

## 論文の内容の要旨

論文題目	Multiple Sequential Data Modeling with Extended Hidden Semi-Markov Models and Its Data Aggregation and Management (拡張隠れセミマルコフモデルによる複数系列データモデリングとデータ収集・管理手法)
学位申請者	成松 宏美

近年、デバイスの発展とデータ収集方式の発展に伴い、解析の対象となるデータや収集方式に変化が来ている。例えば、農業向けのサービスでは、農地に温度や湿度センサをおいて、センサ値をクラウドサーバに送ることで、農地の状態遠隔監視や水量制御の自動調整を行っている。ヘルスケア向けのサービスでは、スマートフォンとスマートウォッチやフィットネスデバイス等のウェアラブルデバイス等で、所有者の状態を計測、管理している。同様のサービスは多々あるが、システム構成及びデータの特性として共通していることがある。1つ目は、センサ等の複数のデバイスのデータを1つまたは複数のアクセスポイント相当の端末を経由してクラウドに送信され、管理されること。2つ目は、1つの端末に収集されたセンシングデータは、1つのグループとして意味をなす集合であることである。農地のセンシングで言えば、中継機がセンサからデータを収集し、一定期間毎にクラウドサーバへデータをアップロードする。また中継機の収集したデータは、同一農地内のデータであるという点でグループとしても意味のあるデータである。ヘルスケアにおいても、スマートフォンがウェアラブルデバイスのデータを収集し、その収集されたデータは、同一所有者のデータであるという点でグループとしても意味のあるデータである。

しかしながらデータの解析、収集の観点で次のような2つの課題がある。1つ目は、複数の系列データをグループとして捉えた解析はあまり想定されていなかったことである。時事刻々と収集されるデータは“系列”という特徴があり、系列データ解析には様々な手法が提案されている。それらの手法の多くは、1つの系列もしくは予め定められた複数系列を対象として提案されている。一方で、現在は、様々な系列データをグループとして収集することができるようになっており、グループ化された多種多様なデータを分析の対象としてモデル化することが求められている。2つ目は、収集用に予めアクセスポイント相当のデバイスを用意する必要があることである。安価なセンサに対して、中継機は高コストであり、設置コストの点で導入障壁が高いと言える。

そこで本研究では、グループ化された系列データの解析と、設置コストを必要

としない効率的な収集を目的として、以下の研究を行う。

#### 【グループ化された系列データの解析】

第2章では、グループ化された系列データの解析における要件の整理とベースとする手法について述べる。グループ化された系列データを想定すると、各センサが観測した生のデータではなく、そこから抽出されたイベント系列が重要であることが多い。また各センサからの抽象化されたイベントデータを抽出するには、それぞれに特化した解析手法が必要なことも多い。そこで、抽象化されたイベント系列に着目し、それらをグループとして扱う際の要件を整理する。導かれる要件は、①イベントの中身が何かとその並び、②イベントの継続時間、③イベントとイベントの間隔、④イベントの重複の4つの要件である。これらを満たすモデルを実現するため、要件を満たす従来手法を比較し、系列順と系列長を考慮可能な隠れセミマルコフモデル(HSMM)に着目する。HSMMに着目した過程と、HSMMにおける学習と認識の手法を述べる。第3章では、HSMMに対して系列間隔を扱えるようにする2つの手法を提案する。一つ目は、系列間隔を時間の間隔確率分布として表現するDuration and Interval HSMM(DI-HSMM)である。二つ目は、系列間隔を系列間隔であるという信号が観測されたものとして、系列長と同様に表現するInterval State HSMM(IS-HSMM)である。それぞれの手法における学習および認識手法を説明し、そのモデルの性能を評価する。さらに第4章では、第3章にて高性能を示したDI-HSMMを拡張し、イベントの重複も考慮できるように拡張する。Overlapped State HSMM(OS-HSMM)を提案し、評価実験によりその有効性を示す。

#### 【系列データの収集】

第5章では、系列データの収集について要件の整理および従来技術の紹介し、要件を満たす手法を提案する。系列データの収集における課題は、想定するセンサやデバイスとクラウドサーバとの中継機であるアクセスポイント相当のデバイスを予め用意する手間とコストである。それらを削減するために、データ収集に必要な役割を事前に端末を用意することなく仮想的に実現することを目指す。これに対し、収集の対象となるエリアに存在する移動端末のストレージを借用し、移動端末間でデータのリレーを行うことで実現する手法を提案する。ここでは、端末の負荷をできる限り削減するために、GPSを用いずにエリアを定義する。またリレー回数は端末の負荷にエリアはデータの蓄積時間に影響する。このトレードオフを解消するために、リレーのタイミング制御手法およびエリアのフレキシブルな制御手法を提案する。またシミュレーションによりベースラインの手法との比較を行い、手法の有効性を確認する。

最後に、第6章ではグループ化された系列データの解析と収集についてまとめ、本実現のさらなる発展の可能性と、他の分野との融合によって実現できることの可能性を述べる。

## 論文審査の結果の要旨

学位申請者氏名 成松 宏美

審査委員主査 笠井 裕之

委員 森田 啓義

委員 植野 真臣

委員 小川 朋宏

委員 大坐 晶 智

近年、デバイスの発展とデータ収集方式の発展に伴い、解析の対象となるデータや収集方式に変化が来ている。例えば、農業向けのサービスでは、農地に温度や湿度センサをおいて、センサ値をクラウドサーバに送ることで、農地の状態遠隔監視や水量制御の自動調整を行っている。ヘルスケア向けのサービスでは、スマートフォンとスマートウォッチやフィットネスデバイス等のウェアラブルデバイス等で、所有者の状態を計測、管理している。同様のサービスは多々あるが、システム構成及びデータの特性として共通していることがある。1つ目は、センサ等の複数のデバイスのデータを1つまたは複数のアクセスポイント相当の端末を経由してクラウドに送信され、管理されること。2つ目は、1つの端末に収集されたセンシングデータは、1つのグループとして意味をなす集合であることである。農地のセンシングで言えば、中継機がセンサからデータを収集し、一定期間毎にクラウドサーバへデータをアップロードする。また中継機の収集したデータは、同一農地内のデータであるという点でグループとしても意味のあるデータである。ヘルスケアにおいても、スマートフォンがウェアラブルデバイスのデータを収集し、その収集されたデータは、同一所有者のデータであるという点でグループとしても意味のあるデータである。

しかしながらデータの解析、収集の観点で次のような2つの課題がある。1つ目は、複数の系列データをグループとして捉えた解析はあまり想定されていなかったこと。時事刻々と収集されるデータは”系列”という特徴があり、系列データ解析には様々な手法が提案されている。それらの手法の多くは、1つの系列もしくは予め定められた複数系列を対象として提案されている。しかしながら、現在は、様々な系列データをグループとして収集することができるようになっており、グループ化された多種多様なデータを分析の対象としてモデル化することが求められている。2つ目は、収集用に予めアクセスポイント相当のデバイスを用意する必要がある点である。安価なセンサに対して、中継機は高コストであり、設置コストの点で導入障壁が高いと言える。

そこで本研究では、グループ化された系列データの解析と、設置コストを必要としない効率的な収集を目的として、以下の研究を行った。

【グループ化された系列データの解析】

グループ化された系列データの解析については、系列データを一度イベントにすることで、当該データのための汎用的な解析手法の実現を目指す。扱う対象は複数のセンサやデバイス等から取得される系列データであるが、観測された生のデータには、それぞれに特化した解析手法が必要なことが多い。しかしながら、グループ化する際に必要となるデータの観点で見ると、全ての生のデータは必要とされず、そこから抽出されたイベント系列だけで表現可能なこともある。抽象化されたイベント系列に着目し、それらをグループ化した複数系列を扱うモデルの要件を整理した。結果、①イベントの中身が何かとその並び、②イベントの継続時間、③イベントとイベントの間隔、④イベントの重複の4つの要件を導いた。4つの要件を満たすモデルを実現するため、本稿では隠れセミマルコフモデル (HSMM) に着目し、その手法を段階的に拡張することで、4つを満たすモデルを提案する。

1つ目の拡張では、系列順と系列長を考慮可能なHSMMに対して、系列間隔を扱えるようにしたIS-HSMMとDI-HSMMを提案、2つ目の拡張では、複数系列を扱う上で重要な、系列重複を扱えるようにしたOS-HSMMを提案した。各手法において、シミュレーションによりモデリング性能と認識性能を評価し、その有効性を確認した。

#### 【系列データの収集】

また系列データの収集にあたっては、想定するセンサやデバイスとクラウドサーバとの中継機であるアクセスポイント相当のデバイスを予め用意する手間とコストの削減を目指し、必要な役割を仮想的に実現することを目指す。センサやデバイスとクラウドサーバとの仲介役を担う中継機は、系列データの収集とグループ化、ある程度蓄積されたデータのサーバへのアップロードを行う。これらの役割を、対象とするエリアに存在するデバイスやモバイル端末の間で情報のやり取りと、当該端末中のストレージを借用することにより実現するアプローチを提案する。移動端末を用いる本アプローチでは、端末間の通信による端末への負荷発生、対象とするエリア内に端末が存在しなくなることによるデータの消失の可能性がある。これらを解決するために、リレータイミングの制御と対象エリアのサイズの制御を行う手法を提案する。シミュレーションにより、通信回数とデータ蓄積時間を評価し、手法の有効性を示した。

本論文で示す提案手法により、これまでに想定しなかったグループ系列やグループ系列からの新たなパタンの発見を生む可能性がある。また、グループ化された系列データの解析と収集の効率化に大きく寄与すると考えられる。以上から、本論文の成果は博士（工学）学位を授与されるに十分な成果であると判断する。