DEIM2020 J8-3(day2 p79)

類似工場推薦のための特徴表現

服部　雄也† 　湯本　高行† 　芦田　真一†† 　井上　直樹†† 　儀川悌次郎†
上浦　尚武†

† 兵庫県立大学大学院工学研究科　〒 671–2280 兵庫県姫路市書写 2167
†† 株式会社 NC ネットワーク　〒 111–0052 東京都台東区柳橋 1 丁目 4-4
E-mail: †ei19h020@steng.u-hyogo.ac.jp, †{yumoto,isokawa,kamiura}@eng.u-hyogo.ac.jp,
††{ashida,n.inoue}@nc-net.or.jp

あらまし　近年，企業経営においては事業継続計画が重要視されるようになっている．特に製造業では，平常時の取引先が操業できなくなった場合の代替の取引先の確保が重要とされている．そこで，本研究では，指定した工場に類似する工場を推薦するための工場の特徴表現を提案する．特徴表現の獲得には，工場検索サイトに登録された加工分類，設備分類のタグ情報を用いる．この情報から工場-分類タグ行列を作成し，LDA や NMF の次元削減を行って，工場の特徴ベクトルを生成する．この特徴ベクトルのコサイン類似度を用いて工場推薦が実現できる．作成した特徴ベクトルの評価は，工場の業態を専門家によってラベルづけした結果に対して，クエリとして与えた工場と同一のラベルづけがされた工場が上位に推薦されるかどうかで行う．

キーワード　LDA，NMF，工場推薦，製造業

1 はじめに

近年，企業経営においては事業継続計画が重要視されるようになっている．特に製造業では，平常時の取引先が操業できなくなった場合の代替の取引先の確保が重要とされている．加えて近年ではインターネットの普及により，代替取引先となる工場を検索サイトを利用し探すことが可能となった．その一方で膨大な検索結果から求める技術や設備を有する工場を迅速に見つけ出すことは困難になっている．たとえば用途が同じ部品であっても形状や使用する材料，大きさによって適切な加工方法が異なるほか，輸送コストに関係する工場の立地，必要な納品数を満足できる規模の工場であるかも発注の際には重要になる．こういった判断基準は長年の知識や経験により培われるものであり，技術に対する知識の浅い人間には判断が難しい．また知識の深い人間であっても，複数の工場を探し出して比較検討するには時間と労力がかかる．もし類似する技術や設備を持つ工場同士を機械的に分類することができれば代替工場を探す検索の労力を少なくすることができ，製造業の円滑化が見込まれる．

そこで本研究では，指定した工場に類似する工場を推薦するための工場の特徴表現を提案する．特徴表現の獲得には，工場検索サイトに登録された加工分類，設備分類のタグ情報を用いる．この情報から工場-分類タグ行列を作成し，LDA や NMF の次元削減を行って，工場の特徴ベクトルを生成する．この特徴ベクトルのコサイン類似度を用いて工場推薦が実現できる．

作成した特徴ベクトルの評価は，工場の業態を専門家によってラベルづけした結果に対して，クエリとして与えた工場と同一のラベルづけがされた工場が上位に推薦されるかどうかで行う．

2 関連研究

Web サイトの情報を用いて企業の業種を分類する佐々木らの研究 [1] がある．しかし，これは Web サイトのテキスト情報から機械学習を行って分類するという点で，タグ情報を利用する本研究の分類手法とは異なっている．また Web サイトのタグ情報を利用する Web ページ分類については丹羽らの研究 [2] がある．しかし，利用するタグ情報がソーシャルブックマークという多くのユーザの手によって編集可能な Folksonomy 形式のタグであること，分類対象が特定の分野に依存しない広く一般的なサイトを取り扱っていることなどが本研究と異なる．本研究では取り扱う加工分類タグ情報はユーザが設定するもののタグ名や種類はユーザによる編集はなく，分類対象となるサイトも製造業の工場検索に限定したものである．

3 手　　法

3.1 問題設定

まずはじめに，本研究で取り扱うデータの構成および問題設定について説明する．本研究では株式会社 NC ネットワークが提供する工場検索サイト“エミダス工場検索”[1]に掲載されている工場 17355 社を対象としている．掲載工場には個々に Web ページが割り当てられており、ページには以下の情報が記載さ

1 : https://www.nc-net.or.jp/

れている．

- 社名
- 工場の住所
- 工場 HP へのリンクアドレス
- 従業員数
- 資本金と年間売り上げ高
- 工場のセールスコメント (自由記述)
- 工場が主に営業先とする業界 (自由記述)
- 工場の主な生産品目 (自由記述)
- 工場の主な加工分類 (複数選択式)
- 工場が保有する設備情報

これらの情報を利用し，指定した工場の代替工場となりえる別の工場を選び出すタスクを問題とする．本研究ではページに複数設定されたタグ形式情報であり，工場が対応可能な技術情報を表す加工分類と，工場の加工技術に直結する工場設備情報の二つに着目し，同一の加工技術，加工設備を有しているか，という観点から類似工場を選択する．加工分類はツリー状にあらかじめ設定されたタグ形式の情報で，工業分野における技術や製造法を示している全 1204 項目の情報である．各工場はこの加工分類の中から自工場に適していると考えられるタグを任意の数だけ設定することが可能である．

加工分類の例を表 1 に示す．

表 1　加工分類の例

大分類	中分類	小分類
量産	カッティング・ブランク	タレパン加工
	金属プレス	打抜き
	プラスチック	射出成形
	鍛造	鉄
試作開発・少量生産	機械加工	汎用フライス加工
	研削加工	平面研削加工
金型製作	バネ金型	板バネ金型
	プレス金型	トランスファー型
部品製造	鋲螺	ネジ製造
表面処理	塗装	粉体塗装

工場の設備情報はその工場が有する加工機械の大きさや型番号のほか、加工機名がタグ形式で登録されている．このうち，加工機を表すタグのなかから「測定器」「顕微鏡」「矯正機/給材機」「バリ取り機」「洗浄機」「付帯設備」「その他」に属する設備は製品の加工とは直結しないと判断し，これらのカテゴリを除いた全 256 項目を分類に用いる．

設備分類の例を表 2 に示す．

3.2　工場のベクトル表現

本節では工場間の類似度を求めるための，ベクトル空間モデルを用いた工場表現について説明する．工場と分類タグの関連を表現するために工場-加工分類行列と、工場-設備分類行列を作成した．工場-加工分類行列および工場-設備分類行列 W は次の式で定義する．

表 2　設備分類の例

大分類	中分類	小分類
機械加工	マシニングセンタ	横型マシニングセンタ
		五軸マシニングセンタ
	フライス盤	NC フライス盤
板金加工	曲げ機	ベンダー
	溶接機	TIG 溶接機
鋳造	ダイカストマシン	ダイカストマシン
樹脂/ゴム加工	RP 造型機	光造形機

$$W = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1m} \\ s_{21} & s_{22} & \cdots & s_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ s_{n1} & s_{n2} & \cdots & s_{nm} \end{pmatrix} \tag{1}$$

工場-加工分類行列は行が工場，列が加工分類を表し，工場-設備分類行列は行が工場，列が設備分類を表す n 行 m 列の行列である．n は工場数，m は加工分類数あるいは設備分類数である．行列の各要素値 s は横軸と対応する分類タグが対象企業に設定されている場合は 1，設定されていない場合は 0 とする．この行列により各工場と設定された加工分類，設備分類の関係を表現する．次にこの行列から特徴ベクトルを作成し，工場分類に利用する．工場の特徴ベクトルの作成手法として以下に示す三つの手法を提案する．

3.2.1　全分類を利用する特徴ベクトル作成

次元削減を行わず，工場-加工分類行列または工場-設備分類行列 W の行 $W_i = (s_{i1}, s_{i2}, \cdots, s_{im})$ を工場の特徴ベクトルとして用いる．すなわち特徴ベクトルの次元数は加工分類の項目総数 m に等しい 1204 次元，もしくは設備分類の項目総数に等しい 256 次元となる．

3.2.2　NMF による特徴ベクトル作成

工場-加工分類行列および工場-設備分類行列を非負値行列因子分解 (NMF) することで工場の特徴ベクトルを作成する．NMF は行列分解の手法の一つで，分解する行列の要素値を非負値に制限して近似的な行列分解を行う手法 [3] である．また NMF を文書分類問題に利用する研究 [4] もある．本研究では工場-加工分類行列 W を，NMF を用いて $W = XH$ と，工場-基底ベクトル行列 X と基底ベクトル-加工分類行列 H の積に，工場-設備分類行列を工場-基底ベクトル行列と基底ベクトル-設備分類行列の積にそれぞれ分解する．この工場-基底ベクトル行列 X の行 $X_i = (x_{i1}, x_{i2}, \cdots, x_{ij})$ を対象工場の特徴ベクトルとして定義することで，分類の次元数を基底ベクトルの次元数 j まで削減する．

3.2.3　LDA による特徴ベクトル作成

Latent Dirichlet Allocation(LDA) [5] を用いて工場-加工分類行列および工場-設備分類行列の次元削減を行い，工場の特徴ベクトルを作成する．LDA はトピックモデルの一種で一つの文章が複数のトピックから確率分布に従い生成されることを仮定したモデルである．LDA による文書生成のグラフィカルモデルを図 1 に示す．

ここで α，β はディリクレ分布のハイパーパラメータであ

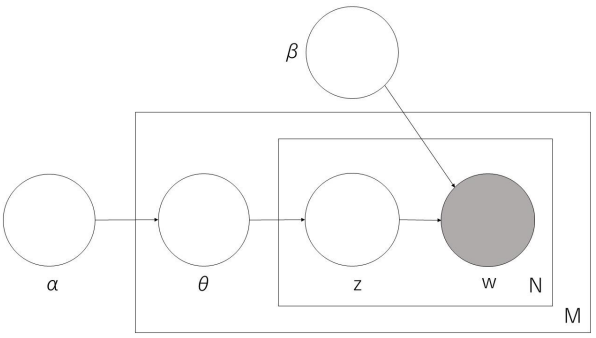

図 1　LDA のグラフィカルモデル

り，M は単語数，N は文書数，w は単語を意味する．z はその単語のトピック，θ はトピック分布のパラメータである．

本研究では一つの工場に設定された加工分類もしくは設備分類の集合を一つの文書として扱い，各分類項目を一つの単語として，複数の分類項目の確率分布で表現される加工分類トピック，設備分類トピックを作成して工場の特徴ベクトルに利用する．LDA によって得られた加工分類トピック，設備分類トピックを特徴ベクトルの次元とし，特徴ベクトルを $X_i = (x_{i1}, x_{i2}, \cdots, x_{ij})$ として定義する．特徴ベクトルの要素値 x は対象工場の加工分類群，設備分類群が各トピックから生成される確率値とする．特徴ベクトルの次元数 j は LDA によって得られた加工分類トピック数，もしくは設備分類トピック数である．これにより分類の次元数を LDA のトピック数まで削減する．また LDA で加工分類や設備分類から潜在的なトピックを可視化することにより，各次元を構成する加工分類群に対してどのような工業分野の加工分類集合が存在しているのか，特定の工場に対しどのような工業分野の特徴が現れているのか，解釈付けを行うことが可能となる．加えて LDA は特徴として一つの要素が複数のトピックに属することを確率的に許容するモデルであるため，幅広い分野で用いられる基礎的な加工技術や加工設備を表すタグを特定のトピックに限定せず，特徴ベクトルの次元として用いることが可能であると考える．

3.3　類似工場の推薦

工場間の類似度の表現にはコサイン類似度を用いる．入力工場の特徴ベクトルを A，比較工場の特徴ベクトルを B とすると，二つの特徴ベクトルのコサイン類似度は以下の式で定義される．

$$cos(A, B) = \frac{A \cdot B}{|A||B|} \tag{2}$$

入力に指定した工場の特徴ベクトルと比較工場の特徴ベクトルからコサイン類似度を計算し，得られたコサイン類似度が高い工場ほど，指定した工場により類似する工場であるとして推薦する．

4　実　　験

4.1　使用するデータ

本節では評価実験で用いたデータの仕様について述べる．評価実験の対象とした工場データは株式会社 NC ネットワークが提供する“エミダス工場検索”に掲載されている工場から抽出した 240 社である．データには各工場が主に対応する加工方法，業態を表すラベルづけが著者以外の一名の専門家の手で行われており，この業態ラベルを正解データとして分類結果に対する評価を行う．業態ラベルの種類は種類である．割り当てられたラベル名とラベルごとの企業数の一覧を表 3 に示す．

表 3　工場の業態ラベル一覧

ラベル名	工場数	ラベル名	工場数
ガラス・セラミック成形	7	ギア製造	15
バネ製造	9	プレス・板金・製缶	39
基板実装・組立	11	機械器具	7
機械加工	40	研磨・ラップ	12
治具・工具・金型部品	10	樹脂・ゴム成形	19
金型	7	樹脂切削	7
旋盤加工	11	鍛造・圧造	11
鋳造・ダイカスト	17	表面処理	18

4.2　実 験 設 定

LDA のトピックを利用した特徴ベクトルを作成するため，前処理として LDA による加工分類トピックと設備分類トピックを作成した．作成するトピックの数は 30 とした．このとき，多くの工場が有する汎用的な加工分類，設備分類が数多くのトピックに現れ特徴ベクトルの次元が類似することを防ぐため，TFIDF の考え方から IDF を利用して分類タグに重み付けを行う．トピック作成に使用する工場ページ総数を N とし，分類タグ w が含まれる工場ページの出現頻度を $Frequency(w)$ として，IDF による重みは次の式で定義する．

$$IDF(w) = log_2 \frac{N}{Frequency(w)} \tag{3}$$

作成したトピックを加工分類および設備分類の確率分布を用いてワードクラウド表現で可視化したものをいくつか例として図 2, 図 3 に示す．

また NMF による特徴ベクトル作成のため，工場-加工分類行列および工場-設備分類行列 W に対して NMF を行い，基底ベクトルを得た．このとき，分解するベクトルの基底数を 30 とした．

4.3　評 価 方 法

クエリとして与えた 1 つの工場に対して残り 239 工場のコサイン類似度を算出し，コサイン類似度が高いものから順位をつけランキングを作成し，得られたランキングについて評価を行う．本論文ではランキングの評価指標として上位 k 件での適合率 ($P@k$) と平均逆順位（MRR）を使用する．上位 k 件での適合率は，任意の順位 k 位までの項目での適合率である．

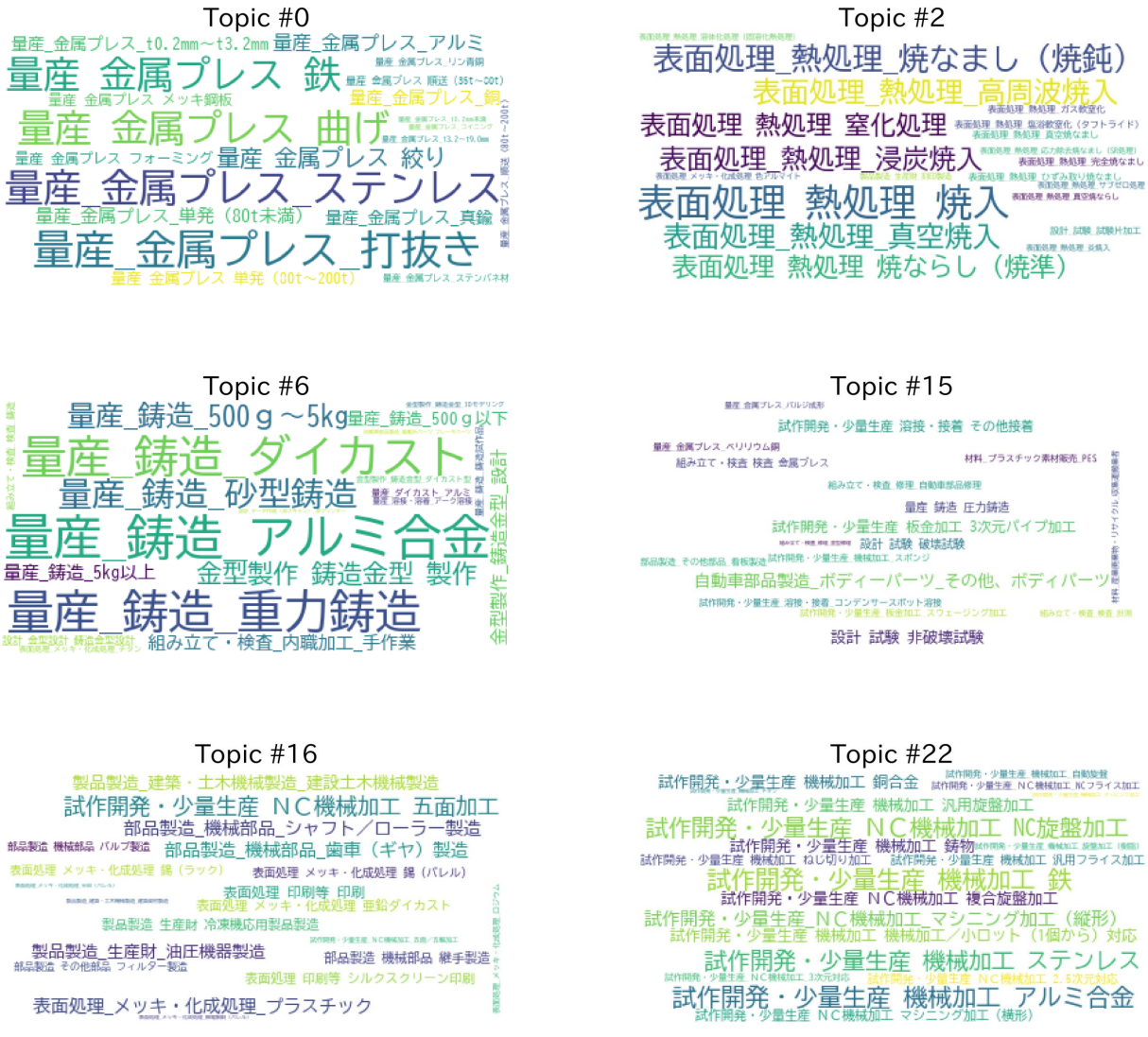

図 2　加工分類のトピック例

分類結果の混同行列を表 4 に示し，適合率の定義式を以下に示す．

表 4　混 同 行 列

分類結果＼正解ラベル	同一	同一ではない
同一	TP	FP
同一ではない	FN	TN

$$precision = \frac{TP}{TP + FP} \tag{4}$$

本論文では $k = 1$，$k = 3$，$k = 5$ のときに，クエリとして与えた工場と同一の業態ラベルを持つ工場が推薦されている場合を正解として適合率を求め，実験に使用した 240 工場それぞれをクエリとして与えた場合の適合率の平均値を評価に使用した．

平均逆順位は得られたランキングで k 件目に目的とする項目がはじめて表れたとき，その順位の逆数の平均を評価値とする評価指標である．本実験では，クエリとして与えた工場と同一の業態ラベルを持つ企業がはじめに現れた順位の逆数の平均を評価値とした．企業数を n，同一の業態ラベルを持つ企業がはじめに現れた順位を k とすると，平均逆順位は次の式で定義される．

$$MRR = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{k_i} \tag{5}$$

4.4　実 験 結 果

4.4.1　加工分類による推薦

工場の加工分類を利用し，次元削減を行わずに全ての加工分類を用いた特徴ベクトル，NMF による次元削減を行った特徴ベクトル，LDA による次元削減を行った特徴ベクトルを用いた工場推薦について，各評価指標の結果を表 5 に示す．

また，LDA のトピックごとの分類を確認するため，LDA トピックを利用した推薦と全加工分類を利用した推薦のそれぞれで工場の業態ラベルごとに上位 5 件の適合率の平均値を求めた結果を表 6 に示す．

図 3　設備分類のトピック例

表 5　加工分類を利用した推薦

分類方法	P@1	P@3	P@5	MRR
全加工分類	0.61	0.56	0.53	0.71
NMF	0.54	0.52	0.50	0.65
LDA	0.57	0.53	0.51	0.67

表 6　業態ラベルごとの P@5（加工分類）

業態ラベル	LDA	全加工分類
ガラス・セラミック成形	0.26	0.34
ギア製造	0.48	0.61
バネ製造	0.38	0.49
プレス板金	0.86	0.82
基板実装・組立	0.53	0.69
機械器具	0.11	0.00
機械加工	0.44	0.58
研磨・ラップ	0.17	0.30
治具・工具・金型部品	0.16	0.18
樹脂・ゴム成形	0.54	0.46
金型	0.14	0.28
樹脂切削	0.17	0.46
旋盤加工	0.22	0.18
鍛造・圧造	0.55	0.58
鋳造・ダイカスト	0.74	0.49
表面処理	0.77	0.74

表 5 について，全ての加工分類を利用した特徴ベクトルを利用した推薦が最も良い結果となり，上位 5 件での適合率が 0.53 で平均逆順位が 0.71 と，LDA を利用した特徴ベクトルによる推薦，NMF による特徴ベクトルによる推薦での評価値を上回った．しかし表 6 から業態ごとの適合率を確認すると，「プレス板金」や「鋳造・ダイカスト」，「表面処理」などのいくつかの業態では LDA による特徴ベクトルを利用した推薦結果の方が高い適合率となっている。

4.4.2　設備分類による推薦

次に，工場の設備分類を利用した場合についても同様に，次元削減を行わずに全ての設備分類を利用した特徴ベクトル，NMF による次元削減を行った特徴ベクトル，LDA による次元削減を行った特徴ベクトルを用いた工場推薦を行った．各評価指標の結果を表 7 に示す．

表 7　設備分類を利用した推薦

分類方法	P@1	P@3	P@5	MRR
全設備分類	0.40	0.38	0.36	0.53
NMF	0.18	0.25	0.27	0.35
LDA	0.31	0.27	0.26	0.45

また，設備分類についても LDA のトピックごとの分類を確認するため，LDA トピックを利用した推薦と全設備分類を利用した推薦のそれぞれで工場の業態ラベルごとに上位 5 件の適合率の平均値を求めた結果を表 8 に示す．

表 7 より，設備分類を用いた特徴ベクトルについても全ての設備分類を利用した特徴ベクトルによる推薦が最も良い結果となり，上位 5 件での適合率が 0.36，平均逆順位が 0.53 と，LDA を利用した特徴ベクトルによる推薦や NMF による特徴ベクトルによる推薦での評価値をどちらも上回った．

表 8　業態ラベルごとの P@5（設備分類）

業態ラベル	LDA	全設備分類
ガラス・セラミック成形	0.11	0.43
ギア製造	0.23	0.49
バネ製造	0.16	0.22
プレス板金	0.55	0.62
基板実装・組立	0.60	0.69
機械器具	0.02	0.14
機械加工	0.34	0.43
研磨・ラップ	0.37	0.50
治具・工具・金型部品	0.24	0.14
樹脂・ゴム成形	0.06	0.16
金型	0.11	0.06
樹脂切削	0.03	0.20
旋盤加工	0.07	0.13
鍛造・圧造	0.20	0.42
鋳造・ダイカスト	0.09	0.16
表面処理	0.13	0.22

5 考　　察

工場の加工分類情報を利用した推薦について，表 5 より全ての加工分類を利用した特徴ベクトルを利用した推薦が最も良い結果となったが，表 6 から業態ごとの適合率を確認すると，「プレス板金」や「鋳造・ダイカスト」，「表面処理」などのいくつかの業態では LDA による特徴ベクトルを利用した推薦結果の方が高い適合率となった。

加工分類から作成した LDA トピックの詳細を図 2 から確認すると，Topic#0 では「金属プレス」，Topic#2 では「表面熱処理」，Topic#6 では「鋳造」など特徴的な加工方法をまとめたトピックが作成されている．これらのトピックは工場の業態ごとの特徴を明確にし，その結果全ての加工分類を利用した特徴ベクトルによる推薦結果よりも特定の業態においては適合率が高くなったものと考える．これより，LDA を用いて類似した加工分類情報を適切にまとめてトピックごとに加工方法の特徴を得られれば，より精度の高い工場推薦につながると考える．

一方で Topic#22 では機械加工のなかの「旋盤加工」，「マシニング加工」，「フライス加工」といった複数の加工方法が混在していて，機械加工のなかでも用途が異なる加工方法同士の違いが明確でない．さらに Topic#15 では「金属プレス」，「パイプ加工」，「検査」など，Topic#16 では「歯車製造」，「表面処理」などの加工分類がまとめられており，一つのトピックにまとめられた加工分類の共通点が乏しく区分が明確でないトピックも存在する．特に「機械加工」に関連する曖昧なトピックが複数存在することで，「機械加工」や「旋盤加工」，「機械器具」とラベルづけされた企業の特徴ベクトルが複数の次元に値を持ち，他の多くの企業と同一の業種であるとして判断され，精度の低下に繋がった．

また工場の特徴として「広い加工方法に対応可能だが，特定の加工方法に対して特に強みがある」という工場の推薦については工場ページに設定された加工分類タグが多岐にわたり，工場の特徴ベクトルが複数の次元に値を持つことでコサイン類似度が下がって上位に推薦されない，あるいは異なる業種ラベルが割り当てられた工場に対して類似すると推薦されたケースがあった．

次に工場の設備分類情報を利用した推薦について，表 7 より，設備分類を用いた特徴ベクトルについても全ての設備分類を利用した特徴ベクトルによる推薦が最も良い結果となった．しかし，表 5 の加工分類を利用した場合に比べてどの手法でも推薦結果の評価値が低くなっている．これは加工分類に比べて工場の設備分類は設備台数の点から分類項目が少ない，汎用的な加工機械が存在し，工場の業種ごとの違いが適切に表現できていないなどの理由が考えられる．

設備分類から作成した LDA トピックの詳細を図 3 から確認すると，Topic#3 では汎用的な機械加工機が中心でマシニング設備がないことから，少量製作が中心の機械加工業者であると予想される．Topic#4 は精密板金加工に用いる加工設備が集まっている板金業者であるなどの特徴が得られた．しかし Topic#10 では旋盤機と光造形機が同時に現れており，これら二種の加工機が両立する業種はない．そのためこのトピックは工場の特徴を表す上で不適切なトピックであると考えられる．また Topic#20 ではトピックに表れている加工機械に関連性が見いだせないなど，トピックの特徴が明確でないものも存在している．特に設備分類から作成した LDA トピックの中には樹脂やゴムの加工業者の特徴を表しているトピックが存在せず，「樹脂・ゴム成形」や「樹脂切削」のラベルの評価値が大きく下がったものと考える．

これらの問題点の改善には，LDA のトピックを利用した特徴ベクトルの次元として適切となるよう，LDA トピックが曖昧であったものをさらに明確に分類する必要がある．本研究では特徴ベクトルの次元数を 30 としたが，分類に適切な次元数を探索し特徴ベクトルの最適化を行うという方法が考えられる．また，今回は加工分類のみ，設備分類のみとそれぞれ独立して特徴ベクトルを作成し工場の特徴を表現したが，これら二つの分類項目を組み合わせることでより推薦の精度をあげることが可能かどうか検証したい．

6 ま と め

工場が対応する加工法を示す加工分類タグ情報から工場と加工分類間，あるいは工場と設備分類間の関係を行列で表現し，LDA や NMF の次元削減を行い特徴ベクトルを作成する手法を提案した．また，作成した特徴ベクトルを用いて類似する工場の推薦を行った．実験の結果，次元削減を行わずに全ての加工分類，設備分類を利用する手法が最も推薦の評価が高い結果となった．しかし，「プレス板金」や「鋳造・ダイカスト」といった特定の加工法に対しては加工分類から作成した LDA のトピックを用いた特徴ベクトルによる推薦のほうが高い適合率を得られた．今後は LDA のトピックを最適化する，あるいは工場の特徴ベクトル作成に加工分類と設備分類の二つを組み合わせる，その他工場の特徴を表現するために適切な特徴量を分

類に取り入れるなどの操作を行い，工場推薦の精度の向上を目指す．

謝辞 本研究は JSPS 科研費 JP17K00429 の助成を受けたものです．ここに記して謝意を表します．

文　　献

[1] 佐々木 稔, 新納 浩幸. ”文書分類手法を用いた企業 Web サイトからの業種分類”. 言語処理学会第 12 回年次大会論文集, 2006

[2] 丹羽 智史, 土肥 拓生, 本位田 真一. ”Folksonomy マイニングに基づく Web ページ推薦システム”. 情報処理学会論文誌 Vol.47 No.5, pp.1382–1392, 2006

[3] Daniel D. Lee, H.sebastion Seung. ”Algorithms for Non-negative Matrix Factorization”. Advances in Neural Information Processing Systems, pp.556–562, 2001

[4] Wei Xu, Xin Liu, Yihong Gong. ”Document Clustering Based On Non-negative Matrix Factorization”. proceedings of the 26th annual international ACM SIGIR, conference on Research and development in informaion retrieval, pp.267–273, 2003

[5] David M.Blei, Andrew Y.Ng, Michael I.Jordan. ”Latent Dirichlet Allocation”. Jounrnal of Machine Learning Research 3, pp.993–1022, 2003