

ライフログ評価環境におけるアプリケーションの 要求条件に基づくデータ品質保証に関する一検討

小口研究室 学部4年 山下 暁香

1 はじめに

データ収集技術とストレージの発達により、ライフログの実現は昔に比べ格段に敷居が低くなったと言え、記録された一つの情報は各種データベース間の連携を通じてマッシュアップすることで様々な価値が生まれる。

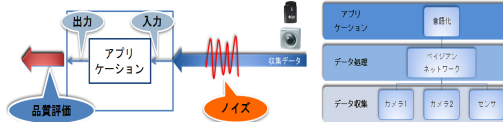


図1: 研究目的と具体的なフレームワーク

本研究では、入力データの品質がアプリケーションに与える影響を評価するために、お茶の水女子大学小林研究室で行われている「人の行動を言語化するアプリケーション」を用いて、評価実験を行う。

このアプリケーションに対して、次の3つを行った（図1）。

1. アプリケーション：言語化できる行動数の増加
2. データ処理：ベイズ判定の適用
3. データ収集：センサデータを同期

2 研究内容

2.1 アプリケーション

人の行動を言語化するアプリケーションとは、別の角度に設置された2台のネットワークカメラからの画像ストリーム情報とセンサデータの情報をベイズネットモデルの条件付き確率の大小比較によって、人の行動を言語化するものである。先行研究でされた、センサデータ同期前のアプリケーションの評価実験では、カメラノードの閾値（これを閾値1とする）を調節しても、言語化のタイミングが悪く、言語化の回数についても誤判定が多かった。

本研究では、既存研究から、言語化対象の行動数を増加させた。

2.2 データ処理

本研究では、収集データの処理の部分に関して、既存研究では実装されていなかったベイズ判定を適用し、データ処理を理論的に裏付けた。

ベイズ判定とは、条件付確率表（CPT）を用いて、結果から原因を予測する判定の事で、ここで言う結果とは、2台のカメラからの動画と物体に取り付けた加速度センサからのデータの「合計3つのノードの情報」、また、原因とは、3つのノードに影響を与えた「人の行動」という意味である。

具体的な判定方法としては、3つのノード結果に対して、どの「人の行動」が原因である確率が一番高いのかを計算し、最大値となった確率を原因である「人の行動」として出力する。この処理は動画データに対して、1フレーム毎にしている。

ただし、この判定式は結局 if-then ルールに帰着するので、現段階では物体毎に CPT を参考に if-then ルールで判定している。

2.3 データ収集

本研究では、誤判定が多かったアプリケーションに対して、センサデータノードを追加した。センサノードの立て方について説明する。まず、時刻順に並んだ加速度データを x 軸、y 軸、z 軸のそれぞれの値について2つの時刻毎に差をとった。この差を先ほどのカメラデータの閾値1に対して、閾値2とする。次に「人の行動が起こる」判定は、閾値2を何回連続して超えるかを数える。この連続数を閾値3とする。評価実験のためにセンサデータに品質の差をつけた。具体的には、

加速度センサ端末 SunSPOT によって1秒間に100回程度取得される加速度データをどの段階までコマ落としてするかをデータの品質とする。アプリの動画は10fpsなので、センサデータの最高の品質を10回/秒とする。この品質を10回/秒から1回/秒まで下げて行く。

以下の実験では10回/秒のデータを用いた。

3 評価実験と考察

閾値1: 10	閾値2\閾値3	1	2	3	4	5
	0.1	322	322	222	222	222
	0.2	222	222	222	222	222
	0.3	222	222	022	022	022
	0.4	122	122	022	022	022
	0.5	122	122	022	022	022
閾値1: 20	閾値2\閾値3	1	2	3	4	5
	0.1	121	121	121	121	121
	0.2	121	121	121	121	121
	0.3	121	121	021	021	021
	0.4	121	121	021	021	021
	0.5	121	121	011	011	011
閾値1: 30	閾値2\閾値3	1	2	3	4	5
	0.1	110	110	110	110	110
	0.2	110	110	110	110	110
	0.3	110	110	010	010	010
	0.4	110	110	010	010	010
	0.5	110	110	010	010	010

図2: 評価結果

2台のカメラでドア、机、いすを撮影し、加速度センサをドアと椅子に設置した環境において、「人がドアを開ける」「人が机を拭く」「人がいすにすわる」という言語化判定を行った。

3.1 閾値1=10 のとき

椅子にすわる時とドアを開ける時の加速度の変化ではドアの動作の方が速く、変化が大きくなるので、閾値2と閾値3を大きくしていくと、いすの言語化回数の方が先に減り始める。一方、ドアの言語化回数は一定となる。

3.2 閾値1=20 のとき

閾値2と3を大きくしていくと、いすの言語化回数は1から0に減少し、言語化されなくなった。閾値2が十分大きくなったときに、ドアの言語化回数が減少した。

3.3 閾値1=30 のとき

ドアの言語化回数は閾値1の時とき1回だけ減少し、その後は一定で変化していない。更に、閾値2と3をさらに大きくしていくと減っていく事が予想される。一方、椅子の言語化回数については、上と同じく、閾値3が3より大きくなると言語化されなくなる。

3.4 机の言語化回数

机の言語化回数については、加速度センサはあまり関与しないので、閾値1に大きく依存し、閾値1を増やしていく毎に、言語化回数が1回ずつ減少した。

4 まとめと今後の課題

センサデータの同期とベイズ判定の適用により、前アプリケーションに見られたタイミングのずれは解消できた。また、適切な閾値を与える事で、ある程度安定した回答率が得られた。

今後は、品質差のある画像データと品質差のあるセンサデータの両方を用いて評価実験を行っていきたい。また、実環境でデータ収集をし、そのデータについても評価実験を行いたい。