

マルチドメイン行動情報の効率的な取り扱いに向けて

土井 裕介 大内 一成

概要：スマートフォンをはじめとして、身のまわりの生活機器のセンサ化により、個々人の生活に関するさまざまな行動情報が抽出可能となってきた。一方、現時点では多くの場合、行動情報取得のための専用のシステムに閉じており、本来開かれた生活空間が持っている、さまざまなベンダによるさまざまな種類の機器が雑多に存在する環境に対応できるものではない。このため、異なるベンダや方式による行動情報の合成により、より高い精度での生活行動の推測を行う、といった応用が困難である。本発表では、マルチドメイン行動情報のやりとりに必要な要求条件を議論し、仮説として XML、あるいはバイナリ XML(EXI) 技術をマルチドメイン行動情報の記述に適用する枠組について考察する。

1. はじめに

昨今、スマートフォン等による生体情報センシング [1] が実現し、行動情報の集積が行われようとしている。一方、集積される情報の形式についての議論は少ない。本稿では、行動情報の形式、特に、複数のセンサや应用到またがるような、マルチドメイン行動情報の集積に必要なデータ形式のあり方について、要求条件を整理する。

センサ情報による行動センシング、ならびに、行動情報にもとづくサービスの高度化については、さまざまな研究が存在する。ように、各種情報から最適な情報の提供を行うアプローチも具体化しつつある [2]。

一見、行動情報によるサービスが今にも爆発的に普及しそうに思えるが、実際にはこれらのプロジェクトは限定的な条件のみにおいて成立するものであり、多くの事業が単独で展開する状況であるとは言いがたい。これらには、さまざまな原因が考えられるが、そのうちのひとつとして、研究やプロジェクトのサイロ化に注目する。

行動情報の抽出においては、何らかのセンサ情報(位置情報や検索クエリなども含む)から対象となる人間の行動を推測する。これらの情報は、例えば研究であれば独自のセンサを製作したり、あるいはスマートフォンのような汎用デバイスに独自のプログラムをインストールして抽出することが多い。また、個々人の生活パターンなども含めた情報の分析を行う際には、行動履歴を含めた過去の情報との突き合わせを行う場合が多いが、こういった情報の蓄積についても同様に個々の研究プロジェクトにおいて独自のシステムを構築する場合が多い。

この結果、行動情報を用いたサービスの構築には多大なコストがかかり、必ずしも事業として成立しづらいものとなっている。一方で、ユーザへの適切な情報提示や、[3]に見られるようなエネルギーを含めた環境の最適制御を含め、行動情報の潜在的な価値は高いと予測する。

そこで、本稿では行動情報の蓄積・流通において、情報の「再利用」あるいは「共有」を促進するための仕組みについて、特にデータ形式に注目して問題を提起する。

2. 関連研究

ユビキタスコンピューティングにおいては、行動認識に限らず多数の研究が存在する [4]。特に Intelligent Environment と呼ばれる領域において、人の行動認識とその情報化は広く議論されている。しかし、これらの情報のやりとりについて、行動認識情報の交換や再利用についての議論は多くない。

Henricksen らは、コンテキストウェアなパーベインブコンピューティングを実現するためのソフトウェアフレームワークを提案している [5]。このフレームワークは Context Management Layer と呼ばれるモデルとレポジトリを中心としたプログラミングモデルである。

同様に Reichle らはコンテキストウェアネスを実現する、MUSIC(Self-Adapting Applications for Mobile Users in Ubiquitous Computing Environment) フレームワークを提案している [6]。興味深いのは、コンテキスト情報の交換として Exchange Layer の存在を明示的に示し、実体としてのデータに対して Contextual Metamodel(UML) および Ontology(OWL) との対照関係を定義することを明示している点である。

¹ 株式会社東芝 研究開発センター

以上の研究と本提案との相違は、以上の研究は実行環境まで含めたフレームワークの提案であるのに対し、本提案においてはデータそのもの、行動情報そのものに着目し、複数のドメインにおいてこの行動情報を流通させるためには、行動情報はどのような性質を持ち、どのような取り扱いが可能であるべきか、という点に注目している。

本提案の着想の元になった研究の一つに、HASC Challenge[7]が存在する。HASC Challengeにおいては特に加速度に関して、計測値とそこから抽象化した人間行動情報を蓄積して、行動情報のコーパスを作成している。このような抽象化を行ったデータの蓄積や交換について、加速度以外についてはどのように扱えば良いのか、そもそも複雑になりがちな、そして健全な競争のためには自由な拡張が可能であるべき行動情報の取り扱いにおいて、どのような形式的な制約条件を与えることで再利用性の高い、複数ドメインにまたがる効率的な情報利用が可能であるかを検討している。

3. 行動情報の表現形式に求められる条件

本節では、行動情報の共有・再利用を促進するための、行動情報の表現形式について議論を行う。

行動情報の共有・再利用を促進するためのデータ形式としてもっとも重要な性質は、当然のことながらデータの再利用性である。一方、データの再利用性は単一の特性ではなく、データの再利用を阻害するさまざまな性質を排除した結果としてあらわれるものであると考える。従って、ここでは、データの再利用を阻害する性質について整理し、うち、特に行動情報の表現において重要と思われる性質について詳細に述べる。

3.1 行動情報の再利用を阻害する性質

行動情報の再利用は、以下のような性質により阻害されると考えられる。

- 主観性: 特定の状況・観測者にとってのみ有効な情報である。測定条件が明示されていない。どこに取り付けたのか、あるいはセンサの型式などの情報が欠けている。
- 相対性の欠如: 情報が置かれるべき空間が定義されない。例えば、照度の測定値が「電圧」で示されるなど。
- 関係性の消失: 生成された行動情報がどのような測定値により生成されたのかが失われている。
- 非正規性: 情報表現が一定しておらず、情報の抽出・収集・計算に手間がかかる。
- 拡張性の欠如 (形式の硬直化): データ形式が拡張できず陳腐化してしまい再利用されない結果、過去の行動情報と現在の行動情報形式との間に断絶が生まれる。
- コスト: データに再利用性を与えるための開発部分が、本来機能からすれば「余分」になるため、再利用性を

与える形式を採用しない

- プライバシ: 利用者の行動情報やセンサデータは、利用者のプライバシーを直接的・間接的に侵害するため、再利用性が与えられない (本稿では論じない)

これらの性質をデータが持つてしまう理由にはさまざまなものが考えられる。一つには、個々のシステムを構築している研究者・技術者にとってデータ再利用のモチベーションが低い、という問題がある。これは、データ再利用が本質的に利他的な考え方であることから単純な方法による解決は難しい。一方、技術的な困難により再利用が阻害される場合については、解決が可能であると考えられる。技術的な理由には2つあり、1つにはセンサの出力のようなプリミティブなデータについて、例えば測定条件などのようなさまざまな付加情報を一つひとつ追記していくことは、機能的に意義が低い上に、性能的に困難である場合が多いことである。

もう1つには、そもそもセンサデータやこれを処理することにより作成した行動情報について、再利用性が高い状態を維持するための技術的な手間がかかること、あるいは、そのためのフレームワークが整備されていないこと、などといった理由が考えられる。これらの問題については、技術的に解決可能な課題であると言える。

また、ここでは特に論じないが、プライバシー等の問題が再利用を阻害する、という側面がある。これについては技術のみでは解決できないため、社会的・文化的側面も含めた考察が必要であるだろう。

3.2 測定条件・系列情報の付与

行動情報の再利用性を高めるためのアプローチの一つとして、測定条件や推測のための系列情報の付与を行う枠組が必要である。

特にセンサデータについて考えると、データの短期の変化だけではなく、長期にわたったトレンドの変化や、センサ自身の測定値のドリフト等、さまざまな問題が考えられる。また、一般にセンサは故障などにより交換され、また更新される。異なるセンサとなった場合は当然のことながら、仮に同一のセンサであっても個体差が発生する可能性がある。このような変化について、最低でも測定条件として残しておき、後から参照できるようにするための仕組みが必要である。

仮にこのような仕組みがなかったとすると、例えばセンサ切り替えの段階で急激な測定値の変化を観測してしまい、医療情報と連動させるなどとした時に誤った診断の元になる可能性がある。直近の情報であれば患者もセンサの交換を覚えているかもしれないが、例えばライフログ的な過去情報も含めての診断となった場合、一般の生活者が数カ月前のセンサ交換を申告することを期待することは困難である。

また、後に述べるように測定値から推測した情報自体についても同様の議論が成立する。例えば、センサから予測された行動情報が具体的にどのような計測情報から成立しているのかについて、後から知ることができれば、例えばアルゴリズムの改善や、問題が発生した際の原因の追求、影響範囲の推定などに用いることができる。著者はこれを「オブジェクトトレーサビリティ」と呼んでおり [8]、また、データベース領域では、Linage あるいは Provenance と呼ばれている [9]。

このような、行動情報本体ではなく、これに付随する情報（メタデータ）の取り扱いのフレームワークを構築する必要がある。

3.3 正規化・拡張性・効率性の両立

行動情報の蓄積においても一つの課題は、どのような形でその情報を表現するかである。特に、再利用性を高めるためにはある程度の情報の正規化が必要であり、特定のスキーマに情報を落とし込む必要がある。一方で、行動情報自体が発展途上の領域であること、および、生活の多様性などから、特定のスキーマで生活を表現することは困難であることが予想される。従って、行動情報のデータ表現においては拡張性が要求される。しかし、拡張性と正規化とはしばしば相反する概念となる。

例えば、時系列データはしばしば CSV の形式で保存、交換される。特定のアプリケーションや目的内での利用においてはこれは手軽で良い方式である。しかし、このような単純な表形式はスキーマ（カラムの意味や順番）の変更が困難である。かといって、個々の研究者や利用者が求めるがままにスキーマを更新してしまったり、データの再利用・共有の際に個々のデータスキーマのつきあわせを行わなければならない。形式的なスキーマにもとづくデータベース技術であればスキーマつきあわせを含めたモデル管理のための研究が存在する [10]。このような成果を利用するためには、データ自体にある程度のスキーマ情報を付随させる必要がある。ただし、特にセンサ情報はしばしば測定頻度が短かかったり測定期間が長かったりすることから大量になりがちであり、時系列情報を効率的に取り扱う方法をあわせもつ必要がある。

なお、いわゆるスキーマレスデータを取り扱う方が良いのではないかと、いう考え方が存在する。本稿では、スキーマレスデータ（CSV やログなど）はその当初の利用ドメインにおいては一様なデータ形式であることから明示的なスキーマ定義が不要であるに過ぎず、本稿が対照とするマルチドメインに向けてはデータ処理のためのロジックとデータが分離できないという問題を指摘し、また、再利用性を高めるためには CSV や Array DBMS 等であっても何らかのスキーマを外部的に定義できると主張する。

3.4 SKAI-Tree

さまざまなセンサから集められたデータは、合成・計算・抽象化を行い行動情報として蓄積される。このとき、再利用可能な情報は必ずしも最終的な行動情報だけではない。図 1 に例として著者らによる睡眠検出センサの場合の計算の流れを示した。このとき、最終的な目的である睡眠状態の検出のみならず、中間の計算結果も他の行動情報として再利用が可能である。

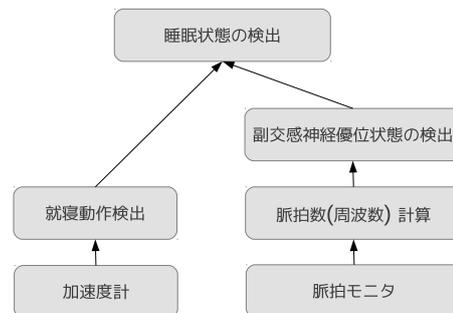


図 1 睡眠検出センサの計算プロセスの例

このような、情報の合成・計算・抽象化の流れについて、行動情報の枠組で再利用するためには一定の正規化が必要である。つまり、再利用を行いたいユーザが、他のアプリケーションで計算した中間結果を利用するためには、中間結果が識別可能であり、アクセス可能であることが求められる。

このような抽象化のプロセスは、木構造として捉えることができる。本稿では、このような木構造を SKAI-Tree (Sensing Knowledge Abstraction Instruction Tree) と呼ぶ。これは、データベースにおいて Provenance と呼ばれる構造を、特に行動情報分析の立場から明確化したものであり、センサ情報の抽象化プロセス自体の抽象化である。

行動情報の再利用性向上のためには、SKAI-Tree を取り扱うための何らかのフレームワークを構築する必要がある。

4. SKAI-Tree を含めた行動情報取り扱い試案

本節では、一つの試案として SKAI-Tree を含めた、行動情報の再利用性を高めるための取り扱い方針について論じる。なお、本節は特定のソフトウェアや実行モデルを提案するのではなく、データに対するラベル付けの方針について論じるに留める。

4.1 データスキーマの取り扱い

再利用性の高い行動情報とするためには、データの構造をある程度規定しつつ、自由に派生できる形にする必要がある。ここでは、表現力の高いスキーマ記述言語として、XML において用いられる W3C XML Schema [11] (以下 W3C Schema) を例として考える。

XML および W3C Schema には多くの批判があるが、形

式的なデータ定義において高い表現力と拡張性を持つスキーマ記述言語である。例えば、XML が持つ名前空間の性質を利用することで、安全な拡張が可能であること、および、型派生を行うことで後方互換性を維持したまま要素や属性の追加が可能であること、などである。特に、W3C Schema においてはさまざまなデータ型拡張方式が存在し [12]。大別すると、オブジェクト志向的な型の定義の拡張（継承）と、文書的な拡張要素の導入（ワイルドカード）の 2 つの方式が考えられる（図 2）。

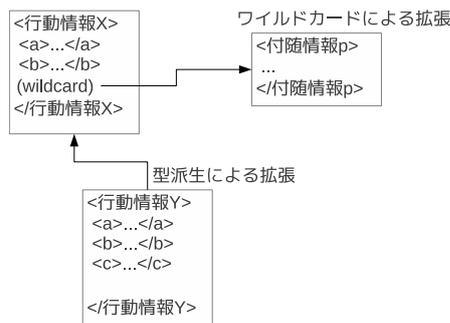


図 2 XML の拡張方式の例

ここで、再利用性を確保するための方策として 2 つの方策が考えられる。一つは再利用性を意図したデータは必ず特定のフォーマット（本稿では仮に SKAI フォーマットと呼ぶ）のエンベロープを付与し、エンベロープの中に置かれたワイルドカード部分に各々のアプリケーションの固有データを含める方式、もう一つは逆に、個々のアプリケーションデータにおけるスキーマ定義を拡張し、再利用性を意図したデータはここに SKAI フォーマットを含めるといった方式である。前者の方式は XML のルート要素が SKAI フォーマットのものになってしまうものの、単純なサブツリーの取り出し操作だけでデータが取り出せるという利点がある。後者はスキーマ自体を変更してしまうため、個々のアプリケーションが持つであろう本来のスキーマに射影するという操作が必要になる。

なお、SKAI フォーマットにはどのような情報を含めるべきか、という点については本稿では答を出さない。少なくとも HASC Challenge[7] にあるような測定条件を格納し、また図 1 に示した睡眠検出の例のように、計算結果に関する情報であれば、この計算/抽象化のデータソースに関する情報を含めることが望ましい。

4.2 行動情報交換方式

センサや、行動情報を出力する機器類、あるいはこれらの蓄積において、現在は独自データ形式や CSV などが多く使われている。また、仮に XML であっても明確なスキーマが与えられるとは限らない。一方、再利用性の高い行動情報を交換するためには、センサの入出力に至るまで何らかの形でスキーマが関連付けられており、第三者による検

証や取り扱いが可能であることが望ましい。

著者の研究のうちの一つに、Internet of Things における情報交換を XML のバイナリ形式の EXI により実現する方法に関するもの [13] がある。この方法では、W3C Schema と機器の情報入出力との組み合わせを事前に定義し、スキーマのうち機器が取り扱う部分のみをエンコーダ・デコーダとして生成することにより、最低限のリソースで動作する XML にもとづく情報交換を実現できる。

この方式を用いることで、SKAI フォーマットにもとづくエンベロープを持つデータの交換を実現できる。EXI の場合は、内部に含む行動情報自体が XML で符号化されていてもよいし、元がバイナリなので MessagePack^{*1}[14] のようなスキーマを持たないバイナリのデータをそのまま格納することもできる。また、仮にセンサ自体がこのような XML Schema により型付けされたデータを入出力しないとしても、センサのコンテキストを理解している中間ノードにおいて単純に SKAI フォーマットエンベロープを追加できる。

また、多量のデータの場合は、個々のエントリではなくデータレコードの系列自体に SKAI フォーマットを付与するなどの工夫が適すると考える。

5. おわりに

本稿では、議論の種として、行動情報をどのように蓄積し再利用するかという観点から行動情報が持つべき性質について論じ、既存のデータ交換技術のうち XML を中心とする技術について、このような性質をどのように満たすかについて述べた。特に、センサ情報から行動情報を算出する課程に多段階の情報が存在し、この中間情報の転用を考えると何らかの「抽象化のプロセスの言語化」が必要であることを主張し、これを SKAI-Tree と呼んだ。

本稿では特に XML を中心としたデータ表現技術のみについて論じた。しかし、実際にはこれを裏付ける SKAI-Tree のアクセス・管理手法や、非 XML 環境におけるデータの扱い、特に高いサンプリング周波数で蓄積される時系列データや、長期間蓄積される時系列データ等の扱いについても今後考える必要がある。また、行動情報にまつわる重大な課題として、プライバシー管理の問題が存在する。本稿を足がかりに、社会に受け入れられ、かつ、利用者の生活品質向上に資する行動情報のありかたについて検討を続けたい。

参考文献

- [1] 大内一成, 土井美和子: スマートフォンによる屋内外生活行動センシング, DICOMO2012 シンポジウム論文集, pp. 173 - 179 (2012).
- [2] 矢野幹樹, 岩崎陽平, 河口信夫: 駅.Locky: 無線 LAN 位

*1 <http://msgpack.org/>

- 置推定を用いた時刻表アプリの開発, 情報処理学会第 72 回全国大会講演論文集, pp. 289 – 290 (2010).
- [3] 久保田和人, 今原修一郎, 熊澤俊光, 田中俊明: 行動適応型省エネシステム BeHomeS における省エネシナリオと行動分析, DICOMO2009 シンポジウム論文集, pp. 1710 – 1716 (2009).
- [4] Endres, C., Butz, A. and MacWilliams, A.: A survey of software infrastructures and frameworks for ubiquitous computing, *Mobile Information Systems*, Vol. 1, No. 1, pp. 41–80 (2005).
- [5] Henricksen, K. and Indulska, J.: A Software Engineering Framework for Context-Aware Pervasive Computing, *Proceedings of the Second IEEE International Conference on Pervasive Computing and Communications (PerCom'04)*, pp. 77–86 (2004).
- [6] Reichle, R., Wagner, M., Khan, M. U., Geihs, K., Lorenzo, J., Valla, M., Fra, C., Paspallis, N. and Papadopoulos, G. A.: A comprehensive context modeling framework for pervasive computing systems, *Proceedings of the 8th IFIP WG 6.1 international conference on Distributed applications and interoperable systems* (2008).
- [7] Kawaguchi, N., Ogawa, N., Iwasaki, Y., Kaji, K., Terada, T., Murao, K., Inoue, S., Kawahara, Y., Sumi, Y. and Nishio, N.: HASC Challenge: Gathering Large Scale Human Activity Corpus for the Real-World Activity Understanding, *Proc. of ACM AH 2011*, pp. 27:1–27:5 (2011).
- [8] 土井裕介: 分散システム上のオブジェクトトレーサビリティに関する研究, 博士論文, 東京大学 (2011).
- [9] Simmhan, Y. L., Plale, B. and Gannon, D.: A survey of data provenance in e-science, *ACM SIGMOD Record*, Vol. 34, No. 3, pp. 31 – 36 (2005).
- [10] Bernstein, P. A. and Melnik, S.: Model management 2.0: manipulating richer mappings, *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, New York, NY, USA, ACM, pp. 1–12 (online), DOI: 10.1145/1247480.1247482 (2007).
- [11] Thompson, H. S., Beech, D., Maloney, M. and Mendelsohn, N.: XML Schema Part 1: Structures Second Edition, W3C Recommendation (2004).
- [12] Warmsley, P.: *Definitive XML Schema*, Prentice Hall (2001).
- [13] Doi, Y., Sato, Y. and Teramoto, K.: XML-Less EXI with Code Generation for Integration of Embedded Devices in Web Based Systems, *In Proc. of IoT2012 (The Third International conference on the Internet of Things)* (2012).
- [14] 古橋貞之: 分散システムのためのメッセージ表現手法に関する研究, 筑波大学大学院修士論文 (2012).