個人認証を見据えた位置情報による識別に関する解析

石井智也 †　鈴木宏哉 †　山口利恵 †　中山英樹 †　山西健司 †

† 東京大学大学院情報理工学系研究科
113-8656 東京都文京区 本郷 7 丁目 3-1
ishii.tomoya@ci.i.u-tokyo.ac.jp susuki.hiroya@sict.i.u-tokyo.ac.jp
yamaguchi.rie@i.u-tokyo.ac.jp nakayama@ci.i.u-tokyo.ac.jp
yamanishi@mist.i.u-tokyo.ac.jp

あらまし スマートフォンを始めとした GPS 機能付きモバイル端末の普及により, 位置情報を活用した機械学習技術が盛んに研究されている. また, なりすましによる不正アクセス被害は現代において深刻な問題となっており, よりセキュアな個人認証技術の確立が重要となっている. 本研究では位置情報から抽出した特徴を用いた Support Vector Machine による個人識別手法を提案する. 実験では提案手法を用いた個人識別の精度を, 実際に GPS によって取得された位置情報を使用して評価し, 得られた結果から個人認証における要素の一つとして位置情報を活用することに関する検討を行った.

Analysis of individual identification using location information for personal authentication

Tomoya Ishii†　Hiroya Susuki†　Rie Shigetomi Yamaguchi†
Hideki Nakayama†　Kenji Yamanishi†

†Graduate school of information science and technology, The University of Tokyo.
7-3-1 Hongo, Bunkyou, Tokyo 113-8656, JAPAN
ishii.tomoya@ci.i.u-tokyo.ac.jp susuki.hiroya@sict.i.u-tokyo.ac.jp
yamaguchi.rie@i.u-tokyo.ac.jp nakayama@ci.i.u-tokyo.ac.jp
yamanishi@mist.i.u-tokyo.ac.jp

Abstract Along with the widespread of GPS-enabled devices, location data are extensively available and a variety of machine learning methods have been developed to cope with them. Recently, unauthorized access with spoofing becomes a serious problem, which makes it necessary to establish a secure personal authentication method. In this paper, we propose an individual identification method using Support Vector Machine with features extracted from location information. We employ a real-world dataset to empirically validate our method and discuss on the effectiveness of secure personal authentication that exploits location information.

1　はじめに

スマートフォン等の携帯端末の普及が進んでいる. 例えば, 総務省の報告 [8] によると, 国内において 6 割を超える世帯にスマートフォンが普及していることが明らかとなっている. スマートフォン等の端末には GPS 機能を搭載しているものが多く, 個人の位置情報の取得は容易になっていると考えられる. また, 人の訪れる場

所は居住地や職場の所在地等の要素が影響するため, 個人ごとに異なる特徴を持つと考えられる. 従って, 個人ごとに取得された位置情報の持つ特徴を用いて, 機械学習によって個人を識別することができれば, 個人認証へ応用できる可能性がある.

本研究では, 位置情報を個人認証の要素として活用する事を目的として, 機械学習における識別アルゴリズムの一つである Support Vector Machine を用いた個人識別手法を提案する. 実験においてカーネル密度推定を用いた識別手法と識別の精度を比較する. また, 抽出した特徴量間の識別精度の違いを調査する.

本論文の構成は以下のようになっている. 2 章で本研究が想定している問題設定について説明し, 3 章で提案手法について述べる. 4 章で行った個人識別の実験と結果について述べ, 5 章で実験結果に対する考察を述べる. 6 章で関連研究について説明し, 7 章でまとめと今後の課題について述べる.

2 問題設定

N 人のユーザ $u_1, ..., u_N$ が存在し, ユーザ毎に位置情報ログが記録されているとする. i 番目のユーザ u_i から日付 d で取得された位置情報ログを $L_d^i = \{\boldsymbol{l}_{d,1}^i, ..., \boldsymbol{l}_{d,n}^i\}$ と表す. 但し, 各 $\boldsymbol{l}_{d,j}^i \in L_d^i$ は $\langle$ 緯度, 経度 $\rangle$ で表された 2 次元のデータ点であり, 訪れた順にソートされているものとする. u_i は上記の形式で表された位置情報ログを $|L^i|$ 日分記録しているものとする. ただし, L^i は u_i の所持しているログの集合を表し, $|L^i|$ は u_i の所持しているログの総数を表すものとする. また, 全ユーザのログの集合を L で表し, その要素数を $|L|$ と表す.

今回扱う問題は, u_i から新たに収集された位置情報ログ $L_{d'}^i = \{\boldsymbol{l}_{d',1}^i, ..., \boldsymbol{l}_{d',n'}^i\}$ が与えられた際に, そのログが u_i から取得されたログであるかどうかを正しく識別することである.

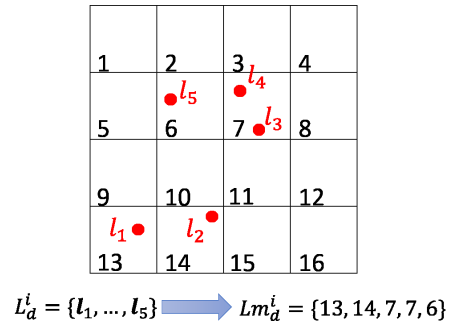

図 1: メッシュ番号への変換

3 提案手法

前節で述べた問題設定に対し, 各ユーザの位置情報ログを特徴ベクトルで表現し, 機械学習を用いてユーザの識別を行う手法を提案する. まず初めに位置情報ログからの特徴抽出について説明し, その後機械学習を用いたユーザの識別を行う方法について述べる.

3.1 位置情報ログからの特徴抽出

まず, 緯度と経度で表された空間をメッシュに分割し, 各メッシュに番号を与える. メッシュ番号の集合は M と表し, j 番目のメッシュの番号は M_j と表す. また, メッシュ番号の総数を $|M|$ と表す. その後, 位置情報ログ内のデータ点をメッシュ番号へと変換する. ある位置情報ログ $L_d^i = \{\boldsymbol{l}_{d,1}^i, ..., \boldsymbol{l}_{d,n}^i\}$ が与えられた時, メッシュ番号に変換した新たなログを $Lm_d^i = \{m_{d,1}^i, ..., m_{d,n}^i\}$ と定義する. ただし, 各 $m_{d,k}^i \in Lm_d^i$ は $\boldsymbol{l}_{d,k}^i \in L_d^i$ を含むメッシュの番号を表す. $|M| = 16$ のときの, 位置情報ログ内のデータ点をメッシュ番号へ変換する際の例を図 1 に示す. なお, 図 1 中の各矩形を一つのメッシュとし, 左下の番号がメッシュ番号を表す. また, 赤丸の点は図 1 中の L_d^i に記録されたデータ点の位置を表すものとする.

本手法では Lm_d^i を, メッシュ番号 $m_{d,k}^i \in Lm_d^i$ を単語とした一つのドキュメントとみなし, 自然言語処理分野においてドキュメントに対して一般的に用いられている特徴量である, Binary Feature(BF), Bag of Words(BoW), TF-IDF 特徴量をそれぞれ使用する. BF はドキュメント内にどの単語が現れたかという情報のみを用い

ており, BoW は単語の出現回数も使用した特徴量である. ログ Lm_d^i に対して BF, BoW を用いた際の特徴ベクトル $\boldsymbol{v}_{BF}$ と $\boldsymbol{v}_{BoW}$ はそれぞれ

$$\boldsymbol{v}_{BF}(j) = \delta(n_{d,j}^i > 0) \quad (1)$$
$$\boldsymbol{v}_{BoW}(j) = n_{d,j}^i \quad (2)$$

を満たす $|M|$ 次元ベクトルとなる. ただし, ベクトル $\boldsymbol{v}$ に対して $\boldsymbol{v}(j)$ は $\boldsymbol{v}$ の j 次元目の値を表すものとする. また, $n_{d,j}^i$ は Lm_d^i に含まれる M_j の個数, δ は括弧内の条件が満たされた場合に 1 を, それ以外の時は 0 を返す関数である. 例として, 図 1 中のログに対し作成した特徴ベクトル $\boldsymbol{v}_{BF}$ と $\boldsymbol{v}_{BoW}$ は

$$\boldsymbol{v}_{BF} = \{0,0,0,0,0,1,1,0,0,0,0,0,1,1,0,0\}$$
$$\boldsymbol{v}_{BoW} = \{0,0,0,0,0,1,2,0,0,0,0,0,1,1,0,0\}$$

となる. TF-IDF では, あるドキュメントには多く出現するが他のドキュメントには多く現れない単語は重要であるとして, そのような単語に対して高い重み付けを行う特徴量である. ログ Lm_d^i に対して TF-IDF を用いて作成した特徴ベクトル $\boldsymbol{v}_{TF-IDF}$ は

$$\boldsymbol{v}_{TF-IDF}(j) = \mathrm{tf}_{\mathrm{d,j}}^{\mathrm{i}} \times \mathrm{idf}_{\mathrm{j}} \quad (3)$$

を満たす $|\boldsymbol{M}|$ 次元ベクトルとなる. ただし

$$\mathrm{tf}_{\mathrm{d,j}}^{\mathrm{i}} = \frac{n_{d,j}^i}{n} \quad (4)$$
$$\mathrm{idf}_{\mathrm{j}} = \log \frac{|L|}{|\{(k,d')|M_j \in Lm_{d'}^k\}|} + 1 \quad (5)$$

であり, $|\{(k,d')|M_j \in Lm_{d'}^k\}|$ は M_j を含むログの総数を表す.

今回は更に, ログの先頭と末尾のメッシュ番号はユーザ毎の特徴が現れると推測し, これらのメッシュ番号も特徴として使用する. 本論文ではこの特徴量を First-Last(FL) 特徴量と呼ぶ. ログ Lm_d^i に対して FL 特徴量を用いた際の特徴ベクトル v_{FL} は

$$\boldsymbol{v}_{FL}(j) = \delta(m_{d,1}^i = M_j) \quad (6)$$
$$\boldsymbol{v}_{FL}(|M|+j) = \delta(m_{d,n}^i = M_j) \quad (7)$$

を満たす $2|M|$ 次元ベクトルと定義する.

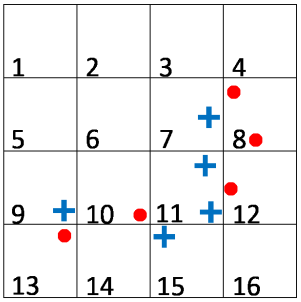

図 2: 境界付近のデータによる問題

3.2 特徴量の平滑化

メッシュを用いて位置空間を表現した場合, メッシュの境界付近に存在するデータ点が問題となる場合がある. 例を図 2 に示す. 図中の赤丸の点と, 青十字の点はそれぞれある位置情報ログに記録されたデータ点であるとする. 両者のログはデータ点の近さを考慮すると類似性が高いと見られるため, 似たような特徴ベクトルで表現されることが望ましい. しかし, 前節で述べた特徴量では, このような位置的な近さという要素は考慮できていない.

この問題に対処するため, 特徴量の平滑化を行う. BF, BoW, TF-IDF 特徴量の平滑化を行う際には, Lm_d^i に含まれる各メッシュに地理的に隣接した 8 個のメッシュの番号を加えた新たなログ $LmSmt_d^i$ を作成し, $LmSmt_d^i$ に対して特徴ベクトルを作成することで平滑化を行う. また, FL 特徴量の平滑化を行う際には, ログの先頭と末尾のメッシュに地理的に隣接した 8 個のメッシュも先頭, 末尾のメッシュとして取り扱うことで平滑化を行う.

3.3 機械学習によるユーザの識別

特徴ベクトルで表現された位置情報ログに対し, 機械学習における識別アルゴリズムを用いることでユーザの識別を行う. 本手法ではそのうちの一つである Support Vector Machine(SVM)[4] を使用する. SVM はベクトル空間上のデータを分離する超平面を学習するアルゴリズムである. 高次元のデータに対しても識別精度が高いことが知られており, 今回使用する特徴量は高

表 1: 使用したデータセット

ユーザ数 (人)	169
位置情報ログ総数	6963
ログ長平均	38.25

次元となりうるため, 本手法では SVM を選択した.

4 評価実験

提案手法による識別精度の評価実験と, その結果について述べる. 始めに使用したデータセットについて述べ, その後に行った実験の手順について説明を行う.

4.1 データセット

本実験では Microsoft より提供されている GeoLife GPS Trajectories[5, 6, 7][1] を使用した. このデータセットは複数のユーザから GPS ロガーによって取得された屋外での移動履歴である. なお, 元のデータセット内には異なるユーザ間で全く同一の位置情報履歴を所有している場合があったが, その際はユーザ番号の大きい方からその位置情報履歴を事前に削除した. その後, 北京市内の緯度 39.669064 以上 40.306683 以下, 経度 116.010534 以上 116.774084 以下の範囲外のデータ点は事前に削除した. また, 実験には元の移動履歴から 180 秒間隔で抜き出したデータ点を使用し, ある一日の中で取得されたデータ点は一つの位置情報ログ内にまとめた. 上記の前処理を行った後, 各ユーザそれぞれに対して記録されたログが最も多い年を調べ, その年の中でのログのみを選出した. 上記の処理を行ったあとの, データセット内のユーザ数, ログの総数, ログの長さの平均について表 1 に示す.

4.2 実験 1:識別手法の比較

提案手法による識別の精度をベースラインの手法と比較する. 今回はベースラインとしてカーネル密度推定を使用した. まず始めに, 比較手法を用いた識別について説明し, その後実験の具体的な手順を述べる.

4.2.1 カーネル密度推定

カーネル密度推定 (KDE) は, 観測したデータ点を用いて新たなデータ点が出現する確率値を推定するための手法である. 今回, KDE を用いた比較手法では, u_i が位置情報ログ $L = \{\boldsymbol{l}_1, ..., \boldsymbol{l}_n\}$ を生成する確率値 $f_i(L)$ を

$$f_i(L) = \prod_{\boldsymbol{l} \in L} \frac{1}{|L_i'|h^2} \sum_{\boldsymbol{l'} \in L_i'} \frac{1}{\sqrt{2\pi}} \exp(-\frac{\|\boldsymbol{l} - \boldsymbol{l'}\|_2^2}{2h^2})$$

と推定した. 但し, L_i' は u_i から既に観測された全てのデータ点の集合, h はバンド幅パラメータである.

今回の実験において KDE を使用して与えられたログに対するユーザの識別を行う際には, そのログに対して最も高い確率値を出したユーザとして識別した.

4.2.2 実験方法

データセット内に 14 個以上のログを所有しているユーザを選出し, 各ユーザから日付が古い順に 14 ログを抜き出して実験に使用した. また, 位置空間をメッシュに分割する方法として Geohash[2] を使用した. Geohash は緯度, 経度の組を一定の大きさのメッシュに対応したハッシュ値へと変換するアルゴリズムであり, ハッシュ値の桁数によってメッシュの大きさを変更することができる. 今回は 7 桁のハッシュ値を用いた. 実験に使用したデータのユーザ数, 位置情報ログの総数, データ点の総数, ログの長さの平均, ログに含まれていたデータ点の属するメッシュの種類の総数を表 2 に示す. また, ある一つの位置情報ログに対し, ログ内のデータ点が属するメッシュを地図上に表示した際の図を図 3 に示す. 図 3 内の青のマーカーはログに記録されていたデータ点であり, 赤い矩形が Geohash によるメッシュを表す. さらに, 今回

[1] http://research.microsoft.com/en-us/downloads/b16d359d-d164-469e-9fd4-daa38f2b2e13

[2] http://geohash.org

表 2: 実験に使用したデータ

ユーザ数 (人)	91
位置情報ログ総数	1274
データ点総数	53459
ログ長平均	41.96
メッシュ種類総数	10459

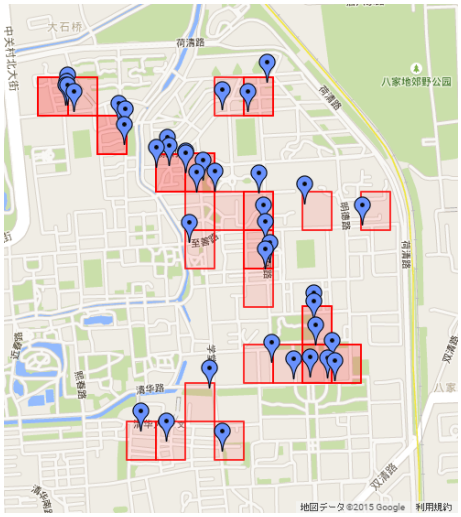

図 3: Geohash でのメッシュ

はデータに含まれているメッシュと, それらのメッシュに隣接しているメッシュのみを考慮した. その際の $|M|$ は 34449 となった. 識別精度の評価は 14 分割交差確認を用いて行った. すなわち, 各ユーザのデータから 13 日分のデータを学習用データとして使用し, 残りの 1 日分をテストデータとして識別を行う作業をテストデータを変更しながら 14 回繰り返した.

今回の実験では SVM の学習に, 機械学習のライブラリである scikit-learn[3] 内の実装を使用した. SVM のカーネルは線形カーネルを使用した. 識別方法としては, one-versus-rest を選択した. SVM の学習時におけるペナルティに対するパラメータは学習用データのみを使用した 13 分割交差確認により $0.01, 0.1, 1.0, 10, 100$ の中から決定した. TF-IDF 特徴量の作成にも scikit-learn による実装を使用した. その際, 式 5 の計算時に $|\{(k, d')|M_j \in Lm^k_{d'}\}|$ が 0 となる場合は $\boldsymbol{v}_{TF-IDF}(j) = 0$ とした. BoW と TF-IDF に関しては学習を行う前に特徴ベクトルのノルムが 1 となるように正規化を行った.

[3] http://scikit-learn.org

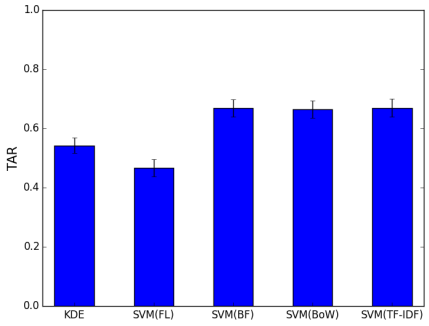

図 4: 実験 1 の結果 (TAR)

KDE では, 学習用データを各ユーザから観測したデータ点として, テストデータに対する確率値を推定した. バンド幅は, 学習用データのみを用いた 13 分割交差確認により $0.0001, 0.001, 0.01, 0.1$ の中から決定した. KDE は scikit-learn による実装を使用した.

実験結果の評価には交差確認により得られた識別結果に対する各ユーザの True acceptance rate(TAR), True rejectance rate(TRR) の平均値を用いた. ある一人のユーザ i に対する TAR と TRR はそれぞれ

$$\mathrm{TAR} = \frac{\mathrm{TP_i}}{\mathrm{TP_i} + \mathrm{FN_i}} \tag{8}$$

$$\mathrm{TRR} = \frac{\mathrm{TN_i}}{\mathrm{TN_i} + \mathrm{FP_i}} \tag{9}$$

と計算した. 但し, $\mathrm{TP_i}$ は正例と識別できた正例の総数, $\mathrm{FP_i}$ は正例と識別した負例の総数, $\mathrm{TN_i}$ は負例と識別できた負例の総数, $\mathrm{FN_i}$ は負例と識別した正例の総数である. また, ここでの正例とはユーザ i のデータを指し, 負例とはユーザ i 以外のデータを指す.

4.2.3 実験 1 の結果

図 4, 5 に実験 1 の結果を示す. なお図 4, 5 の縦軸は TAR, TRR を表す. また, エラーバーは標準誤差を表す. 横軸は使用した手法を表し, SVM においては括弧内に使用した特徴量を記した. 実験 1 の結果より, BF, BoW, TF-IDF を用いて学習した SVM は KDE よりも TAR, TRR 共に上回っていることが確認された. FL

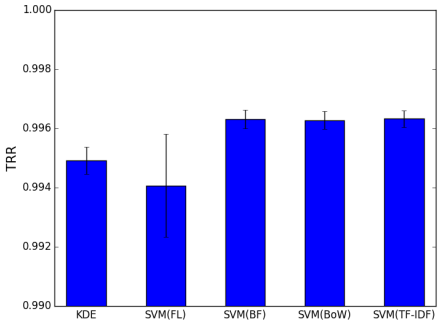

図 5: 実験 1 の結果 (TRR)

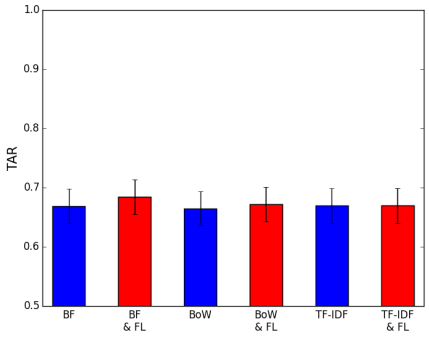

図 6: 実験 2 の結果 (TAR)

特徴量のみを使用して学習した SVM は KDE よりも TAR, TRR 共に下回っていた. また, この実験では TF-IDF を用いた場合が最も TAR の平均値と TRR の平均値が高く, TAR の平均値は 0.6695, TRR の平均値は 0.9963 であった.

4.3 実験 2:FL 特徴量の効果

BF, BoW, TF-IDF 特徴量に FL 特徴量を組み合わせた際の識別精度の変化について調査を行った. 本節にてその実験手順と結果を述べる.

4.3.1 実験方法

BF, BoW, TF-IDF を用いて学習した SVM の識別精度と, それらに FL 特徴量を追加した際の SVM の識別精度を比較する.

BF と FL を組み合わせる際には, BF を用いて作成した特徴ベクトルと FL を用いて作成した特徴ベクトルを接続し, 学習に使用した. BoW 及び TF-IDF に FL を組み合わせる際には, 2 つの特徴ベクトルを接続したのち, ノルムが 1 となるように正規化したベクトルを学習に使用した. SVM の学習手順, 及び識別精度の評価方法は実験 1 と同じである.

4.3.2 実験 2 の結果

図 6,7 に実験 2 の結果を示す. なお図 6, 7 内のエラーバーは標準誤差を表す. 横軸は使用した特徴量を表す. また, 名前の末尾に FL と記載された特徴量が, FL と組み合わせた際の特徴量

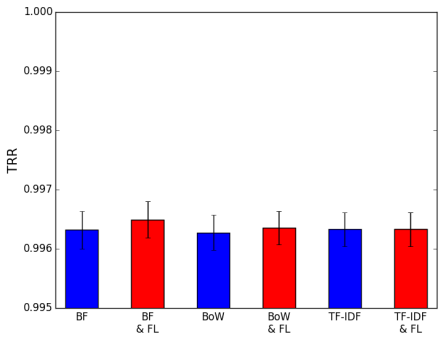

図 7: 実験 2 の結果 (TRR)

を表す. 結果より, BF, BoW においては FL と組み合わせた場合の方が TAR, TRR が向上していることが確認された. しかし, TF-IDF と FL を組み合わせた際には TAR, TRR は全く変化しなかった. また, この実験では BF と FL を組み合わせた場合が最も TAR の平均値と TRR の平均値が高く, TAR の平均値は 0.6845, TRR の平均値は 0.9965 であった.

4.4 実験 3:平滑化の効果

平滑化を行うことによる, 識別精度への影響を調査した. 本節にてその実験手順と結果を述べる.

4.4.1 実験方法

FL, BF, BoW, TF-IDF を用いて学習した SVM による識別精度と, それらの特徴ベクトルに平滑化を行った特徴ベクトルの識別精度を比較する. さらに, 平滑化を行った BF, BoW,

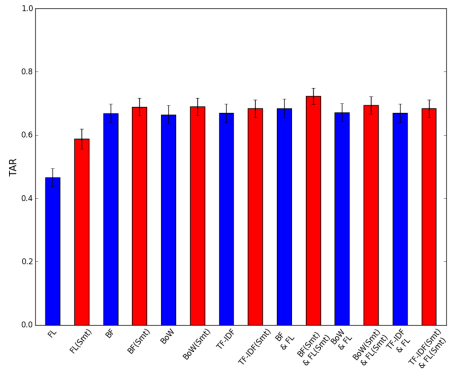

図 8: 実験 3 の結果 (TAR)

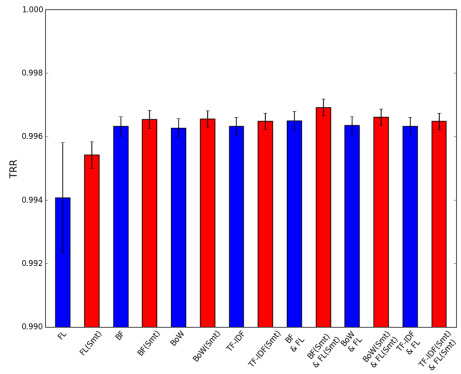

図 9: 実験 3 の結果 (TRR)

TF-IDF に平滑化を行った FL を組み合わせた特徴ベクトルで学習した SVM の識別精度も調査する．SVM の学習手順, 及び識別精度の評価方法は実験 1 と同じである．

4.4.2 実験 3 の結果

図 8, 9 に実験 3 の結果を示す．なお図 8, 9 内のエラーバーは標準誤差を表す．横軸は特徴量の名称を表す．また, 括弧内に Smt と記載された特徴量が, 平滑化を行った際の特徴量を表す．

実験の結果より, FL, BF, BoW, TF-IDF 全ての特徴量において平滑化を行うことで TAR と TRR が向上したことが確認された．また, 平滑化後の特徴量同士を組み合わせた場合の方が, 平滑化前の特徴量を組み合わせた場合よりも TAR と TRR が向上することが確認された．この実験において, BF(Smt) と FL(Smt) を組み合わせた場合が最も TAR の平均値と TRR の平均値が高く, TAR の平均値は 0.7229, TRR の平均値は 0.9969 であった．

5 考察

5.1 実験結果に対する考察

今回実験で使用したデータでは, 提案手法を用いた場合において TAR の平均値が最大で 0.7229, TRR の平均値が 0.9969 の精度で個人識別を行えることが確認された．実験を通して, BF, BoW, TF-IDF 間に識別精度において大きな差は見られなかった．これは, 今回使用した位置情報ログが屋外での測位に限定されているため, 記録されたログが移動時であることが多かったからではないかと考えられる．すなわち, 移動中であるためユーザはあまり一つのエリアに留まらず, 目的地が屋内であった場合には測位がされないため, BoW や TF-IDF で用いていたメッシュへ訪れた回数という情報が余り効果が無かったのではないかと考えられる．

実験 3 において平滑化を行うことによる識別精度の向上が確認された．しかし, 隣接したメッシュだけでなくより遠くに離れたメッシュにも重みを加えるなどの方法で, 更に識別精度が向上する可能性がある．

今回は問題設定として位置だけを考慮し, 時間情報は使用していないが, 時間情報も考慮することでより識別精度を向上できる可能性はあると考えられる．また, 屋内での測位も行われている位置情報ログを用いた場合には, 更に別の特徴を追加することで識別精度が向上することも考えられる．

5.2 個人認証への応用に対する考察

今回の実験結果で確認された提案手法を用いた際の識別精度では, 認証の要素として単体で使用するには十分ではないと考えられる．しかし, 多要素認証の要素の一つとして活用するという方法が考えられる．例えば, スマートフォンに搭載されている他のセンサの情報と組み合わせることで, 位置情報で識別できないパターンと他のセンサで識別できないパターンを相互

に補い, 識別精度の向上が行えるのではないかと考えられる.

6 関連研究

位置情報を用いた個人特定に関する研究が Rossi & Musolesi と Rossi らによって行われている [1, 2]. 著者らは位置情報データを用いた個人特定に関する解析を行っており, その中で機械学習による手法を用いている. しかし, 個人特定に対してプライバシの観点から解析を行った研究であるため, 個人認証への応用は目的としていない.

Shi らによって多要素を用いた端末の不正操作検知に関する研究 [3] が行われており, その中で要素の一つとして位置情報が使用されている. 著者らは位置情報に対しては混合ガウスモデルの学習を行い, 新規の位置データに対する確率値を用いて本人らしさを表すスコアを算出している. Shi らの研究は端末が盗難された際の検出を目的としているため, 本論文での問題設定とは異なる. また, Shi らの研究ではスコアの算出時に用いているデータ点は一点のみであり, 一日の間に蓄積された複数のデータ点を使用した我々の研究とはデータの扱いが異なる.

7 まとめと今後の課題

本論文では位置情報による個人識別を行うために, 位置情報ログから抽出した特徴量を用いた SVM による個人識別手法を提案した. 実験では屋外での位置情報が記録された 91 人分の位置情報ログを用いて個人識別を行い, TAR の平均値が 0.7229, TRR の平均値が 0.9969 の精度で個人を識別できることが確認された.

今後の課題として, より大規模なデータセットに対して本手法を適用し, 識別の精度を評価することが必要であると考えられる. また, 時間情報も考慮するなどして使用する特徴量に更に工夫を加え, 識別精度の向上を行いたいと考えている.

謝辞

本研究は JST CREST 及び科研費 23240019, 次世代個人認証技術講座 (三菱 UFJ ニコス寄附講座) の助成を受けて実施された. ここに謝意を表する.

参考文献

[1] Luca Rossi and Mirco Musolesi. It's the way you check-in. In *Proceedings of the second edition of the ACM conference on Online social networks*, pp. 215–226. ACM Press, 2014.

[2] Luca Rossi, Matthew Williams, Christoph Stich, and Mirco Musolesi. Privacy and the city: User identification and location semantics in location-based social networks. In *Proceedings of the 9th International AAAI Conference on Web and Social Media*, 2015.

[3] Elaine Shi, Yuan Niu, Markus Jakobsson, and Richard Chow. Implicit Authentication through Learning User Behavior. *Information Security*, Vol. 6531, pp. 99–113, 2011.

[4] Vladimir N. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, 1998.

[5] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, and Wei-Ying Ma. Understanding mobility based on GPS data. In *Proceedings of the 10th international conference on Ubiquitous computing*, No. 49, p. 312, 2008.

[6] Yu Zheng, Xing Xie, and Wy Ma. GeoLife: A Collaborative Social Networking Service among User, Location and Trajectory. *IEEE Data Engineering Bulletin*, Vol. 33, No. 2, pp. 32–40, 2010.

[7] Yu Zheng, Lizhu Zhang, Xing Xie, and Wei-Ying Ma. Mining interesting locations and travel sequences from GPS trajectories. In *Proceedings of the 18th international conference on World wide web*, p. 791. ACM Press, 2009.

[8] 総務省. 平成 25 年通信利用動向調査. `http://www.soumu.go.jp/johotsusintokei/statistics/pdf/HR201300_001.pdf`, accessed on 24th August, 2015.