

論文要旨

Evaluation on Lifelog Analysis Application based on Data Quality Evaluation Framework and Proposal of Practical Realization on Cloud Environment.

(データ品質評価フレームワークを用いたライフログアプリケーションの評価とクラウド環境における実装の提案)

山下 暁香

近年、動画データを取得できるネットワークカメラ、加速度やGPS情報を取得できるスマートフォンといった様々なセンサを搭載した端末の小型化、高性能化により、実世界のデータを収集する事が技術的に容易になった。また、ストレージの大容量化やクラウドなどの普及により、個人が大量のストレージを持つ事が十分可能となり、低廉、もしくは無料で大量の情報を蓄積、検索、共有する事が可能になった。これらのデータ収集技術とストレージの発達により、人の行動をデジタルデータとして記録する「ライフログ」の実現は以前と比べ格段に敷居が低くなったと言え、これを受け、様々な種類のライフログ解析アプリケーションが開発されてきた。ライフログ解析アプリケーションとは、実世界で収集された様々なセンサデータに対して、各種解析処理を加える事によって、収集されたセンサデータを人の行動履歴や健康状態といった、人間にとって有用な情報に変換するシステムの事である。近年のデータ収集技術とストレージの発達により、様々な種類のライフログ解析アプリケーションが開発されてきたが、これらのライフログ解析アプリケーションに入力されるデータの品質については、あまり詳しく考慮されてこなかった。ライフログ解析アプリケーションの振舞いが異なれば、そのアプリケーションが要求するデータの品質も異なる。ライフログ解析アプリケーションに関する既存研究や関連研究においては、ライフログ解析アプリケーションの性質を考えず、とにかく高品質で大容量のデータを収集し、そのデータを蓄積し、処理するような手法が提案されてきた。これは、結果として、ストレージの圧迫、また、データ処理端末の処理速度の低下、更には、電力の浪費を引き起こす。しかし、本研究が提案するように、それぞれのライフログ解析アプリケーションが要求する入力データの品質が明らかになれば、収集データを効率的に蓄積する事が可能になる。どのような処理をするライフログ解析アプリケーションがどの程度の品質の入力データを必要とするのかという点を明確にすることは重要な課題である。本研究では、ライフログ解析アプリケーションにおける「データの品質」を詳細に考慮しているという点に特色があり、他の研究内容とは異なり独創的であると言える。

本研究の目的は、ライフログ解析アプリケーションにおけるデータ品質保証である。具体的には、センサ空間から吸い上げたデータを入力とし、データ処理の結果として、何らかの出力を返すライフログ解析アプリケーションにおいて、入力データの品質の差が出力結果に及ぼす影響、さらにライフログ解析アプリケーションが必要とする最低限の品質を定量的な指標で評価した。

本研究の評価実験では、実際にセンサ空間内の人の行動を言語化した情報をユーザに提供する「言語

化アプリケーション」を実装、改良し、この言語化アプリケーションをライフログ解析アプリケーションの代表例として利用し、シミュレーションではなく、実環境にて実機実験を行った。言語化アプリケーションは、動画データと加速度データを入力とし、画像処理によって人の行動を分別し、その結果を言語化してユーザへ出力するシステムである。出力の例として「人がドアを開けた」、「人が椅子に座った」、「人が冷蔵庫を開けた」などがある。この言語化アプリケーションの実用例としては、独り暮らしの高齢者見守りシステム、異常検知などが考えられる。

評価実験に用いた言語化アプリケーションの動作環境では、2つの異なる角度に設置された2台のネットワークカメラによって撮影された動画データの各フレームに対する画像処理と、ドアや椅子などの動きを伴う物体に取り付けた加速度センサ端末から取得された加速度データの変化量によって、人がドアを開けた時や椅子に座った時の物体の動きを検知する。データの処理方法としては、Bayesian Classifier と HMM (Hidden Markov Model) の 2 つの性質の異なる確率モデルを用いて比較、評価した。Bayesian Classifier は入力データのフレーム毎にデータ処理を施しているのに対し、HMM は入力データのある程度の長さの系列に対してまとめてデータ処理を施している。

データ品質評価実験の結果、ライフログ解析アプリケーションの入力データの品質が全て 100%と、最高の品質である必要がない事を示した。単位時間当たりの入力データ量の品質として、動画データが100%の品質であれば、加速度データは 30%で十分、また、動画データと加速度データ双方のコマ落ちの場合、それぞれ 70%のデータ品質で言語化アプリケーションの最低限の正答率が得られる事がわかった。また、Bayesian Classifier と HMM のデータ処理方法の違いにより、無線 LAN の通信品質が言語化アプリケーションの正答率に 60%の差が生じる事を明らかにした。無線 LAN のデータ品質評価実験では、センサ空間から収集した動画データを、アクセスポイント (AP) を通して無線 LAN で処理用 PC に送信する場合の到達データの品質と、その品質の動画データが言語化アプリケーションの正答率に与える影響を評価した。通信用の AP においては、メインの動画データの通信の帯域を奪う要因として Android 端末が複数台接続されており、動画データにパケットロスをもたらす。この無線 LAN 通信品質評価実験の結果として、背景端末である Android 端末が 4 台の時に、動画データのパケットロス率が 60%になり、この時、Bayesian Classifier で処理する場合の言語化アプリケーションの正答率が 60%、HMM では 0%になった。このことから、動画データのみといった、単一の入力データ品質が劣化する際は Bayesian Classifier を用いた方が入力データの品質劣化に強いという考察が得られた。