



Ling, B. and Intille, S. S. :

## Activity Recognition from User-Annotated Acceleration Data

International Conference on Pervasive Computing, pp.1-17 (2004)

### ウェアラブル行動認識とは

表題の論文は「ウェアラブル行動認識において、研究者が監視していない状況で被験者が自由に活動してデータを採取しながら被験者自身でアノテーションを付与する方法を考案し、自然な環境で採取したデータで行動認識精度を評価する」ことを最初に実証した論文である。「ウェアラブル行動認識」とは、種々のセンサを搭載したスマートフォンやウェアラブル（装着型）デバイスを用いて、時間や場所を問わず人間の行動をセンシングし、収集した加速度や角速度、脈拍、位置などのデータから、「歩行している」「タイピングしている」などの行動を認識する技術である。そのような技術やセンサデータは、健康管理や労働作業、医療、スポーツ、エンタテインメント、認証などの分野において、これまでは手作業や人間の勘、人海戦術で行われてきた作業の支援や、実現困難であった高度な支援への応用が期待されている。

センサデータから人間の行動を推定する処理は、正解の行動情報（アノテーション）が付与されたセンサデータのセットを用いてユーザの行動を解釈するモデルを事前に構築し、未知のセンサデータを識別する。そのため、センサデータの網羅性とアノテーションの正確性が重要となる。

### 実験室環境と実環境

本論文で問題提起しているのは、データの採取方法とアノテーションの方法の妥当性である。研究者が被験者と直接的に指示や監視をしてデータを採取する環境を実験室環境（Laboratory settings）と呼び、逆に研究者は帯同せずに被験者が自由に行動できる環境を実環境（Naturalistic environment）や非実験室環境（Non-laboratory settings）と呼んでいる。研究は実環境で採取されたデータによって評価されるべきであるが、多くの論文では実験室環境で行われる。

表-1 は本論文に掲載されている行動認識の関連研究のデータ採取環境である。データタイプの列には「L」（実験室環境）か「N」（実環境）が記載されており、ほとんどが「L」である。このなかで、F. Foerster らの論文では、実験室環境のデータを学習して別の実験室環境のデータを認識すると 95.8% の精度が得られたが、同じモデルで実環境のデータを認識すると 66.7% に低下したと報告している。この結果は、実験室環境のデータを学習しても実環境では正しく動かないことを示しており、実環境で採取したデータでモデルを構築し、評価すべきであることが分かる。

## アノテーションのジレンマ

自然なデータを採取するだけであれば、センサを装着した被験者を自由に活動させればよい。しかし、画像や音声、テキストは人間が理解できるため、データだけが記録されたとしても後からそれらを見てアノテーションを付与できるのに対し、加速度や地磁気を見てアノテーションを付与することはきわめて困難である。そのため、正確なアノテーションを採取するにはセンサデータの収集と同時に研究者や被験者、第三者が正解の行動を記録する必要がある。

アノテーションの採取方法は3つある。常時監視、自己申告、経験サンプリングである。常時監視は研究者が被験者を監視して行動を記録する方法で、正確なアノテーションを採取できるが、コストが高くスケールできない。自己申告は被験者が自分の行動を記録する方法で、複数の被験者から同時並行でデータを採取できるが、アノテーションの正確性は被験者次第である。経験サンプリングはメール等で定期的に被験者に行動を聞く方法で、行動認識に限らず幅広い分野で用いられているが、発生頻度の低い行動のアノテーションを得るには偶然その瞬間にサンプリングされるまで長期間のデータ採取が必要となる。

実環境の自然なデータが欲しいが、本当に自然だ

と真実を把握できないというジレンマがある。自己申告や経験サンプリングでは十分な質や量のアノテーションを収集できないため、多くの研究では常時監視によるデータ採取と研究者によるアノテーションが行われてきた。

## 被験者によるアノテーション

本論文では、被験者を監視せず自由に活動させつつも、被験者による正確なアノテーションを実現している。この方式を Semi-naturalistic と呼んでおり、完全な実環境ではないが、きわめて実環境に近い手法でデータ採取している。具体的には、認識対象の行動（本論文では20種類）を含むタスクを被験者に行わせ、その開始時刻と終了時刻を被験者に記録させる。タスクの指示は、たとえばPCを使っているデータが欲しい場合に「PCを使え」と指示するのではなく、「人口最大の都市をWebで検索しろ」というように、採取すべき行動に被験者が気づいて付度し、忠実にその動作を行うことのないようにしている。複数のタスクをこなすことで、最終的に20種類の動作を収集できるようにしている。データ採取中、研究者と被験者は接触せず、カメラで監視もしない。被験者はタスクをこなしてさえいれば自由である。ただし、開始と終了の記録時刻のずれ

表-1 関連研究のデータ採取環境（論文の Fig.1 より抜粋）

文献	精度	対象の行動	被験者数	データタイプ
S. W. Lee et al. 2002	92.85 ~ 95.91%	歩行	8	L
J. Mantyjarvi et al. 2001	83 ~ 90%	歩行, 姿勢	6	L
F. Foerster et al. 1999	95.8%	歩行, 姿勢, タイピング, 会話, 自転車	24	L
F. Foerster et al. 1999	66.7%	歩行, 姿勢, タイピング, 会話, 自転車	24	N
K. Aminian et al. 1999	89.3%	歩行, 姿勢	5	L
R. Herren et al. 1999	N/A	歩行速度, 傾斜	20	L
M. Uiterwaal et al. 1998	86 ~ 93%	歩行, 姿勢, 遊び	1	N
N. Kern et al. 2003	65 ~ 95%	歩行, タイピング, 階段, 握手, 筆記	1	L
G. S. Chambers et al. 2002	96.67%	3種のカンフー動作	1	L
K. V. Laerhoven et al. 2000	42 ~ 96%	歩行, 姿勢, 自転車	1	L
C. Randell et al. 2000	85 ~ 90%	歩行, 姿勢	10	L

は出るので、記録前後 10 秒のデータは捨てている。このようにして、きわめて実環境に近いデータを採取しつつも、正確なアノテーションを実現している。

20 人の被験者から採取したデータを用いて、個人特化学習（学習データとテストデータの被験者が同じ）と Leave-One-Subject-Out Cross Validation (LOSO-CV)（特定の被験者 1 名以外のデータで学習し、その 1 名の被験者のデータでテストする）で評価したところ、LOSO-CV の方が高精度であった。実験室環境のデータを用いた先行研究では、学習データとテストデータは同質であるため個人特化学習の方が高精度となることが常識とされてきたが、実環境のデータの場合、本人のデータのみを学習するよりも多くの他人のデータを学習したほうが汎化され、高性能であることを示している。このようにそれまでの常識を覆した点が本論文の最大の功績で

ある。その功績が称えられ、本論文は発表の 10 年後に開催された当該研究分野最大の国際会議である UbiComp 2014 において、10-Year Impact Award を受賞している。

(2018 年 10 月 29 日受付)

.....  
村尾和哉（正会員） [murao@cs.ritsumei.ac.jp](mailto:murao@cs.ritsumei.ac.jp)

2010 年大阪大学大学院情報科学研究科博士課程後期課程修了。博士（情報科学）。日本学術振興会特別研究員 PD，神戸大学大学院工学研究科助教，立命館大学情報理工学部助教を経て、現在、同准教授。ユビキタス・モバイル・ウェアラブルコンピューティングの研究に従事。

