Vol.2018-DBS-167 No.5
Vol.2018-IFAT-132 No.5
2018/9/12

ジオタグツイートの多言語性と評判に基づくVenue推薦

中岡 佑輔[1,a)]　パノット シリアーラヤ[1,b)]　王 元元[2,c)]　河合 由起子[1,3,d)]　秋山 豊和[1,e)]

概要：本研究では，ジオタグツイートの発信場所と言語さらにレビューの評判情報から群衆（国民）の嗜好性を抽出し，類似性からツイートの少ない低密度地域でも国民性に合わせたVenueを推薦可能なシステムの実現を目指す．提案手法はCF手法に基づき，ツイート数（レビュー数）の少ないVenue（アイテム）の評価値を，そのVenueのジャンル（例えばインド料理店）に対する他の国民の嗜好性との類似度から算出する．さらに各Venueに対するレビューとVenueに対するツイートの極性を評判情報として抽出し，ジャンルに基づき選出したVenueを評判情報に基づき選出する．なお，言語ごとにツイート数の多い地域では従来手法の出現頻度よりVenueの評価値を算出し，国民ごとにVenueを抽出推薦する．本稿では，特に多様な国民性が共存するヨーロッパを対象とし，ツイートの時空間と言語，評判情報に基づく群衆の嗜好性抽出およびVenue推薦手法について述べ，欧州の複数の都市に対するフランス語を母国語とする被験者89名を対象とした評価実験を行い，提案手法より抽出したVenue推薦精度の有用性を検証する．

キーワード：ジオタグ付ツイート分析，ツイート多言語分析，Venue推薦，AMTurk検証

Yusuke Nakaoka[1,a)]　Panote Siriaraya[1,b)]　Yuanyuan Wang[2,c)]　Yukiko Kawai[1,3,d)]
Toyokazu Akiyama[1,e)]

1. はじめに

近年，ユーザの行動分析および可視化に関する研究において，ジオタグ付きのソーシャルネットワークサービス（SNS）データ分析に関する研究開発が盛んに行われている．都市に存在する店舗や施設などでCheck-inするユーザの移動軌跡を分析し，その都市の特徴を抽出する手法[1]や，タクシーに設置したGPSから取得した人々の移動パターンと地域に存在する施設のカテゴリ情報を用いて地域の機能性を発見する手法[2]が実証されている．これまで著者らも，ユーザ行動分析としてデータ発生位置とコンテンツで言及されている位置との差異，発生時間とコンテンツ言及時間との差異分析，さらに位置と時間の関係性を考慮した時空間差異分析および可視化に関する研究を行ってきた[4]．これにより，ユーザの関心を時空間の観点から俯瞰することが可能となったが，ユーザ特性（年齢や性別，人種）までは考慮しておらず，群衆の嗜好性に基づいた情報推薦までには至っていなかった．また，ジオタグツイートがツイートに占める割合は数パーセントと低く，都市部以外では適応が困難という根本的問題が残る．

そこで，本研究では，ジオタグツイートから時空間情報となる場所と時間以外に，発信ユーザが登録する母国語および内容に記述されている言及言語の言語情報を考慮することで，発信位置（国）と言語（国）との同一性から群衆（国民）のジャンルに対する嗜好性を抽出し，それら各国民間の類似性からジャンルをランキングし，さらにジャンルのVenueに対するレビュー数の評価値よりVenueを推薦する．これによりツイートの少ない地域も含めたいずれの場所でも嗜好性の高い情報の推薦が可能となる．例えば，フランス人のツイートが少ない「ローザンヌ」において，類似度の高いスペイン人の嗜好や類似度は低いがツイートの多いイタリア人の嗜好も考慮することで，フランス人にとって指向性の高いVenue推薦が可能となる．

本論文では，対象領域を多言語性の高いヨーロッパ19カ国とし，指定言語に応じたVenue推薦システムを構築し，フランス語を母国語とする被験者89名から有用性を検証

[1] 京都産業大学, 〒 603–8555 京都市北区上賀茂本山
[2] 山口大学, 〒 755–8611 宇部市常盤台 2–16–1
[3] 大阪大学, 〒 5653–0871 吹田市山田丘 2 番 8 号
[a)] g1444936@cc.kyoto-su.ac.jp
[b)] k6180@cc.kyoto-su.ac.jp
[c)] y.wang@yamaguchi-u.ac.jp
[d)] kawai@cc.kyoto-su.ac.jp
[e)] akiyama@cc.kyoto-su.ac.jp

する．具体的には，まず取得したツイートから Venue 名を抽出し，Venue 名と発信位置から Venue の属性情報となるジャンル名を取得する．ジャンル名は「BAR」や「CAFE」など 100 種類程度の統一形式となるため，数十万以上の固有の Venue 名を用いた言語国の類似度抽出（次のステップ）で生じるコールドスタート問題を回避できる．次に，発信位置（国）ごとに同一の言語（国）のツイートを分類し，それらのジャンル名の出現頻度を算出し，各言語国間の相関係数を類似度として算出する．また，ユーザ指定の地域内のツイートの Venue の出現頻度をツイートから算出し，同時に Venue に対する google と Foursquare の複数の rating より評価値を算出し，最後に類似度，出現頻度，評価値より算出したスコアの高い Venue をマップ上に提示する．

本論文では，ジオタグツイートの時空間ならびに言語分析に基づく群衆の嗜好性抽出および Venue に対する評判に基づいた Venue 推薦手法を提案し，欧州の 13 ヶ月分のジオタグツイートを用いて構築した Venue 推薦精度を検証する．

2. 関連研究

大量のジオタグツイート（以下，ツイート）に対する時空間分析に関する研究が，国内外で広く取り組まれている．

Qu ら [3] は，レストランや店舗などの特定の店舗で Check-in した際に発信されるツイートを分析し，ユーザの移動軌跡を抽出し，そのレストランや店舗などのトレードエリアの発見を行った．また，一定領域の分析結果を地図の LOD に同期し可視化することで効果的な時空間解析が実証されている [5]．さらに，地域に特色のある語と位置情報に新たな地域ユーザを手がかりとして付け加えた口コミ収集の提案 [8] や，観光客に関する情報を抽出する研究の 1 つとして Twitter に投稿されたツイートの位置情報と本文を用いることで，ユーザの観光地での訪問動向より訪問目的を推定する手法の提案 [9] などの研究が行われている．

これまで著者らも，ユーザ行動分析として日米両国の数ヶ月間のツイートを分析し，データ発生位置とコンテンツ内容位置との差異，発生時間と内容時間との差異の分析，さらに位置と時間の関係性を考慮した時空間差異の分析および可視化に関する研究を行ってきた [6]．また，ツイートの時間と場所と言語に基づき分析し，ユーザ行動に対する場所と言語の相違の可視化に関する研究を行ってきた [7]．

以上，既存研究を含めジオタグの時間および位置情報分析に関する研究は広く行われているが，これらに加えて言語情報から群衆（国民）の特性を抽出し，さらに群衆間の類似性および位置特性に基づき任意の場所のいずれにおいても Venue（地物）推薦を可能にする研究開発は稀である．

3. 位置と言語分析に基づく Venue 推薦手法

本章では，任意の場所における言語（国民）の嗜好性抽出ならびに Venue 推薦，可視化手法について述べる．Venue 推薦システムの処理の概要（ステップ）を以下に示す．

(1) 各言語国の Venue のジャンルに対する出現頻度算出
(2) 言語国間のジャンルに基づく類似度抽出
(3) Venue の rating に基づく評価値算出
(4) 任意地域の各言語国の Venue に対する出現頻度算出
(5) 任意地域の各言語国のジャンルに対する評価値算出
(6) Venue 数が閾値以上の場合は (3) と (4) より Venue のスコア算出
(7) Venue 数が閾値未満の場合は (3)，(4) と (5) より Venue のスコア算出
(8) マップ上に任意地域の言語毎の Venue をスコアの高い順に推薦提示

3.1 発信場所と言語に基づく Venue 抽出

まず，ジオタグツイートの発信位置，発信時刻，母国語および言及言語を抽出し，任意の期間と地域と言語に基づきツイートを分類する．ここで母国語とは，ユーザがツイート利用登録時に設定する言語とし，言及言語はツイートの内容に用いられている言語とする．この母国語と言及言語より，任意の言語 l は $\{$ 母国語 $_l\} \vee$（言及言語 $_l \subseteq \overline{\text{母国語}_l}$）として分類される．例えば，フランス人の嗜好性抽出では，任意の言語 $l_{\text{フランス}}$ は，母国語がフランス語の全てのツイートおよび母国語がフランス語以外で言及言語がフランス語のツイートが分類される．

次に，分類された言語ごとの Venue 辞書を作成する．Venue 辞書は，言語，緯度経度，地物名，属性情報のタプルであり，ツイートの定式文となる “I’m at” とマッチングしたツイートの定式文以降に記載される単語を地物名（Venue）として抽出する．属性情報は，抽出した Venue 名を用いて Swarm API[*1]から取得したカテゴリとジャンルとし，ジャンルはカテゴリの下位層になる．例えば，カテゴリは「公共施設」や「フード」などで，「フード」の下位層のジャンルには「中華」や「喫茶店」などが含まれる．

各言語の Venue 辞書に基づき，全言語 L に対して言語 l_x の言語国の都市 p でのみ発信された各ジャンル j に対する嗜好性となる評価値を出現頻度 $TF_{\{x,j\}}=$（l_x におけるジャンル j 出現回数）／（l_x におけるジャンル総出現回数）から算出する．例えば，l_x=フランス語の母国フランスの都市 p=パリ周辺で発信されたツイートのジャンル j=カフェの出現頻度から，フランス人（この場合はパリ人）のカフェに対する嗜好性となる評価値が算出される（ステップ 1）．

算出した言語 l_x のジャンル j に対する評価値 $TF_{\{x,j\}}$ と

[*1] https://developer.foursquare.com/

表 1　ジオタグツイートの収集数と分類結果 (下線：対象都市の母国語).

言語	#Tweet 総数	“I’m at” を含む (%)	#Venue 総数 (%)	ロンドン	ローマ	パリ	バルセロナ	ベルリン	リスボン	アムステルダム
全言語	25,993,771	1,231,980(4.7%)	342,992(1.3%)	-	-	-	-	-	-	-
イタリア	2,251,204	98,488(3.6%)	36,940(1.6%)	2,914	6,203	369	1,706	81	39	153
フランス	2,430,737	36,163(1.4%)	29,851(1.2%)	1,568	363	16,445	797	5	157	209
スペイン	4,801,999	40,367(0.8%)	34,813(0.7%)	3,624	3,419	868	20,614	117	240	464
ドイツ	2,041,920	216,242(8.6%)	55,414(2.7%)	1,454	367	211	820	873	44	276
ポルトガル	881,874	24,585 (2.8%)	22,359 (2.5%)	634	115	479	373	131	2,127	313
オランダ	1,671,522	257,383(15.4%)	269,413 (16.1%)	197	67	368	261	68	101	3,165
合計	14,079,256	673,228(4.8%)	448,790 (3.2%)	10,391	10,534	18,750	24,571	1,275	2,708	4,580

他言語 l_y の評価値 $TF_{\{y,j\}}$ より，x 国と他国 y 間の類似度 $sim(x,y)$ を下記の相関係数より算出する（ステップ 2）.

$$\frac{\sum^{J}(TF_{\{x,j\}}-\overline{TF_{\{x,j\}}})(TF_{\{y,j\}}-\overline{TF_{\{y,j\}}})}{\sqrt{\sum(TF_{\{x,j\}}-\overline{TF_{\{x,j\}}})^2\sum(TF_{\{y,j\}}-\overline{TF_{\{y,j\}}})^2}} \quad (1)$$

最後に，任意の地域 p の Venue v を含むツイートを取得し，下記の式（2）より Venue v に対する出現頻度を算出する（ステップ 4）.

$$\frac{p\text{ で発信された }l_y\text{ 言語の Venue }v\text{ の出現回数}}{p\text{ で発信された }l_y\text{ 言語における Venue 総数}} \cdot \log\frac{\text{言語総数 } L}{\text{Venue } v\text{ の出現した言語数}} \quad (2)$$

3.2 Rating に基づく評判情報算出

前節より，各言語ごとのジャンルに対する嗜好性をツイートに出現する Venue 名の出現頻度より算出した．これにより人気の Venue を抽出できるが，その Venue を利用した結果の評判は考慮されていない．そこで，Venue に対する評判としてツイートの内容からポジティブかネガティブかを判定する手法が考えられる．これは任意の言語のツイート数が多い場所では有効性が高いと考えられるが，ツイート数の少ない場所では有意な差で有効性を示すことが困難である．そこで，本稿では Venue に対する評判として公開されているユーザの rating を用いて，評価値を下記の式 (3) より算出する（ステップ 3）.

$$\sum_{s}\frac{AvgR_{\{v,s\}}}{MaxR_s} * \frac{\#R_{\{v,s\}}}{\#AllR_v} \quad (3)$$

s はユーザによる rating サービスを示しており，本稿における実装では google と Forsquare を用いた．v は Venue であり，$R_{\{v,s\}}$ は任意の rating サービス s における v の rating となる．これを $MaxR_s$ は rating サービスの最大値で正規化する．また，rating 数を考慮し，$\#R_{\{v,s\}}$ は s における Venue v で rating された数を rating 総数で除算する．

*2　全カテゴリ中の重複省いた数で括弧は Tweet 総数に対する割合
*3　言語国と発信都市の国が同一
*4　全言語約 34 万 Venue に対するカテゴリ（ジャンル）取得は API 制限より本稿では未実施

3.3 ツイート数の少ない地域における各言語との類似性に基づいたジャンル抽出

地域 p におけるツイート数が閾値未満の場合は，言語 l_x にとっては訪問頻度の少ない地域であり，これは未知のアイテム推薦と捉えられる．そこで，他言語とのジャンルの類似性（ステップ 2）を考慮することで，他言語の l_y におけるジャンル j に対する評価値 $TF_{\{y,j\}}$ を用いて下記の式（4）より言語 l_x のジャンル j に対する評価値を抽出する（ステップ 5）.

$$\sum^{D}\left(sim(x,y)\ \cdot\ TF_{\{y,j\}}\right)\Big/\sum^{D}TF_{\{y,j\}} \quad (4)$$

D は言語数であり，式（4）は場所 p における言語 l_x のジャンル j に対する推薦度を算出しており，第一項目は，各言語 l_y との類似度 $sim(x,\ y)$ に言語 l_y のジャンル j に対する評価値を乗算した値の総和を全言語の類似度の総和で割った値である．第二項は場所 p におけるジャンル j に対する l_x の評価値であり，これを加算する．

提案手法より例えば，任意の地域でフランス人の訪問数が少なくツイート数が閾値以下の場合，フランス人のジャンル j に対する評価値は，まず，スペイン人との類似性が 0.8 で評価値が 0.6 の場合 0.48 が算出され，同様にイタリア人との類似性 0.5 と評価値 0.4 から 0.24 が算出され，それら総和 0.72 を類似度の総和で割った値 0.55 がジャンル j に対する評価値として算出される．

3.4 Venue 抽出および提示

地域 p におけるツイート数が閾値未満でツイート数の少ない地域では，まず，前節のステップ 5 より抽出された全ジャンルのうち推薦度の高いジャンル j を用いて場所 p の周囲 r 内における同一ジャンルの全言語の Venue を Venue 辞書より選出する．次にそれら Venue のうちステップ 4 より算出したその言語における出現頻度の高い Venue の上位 n 件を取得し，最後にステップ 3 より抽出した評判となる評価値の高い順にランキング付けて Venue を抽出する（ステップ 7）．ただし，Venue 辞書の p における Venue 数が少ない場合は，ジャンル j と位置情報 p と r を用いた Swarm API の逆引きによる Venue 名検索，またはジャンル名 j と位置情報 p と r を用いた Web 検索より Venue 情報を取得

Vol.2018-DBS-167 No.5
Vol.2018-IFAT-132 No.5
2018/9/12

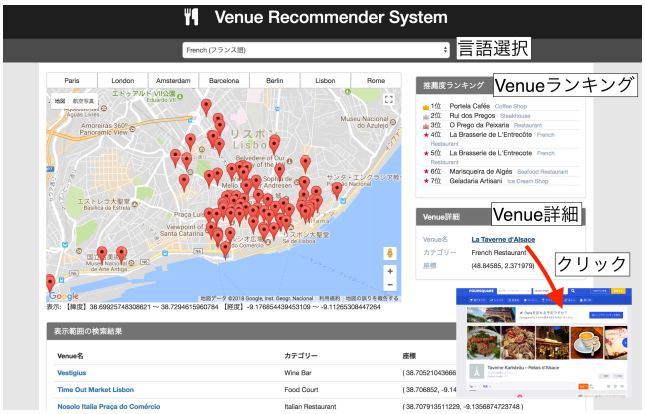

図 1　Venue 推薦システムのインタフェース

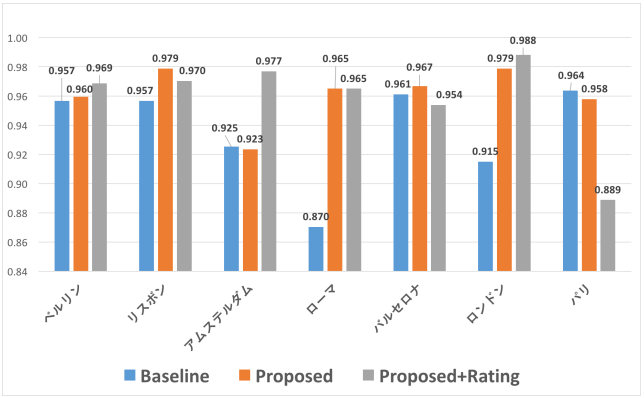

図 2　各都市のフランス語の話し手に対する nDCG@10 に基づいた Venue 推薦手法の比較

する．

ツイート数が閾値以上の場合は，ステップ 4 より抽出したその言語における出現頻度の高い Venue の上位 n 件を取得し，最後にステップ 3 より抽出した評判となる評価値の高い順にランキング付けて Venue を抽出する（ステップ 6）．

最後に，Venue 辞書から抽出した緯度経度に基づき地域 p における言語 l_x に対するお勧めの Venue として，地図上にピンをプロットする（ステップ 8）．

4. 実験

我々の提案する推薦手法の実現可能性を調査するために，Twitter からジオタグ付ツイートを，2016 年 4 月 1 日から 2017 年 4 月 30 日までの 13 ヶ月間で収集した．欧州 7 都市 (ロンドン，ローマ，パリ，バルセロナ，ベルリン，リスボン，アムステルダム) に着目し，その都市の母国語となる 6 言語 (イタリア語，フランス語，スペイン語，ドイツ語，ポルトガル語，オランダ語) を収集対象とした．

各対象都市の中心から半径 20km 以内のジオタグ付ツイートの分類結果を表 1 に示す．全体で約 2,600 万のツイートから延べ約 34 万件の Venue が抽出された．各対象 7 都市ごとで各 6 言語で発信されたユニークな Venue 総数は，7 万 3 千件であった（表の右半分）．表よりツイートに使用される言語と母国語が同じ場合 (例えばローマのイタリア語，パリのフランス語など) は高密度であったが，ツイートに使用される言語と母国語が異なる場合 (パリのイタリア語，ローマのポルトガル語など) は低密度であることが分かる．また，ロンドンなどの都市では異なる言語のツイートに対してかなりの密度があることも確認された．特筆すべきは，ベルリンにおけるフランス語では Venue が 5 件しかなく，本研究の他言語ユーザにとってツイート数の少ない低密度地域であり，Venue 推薦精度検証の対象言語および地域として有用である．

各都市における Venue は，ツイートの地理座標 (緯度経度) データをもとにした”I’m at”を含むツイートにより判別し，全体で 342,992 件の異なる Venue をデータセットとして取得した．また，Venue のジャンルに関しては，Foursquare の Swarm API より取得した．

さらに，rating による評判情報の算出に必要な Venue に対する rating は，Google Place API[*2]と Foursquare の 2 つの位置情報サービスを用いた．取得した rating 数は，ベルリンの総数は 66（1 店舗あたりの平均数 6.8），リスボンは 44（平均数 2.2），アムステルダムは 157（平均数 7.9），ローマは 94（平均数 4.7），バルセロナは 222（平均数 11.1），ロンドンは 353（平均数 17.7），パリは 349（平均数 17.5）となった．

推薦された Venue を可視化するために，インタラクティブなオンラインマップシステム[*3]を開発した (図 1)．このシステムでは，始めにユーザが言語を選択することで，地図上にマーカーが表示され，その都市の言語のユーザに対して Venue が推薦される．地図の右側には，評価スコアによる Venue のランキングが表示され，下側にはマップの中心から近い順にソートされた Venue の一覧が表示される．また，ユーザは地図上のマーカーにマウスを合わせることで，Venue の情報を確認することができる．

4.1 フランス語を母国語とするユーザによる評価方法

推薦手法の精度を検証するために，Twitter データを取得した都市に訪問したことがあるかまたは住んだことのあるユーザに対して提案手法と比較手法により推薦したレストランの好みを評価する実験を行った．本実験では欧州在住の被験者を対象とする必要があるため，クラウドソーシングである Amazon Mechanical Turk(AMTurk) を用いた．AMTurk は多くの研究分野 [10] の調査研究において，良質なデータを収集するのに効率的であることが示されている．各都市の被験者に Venue のリスト (それぞれの推薦手法による上位 10 件の Venue) を提示し，各 Venue に行きたいか否かを 7 段階のリッカート尺度で評価するように依

*2　https://developers.google.com/places/
*3　http://yklab.cse.kyoto-su.ac.jp/~sirakazu/VenueRecommender/

表 2　4 都市における抽出された Venue 上位 10 店舗（上段：baseline，中段：言語のみ考慮した Proposed，下段：評判情報も考慮した Proposed-rating）

都市	手法	Venue「Food」店舗（上位 10 店舗）
リスボン	baseline	Rui dos Pregos，Portela Cafés，Nosolo Italia PraÇa do ComÉrcio，Restaurante & Bar Terreiro do Paço，A Brasileira，O Prego da Peixaria，La Brasserie de L'Entrecôte，La Brasserie de L'Entrecote，Geladaria Artisani，Marisqueira de Algés
	Proposed	Taberna da Rua das Flores，Bella Lisa Rossio，A Brasileira，Cafetaria Museu do Teatro，Restaurante & Bar Terreiro do Paço，Blend Bairro Alto，Lisboa Bar，Lisbona Bar，Quasi Pronti，Nosolo Italia PraÇa do ComÉrcio
	Proposed +rating	Lisboa Bar，Taberna da Rua das Flores，Bella Lisa Rossio，Cafetaria Museu do Teatro，Restaurante & Bar Terreiro do Paço，Lisbona Bar，Quasi Pronti，Blend Bairro Alto，A Brasileira，Nosolo Italia PraÇa do ComÉrcio
アムステルダム	baseline	Starbucks，Hard Rock Cafe Amsterdam，Burger Bar，Wok to Walk，Restaurant Café In de Waag，Vondelpark3，Herengracht Restaurant Bar，The Breakfast Club，Wok to Go，Green House Centrum
	Proposed	Restaurant Café In de Waag，Starbucks，De Koffieschenkerij，Café t Papeneiland，Café 't Smalle，McDonald's，Burger King，Vondelpark3，Hard Rock Cafe Amsterdam
	Proposed +rating	De Koffieschenkerij，Café 't Smalle，Hard Rock Cafe Amsterdam ，Café 't Papeneiland，Café de Barones，Restaurant Café In de Waag，Vondelpark3，Starbucks，McDonald's，Burger King
バルセロナ	baseline	McDonald's，100 Montaditos，La Paradeta Passeig de Grcia，Hard Rock Cafe Barcelona，Central Café，El Tastet de la Mar，El Merendero de la Mari，Marco Aldany，Hidden Café Barcelona，Gran Café
	Proposed	100 Montaditos，McDonald's，La Paradeta Passeig de Gràcia，Marco Aldany，Central Café，Hard Rock Cafe Barcelona，Gran Café，Hidden Café Barcelona，El Tastet de la Mar，El Merendero de la Mari
	Proposed +rating	El Tastet de la Mar，Hidden Café Barcelona，La Paradeta Passeig de Grcia，Hard Rock Cafe Barcelona，El Merendero de la Mari，Gran Café，McDonald's，100 Montaditos，Central Café，Marco Aldany
ロンドン	baseline	Starbucks，Côte Brasserie，Comptoir Libanais，Costa Coffee，The Breakfast Club，Burger & Lobster，Bill's Restaurant，Adelaide Ice Cream And Hot Dogs，Paul，Cote Brasserie
	Proposed	Côte Brasserie，The Breakfast Club，Cote Brasserie，Starbucks，Comptoir Libanais，Bill's Restaurant，Paul，Burger & Lobster，Adelaide Ice Cream And Hot Dogs，Costa Coffee
	Proposed +rating	Côte Brasserie，The Breakfast Club，Côte Brasserie，Burger & Lobster，Adelaide Ice Cream And Hot Dogs，Paul，Bill's Restaurant，Starbucks，Comptoir Libanais，Costa Coffee

頼した．回答者には，AMTurk が定めた所要時間に対する報酬方針に相当する 0.15 ドルを報酬として支払った．また，(1) 予想される AMTurk の被験者数，(2) 母国と異なる都市に居住している被験者数から，本稿ではフランス語に焦点を当てた．その結果，フランス語を母国語とする被験者は，89 名であった (ベルリン 13 名，リスボン 11 名，アムステルダム 12 名，ローマ 16 名，バルセロナ 13 名，ロンドン 10 名，パリ 14 名)．

異なる欧州都市のフランス語のツイートから識別した Venue 数は，ベルリン (5)，リスボン (157)，アムステルダム (209)，ローマ (363)，バルセロナ (797)，ロンドン (1,568)，パリ (16,445) となった．この結果より，提案手法はツイート数の多い都市と少ない都市によって推薦手法が異なるため，本実験では判定に用いるツイート数の閾値を対象都市の半数以上の 4 都市となる 500 として，多い都市では式 (2) (ステップ 6) と少ない都市では式 (4) (ステップ 7) に基づいて Venue 抽出を行った．

提案された Venue の推薦手法の順位付け精度は，各手法によって推薦された Venue に対するフランス語を話し手とする被験者の評価値の平均に基づいて，正規化減価累積利得 (nDCG@10) より評価した．

4.2　言語と評判情報による Venue 推薦の検証

提案手法を含め 3 つの異なる手法より比較検証を行った．1 つ目の手法は，各都市における言語を考慮しない全てのツイート数より TF 値を算出することで言語に依存しない人気のある Venue を推薦する手法とし，これを baseline とした．2 つ目の手法は，各都市における言語の違いを考慮した提案手法による Venue 推薦を Proposed とした．3 つ目の手法は，各都市における言語の違いに加えて，Google と Foursquare から取得した Venue に対するユーザの評判情報も考慮した推薦を Proposed+rating とした．

図 2 に異なる欧州の 7 都市におけるフランス語の話し手の Venue 評価の結果を示す．全体として，パリ以外では Proposed（言語の違いのみ考慮) および Proposed+rating（言語および評判情報考慮）は言語の違いを考慮しない人気性のみに基づいた baseline より評価が上回った．なお，パリで baseline が最も評価が高かった理由は，今回フランス語を対象としていることが要因と考えられる．

パリを除いた 6 都市では，評判情報を考慮した場合と考慮しない場合での明確な違いは見られなかった．具体的にはベルリン，アムステルダム，ロンドンでは評判情報を考慮した Proposed+rating が最も良い結果であったが，他のリスボン，バルセロナでは Proposed が最も良い結果となった．これはリスボンとバルセロナはベルリン，アムステルダム，ロンドンと比較してフランス語のツイート総数と比較して rating 総数が少なかったことが要因として考えられる．

以上より，母国語の都市では単純なツイートに含まれる Venue 名の TF による推薦手法が最良となり，母国語でない都市のうち rating 数の少ない都市では言語情報に基づく Venue 推薦が最良となり，それ以外では提案手法の言語情報と評判情報に基づく Venue 推薦が有用であることが示された．

4.3 Venue 検証

表 2 に各手法を用いて抽出された Venue の推薦結果例を示す．検索結果例として，Venue 数が少なく rating の割合も少ない場所となるリスボン，Venue 数が少なく rating の割合の多い場所となるアムステルダム，Venue 数が多く rating の割合の多くないバルセロナ，Venue 数も rating の割合も多いロンドンの 4 都市における上位 10 店舗をランキングの高い順に示す．

リスボンでは，baseline では店舗数の多い Venue が上位となる傾向であったが，Proposed および Proposed+rating では，その都市のチェーン店ではなく一つしかないユニークな Venue が上位にランキングされた．アムステルダムは，Starbucks や Hard Rock Cafe といった他国にも多く存在する店舗が baseline の上位になり，提案手法ではいずれも低い順位となり，特に Proposed+rating では，ユニークでない Venue の順位が下がった．

バルセロナとロンドンでは，baseline で上位の McDonald's や Starbucks，100Motaditos や Costa Coffee の順位が提案手法により下がり，代わりにユニークな Venue が上位にランキングされた．以上より，提案手法により各国でユニークな Venue を上位に推薦できることを確認できた．

5. おわりに

本論文では，ジオタグ付ツイートの言語特性と評判情報を利用した Venue 推薦システムを提案した．ジオタグ付ツイートの場所と言語，ツイート発信者の母国語に関する情報を抽出し，さらに Venue に対する複数の rating の評判情報を用いることで，ユーザの言語と場所に基づいた Venue 推薦を実現した．ユーザ評価による実験では，baseline の TF 法より提案手法の言語と評判情報を用いた提案手法が訪問先の都市に全てにおいて優れた結果となった．また，rating 割合の少ない場所では言語による推薦手法が優位であることが明らかとなった．今後，raiting の評判だけでなく，ツイートの内容から評判情報となるポジティブなコメント分析を行う予定である．

謝辞 本研究の一部は，総務省 SCOPE（受付番号 171507010），JSPS 科研費 16H01722，17K12686 の助成を受けたものである．ここに記して謝意を表す．

参考文献

[1] T. Hu, R. Song, Y. Wang, X. Xie, J. Luo: Mining Shopping Patterns for Divergent Urban Regions by Incorporating Mobility Data，Proc. of the 25th ACM International on Conference on Information and Knowledge Management (CIKM2016)，pp．569-578 (2016)．

[2] J. Chen, S. Yang, W. Wang, M. Wang: Social Context Awareness from Taxi Traces: Mining How Human Mobility Patterns Are Shaped by Bags of POI，Adjunct Proc. of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers (UbiComp/ISWC'15 Adjunct)，pp．97-100 (2015)．

[3] Y． Qu, J． Zhang: Trade Area Analysis using User Generated Mobile Location Data, Proc. of WWW2013, pp. 1053-1064 (2013).

[4] É． Antoine，A． Jatowt，S． Wakamiya，Y． Kawai，T． Akiyama: Portraying Collective Spatial Attention in Twitter，Proc. of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD2015)，pp．39-48 (2015)．

[5] A. Magdy, L. Alarabi, S. Al-Harthi, M. Musleh, T. M. Ghanem, S. Ghani, M. F. Mokbel: Taghreed: A System for Querying, Analyzing, and Visualizing Geotagged Microblogs, Proc. of SIGSPATIAL2014, pp. 163-172 (2014).

[6] S. Wakamiya, A. Jatowt, Y. Kawai, T. Akiyama: Analyzing Global and Pairwise Collective Spatial Attention for Geo-social Event Detection in Microblogs, Proc. of WWW2016, pp. 263-266 (2016).

[7] M. S. Mohd Pozi, Y. Kawai, A. Jatowt, T. Akiyama: Sketching Linguistic Borders: Mobility Analysis on Multilingual Microbloggers, Proc. of WWW2017, pp. 825-826 (2017).

[8] 長島里奈， 関洋平， 猪圭: 地域ユーザに着目した口コミツイート収集手法の提案， WebDBForum (2016)．

[9] 野沢悠哉，遠藤雅樹， 江原遥， 廣田雅春， 横山昌平， 石川博: マイクロブログを用いたユーザの訪問目的と動向の推定， WebDBForum (2016)．

[10] F. R. Bentley, N. Daskalova, and B. White, "Comparing the reliability of amazon mechanical turk and survey monkey to traditional market research surveys," in Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. ACM, 2017, pp. 1092-1099.