体験記録映像を用いたユーザ行動モデル作成の検討
Modeling of Life Patterns from Lifelog Movie Data

山城　貴久 † **平野 靖** ‡ **梶田 将司** ‡ **間瀬 健二** ‡
Takahisa Yamashiro Yasushi Hirano Shoji Kajita Kenji Mase

1. はじめに

計算機や種々のセンサの小型化によって、センサを搭載した情報端末を持ち歩き、センサデータを常に記録することが可能となりつつある。ユーザと行動を共にするセンサが記録したデータには、日常生活における行動パターンが反映されることが予想される。そのパターンからユーザの行動をモデル化できれば、駅に着く時刻に合わせた時刻表の提示のような、ユーザの行動を先読みしたサービスを実現することができる。このような体験記録に関する研究には相澤ら [1] や Clarkson[2] によるものがあるが、ユーザの行動モデルや行動予測に関する研究は十分に行われていない。

そこで、本研究ではまず、ある被験者 (大学院生) の 20 日間にわたる日常生活の記録を PDA(Personal Digital Assistant) を用いて収集し、体験記録映像データベースを作成した。次に、当該被験者の日常生活における主要な場所を 5 つ選び、それらの場所について音声・画像特徴量に基づいた識別実験を行うとともに、ユーザ行動をベイジアンネットによってモデル化し、行動予測実験を行った。

2. 体験記録映像データベースの作成

本研究では、カメラおよびマイクを備えた PDA を用いて体験記録映像の収集を行った。映像の収集条件を表 1 に示す。記録装置は、首から紐でぶら下げる形で固定した（図 1)。撮影は、大学内での一般的な活動の中で行い、平日 20 日間で行った。3 節で示す特徴量による場所識別実験でに用いるデータを取り出す手がかりとして、大学内で活動の中心となる 5 つの場所の出入り口で、撮影装置のコントローラのボタンを押してインデックスを記録した。

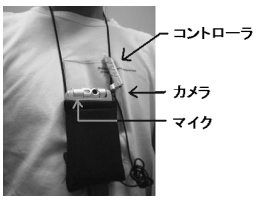

図 1: 記録装置固定状況

表 1: 体験記録映像の収集条件

映像フォーマット	MoviePlayer 形式 (MPEG4 準拠)
サンプリングレート	10fps
サイズ	160 × 112 ピクセル
音声フォーマット	MPEG4 オーディオ
サンプリングレート	24kHz（モノラル）

†名古屋大学大学院情報科学研究科社会システム情報学専攻
‡名古屋大学情報連携基盤センター

3. 音声・画像特徴量による識別実験

体験記録動画の特徴量から記録場所の識別をする可能性を調べるために、音声、画像の特徴量を用いてクラスタリング処理を行った。撮影時にインデックス付けを行った 5 つの場所（駅、研究室、生協、講義室、カフェ）の入口と出口（計 10 種類）における記録映像を入力として用いた。

3.1 色ヒストグラムによる識別実験

各場所で撮影されたフレームから、RGB の 3 色について階級数 16 の色ヒストグラムを求め、k-means 法によるクラスタリングを行った。出入り口を通過する瞬間のフレームを中心として、その前後 10 枚ずつのフレームをサンプル画像として色ヒストグラムの平均を求め、クラスタリングの初期重心とした。ヒストグラム間の距離としては、各スペクトルの差分の 2 乗和を用いた。また、サンプリング時間間隔 (St) を 100ms から 1000ms まで、100ms ごとに変化させ、10 種類の処理を行った。なお、初期重心の学習に用いたデータと同じデータをテストサンプルとして用いた。

その結果、St を 500ms とした時に最も正確に識別され、St が 500ms より小さい、または大きい場合には識別能力徐々に低下することがわかった。St を 500ms とした時の結果を表 2 に示す。

3.2 ケプストラムによる識別実験

体験記録映像の音声について、音声処理において一般的に用いられる特徴量であるケプストラムを計算した。各場所において、8192 点のサンプリングデータ（約 341ms）を用いてケプストラムを計算し、k-means 法によるクラスタリングを行った。ケプストラム間の距離としては、1 次以上のスペクトル差分の 2 乗和を用いた。なお、初期重心の学習に用いたデータと同じデータをテストサンプルとして用いた。

結果を表 3 に示す。ただし、音声の特徴量は同じ場所であれば、そこを通過する方向にはほとんど依存しないと考えられるため、出口と入口を同一クラスタとして扱った。

3.3 考察

今回の実験では、ヒストグラム、ケプストラムによる識別能力がそれぞれ 7 割程度であることがわかった。これは、同伴者の映りこみや発話状況、時間帯による明るさの違いなどによる例外的なデータが含まれるためだと考えられる。今後、複数の特徴量を相互補完的に用いることで、そのような例外に対して対応可能な識別方法を検討していく必要がある。

4. 確率的行動モデル作成および行動予測実験

次に、特徴量などによって得られるユーザ行動に関する情報を用いて、ユーザの行動をモデル化し、ユーザの

表 2: 色ヒストグラムによる識別結果 (St=500ms)

	識別結果												
	場所	Si	Li	SEi	Ci	Ki	So	Lo	SEo	Co	Ko	総数	正解率 (%)
正解	Si	6	0	0	0	1	1	0	0	0	3	11	54.6
	Li	0	38	4	1	1	0	9	0	13	1	67	56.7
	SEi	0	0	12	0	0	0	0	0	0	1	13	92.3
	Ci	0	0	1	4	0	0	0	0	0	0	5	80.0
	Ki	0	0	0	0	5	0	0	0	0	0	5	100.0
	So	1	0	2	0	0	6	0	1	0	1	11	54.6
	Lo	0	0	2	0	0	0	59	0	7	0	68	86.8
	SEo	2	0	0	0	0	3	0	8	0	1	14	57.1
	Co	0	0	0	0	0	0	1	0	5	0	6	83.3
	Ko	0	0	1	0	0	1	0	0	0	3	5	60.0
												全体	71.4

ただし、S は駅、L は研究室、SE は生協、C はカフェ、K は講義室を表し、添え字は i が入口、o が出口を表す。また、表中の左側の数字は計算機による識別の結果を、全体の正解率は、全正解数/全サンプル数を表す。表 3 も同様。

表 3: ケプストラムによる識別結果

	識別結果数							
		S	L	Se	C	K	総数	正解率 (%)
正解	S	18	2	2	0	0	22	81.8
	L	26	100	9	0	0	135	74.1
	Se	1	3	22	0	0	26	84.6
	C	0	3	1	7	0	11	63.6
	K	0	0	1	0	9	10	90.0
							全体	77.0

行動を予測することを検討した。センサ情報からユーザの状況を認識する技術としては、音声・画像特徴量を用いるものの他に、GPS、無線 LAN を用いた位置推定など様々なものが研究されているが、認識の不確実性を完全に無くすことは難しい。本研究ではこれらの技術から得られる不確かな知識を利用するために、ベイジアンネットによるモデルを仮定した。モデルのグラフを図 2 に示す。グラフ中の Time は時間、Activity は活動、Location は現在の場所、NextLocation は次に向かう場所、Day は曜日をあらわす。なお、5 つの場所と活動には表 4 のような対応関係がある。

4.1 確率推論

Location、Time、Day ノードの状態が条件として与えられた時、モデルから NextLocation の状態に関する確率分布が得られる。NextLocation が i である確率 $P(Ni)$ は、式 (1) によって求められる。ただし、Ni は NextLocation の状態、L は Location の状態、D は Day の状態、T は Time の状態を表す。

$$P\{Ni\} = \sum_k [P\{Ni|A_k, L, D\} \times P\{A_k|L, T\}] \quad (1)$$

4.2 行動予測実験

各ノードの条件付確率表 (CPT) を 15 日分のデータから計算し、5 日分のデータに対して最も確率の高い NextLocation の状態を推論し、実際の行動との比較実験を行った。その結果、正解率は 63.2%であった。

全体の予想精度は十分とは言えないが、月曜日の講義の後にはカフェに行くといった週間の講義等の予定に強く依存した行動については、高い確率で正解が得られた。今回使用したデータでは各曜日について、学習データ 3 例、テストデータ 1 例しか含まないため、例外的な行動の影響が強く出てしまった。今後、体験記録映像をさらに収集することで、予想精度は向上すると考えられる。

表 4: 大学内の場所と活動の関係

場所	活動
研究室	研究活動 ミーティング
生協	昼食 買い物 (雑貨)
駅	登校 帰宅
講義室	講義
カフェ	買い物 (パン)

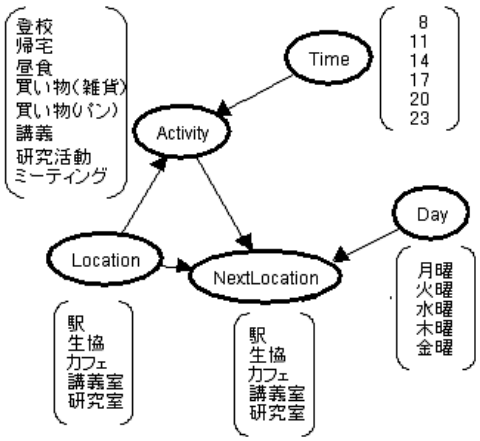

図 2: ベイジアンネットによる行動モデル

5. まとめ

本研究では 20 日間にわたって体験記録映像の撮影を行った。次に画像・音声特徴量を用いて撮影場所を識別することを検討した。その結果、ヒストグラム、ケプストラムを用いた方法ではそれぞれ 71.4%、77.0%の識別能力があることがわかった。さらに、ユーザ行動をベイジアンネットを用いてモデル化し、行動予測実験を行った結果、正解率は 63.2%であった。

今後、体験記録データのさらなる蓄積、特徴量による場所識別精度の向上、行動モデルの改善、さらに行動予測に基づくアプリケーション開発などを行っていく予定である。

謝辞

本研究の一部は，文科省 21 世紀 COE プログラム「社会情報基盤のための音声映像の知的統合」によった.

参考文献

[1] 相澤清晴, 石原健一郎, 椎名誠, “ウェアラブル映像の構造化と要約:個人の主観を考慮した要約生成の試み”, 電子情報通信学会論文誌, Vol.J86-D-II No.6 pp. 807-815,2003

[2] Brian Patrick Clarkson “Life Patterns: structure from wearable sensors”,Ph.D thesis,MIT MediaLab,September,2002