Markov Network: HMnet) を用いて時系列をモデル化する方法を提案している。HMnet は多数の left-to-right 型隠れマルコフモデルの適切な状態を共有したもの，とみなすことができるため，高精度な時系列のモデル化が可能となる。また，局所的な類似性を反映したモデル化が可能となるため，その状態共有構造からクラスタを抽出することで高精度なクラスタリングが可能となる。

しかし，HMnet の状態共有構造を自動的に獲得するアルゴリズムである逐次状態分割法は，音声認識のための音響モデル学習用に開発されたため，任意の時系列を HMnet によってモデル化することはできない。そこで本論文では，第 2 章において逐次状態分割法を一般化し，任意の時系列データの HMnet によるモデル化を可能にする。また，一般化逐次状態分割法を用いた時系列データのモデル化やクラスタリングの有効性を確認するため，第 3 章以降において音声認識の各分野に適用し，それぞれの分野で従来から提案されているモデル化法やクラスタリング法に比べてよい性能を示すことを確認する。

第2章　逐次状態分割法の一般化

HMnet 構成法のひとつである逐次状態分割法(Successive State Splitting: SSS)は，鷹見らによって 1992 年に提案された。この方法は尤度最大基準で HMnet の状態共有構造を自動的に学習することが可能である。しかし SSS は音声認識システムで用いられる音響モデルの学習用として開発され

たため，そのまま任意の時系列データのモデル化に用いることはできない。そこで本章では SSS を一般化し，任意の時系列データを HMnet によってモデル化可能にする。

SSS は 1 状態の HMM を初期モデルとし，状態を逐次分割していくことで大規模な HMnet を自動学習するアルゴリズムである。このアルゴリズムのうち，どの状態を分割するかを決定するための評価値と，状態分割時に学習サンプルをどちらの状態へと割り当てるかを決定する関数を個別の問題ごとに定義することで，SSS を一般化した。また，これらを適切に定義することで，SSS や SSS の改良型のアルゴリズムがすべて一般化逐次状態分割法の枠組みで記述可能であることを示した。

第3章　音響類似性に基づいた音素 HMnet の構成法

音声認識システムでは，音響モデルのモデル化の単位として音素がよく用いられている。しかし音素はまわりの音素の影響によって音響的特徴が変形してしまうことが知られている。そこで従来から，音素環境別に HMM を学習する(音素環境依存モデルと呼ばれる)ことで，音響的特徴が大きく異なる音素を 1 つの HMM で表現することを回避し，音声認識率において有効性を示してきた。しかし，音素環境依存モデルを構築する時には，あらかじめどの環境要因について考慮するかを決定する必要がある。これが適切でなかった場合はよいモデルは得られないが，常に適切に環境要因を与えるのは非常に困難である。そこで本章では，一般化逐次状態分割法を用いて HMnet を構成することで，考慮する環境要因を与えずに音素環境依存モデルを構築する方法(SSS-free と呼ぶ)を提案する。SSS-free では音響類似性にのみ従って状態分割が行われるため，すべての環境要因を考慮したのと同等のモデルが獲得される。

男性 4 名，女性 6 名の発声した音声を用いて音響モデルを構成し，音声認識率で性能を評価したところ，従来の音素環境依存モデルに比べて平均で 21%程度誤認識率を削減することができ，SSS-free による音響モデルの有効性が示された。

第4章　離散型 HMnet を用いた言語モデルの自動獲得

音声認識等の分野では，日本語のモデル化法として *n*-gram を用いるのが一般的である。*n*-gram は(*n*-1)単語の条件付き確率で単語系列をモデル化しようとする方法であり，通常は *n* として 2 や 3 が選択される。そのため，あまり遠くに離れた単語同士の共起関係を表現することはできない。一方，HMnet を用いて言語モデルを構築した場合，HMnet は複数本のパスで構成されていることから，パスに分岐が存在しなければ遠く離れた単語同士の共起関係を表現することが可能である。そこで本章では，HMnet を用いて単語系列をモデル化することで言語モデルを構築し，*n*-gram に比べてよい性能が得られることを示す。

単語系列をモデル化する時は，HMnet は離散的な単語を出力するために SSS を用いて学習することはできない。そこで一般化逐次状態分割法を用いて離散型の HMnet を学習し，その性能を評価した。有限状態オートマトンによって生成される人工的な言語を対象とした実験では，HMnet の状態数を適切に選択することで *n*-gram よりもよい性能が得られた。またその時の HMnet の構造は，真の文法である有限状態オートマトンの構造によく似たものが得られた。更に新聞記事を学習サンプルとして用いて HMnet によるモデル化を行い，新聞記事の読み上げ音声の認識を行ったところ，*n*-gram を用いた音声認識システムに比べて 23%程度単語の誤り率を削減することができた。

第5章　離散型 HMnet における最適な状態数の自動決定法

前章において，離散型 HMnet は適切な状態数を選択することでよい言語モデルとなることを示した。しかしこの方法では，あらかじめ HMnet の総状態数を与えておく必要がある。そこで本章では適切な状態数を自動的に決定する方法を提案する。適切な状態数とは，記憶容量や計算時間等に制限

がなければ，単語の予測性能が最大になる状態数である。しかし一般に単語の予測性能を計算するためには未知サンプルが大量に必要となるため，直接計算によって求めるのは現実的ではない。そこで，学習サンプルのみから，未知サンプルに対する予測性能を推定する方法を提案する。

この方法では，学習サンプルの一部を隠して HMnet の学習を行っても，その構造や出力確率分布に変化はない，という近似を導入することで学習サンプルの一部を未知サンプルとみなし，それを用いて単語の予測性能を推定する。前章で用いた人工言語のモデル化実験において，予測性能の推定を行ったところ，実際の単語の予測性能とほぼ同じ推定値が計算され，その有効性が示された。また比較のため，従来から統計的モデルの規模選択によく用いられてきた情報量基準のひとつである最小記述長(Minimum Description Length: MDL)基準についても同じ実験を行ったところ，MDL は実際の予測性能とは異なる値を見せ，単語の予測性能の推定法が，より適切に最適な状態数を自動決定可能であることが示された。

第6章　不特定話者 HMnet の構造に基づいた話者間距離の計算法

音声認識システムにおいて話者による音響的な特徴の違いを吸収することは，重要な技術である。話者適応と呼ばれるこれらの技術では，話者の選択やクラスタリング等を行う時に話者間の距離がよく用いられている。しかし，これらで用いられている話者間距離の定義は様々であり，またその妥当性についても充分には検討されていない。ここで，SSS-free を用いて不特定話者による音声を HMnet を用いてモデル化することを考える。SSS-free は音響的な類似性に従って状態分割を行うため，獲得された HMnet の状態共有構造には話者による音響的特徴の類似性が反映されていると考えられる。そこで HMnet の状態共有構造から話者間距離を計算する方法を提案し，その有効性を示す。

HMnet において音響的な特徴が似ているサンプルは同じ状態を通る。そこで，各話者に対応するパスがどの程度状態を共有しているかに着目して，話者間距離の計算を行った。得られた話者間距離の性能を評価するため，目標とすべき話者間距離として音声認識率に基づく話者間距離を計算し，それとの相関係数を計算した。その結果，従来からよく用いられている母音中心間距離に比べて高い相関が得られ，HMnet の構造に基づいた話者間距離の有効性が示された。更に得られた話者間距離を用いて話者のクラスタリングを行い，クラスタごとに音響モデルを学習して音声認識率で評価を行ったところ，HMnet の構造に基づく話者間距離はより少ないクラスタ数で高い認識精度が得られた。

第7章　HMnet を用いた最適な認識単位の自動獲得

音声認識システムでは音響モデルの単位として音素がよく用いられている。しかし，統計的なモデル化という観点から考えると，必ずしも音素単位でモデル化することが最適であるという保証はない。一方，SSS-free を用いて音声をモデル化する場合，音響的な類似性に従って状態分割が行われるため，HMnet の状態共有構造に音声の音響的な類似性によるクラスタ構造が反映されていると考えられる。そこで本章では，SSS-free を用いて HMnet を学習しながら，その構造から音素にかわるモデル化の単位を自動獲得する方法を提案する。

HMnet の終了状態から開始状態へ戻る状態遷移を追加することで，音声を音響的な類似性に従ってクラスタリングする方法を提案する。この方法では，音声をクラスタに分割するステップと，分割された認識単位を用いて HMnet を学習するステップを交互に繰り返すことで，認識単位と HMnet の最適化を同時に行う。得られた認識単位は数音素程度の長さのものがほとんどであり，また話者によってその傾向に多少の差が見られた。この差は発声速度や発話のくせ等によるものだと考えられる。得られた認識単位を用いて音響モデルを学習し，音素単位の音響モデルと音声認識精度によって性能を評価したところ，平均で 3.5%程度の認識精度の向上が得られた。

第8章　結論

本論文では，任意の時系列データの高精度なモデル化やクラスタリング法として HMnet を用いた方法を提案した。HMnet の状態共有構造を自動獲得できるアルゴリズムである逐次状態分割法は，音声認識システムにおける音響モデルの学習用に特化してしまっているため，任意の時系列データを HMnet を用いてモデル化することはできない。そこで本論文では逐次状態分割法を一般化し，任意の時系列データのモデル化を可能にした。また，HMnet による高精度なモデル化法やクラスタリング法の有効性を確認するため，音声認識の各分野に一般化逐次状態分割法を適用し，それぞれの分野において従来から提案されている方法とくらべて，よい性能であることを示した。

従来から時系列データのモデル化やクラスタリングは非常によく行われていたが，それらはいずれも個別のケースごとにモデル化のアルゴリズムやクラスタリング法，またそこで用いられる距離尺度等を定義する必要があった。本論文で提案した HMnet による高精度なモデル化法やクラスタリング法を適用することで，今までモデル化等が困難であったケースにおいても高精度なモデル化等が行えるようになり，その結果，時系列データ解析の適用範囲が格段に拡がることとなった。今後，現在すでに専用のモデル化法やクラスタリング法が開発されている分野はもちろん，モデル化等が困難であった分野においても時系列データ解析が行われることで，新たな発見や発明等が行われることが期待される。

論文審査結果の要旨及び学力確認結果の要旨

論文提出者氏名	鈴木　基之
論　文　題　目	隠れマルコフ網を用いた時系列データのモデル化に関する研究
論文審査及び学力確認担当者	主査　教　授　牧野　正三　　教　授　阿曽　弘具 教　授　鈴木　陽一(情報科学研究科)　助　教　授　伊藤　彰則

論文審査結果の要旨

音声認識が種々の分野で実用に供されているが，さらなる精度の向上には準定常時系列データである音声信号の高精度なモデル化が必要である。しかし，準定常時系列データは，長さが不定であるため，従来方法では高精度なモデル化やクラスタリングが困難である。著者は，この問題の解決に取り組み，モデル化には隠れマルコフ網が有効なことを示し，さらに隠れマルコフ網の構築方法として一般化逐次状態分割法を提案した。また提案方法を音声言語処理の各分野に適用し，有効なことを示した。本論文は，その研究成果についてまとめたもので，全文8章よりなる。

第1章は序論であり，本研究の背景及び目的を述べている。

第2章では，準定常時系列データに対して，隠れマルコフ網が有効なモデルであることを示し，隠れマルコフ網の構築方法として一般化逐次状態分割法を提案している。この方法は，汎用性の高い方法と評価できる。

第3章では，音素モデルの構築に従来方法では問題があることを示し，提案方法を用いた隠れマルコフ網に基づく音素モデルがその問題を解決し，高精度なモデルが作成できることを示している。また認識実験により，従来方法より高い認識率が得られることを示している。これは実用上重要な成果である。

第4章では，音声認識に用いる言語モデルの構築に提案方法が有効なことを示している。与えられた認識対象の文から隠れマルコフ網に基づく言語モデルを自動的に構築できることを実験的に示している。この方法は言語モデル構築法として興味深い方法である。

第5章では，隠れマルコフ網の状態数が自動的に決定できることを示している。前章で構築した言語モデルの状態数決定に隠れマルコフ網の性質を利用する方法を提案し，情報量基準を用いた方法に比べ，より適切な状態数を決定できることを示している。

第6章では，隠れマルコフ網が不定長データのクラスタリングに利用できることを示している。話者間距離を隠れマルコフ網の共有状態に基づいて定義し，それに基づいて話者クラスタリングを行い，クラスタごとに音素モデルを構築し認識を行ったところ，従来方法より少ないクラスタ数で高い認識精度が得られた。この方法は不定長データに対するクラスタリング法として高く評価できる。

第7章では，隠れマルコフ網を利用して認識単位を自動獲得する方法を提案している。得られた認識単位を用いた場合，従来から用いられている音素を認識単位とする場合に比べ，高い認識精度が得られることを示している。

第8章は結論である。

以上要するに本論文は，音声信号に代表される準定常時系列データに対する高精度なモデル化やクラスタリング法として隠れマルコフ網が有効なことを示し，その構築法として一般化逐次状態分割法を新たに提案し，音声言語処理に適用してその有効性を示したものであり，音声情報処理工学および電気通信工学の発展に寄与するところが少なくない。

よって，本論文は博士(工学)の学位論文として合格と認める。

学力確認結果の要旨

平成15年12月4日，審査委員ならびに関係教官出席のもとに，学力確認のための試問を行った結果，本人は電気・通信工学に関する十分な学力と研究指導能力を有することを確認した。

なお，英学術論文に対する理解力から見て，外国語に対する学力も十分であることを認めた。

論文審査結果の要旨

音声認識が種々の分野で実用に供されているが，さらなる精度の向上には準定常時系列データである音声信号の高精度なモデル化が必要である。しかし，準定常時系列データは，長さが不定であるため，従来方法では高精度なモデル化やクラスタリングが困難である。著者は，この問題の解決に取り組み，モデル化には隠れマルコフ網が有効なことを示し，さらに隠れマルコフ網の構築方法として一般化逐次状態分割法を提案した。また提案方法を音声言語処理の各分野に適用し，有効なことを示した。本論文は，その研究成果についてまとめたもので，全文8章よりなる。

第1章は序論であり，本研究の背景及び目的を述べている。

第2章では，準定常時系列データに対して，隠れマルコフ網が有効なモデルであることを示し，隠れマルコフ網の構築方法として一般化逐次状態分割法を提案している。この方法は，汎用性の高い方法と評価できる。

第3章では，音素モデルの構築に従来方法では問題があることを示し，提案方法を用いた隠れマルコフ網に基づく音素モデルがその問題を解決し，高精度なモデルが作成できることを示している。また認識実験により，従来方法より高い認識率が得られることを示している。これは実用上重要な成果である。

第4章では，音声認識に用いる言語モデルの構築に提案方法が有効なことを示している。与えられた認識対象の文から隠れマルコフ網に基づく言語モデルを自動的に構築できることを実験的に示している。この方法は言語モデル構築法として興味深い方法である。

第5章では，隠れマルコフ網の状態数が自動的に決定できることを示している。前章で構築した言語モデルの状態数決定に隠れマルコフ網の性質を利用する方法を提案し，情報量基準を用いた方法に比べ，より適切な状態数を決定できることを示している。

第6章では，隠れマルコフ網が不定長データのクラスタリングに利用できることを示している。話者間距離を隠れマルコフ網の共有状態に基づいて定義し，それに基づいて話者クラスタリングを行い，クラスタごとに音素モデルを構築し認識を行ったところ，従来方法より少ないクラスタ数で高い認識精度が得られた。この方法は不定長データに対するクラスタリング法として高く評価できる。

第7章では，隠れマルコフ網を利用して認識単位を自動獲得する方法を提案している。得られた認識単位を用いた場合，従来から用いられている音素を認識単位とする場合に比べ，高い認識精度が得られることを示している。

第8章は結論である。

以上要するに本論文は，音声信号に代表される準定常時系列データに対する高精度なモデル化やクラスタリング法として隠れマルコフ網が有効なことを示し，その構築法として一般化逐次状態分割法を新たに提案し，音声言語処理に適用してその有効性を示したものであり，音声情報処理工学および電気通信工学の発展に寄与するところが少なくない。

よって，本論文は博士(工学)の学位論文として合格と認める。