

業績と行動の関係マイニングのための手法検討

江崎 健司^{1,a)} 高屋 典子^{1,b)} 澤田 宏^{1,c)}

概要：本論文では行動のログと業績のログを用いて、業績と行動の関係を明らかにする手法を検討する。業績と行動の関係が明らかになれば、職場活性化やそれを促すコンシェルジュソフトへの活用が実現される。関係の例としては、「業務を行う際に作業の目的を把握したりするためにハブとなる人物を多部署で知っている傾向が高業績者に多い」といったことがある。このような知見を獲得するためにデータから適切な手法を用いて業績と行動の関係を抽出する必要があるため、ベース技術としてL1正則化を用いた回帰モデルを用いる。関係を抽出するための手法を検討するために、業績を被説明変数としたいくつかの回帰モデルをデータに適用し、手法ごとに回帰モデルの予測精度の評価と抽出できた業績と行動の関係を考察する。

キーワード：予測モデル，特徴選択，オフィス行動，ユーザモデル，業績，モデル比較

1. はじめに

近年、業務の電子化やセンサの発達による行動ログの取得が進みオフィスにおける行動のデータが取得可能になってきている。そのため取得したオフィスにおける行動のデータを有効活用し業績を改善，向上させる施策に取り組むことが重要になっている。

それを受けログ化できるようになったオフィス内の行動やコミュニケーションが、業績や成果とどのような関係にあるのかを明らかにする取組が行われている。オフィス行動の分析は種々の取り組みが行われており以下に例を2つ示す。1つ目の例はPCの作業ログを収集し業務効率が良い人が行っている行動を抽出して類似業務へその行動を水平展開したものがあ[1]。2つ目の例は業務中の対面コミュニケーションのログから業績が良いチームと悪いチームを比較し、良いチームが行っているコミュニケーションをほかのチームでも促す施策を打ち業績を向上させる施策がある[2]。しかしこれらの例では、データを人手で見ながら企画や分析の担当者が経験で施策を決定する方法が主であった。

一方で最近ではマイニング技術によってデータから業績に効果がある行動を抽出する業績と行動の関係マイニングの研究が進められている。得られた行動の業績への効果を基に、実際の現場のデータに基づくシステムを構築すること

ができる。しかし、マイニング技術を用いる場合、業績と行動が持つ構造に関する仮説が正しい場合でないと、所望の行動の業績への効果を学習によって抽出することができない。例えば、ユーザごとに行動の業績への効果が異なるのに、全体を同一の構造として学習すると効果を正しく抽出できない。

このため現場に応じたシステムをデータマイニングの技術を用いて構築する事を目指す場合、行動の業績への効果を明らかにすると同時に、業績と行動が持つ構造も明らかにする必要がある。この際データマイニングの技術を用いて明らかにするべきものは次の2つである。まず、業績と行動の構造を明らかにすることでどんなシステムにするべきかを明らかにする。例えば一人ひとり違うシステムなのか、どのように情報を提示するシステムにするべきかなど、である。次に、行動の業績への効果を明らかにすることでそのシステムがどのようなパラメータを持ち制御されるかを明らかにする。例えば個人ごとにどんな情報を提示するべきかなど、である。

そこで本論文では、業績と行動が持つ構造に関する仮説を複数設定し次の2つを行い目的を達成できると考える。まず業績と行動が持つ構造の尤もらしい仮説を明らかにする。そのために次の指標を用いてデータから最適な仮説を選択する。それはデータから学習したモデルについてテストデータの予測精度が最も高いことを指標とする。次に行動が業績に及ぼす効果を明らかにする。そのために次の解釈を用いる。それはデータから、L1正則化を用いた回帰モデルの学習した重みの係数を行動が業績に及ぼす効果と解釈する。

¹ NTT サービスエボリューション研究所
NTT Service Evolution Laboratories

a) esaki.kenji@lab.ntt.co.jp

b) takaya.noriko@lab.ntt.co.jp

c) sawada.hiroshi@lab.ntt.co.jp

今回想定するシステムとして、高業績者の行動特徴を明らかにしそれを促すシステムとする。

そこでどんなシステムにするべきかを検討するのに必要な仮説は、個人ごとに違うのか？相乗効果が存在するのか？についてである。まず個人ごとに業績と行動の関係が違つかどうかでシステムをパーソナライズする価値があるかを明らかにすることができる。つぎに相乗効果があるかによって、二つ以上の行動を考慮したシステムを作成する価値があるかを明らかにすることができる。

さらに、予測できたモデルの特徴選択結果によってある仮説に基づきユーザが高業績を得るためにどのような行動をとることが有効かを明らかにする。

本論文の構成としては次のとおりである。まず2章で従来研究について述べる。そして、仮説と各仮説に基づくモデルについて3章で述べる。また、業績に及ぼす効果を明らかにしたい行動については4.1節と4.2節で述べる。さらに、今回実験で用いたデータから導かれた結果について5章で考察する。

2. 従来研究

本章で行動・コミュニケーションと業績の関係を取り扱った従来法を紹介する。

2.1 人手を用いた相関分析による行動特徴抽出

ビジネス顕微鏡[®] [3]^{*1}を用いた業績向上の施策立案のための行動特徴抽出の研究がある [2]。これはビジネス顕微鏡[®] に内蔵された各種センサを用いて業務者の行動を取得した後で特徴量化し、異なるチーム間で傾向が異なる特徴量を人手で抽出する取組であった。人手で業績の差異と特徴量の関係を把握する必要があるため、一次元で表現された特徴と業績の関係しかマイニングできないものの、休憩中の加速度が高い、つまり休憩中に人が動き動作を含めた会話などが活発に起こっていることを表現する、場合に業績が上がる傾向があることを発見した。

2.2 主成分と満足度との相関分析によるパタン抽出

メール等の情報を利用して業務種類を分類した時系列特徴量と業務の満足感に関連する業務パタンを抽出するための研究がある [1]。これは業務の時系列データをあらかじめ主成分分析によってパターン化し、そのパターンと業務の満足度との相関分析を行うことで、業務の満足度が高いユーザがどのような行動パタンを取りやすいかを発見する方法を取っている。この方法を用いることで業務の満足度と負の相関のあるパタンの中に、典型的なシーケンスパターンがあり典型的な仕事ばかりしていると業績が悪化することが示唆された。

*1 ビジネス顕微鏡[®] / Business Microscope は株式会社日立製作所の登録商標です。

表 1 各手法ごとの特徴まとめ

Table 1 list of performance with each method

手法	自動抽出	可読性	相乗効果	個別異質性
2.1	-		-	-
2.2		-	-	-
2.3			-	-

表 2 本論文で用いる各モデルごとの特徴まとめ

Table 2 list of performance with each method

手法	相乗効果	個別異質性
基本モデル	-	-
多次元モデル		-
混合モデル	-	
潜在変数モデル		

2.3 回帰分析による行動特徴抽出

実験環境でのチームのコミュニケーションとそのチームが達成したタスクの精度の関係を明らかにする研究がある [4]。これはチームのコミュニケーションをグラフ構造でデータ化しコミュニケーションの傾向を特徴量化し回帰分析を行うことでタスクの達成度に効果がある行動特徴を抽出する方法をとっている。この方法を用いることでタスクの達成度に効果がある特徴量は、チーム内の最小次数であり、チーム内の最小次数が大きければ大きいほどタスクの達成度が高くなる傾向があることを明らかにした。チーム内の最小次数は直感的にチーム内のメンバが密につながり連携がうまくいっていると考える事ができるため、チームが連携する重要性を示唆したといえる。

2.4 既存研究のまとめと比較表

表 1 に既存手法を比較したものをしめす。この表の列が特徴を表しており各手法が対応している場合には、対応していない場合には - をつけている。特徴の自動抽出とは人手を介さず有効な特徴を把握し施策に活用することが可能かである。一般的に言われる機械学習の手法は自動抽出に対応しており、人手で相関係数を一つずつ調べる方法などは対応していないと考える。特徴の可読性とは、得られた結果から有効な特徴を紐解き施策に活用することが可能かである。特徴の相乗効果とは、他の特徴に依存して業績の効果が変化することをモデル化可能かである。特徴の個別異質性はモデルがユーザごとに可変かということである。

3. 本論文で用いる各種モデル

本論文では従来研究である 2.3 節の回帰モデルを用いた業績と行動の関係マイニングに対して、4つの仮説に基づき拡張した回帰モデルを利用しデータに適用する。これはどの仮説がデータに含まれる仮説かを検証するためである。今回用いる仮説は、「各行動の業績への効果は独立であ

るか、相乗効果を持つか」という相乗効果に関する仮説と、「全体で業績と行動の関係が同一であるか、個人ごとに業績と行動の関係はことなるか」という個別異質性に関する仮説の組み合わせを検証する。この2つの仮説の組み合わせによって4つの仮説は下記でそれぞれに当てはまるモデルを考える。

各行動の業績への効果が独立で、全体で業績と行動の関係が同一であるモデルを基本モデルとする。これは2.3節の従来研究と同じ仮説を検証し、仮説に基づき高業績者がおこなっている行動特徴を抽出することになる。

各行動の業績への効果が相乗効果を持ち、全体で業績と行動の関係が同一であるモデルを特徴量の相関を考慮するモデルとする。

各行動の業績への効果が独立で、個人ごとに業績と行動の関係が異なるというモデルを混合モデルとする。

各行動の業績への効果が相乗効果を持ち、個人ごとに業績と行動の関係が異なるというモデルを潜在変数モデルとする。

本論文のアウトプットは高業績者の行動特徴を抽出することであるため、業績者が行っている行動特徴は事前に特徴量化されているものとする。今回用いる行動の特徴量は次章で示すが、例えば上司と一日何回会話をするのかを一月分で平均を取ったものを用いる。各業績者を j とし、それぞれの特徴次元を i とする。この特徴量を x_{ij} とし入力として用い、全特徴量数は N_x とする。また業績が高い人が特に行っている行動特徴を抽出するため、各個人の業績を y_j とし入力として用いる。

次節から示すモデルについて共通する枠組みとしては次の通りである。各モデルは仮説に基づき行動特徴から業績を予測できるよう行動の業績への効果を重みとして学習を行う。その予測値を y'_j とし、例えばL1正則化を用いた線形回帰モデル $y'_j = \sum_i w_i x_{ij}$ に各個人の行動特徴量と業績を入力することで、業績に効果のある行動特徴量のみが重みを持ち、効果のない行動特徴量の重みは0を取るよう重み w_i が学習される。学習した行動特徴量の重みが正の場合は、業績の高い人が特に行っている行動であり、その行動をとることが高業績の条件となる。一方学習した行動特徴量の重みが負の場合は、業績の低い人が特に行っている行動であり、その行動をとらないことが高業績の条件となる。

3.1 基本モデル

基本モデルとしては高業績者の行動特徴を抽出する方法としてL1正則化を用いた線形回帰モデルを用いる。線形回帰モデルを用いる場合は業績が各行動特徴の業績への効果の累積で表現されるということを仮定している。この場合 $y'_j = \sum_i w_i x_{ij}$ で入力 x_{ij} と y_j が与えられた場合に w_i は以下の目的関数 $J1$ を最小化する解とし得られる。

$$J1 = \sum_j (y_j - \sum_i w_i x_{ij})^2 + r * \sum_i |w_i| \quad (1)$$

ただし、 r は正則化項を表し誤差と正則化の影響をコントロールするための定数である。

3.2 特徴量の相関を考慮するモデルへの拡張

前述した基本モデルでは、ある行動をとればとるほど業績が上がるというモデルであった。これはモデルに線形性を仮定し一次元でモデル化しているからであり、モデルを特徴量の相関を考慮できるようにすればより行動と業績の関係を複雑にすることが可能である。まず本論文では2つの特徴量の相関を考慮するように拡張する場合を考える。この場合 $y'_j = \sum_i w_i x_{ij} + \sum_i \sum_{i'} w_{ii'} x_{ij} x_{i'j}$ で入力 x_{ij} と y_j が与えられた場合に w_i は以下の目的関数 $J2$ を最小化する解として得られる。

$$J2 = \sum_j (y_j - \sum_i w_i x_{ij} - \sum_i \sum_{i'} w_{ii'} x_{ij} x_{i'j})^2 + r * \sum_i |w_i| + r' * \sum_i \sum_{i'} |w_{ii'}| \quad (2)$$

ただし、正則化項は1つの行動特徴と2つの行動特徴の相関を考慮する場合で異なるとし、それぞれを r, r' とした。

3.3 特徴量の相関を考慮するモデルへの拡張：Factorization Machines の利用

前述した特徴量の相関を考慮するモデルでは、学習すべき重みは特徴量次元の二乗という膨大な数になってしまうというデメリットがある。相乗効果をモデリングすることが可能でありつつ、パラメータを学習できる範囲に抑える方法として Factorization Machines が提案されている [5]。これは二つ以上の特徴量を用いる項の重みを行列分解し、各次元の重みがK次元ベクトルの積で表現されることを仮定したモデルである。これを v_{ik} とした。すなわち、ある特徴量が与える相乗効果はすべての特徴量で同一であることを仮定するものである。この場合 $y'_j = \sum_i w_i x_{ij} + \sum_i \sum_{i'} \sum_k v_{ik} v_{i'k} x_{ij} x_{i'j}$ で入力 x_{ij} と y_j が与えられた場合に w_i は以下の目的関数 $J3$ を最小化する解として得られる。

$$J3 = \sum_j (y_j - \sum_i w_i x_{ij} - \sum_i \sum_{i'} \sum_k v_{ik} v_{i'k} x_{ij} x_{i'j})^2 + r * \sum_i |w_i| + r' * \sum_i \sum_k |v_{ik}| \quad (3)$$

ただし、正則化項は1つの行動特徴と2つの行動特徴の相関を考慮する場合で異なるとし、それぞれを r, r' とした。

3.4 混合モデルへの拡張

前述した基本モデルでは、単一のモデルで業績と行動特

徴の関係をモデル化しようとしていた．しかしこのモデル化ではある行動をすると業績が上がるか下がるかが唯一つの関係として決まっていることを示している．そのため個人ごとにモデルが違う混合モデルを考える．混合モデルを実現する場合，人によって異なるモデル化をするため $p(y'_j|x) = \sum_k p^{i_k} * p(y'_j|w_k, x_j)$ となりユーザによって重みが異なる，例えば代表的なK個のキャラクタへの帰属度でユーザをモデル化すると重みはk種類になる，ことが期待できより正確なモデリングが可能になる．この場合 $p(y'_j|x) = \sum_k p^{i_k} * p(y'_j|w_k, x_j)$ で入力 x_{ij} と y_j が与えられた場合に w_i は以下の目的関数 $J4$ を最小化する解として得られる．

$$J4 = \sum_j -\log p(y_j|x_{ij}) + r * \sum_i \sum_k |w_{ki}| \quad (4)$$

$$p(y'_j|x_j) = \sum_k p^{i_k} * p(y'_j|w_k, x_j) \quad (5)$$

$$p(y'_j|w_k, x_j) = N(y'_j | \sum_i w_{ki} x_{ij}, \beta) \quad (6)$$

ここで， x_j は業績者 j に対する i 個の行動特徴を表すベクトルとする．また w_k は各行動特徴に関する重み， $N()$ は正規分布， β は分散パラメータである．

3.5 潜在変数モデルへの拡張

前述した基本モデルでは，ある行動が業績に与える効果はそれぞれ独立であり，どんな時でも行動が与える効果は一定であった．しかし，各行動は非独立で，人によって行動がもつ業績への効果は異なるモデルを考える必要がある．それを実現するために潜在変数への拡張を行う．そのために， $y' = wx$ ではなく y' と x に共通する潜在変数を考えて $y'_j = \sum_k w_{yk} z_k$ ， $x_{ij} = \sum_k w_{xik} z_k$ のモデルを学習することとする．この場合 $y'_j = \sum_k w_{yk} z_k$ ， $x_{ij} = \sum_k w_{xik} z_k$ で入力 x_{ij} と y'_j が与えられた場合に w_i は以下の目的関数 $J5$ を最小化する解として得られる．

$$J5 = \sum_j (y_j - \sum_k w_{yk} z_k)^2 + b * \sum_j \sum_i (x_{ij} - \sum_k w_{xik} z_k)^2 + r * \sum_k |w_{yk}| + r' * \sum_i \sum_k |w_{xik}| \quad (7)$$

ただし，b は行動特徴量の近似誤差に関する重みである．

3.6 モデルの学習

各モデルはデータから任意の学習法を用いて学習することができる．本論文では，確率的勾配法を用いて各モデルの重みを学習するよう実装した．それぞれのモデルの更新式は付録に示す．なおこの場合各ユーザのデータごとに重みを更新した．

4. 定量評価実験

本論文では，業務者の行動特徴と業績の関係を明らかに

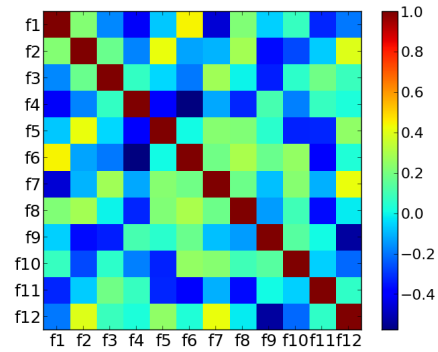


図 1 特徴量の相関まとめ (対面データ)

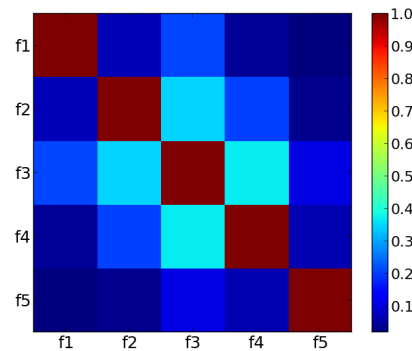


図 2 特徴量の相関まとめ (オープンデータ)

するため前章で用いた手法を使って高業績者の行動特徴を抽出する実験を行う．本章では，各手法がどのような行動特徴を抽出可能かを検証するために2つのデータセットを用いる．まずはそのデータセットの概要と業績に効果があると想定される行動特徴量を示す．

4.1 日報と対面データ

本データはあるオフィスにおける対面のコミュニケーションをビジネス顕微鏡[®] というセンサで収集したものである．このデータによりだれとだれがいつ話したかという情報を得る事ができ，このコミュニケーションの仕方と業績の関係を明らかにすることができる．日報は【ユーザID：日付：業績種類：業績値】の形式で与えられ，対面データは【日付・時間：ユーザID1：ユーザID2：ユーザID3】の形式で与えられる．日報はユーザが出社した日ごと，対面データは1秒ごとにデータが収集されている．

4.1.1 対面データの特徴量

対面データからは次に示す12個の特徴量を作成した． $f1$ ：ユーザが何人とコミュニケーションしたか， $f2$ ：ユーザが上司と何回コミュニケーションしたか， $f3$ ：ユーザが何人のニッチなユーザと頻繁にコミュニケーションしたか(5人以下のユーザとしかコミュニケーションしていないユーザをニッチとした)， $f4$ ：ユーザが2ホップ先にいるユーザの網羅率(網羅率：2ホップ先にいるユーザ数/1ホップ先のユーザ数)， $f5$ ：ユーザが活動する時間の分散，

f_6 : ユーザが同じ部署の人と会話をした割合, f_7 : ユーザが会議に出席した割合, f_8 : ユーザが他部署の人と会議に出席した割合, f_9 : ユーザが自部署のリーダーと会議に出席した割合, f_{10} : ユーザが出席している会議の人数が増加した割合 f_{11} : ユーザが会話したユーザの中で会話回数が一回の人数 f_{12} : ユーザがペアとコミュニケーションした回数

4.1.2 対面データ特徴量の相関

各特徴量の相関係数をまとめる. 図1に縦軸と横軸が対面データ特徴量の各次元, 値が二つの特徴量の相関係数としたものを示す. 図の中の黄色が0.4以上の正の相関を持つ二つの特徴量, 濃い青が-0.4以下の負の相関を持つ二つの特徴量であり, 他の色はそれ以下の相関を持つ二つの特徴量である. 図をみるとほとんどの相関係数が0.5以下であるためほぼ独立な変数が作成できていることがわかる. ただし自己相関は1.0であるが図中では赤で示した.

4.2 オープンデータ(投稿・コメントとイイネ)

本データは AskMe という Q A サイトにおける投稿とそれに対するコメントとそれについてイイネという関係を収集したものである [6]. このデータによりどんな投稿やその投稿がどんなコメントをされたのかという特徴とイイネという評価の関係を明らかにすることができる. 投稿・コメントデータは【ポストID: ユーザID: 日時: 投稿カテゴリ: コメント数: コメント者】の形式で与えられる. イイネデータは【ポストID: 日時: ユーザID】の形式で与えられる.

4.2.1 投稿の特徴量

f_1 : 投稿が何人にコメントされているか, f_2 : 投稿がどれくらいの頻度でコメントされているか, f_3 : 投稿がどのカテゴリに属するか, f_4 : 投稿ユーザが過去何人にコメントされているか, f_5 : 投稿ユーザがどれくらいの頻度で投稿しているか,

4.2.2 投稿の特徴量の相関

各特徴量の相関係数をまとめる. 図2に縦軸と横軸が対面データ特徴量の各次元, 値が二つの特徴量の相関係数としたものを示す. 図の中の薄い青が0.3以上の正の相関を持つ二つの特徴量である. 図をみるとほとんどの相関係数が0.3以下であるためほぼ独立な変数が作成できていることがわかる. ただし自己相関は1.0であるが図中では赤で示した.

4.3 モデルへのデータの適用

各仮説に基づくモデルをデータへ適用するが, 回帰モデルの被説明変数としてそれぞれ日報とイイネ数を与える. 回帰モデルの説明変数としてはそれぞれ定義した特徴量を用いる.

表3 予測精度比較(対面データ)

Table 3 prediction

手法	予測精度
基本モデル	8.27
多次元モデル	22.6
FMs	10.3
混合モデル	15.5
潜在変数モデル	14.3

表4 予測精度比較(オープンデータ)

Table 4 prediction

手法	予測精度
基本モデル	21.3
多次元モデル	39.5
FMs	16.4
混合モデル	13.4
潜在変数モデル	13.8

4.4 評価

今回前章で述べた仮説をそれぞれデータに当てはめて評価を行う. データを学習データとテストデータに分け, 学習データからモデルを学習しテストデータを予測する. この予測精度が最も高かったモデルに対応する仮説がデータのもつ業績と行動の関係であると仮定する. これにより, 最も予測精度が高かったモデルに対応する仮説に基づき業績改善のための施策を構築することがもっとも妥当と判断することができる.

4.4.1 データセット

全データを100分割し, 99セットを学習データに1セットをテストデータに利用してテストデータの値とモデルの予測値の二乗誤差で予測精度を評価する.

4.4.2 予測精度

予測をするにあたり, テストデータのユーザの業績を, そのユーザの特徴量と学習データから学習した重みを用いて予測した業績の二乗誤差の平均を予測精度とした. 式は以下の通りである.

$$\sum_j (y_j - y'_j)^2 \quad (8)$$

ただし, j は業績者を, y_j は業績者 j の業績を, y'_j は業績者 j の特徴量からモデルが予測した業績である.

4.4.3 特徴量選択

各手法が学習し業績を説明するために用いた各データの特徴量を表に示す.

5. 考察

5.1 対面データ

業績と行動の構造に対する仮説としては基本モデルが表す仮説が当てはまり, 業績に効果のある行動は, 良くコミュニケーションをとる行動(間隔があかずに対面する,

表 5 特徴量選択の比較 (対面データ)

手法	非 0 の重み
基本モデル	f1,f2,f3,f5,f6,f10,f11
多次元モデル	f1,f4,f6,f8,f,f,10
FMs	f1,f2,f4
混合モデル	(f1,f2,f12) (f1,f2,f10)
潜在変数モデル	f1,f2,f3,f4,f5,f6,f11

表 6 特徴量選択の比較 (オープンデータ)

手法	非 0 の重み
基本モデル	f1,f3,f4
多次元モデル	f1,f2,f3,f4,f5
FMs	f1,f2,f4,f5
混合モデル	(f1,f2,f4) (f1,f2,f5)
潜在変数モデル	f1,f2,f4,f5

コミュニケーションの相手にバリエーションがある)が業績を高める効果があった。ここから言えることは、対面データの環境では個人差や行動の複雑さは少なく、よくコミュニケーションするというシンプルな行動が必要な業務であることがわかる。この結果を用いると作成すべきシステムは、全員がアプリケーションの内容が同一であり、コミュニケーションを相互に促進するような仕様である。今回のデータは全体の方向性が一貫するような場合に最大の効果をもたらしそうと推測されるため妥当であると考えられる。

5.2 オープンデータ

業績と行動の関係に対する仮説としては混合モデルが表すがあてはまり、業績に効果のある行動は、よくコメントされているか、が業績を高める効果があった。ここから言えることは、オープンデータの環境では個人差は生じているが j 、行動の複雑さは少なく、人によって異なるもののコメントされることで業績も伸びることがわかる。これはデータの取得元である環境がカテゴリごとにイイネなどがつく条件が異なっていることを反映しているため妥当ではないかと考えられる。この結果を用いると作成すべきシステムは、投稿するカテゴリごとに異なるシステムであり促進すべき内容が変化する一方促す行動はそれぞれ独立であってよい(個別に行動を表示し連動性を意識させない)ことがわかる。これもコメント数を伸ばす工夫が線形にイイネに聞く可能性が高い対象であることから妥当であると考えられる。

6. むすび

本論文では行動のログと業績のログを用いて高業績者が行っている行動の特徴やコミュニケーションの特徴を抽出し、それを用いて業績と行動の関係を明らかにし、その

上で業績を改善や向上させる知見を抽出した。仮説を検証し高業績者の行動特徴を抽出するために、またいくつかの仮説に基づくモデルを比較し、業績と行動・コミュニケーションの関係を抽出するために有効なモデルを比較・考察した。今回のデータに対しては、行動と業績が持つ構造についての仮説を明らかにし、特にオープンデータにおいては混合モデルが表す仮説が有効であると検証できた。検証した仮説としては投稿者ごとに行動が及ぼす業績への効果が異なるものの、各行動の相関は考慮しなくてもよい仮説であった。

7. 謝辞

本研究は、NTT データ株式会社様との議論とアドバイスをもとに推進することができました。この場をおかりして御礼申し上げます。

参考文献

- [1] 岡田将吾, 神谷祐樹, 佐藤裕作: センサ環境を利用したオフィスワーカーの行動パターン分析, 人工知能学会全国大会 (2013).
- [2] 渡邊純一郎, 藤田真理奈, 矢野和男, 金坂秀雄, 長谷川智之: コールセンタにおける職場の活発度が生産性に与える影響の定量評価, 情報処理学会論文誌 (2013).
- [3] ビジネス顕微鏡: <http://www.hitachihitec.com/jyohou/business-microscope/solution/microscope.html>
- [4] Lynn Wu, Benjamin N. Waber, Sinan Aral, Erik Brynjolfsson, Alex Pentland: *Mining Face-to-Face Interaction Networks Using Sociometric Badges*, *Predicting Productivity in IT Configuration Task*, MIT Press (2008).
- [5] Steffen Rendle: *Factorization Machines with libFM*, TIST (2012).
- [6] Metafilter Infodump (askme data): <http://stuff.metafilter.com/infodump/>

8. 付録

本付録では 2 章で述べた各モデルをデータから学習するさいに用いる重みの更新式を示す。softmax(x,) はソフトマックス関数であり、 x が $-$ から 0 の間にある場合は 0 正の場合は 1 を引き、負の場合は 0 を足した値を x とする。

8.1 基本モデルの重み更新式

$$w_i^{new} = softmax(w_i', r) \quad (9)$$

$$w_i' = w_i^{old} - g * -2 * (y_j - \sum_i w_i x_{ij}) * x_{ij} \quad (10)$$

8.2 特徴量の相関を考慮するモデルの重み更新式

$$w_i^{new} = softmax(w_i', r) \quad (11)$$

$$w_i' = w_i^{old} - g * -2 * (y_j - \sum_i w_i x_{ij} + \sum_i \sum_{i'} w_{ii'} x_{ij} x_{i'j}) * x_{ij} \quad (12)$$

$$w_{i'i'}^{new} = \text{softmax}(w_{i'i'}', r') \quad (13)$$

$$w_{i'i'}' = w_{i'i'}^{old} - g * -2 * (y_j - \sum_i w_i x_{ij} + \sum_i \sum_{i'} w_{i'i'} x_{ij} x_{i'j}) * x_{ij} * x_{i'j} \quad (14)$$

8.3 特徴量の相関を考慮するモデル: Factorization Machines の利用 の重み更新式

$$w_i^{new} = \text{softmax}(w_i', r) \quad (15)$$

$$w_i' = w_i^{old} - g * -2 * (y_j - \sum_i w_i x_{ij} + \sum_i \sum_{i'} \sum_k v_{ik} v_{i'k} x_{ij} x_{i'j}) * x_{ij} \quad (16)$$

$$v_{ik}^{new} = \text{softmax}(v_{ik}', r') \quad (17)$$

$$v_{ik}' = v_{ik}^{old} - g * -2 * (y_j - \sum_i w_i x_{ij} + \sum_i \sum_{i'} \sum_k v_{ik} v_{i'k} x_{ij} x_{i'j}) * \sum_{i'} v_{i'k} x_{ij} x_{i'j} \quad (18)$$

8.4 混合モデルの重み更新式

各ユーザが最大の所属確率を持つクラスの重みにたいして更新式する .

$$w_{ik}^{new} = \text{softmax}(w_{ik}', r) \quad (19)$$

$$w_{ik}' = w_{ik}^{old} - g * -2 * (y_j - \sum_i w_{ik} x_{ij}) * x_{ij} \quad (20)$$

8.5 潜在変数モデルの重み更新式

$$w_{yk}^{new} = \text{softmax}(w_{yk}', r) \quad (21)$$

$$w_{yk}' = w_{yk}^{old} - g * -2 * (y_j - \sum_k w_{yk} z_{kj}) * z_{kj} \quad (22)$$

$$w_{x_{ik}}^{new} = \text{softmax}(w_{x_{ik}}', r') \quad (23)$$

$$w_{x_{ik}}' = w_{x_{ik}}^{old} - g * -2 * (x_{ij} - \sum_k w_{x_{ik}} z_{kj}) * z_{kj} \quad (24)$$

$$z_{kj}^{new} = \text{softmax}(z_{kj}', r) \quad (25)$$

$$z_{kj}' = z_{kj}^{old} - g * -2 * (y_j - \sum_k w_{yk} z_{kj}) * w_{yk} \quad (26)$$

$$-g * s * -2 * (x_{ij} - \sum_k w_{x_{ik}} z_{kj}) * w_{x_{ij}}$$