

# 教師ありサービスオペレーション推定のための 対話型ラベル付与システムの開発

○加藤狩夢 (筑波大学) 大隈隆史 (産業技術総合研究所) 蔵田武志 (筑波大学) (産業  
技術総合研究所)

## Developing an interactive label attaching system for supervised service operation estimation

\* K. Kato (Tsukuba Univ.), T. Okuma (National Institute of Advanced Industrial Science and  
Technology) and T. Kurata (Tsukuba Univ.) (National Institute of Advanced Industrial Science and  
Technology)

**Abstract**— Preparing training data is the most expensive work in supervised learning methods. Therefore, we have been developing an interactive label attaching system that helps users make training data for a supervised learning method, named Service-Operation Estimation (SOE), which makes human behavior observation for service improvement effective. This system is expected to contribute to reducing costs of creating training data. Through our preliminary evaluation, we figured out some issues of the current implementation of our proposing system.

**Key Words:** Service-process analysis, Service operation, Machine learning, User interface

### 1 はじめに

ディープラーニング等の機械学習を用いたデータ分析に関する研究開発やその実用化が盛んに行われている。しかし、教師あり学習において、訓練データに含まれる正解ラベルの準備は人間が行う作業であり、その準備作業コストが機械学習の適用の大きな障害となっている。実際、訓練データの準備を支援する企業<sup>1)</sup>が現れるなど、この問題の軽減や解決が望まれている。

筆者らはこれまでサービスオペレーション推定 (SOE: Service Operation Estimation) の開発を進めてきた<sup>2)</sup>。SOEは、従業員の作業内容 (SO) を行動計測データと業務データを用いた教師あり学習に基づいて推定する手法である。これまでは、タイムスタディやワークサンプリングといった人海戦術で収集した分量的に限られたデータを用いた業務分析が主流であったが、SOEのようなアプローチによりビッグデータを用いた新たな視点からの業務分析や改善支援が可能となる。

福原らは、POSデータと測位データからレストラン従業員のスキルに関する指標を設計し、QCサークルの改善活動によるスキルの変遷を可視化した<sup>3)</sup>。例えば、この指標の要素に推定された作業内容を加えることにより、スキル分析を高度化することができる。明官らは、物流倉庫内のピッキング作業の改善案をシミュレーションにより事前評価している<sup>4)</sup>。このシミュレーションのために、測位データとWMS (Warehouse Management System) データに基づいて従業員の作業モデルが構築されているが、推定された作業内容を加えることで、作業モデルの精緻化が期待できる。

### 2 訓練データ作成に関する問題点

SOEにおける訓練データは、従業員データと作業内容の正解ラベルの組み合わせから成る。従業員データは行動計測データと業務データから構成される。Fig1は、日本食レストランでの従業員行動計測実験で用いられた作業動線を記録する歩行者自律航法 (PDR: Pedestrian Dead Reckoning) センサモジュールと、音声データを記録するボイスレコーダー・骨伝導マイクの外観

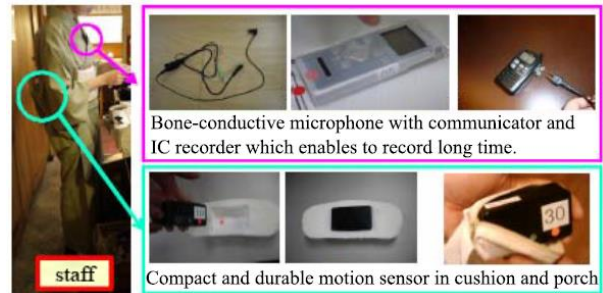


Fig.1: A human behavior sensor, microphone and recorder<sup>5)</sup>.

である。これらの機器を用いて行動計測を行う。一方、映像データは行動計測には用いていない。これは、プライバシー保護や、広域空間もしくは入り組んだ場所での網羅的な撮影の困難さに起因して、カメラによる従業員撮影を前提としていないためである。業務データは業務スケジュール、シフトデータ等もあるが、今回はPOS (Point of Sale) データを用いている。

複数人分の従業員データを任意の時間間隔で区切り、その各区間に作業内容の正解ラベルを紐付けることで訓練データが作成される。SOEは教師あり学習に基づいた手法であり、この紐づけ作業、つまり、正解ラベル付与作業に最も高い人的コストを必要とする。前述のように、カメラによる従業員撮影を前提としていないため、作業内容の正解ラベル付与に映像データを用いることができない。そのため、正解ラベル付与作業には、各計測データや業務データの確認、作業の前後関係の確認等が伴うこととなる。結果的に、一般物体認識や音声認識等のための正解ラベル付与と比較し、作業内容の正解ラベル付与はより負荷が高い作業になってしまう。

加えて、作業プロセスの変化も問題となる。作業ラベルが同じでも、場所、時間、従業員、顧客等の様々な要因で作業プロセスが変化する。例としてある日本食レストランでの挨拶・案内作業を挙げる。この作業

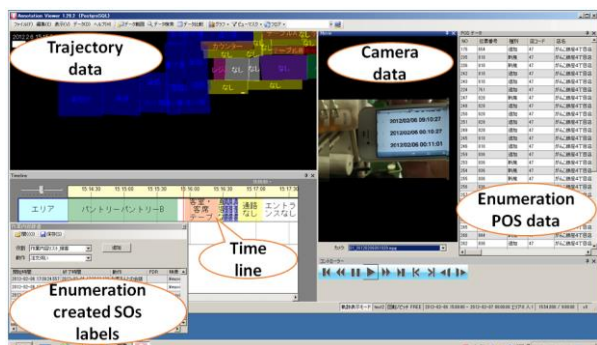


Fig.2: Service-process analysis support system's main screen

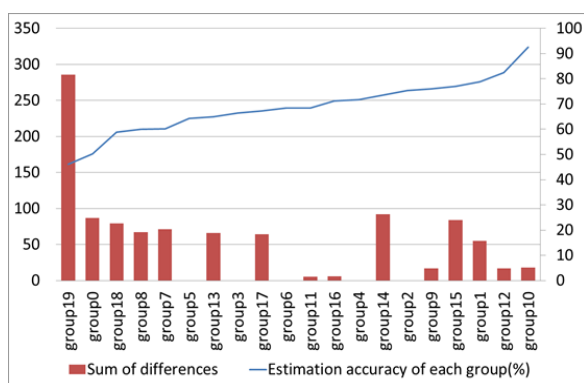


Fig.3 : Sum of differences between estimation data and training data by SO and cross-validation results

は玄関にきた顧客に挨拶して、席まで案内するというプロセスによって構成されているが、出入口の数、案内する席の空き具合によって案内ルートは様々に変化する。顧客が常連客の場合、雑談等が入りプロセスがさらに複雑化する。このような作業プロセスの変化も正解ラベル付与の高コスト化に拍車をかけ、訓練データ作成の大きな障害となっている。

### 3 先行事例

正解値を効率的に作成・収集する手法やシステムは既に実用化されている。文字変換ソフトの分野では Microsoft Office IME, デジタル写真管理ソフトの分野では Picasa, 検索エンジンの分野では Google や Microsoft のオートコンプリート機能等が挙げられる<sup>6)</sup>。これらはどれもユーザが作成した正解値を用いて推定精度を高め、フィードバックするという機能を用いることで正解値を効率的に確保した事例である。

アカデミックの分野では、Zhang 氏らによる顔認識を用いた写真管理手法がある<sup>7)</sup>。この論文では、Bayesian Framework を用いて写真に写る顔に対して人物を推定し、その名前候補をユーザに与えることで、家族アルバムの管理システムを作成しており、デジタル写真管理ソフトの礎を築いた。

また、後藤らによる PodCastle<sup>8)</sup>と呼ばれるシステムで

は、Web 上の日本語のポッドキャストを音声認識によって自動的にテキスト化すると共に、クラウドソーシングの考え方に基いて、誤識別箇所を閲覧者に修正してもらうことで、音声認識率を向上させている。

しかし、外食や介護等のサービス分野において、訓練データを効率的に作成することを目的としたシステムはまだ存在しない。

### 4 パイロットスタディに基づく仮説と提案

本研究では、正解ラベル付与担当者が業務分析支援ツール(Fig.2)と呼ばれるシステムを用いて作業内容の正解ラベル付与を行うことを想定している。このツールは、計測時に得た行動計測データ、業務データ等を時刻同期させて可視化する。正解ラベル付与担当者は、各種データを確認しながら、「従業員Aは時刻 $t_1$ から時刻 $t_2$ の間にSO<sub>i</sub>に従事した」といった、従業員ID、時間、作業内容ラベルに関する正解値を付与する。

正解ラベルを決定するには、注意深く各作業プロセスの前後関係を確認する必要がある。ところが、聞き取りが難しくなるため、音声データの早送り再生は現実的ではなく、作業内容の確認には多大な時間的コストがかかることになる。

機械学習で高い推定精度を得るには、訓練データの分布が母集団の分布と比較し、偏りが少ないことが望ましい。SOE もまた、訓練データの分布に偏りがなければ、推定対象データが訓練データの分布から外れる可能性が低くなるため、その訓練データによって学習した識別関数の推定精度は高くなる。

しかし、前述したように、作業プロセスは、場所、時間、従業員、顧客等の様々な要因によって変化しており、それが各種データに影響を与え、母集団の分布を複雑なものにしていると考えられる。より良い学習結果を得るには、プロセスが変化する要因ごとに作業内容の分布を分析することが重要だと著者らは考えた。

SOE に対して交差検証を行うため訓練データを 20 個のグループに分けて学習・推定を行った (Fig.3)。その結果、グループによって推定精度に大きな差が生じた。

ここで、正解ラベルの時間的偏在が推定精度の差を生む主要因の1つであるという仮説を立てる。その検証のため、まず、推定対象データと学習データを作業内容・時刻別にグループ化し、データ数の差分を算出した。次に、推定対象データ数が学習データ数よりも多い場合に限り差分を集計した。集計結果の値が大きければ大きいほど、推定対象データが存在する時間帯の作業内容を学習していないことを意味する。Fig.3 に本検証の結果を示す。集計結果と推定精度に相関が見られたため、用いたデータに限って言えば、前述の仮説が検証されたとと言える。

### 5 対話型ラベル付与システム

本研究では業務分析支援システムと連携する SOE のための対話型ラベル付与システムを開発する。本システムによって、正解ラベル付与にかかる作業量の軽減を目指している。Fig.4 は本システムの概念図である。本システムは、正解ラベルの時間的偏在推定、SOE、正解ラベル付与インターフェースの3つの機能から構

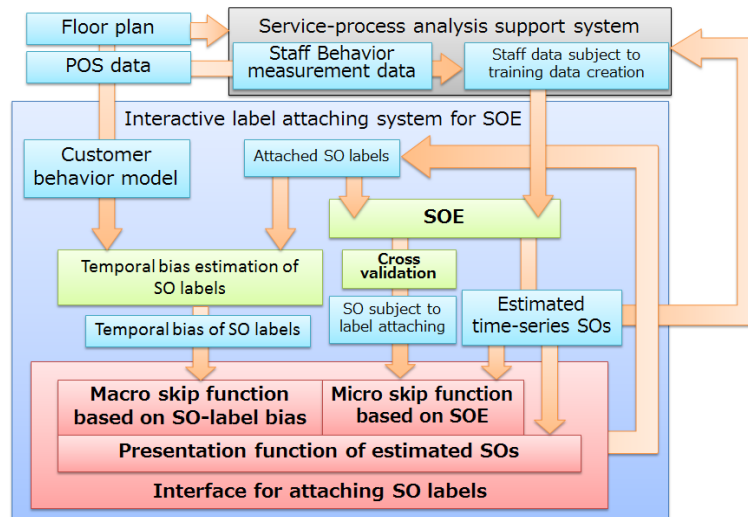


Fig.4: Interactive label attaching system's conceptual diagram

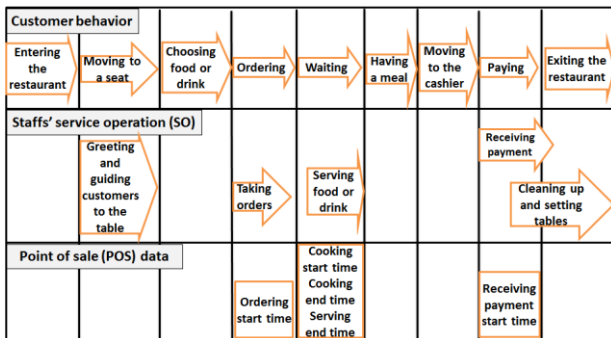


Fig.5: Customer behavior model and correlation between

成される。正解ラベルの時間的偏在推定は、付与済み正解ラベルの時間的な分布の偏りを算出する。SOEは付与済み正解ラベルを用いて（再）学習し、正解ラベル未付与の従業員データを対象に正解ラベル（作業内容）を推定する。正解ラベル付与インターフェースは、上記の機能が出力した結果を用いた機能であり、正解ラベルの偏りに基づくマクロスキップ機能、推定正解ラベルに基づくミクロスキップ機能、推定正解ラベル提示機能の3つの機能を含んでいる。

### 5.1. 正解ラベルの時間的偏在推定

正解ラベルの時間的偏在を算出するには、正解ラベルの時間的な分布の母集団を生成する必要がある。母集団の分布生成の為、顧客行動モデルを設計し、モデルに基づく各作業の発生確率分布を生成した。従業員の各作業とPOSデータは、顧客の入店から退店までの行動に連動すると仮定した。Fig.5は顧客が入店から退店までに行う行動を元に、それに関連する複数の従業員が行う作業とPOSデータに記録された内容を時系列で図にしたものである。

作業の発生確率分布は、POSデータの注文発生時刻と会計発生時刻に基づく。Fig.6は、昼の時間帯の作業「挨拶・案内」をモデル化した例である。「挨拶・案内」

内」は注文発生時刻よりも前に行われるため、注文発生時刻から注文伺いの平均時間・品定め時間・玄関から席までの移動時間だけ戻ることによって「挨拶・案内」の発生時刻と終了時刻を算出する。注文伺いの平均時間は、付与済みの正解ラベルの平均値を用いており、品定め時間は経営陣へのヒアリングにより昼は3分、夜は5分とした。玄関から席までの移動時間は、POSデータ内のテーブル番号とCADデータから玄関から席までの距離を算出し、国際交通安全学会112プロジェクトチームによる人の歩行速度を参考に移動時間を算出した<sup>9)</sup>。

以上の結果から作業の発生確率分布を定義しPOSデータを組み合わせることで母集団を生成した。そして、付与済みの正解ラベルの時間的な分布と生成した母集団の分布との差分から正解ラベルの時間的偏在を算出する。Fig.7に「挨拶・案内」の時間的偏在の算出結果を示す。図から、作成済みの正解ラベル「挨拶・案内」が13時に集中していることが分かり、母集団の分布と比較すると、15時の正解ラベルが少ないのが分かる。

### 5.2. 正解ラベル付与インターフェース

本インターフェースは出力された結果から時間帯・ $SO_x$ を絞り込むスキップ機能を提供することで、従業員の作業内容を確認する手間を省くことが可能になる。まず、SOEの交差検証結果から $SO_x$ を選定する。次に $SO_x$ の時間的偏在から偏りが最も大きい時間帯を把握しスキップする。これが、正解ラベルの偏りに基づくマクロスキップ機能である。

一方、推定正解ラベルに基づくミクロスキップ機能では、SOEの推定結果と、偏りの大きな時間帯の $SO_x$ を用いてスキップする。

Fig.8は推定正解ラベル提示機能を示している、スキップはこの提示機能を確認しながら行われる。赤い縦線が業務分析支援システムに表示されているカレントタイムである。スキップは登録されたマクロキーによって行われ、正解ラベル付与対象の $SO_x$ にスキップすることが可能である。Fig.8での $SO_x$ は「挨拶・案内」である。それは、カレントタイムから15:15:50の間で



の推定値および正解値として示される 1 から 8 の数値は Table1 における正解ラベルの ID に対応する。対角線上にある値が、推定に成功したサンプルの数であり、この値が高く、それ以外が低いほど推定精度が高いと判断できる。以上の結果から活用できなかった原因は提案 UI に用いた SOE の性能が不十分であったためだと考察される。ミクロスキップ機能の対象となった「挨拶・案内」の適合率は 62.14% であり、この条件ではミクロスキップ機能が十分に機能しないことが確認されたといえる。今後、ミクロスキップ機能が有効に働くために必要な SOE の適合率について検証を進める必要がある。一方で、訓練データ数が少ない間ほどのような SOE 手法を実現しても十分な適合率が得られる保証はないと考えられることから、推定結果の適合率が低い場合にも効率化を実現できるインターフェースについて検討をする必要がある。

正解ラベルを追加によってどのように推定精度が変化したのかを Table3 に示す。各正解ラベルの推定精度は適合率と再現率から調和平均をとることで算出している。予備実験では「移動・物を運ぶ」以外の推定精度が低下した。正解ラベルを追加した後の学習結果と追加する前の学習結果の差分を Table4 に示す。表から、「片付け・セッティング」を「挨拶・案内」として誤認識する場合と、「配膳」と「注文伺い」の相互の誤認識が推定精度低下の大きな要因となっていることが分かる。

被験者のインタビューから、正解ラベル付与時に、顧客がいる状態での細かな片づけをおこなっていることが推測されるデータに「片付け・セッティング」のラベルを多く付与したことが分かった。このとき、顧客へ挨拶しながら「片付け・セッティング」を実施しているシーンへのラベリングが多く行われた可能性がある。予備実験開始前に付与済みの訓練データでは「片付け・セッティング」のほとんどが無言で、食器の音だけが記録されていた。そのことから、従業員の発話が含まれているにもかかわらず「片付け・セッティング」というラベルを付与されたデータは、交差検証時には「挨拶・案内」だと認識され、精度低下を引き起こしたと考えられる。

また、「注文伺い」は業務データの寄与率が高い作業内容であり、POS データが推定精度に多大な影響を与えていることが分かっている。Fig.10 は、特徴カテゴリ毎に寄与率を算出した図である。寄与率は、Random Forest によって得られる特徴の重要度に基づいており、本実験で用いた特徴を 12 種類のカテゴリに分類して集計を行い、作業内容毎に正規化した数値である。この図から各作業内容の推定に各カテゴリがどの程度寄与していたのかが分かる。POS データから得られる特徴を業務として分類した結果、「注文伺い」は他の SO と比べて業務の特徴が強く影響していることが分かる。しかし、POS データに存在しない「注文伺い」は実際の業務上、よく発生する。そのような「注文伺い」は POS データのある「注文伺い」よりも特徴を捉えづらく、推定が難しい。そのため、顧客に対して品物を説明して配る「配膳」と POS データのない「注文伺い」を現状の特徴量を用いた Random Forest で分離することが難しく、それが推定精度低下につながった

と考えられる。

本実験によって、提案された UI は想定以上に SOE の性能に依存することが明らかとなった。今後は、新たな特徴量を導入するなどして SOE の推定精度を向上した上で、推定精度を加味しながら提案 UI と今後提案する UI の切り替え利用等を検討する。

## 7 おわりに

本研究では、正解ラベルを効率的に付与できるシステムを設計・実装した。予備実験の結果、スキップ機能は高速再生等ある程度の文脈を把握できる設計が必要であることが確認された。また、フィードバックされる推定結果の適合率が低い場合、提案 UI のスキップ機能が活用できないことが確認された。さらに、提案 UI を設計時の意図通りに機能するような実装のためには SOE 機能の性能向上が必須であることが確認された。

今後は、予備実験を基にした UI の検討・実装、SOE の性能向上を行い、より厳密な評価実験を行う計画である。

## 参考文献

- 1) Training Data as a Service, <https://spare5.com/>
- 2) R Tenmoku, R. Ueoka, K. Makita, T. Shinmura, M. Takehara, S. Tamura, S. Hayamizu, T. Kurata, Service-Operation Estimation in a Japanese Restaurant Using Multi-Sensor and POS Data, APMS 2011 conference, Parallel 3-4: 1 (2011)
- 3) T. Fukuhara, R. Tenmoku, R. Ueoka, T. Okuma, T. Kurata, Estimating Skills of Waiting Staff of a Restaurant based on Behavior Sensing and POS Data Analysis: A Case Study in a Japanese Cuisine Restaurant, AHFE2014, 4287/4299 (2014)
- 4) T. Myokan, M. Matsumoto, T. Okuma, R. Ichikari, K. Kato, D. Ota, T. Kurata, Pre-evaluation of Kaizen Plan Considering Efficiency and Employee Satisfaction by Simulation Using Data Assimilation -Toward Constructing Kaizen Support Framework-, ICServ 2016 conference, 331/337(2016)
- 5) M. Takehara, K. Kato, S. Tamura, R. Tenmoku, T. Kurata, S. Hayamizu, Improvement of service-operation estimation using voice activity detection of employee's speech, 電子情報通信学会論文誌 D, **J97-D-10**, 1563/1571(2014)
- 6) J. Pereira, L. Chen, Server-deployed cache list management for presenting an auto-complete list, US, Patent US20060242109 A1(2006)
- 7) L. Zhang, L. Chen, M. Li, H. Zhang, Automated annotation of human faces in family albums, Proceedings of the 11th ACM International Conference on Multimedia, 716/723(2003)
- 8) M. Goto, J. Ogata, PodCastle: Recent advances of a spoken document retrieval service improved by anonymous user contributions, Proceedings of the 12th Annual Conference of the International Speech Communication Association, 3073/3076(2011)
- 9) A. Tsujimura, Y. Nagayama, M. Nagamachi, K. Takizawa, H. Suzuki, T. Morita, M. Nakai, 地域文化特性と運動行動 - ソーシャル・スピードのインデックス化-, IATSS Review, **5-4**, 231/24(1979)
- 10) L. Breiman, Random Forest, Machine Learning, **45-1**, 5/32(2001)

Table 2: The true or false matrix

|                  |   | SO labels |       |       |       |       |       |       |       | Precision (%) |
|------------------|---|-----------|-------|-------|-------|-------|-------|-------|-------|---------------|
|                  |   | 1         | 2     | 3     | 4     | 5     | 6     | 7     | 8     |               |
| Estimated result | 1 | 158       | 1     | 2     | 0     | 0     | 50    | 20    | 0     | 68.40         |
|                  | 2 | 9         | 592   | 96    | 13    | 18    | 77    | 24    | 54    | 67.04         |
|                  | 3 | 2         | 81    | 709   | 25    | 46    | 47    | 76    | 155   | 62.14         |
|                  | 4 | 0         | 2     | 23    | 155   | 0     | 27    | 14    | 124   | 44.93         |
|                  | 5 | 2         | 18    | 0     | 0     | 358   | 10    | 42    | 6     | 82.11         |
|                  | 6 | 79        | 52    | 66    | 9     | 8     | 866   | 33    | 85    | 72.29         |
|                  | 7 | 58        | 41    | 59    | 37    | 11    | 95    | 1211  | 214   | 70.16         |
|                  | 8 | 9         | 92    | 259   | 143   | 11    | 108   | 152   | 1806  | 70.00         |
| Recall(%)        |   | 49.84     | 67.35 | 58.40 | 40.58 | 79.20 | 67.66 | 77.04 | 73.90 | 68.56         |

Table 3: Accuracy of each SO and the difference

| SOID | without the newly added training data | with the newly added training data | difference |
|------|---------------------------------------|------------------------------------|------------|
| 1    | 57.66                                 | 55.37                              | -2.30      |
| 2    | 67.20                                 | 61.59                              | -5.60      |
| 3    | 60.21                                 | 54.28                              | -5.93      |
| 4    | 42.64                                 | 45.22                              | 2.58       |
| 5    | 80.63                                 | 76.99                              | -3.64      |
| 6    | 69.90                                 | 66.02                              | -3.88      |
| 7    | 73.44                                 | 66.31                              | -7.13      |
| 8    | 71.89                                 | 65.37                              | -6.53      |

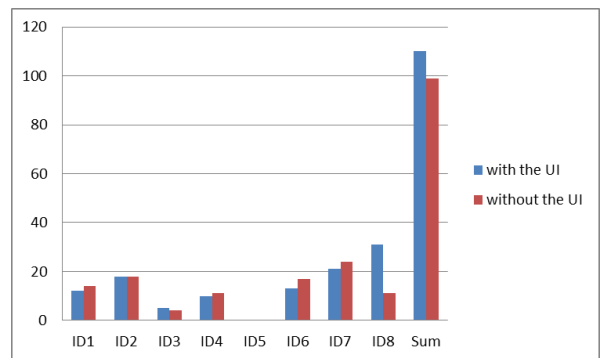


Fig.9: Accumulated number of created SO labels

Table 4: The difference of matrix

|                  |   | SO labels |    |     |     |     |     |     |     |
|------------------|---|-----------|----|-----|-----|-----|-----|-----|-----|
|                  |   | 1         | 2  | 3   | 4   | 5   | 6   | 7   | 8   |
| Estimated result | 1 | 7         | 3  | -1  | 0   | 0   | 8   | 7   | 0   |
|                  | 2 | 20        | 84 | 12  | 7   | -7  | 3   | 57  | 95  |
|                  | 3 | 1         | -5 | -28 | -8  | -13 | 21  | 15  | 116 |
|                  | 4 | 0         | 4  | -13 | 20  | 0   | -1  | -2  | 9   |
|                  | 5 | -2        | 0  | 0   | 0   | 15  | 4   | 7   | 36  |
|                  | 6 | -7        | 21 | 57  | 31  | 7   | 254 | 103 | 68  |
|                  | 7 | 4         | 7  | 2   | -9  | 1   | 98  | 39  | 57  |
|                  | 8 | 1         | 48 | 26  | -11 | 18  | -6  | 47  | 2   |

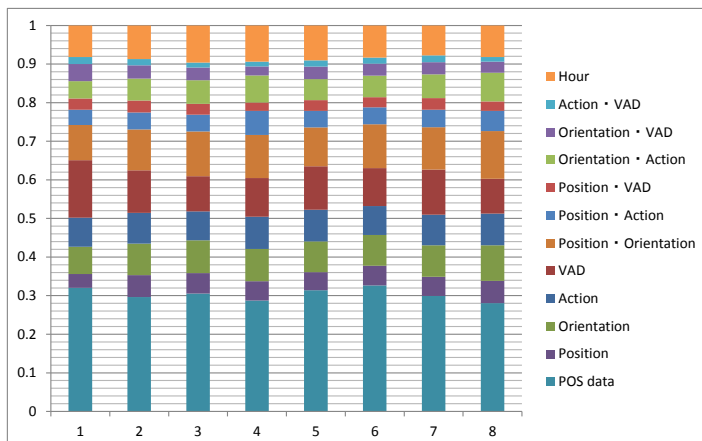


Fig.10: Contribution ratio of each category data