Vol.2022-MPS-137 No.3
2022/3/3

画像投稿SNSにおけるハッシュタグの投稿関連度予測

神野 悦太郎[1,a)]　北 栄輔[1,b)]

概要：画像投稿 SNS でインフルエンサー・マーケティングをする際には，投稿に大量のハッシュタグを付加することが多い．ハッシュタグを自力で探すのは難しく，時間のかかる作業であるため，ハッシュタグを自動で推薦するシステムが作成されてきた．従来のシステムでは，ハッシュタグをキーワードとして検索して取得した共起ハッシュタグを用いて推薦を行うため，投稿と直接関係のないハッシュタグを複数推薦してしまうという問題があった．そこで本研究では，ハッシュタグごとに Instagram から取得した共起ハッシュタグの共起数ランキング上位 50 個に対して，逆共起カウント数，逆共起ランキング値，及びコメント間類似度を適用してハッシュタグとの関連度を評価する数理モデルを提案する．キーワードとの関連度は，実際の Instagram ユーザに行ったアンケート調査の結果を正解値として使用した．逆共起カウント数と逆共起ランキング値は，正解値との相関を示した．ランダムフォレストを使用した関連度予測モデルは SVM と比較して高い正解率をみせ，推薦システムへの搭載が期待できる結果となった．

Predicting Post Relevance of Hashtags in Image Posting SNS

1. はじめに

Instagram[*1]をはじめとする画像投稿 SNS を使ったマーケティングに取り組む企業や個人が増えている．知名度があって多数のフォロワー数を持つインフルエンサーのみならず，フォロワーが少なくても，得意とする分野・ジャンルに特化した発信をするマイクロインフルエンサーも数多く存在している．

Instagram や，Twitter[*2]をはじめとし，数多くの SNS や動画投稿プラットフォームの投稿には，ハッシュタグが付加されている．Instagram でインフルエンサー・マーケティングをする際には，投稿に大量のハッシュタグを付加することでより多くのターゲットに投稿画像を閲覧してもらえるようにすることが多い．特にマイクロインフルエンサーは，得意とする分野・ジャンルを好むターゲットにアプローチする必要があり，画像への関連度が高いハッシュタグを適切に選択する必要がある．

ハッシュタグを自力で探すのは難しく，時間のかかる作業であるため，ハッシュタグの推薦技術の研究が行われている [1], [2], [5]．山崎ら [1] らは，人気度を向上させるハッシュタグ推薦技術を開発している．Wang ら [2] は，投稿の人気度とユーザの人気度の両方を考慮したハッシュタグの推薦を行っている．蒲原ら [5] は，そのコンテンツに最もふさわしいと考えられるハッシュタグを推薦することを目的としたシステムを実行し，メディアに対して推薦されたハッシュタグが正しいかアンケートを行っている．我々も，Instagram へ投稿しようとしている画像を入力とし，推薦するハッシュタグを出力とするアプリの開発を行っている．入力画像から，オブジェクト検出アルゴリズムである YOLO[3] などを用いてキーワードを取得し，そのキーワードを使って Instagram でハッシュタグ検索を行い，キーワードとなっているハッシュタグと共起しているハッシュタグを共起カウント数の多い順にランキングで提案するものである [4]．しかし Instagram では，投稿画像と直接関係のないハッシュタグ（#instagram，#like4like，#f4f など）を投稿のコメントに付加するユーザが数多く存在するため，この共起カウント数ランキングにはキーワードとなるハッシュタグと直接関係のないハッシュタグが複数混じることがある．そこで本研究では，マイクロインフルエンサーが得意とする分野・ジャンルを好むターゲットにアプローチするために，自動でキーワードとなるハッシュタグへの関連度が高いハッシュタグを予測することを目的と

1 名古屋大学大学院 情報学研究科
Graduate School of Informatics, Nagoya University
a) etsu@nagoya-u.jp
b) kita@i.nagoya-u.ac.jp
*1 https://www.instagram.com/
*2 https://twitter.com/

ⓒ 2022 Information Processing Society of Japan

する．

キーワードとなるハッシュタグを使って検索をして取得した共起ハッシュタグの共起カウント数ランキング上位 50 位までのハッシュタグに対して，キーワードとなるハッシュタグとの関連度を予測するモデルを作成する．そのために，本研究では逆共起カウント数，逆共起ランキング値，及びコメント間類似度を提案する．関連度は，実際に Instagram を使用しているユーザにアンケート調査を行い，その結果を正解値として用いることで有効性を評価する．

2. 提案手法

この章では，まずハッシュタグ推薦システムの概要を示す．その後，本研究で探索対象とするハッシュタグ，予測モデルに用いる変数の定義，ユーザアンケート，予測モデル作成の実施方法について述べる．

2.1 ハッシュタグ推薦システムの概要

入力画像 X に対して，物体認識技術を使用して画像を説明しているラベル群 $\mathbf{V} \in v_j$ を抽出する．

次に取得した $\mathbf{V}$ を用いて画像投稿 SNS で投稿画像を検索する．検索して取得した投稿画像 I_k に付与されている共起ハッシュタグを $p_{k,l}$ とする．全ての I_k に対して以上の操作を実行すると，ハッシュタグ群 $P \in p_{k,l}$ が取得できる．これを，ハッシュタグプールと定義する．ハッシュタグプールに対して，本研究における投稿関連度予測モデルを適用させ，関連度の高いと判断されたハッシュタグを提案ハッシュタグとして出力する．

2.2 探索対象とするハッシュタグ

YOLO のデータセットとして使用される Pascal VOC 2012[*3]には person，bird, cat, cow など 20 の分類クラスが存在する．そのそれぞれをハッシュタグとし，Instagram でハッシュタグ検索をかけ，その投稿数を取得する．20 ハッシュタグのうち，投稿数の多い順であった dog，cat，car，train，horse の中から train を除いた 4 つを本研究で探索対象とするハッシュタグとする．

2.3 予測モデルに用いる変数の定義

本研究では，類似度予測モデルに次の変数を用いる．

- 逆共起カウント数
- 逆共起ランキング値
- コメント間類似度
- ハッシュタグ投稿数

本研究では，ハッシュタグ推薦の研究分野において一般的に使われている共起ハッシュタグそれぞれについて 4 変数を計算する．以下，まず共起カウント数について説明した後，それぞれの変数の導出過程について説明する．

[*3] http://host.robots.ox.ac.uk:8080/pascal/VOC/voc2012/

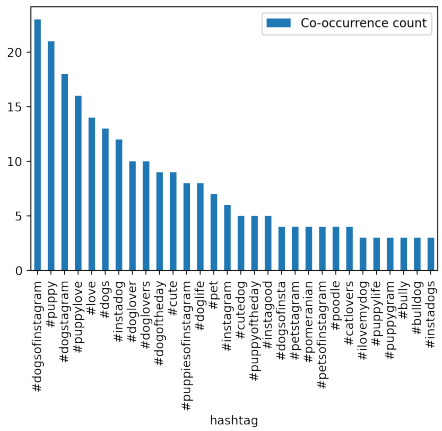

図 1 ハッシュタグ'dog' の共起カウント数ランキング
Fig. 1 Ranking of co-occurrence counts for the hashtag 'dog'

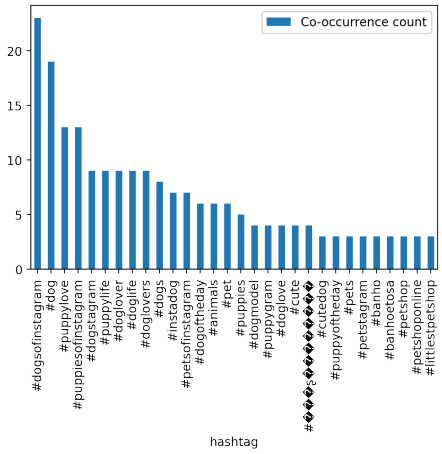

図 2 ハッシュタグ'puppy' の共起カウント数ランキング
Fig. 2 Ranking of co-occurrence counts for the hashtag 'puppy'

2.3.1 共起カウント数

共起カウント数の導出課程は以下である．

(1) 探索対象のハッシュタグ k をキーワードとして Instagram でハッシュタグ検索を行い，最大 100 個の最新投稿を取得する．
(2) 取得した投稿に付加されているコメント文の中のハッシュタグを抽出する．
(3) ハッシュタグごとにその出現数をカウントし，これを共起カウント数として定義する．

例としてハッシュタグ'dog' を使用して取得した共起カウント数のランキング上位 30 個を図 1 に示す．横軸はハッシュタグで，縦軸は共起カウント数である．これにより，ハッシュタグ'dog' の共起カウント数が最大なものはハッシュタグ'dogsofinstagram' であることがわかる．

2.3.2 逆共起カウント数

逆共起カウント数の導出課程は以下のとおりである．

(1) 探索対象のハッシュタグ k をキーワードとして共起カウント数を取得する．

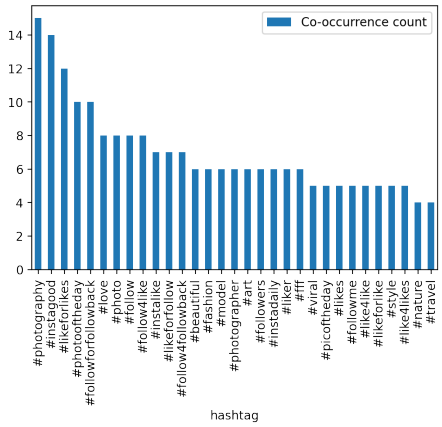

図 3 ハッシュタグ'instagram' の共起カウント数ランキング
Fig. 3 Ranking of co-occurrence counts for the hashtag 'instagram'

(2) 共起カウント数のランキング上位 50 ハッシュタグ $h_{(1...50)}$ それぞれをキーワードとして共起カウント数を取得する．
(3) (2) で取得した共起カウント数の中で，ハッシュタグ k の共起カウント数をハッシュタグ $h_{(1...50)}$ それぞれの逆共起カウント数として定義する．

例えば，図 1 の中に存在するハッシュタグ'puppy' についての共起カウント数のランキング上位 30 個は図 2 となる．図 2 の中に，元の探索対象のハッシュタグ'dog' は存在し，その共起カウント数は 19 である．よって，探索対象のハッシュタグ'dog' におけるハッシュタグ'puppy' の逆共起カウント数は 19 となる．

同様に，図 1 の中に存在するハッシュタグ'instagram' についての共起カウント数のランキング上位 30 個は図 3 となる．ランキング上位 30 個に入らなかったため，図 3 の中に，元の探索対象のハッシュタグ'dog' は存在していないが，共起カウント数は 1 であった．よって，探索対象のハッシュタグ'dog' におけるハッシュタグ'instagram' の逆共起カウント数は 1 となる．

2.3.3 逆共起ランキング値

逆共起ランキング値の導出課程は以下のとおりである．

(1) 探索対象のハッシュタグ k をキーワードとして共起カウント数を取得する．
(2) 共起カウント数のランキング上位 50 ハッシュタグ $h_{(1...50)}$ それぞれをキーワードとして共起カウント数を取得する．
(3) (2) で取得した共起カウント数を降順に並べ替えた場合の，ハッシュタグ k のランキング順位を，ハッシュタグ $h_{(1...50)}$ それぞれの逆共起ランキング値として定義する．

図 2 の中に，元の探索対象のハッシュタグ'dog' は存在し，それは 2 番目に多いハッシュタグである．よって，探索対象のハッシュタグ'dog' におけるハッシュタグ'puppy' の逆共起ランキング値は 2 となる．

ランキング上位 30 個に入らなかったため，図 fig:instagram の中に，元の探索対象のハッシュタグ'dog' は存在していないが，共起カウント数は 1 であったハッシュタグ'instagram' は降順で 157 番目のハッシュタグであった．よって，探索対象のハッシュタグ'dog' におけるハッシュタグ'instagram' の逆共起ランキング値は 157 となる．

2.3.4 コメント間類似度

ハッシュタグ検索して得た投稿のコメントを結合したものをコメント集合 D とする．キーワードとしたハッシュタグ k でハッシュタグ検索して得たコメント集合 D_k と，そのキーワードに関する共起カウント数上位 100 ハッシュタグ $h_{(1...100)}$ それぞれでハッシュタグ検索して得たコメント集合 D_{h_i} を用意する．これらのコメント集合に対して，文書中に含まれる単語の重要度を評価する手法の 1 つである TF-IDF を適用し，単語に重み付けをする．

次に，TF-IDF を用いて重み付けされたハッシュタグ k でハッシュタグ検索して得たコメント集合 D_k の単語ベクトル $a = (a_1, a_2, \cdots, a_n)$ と，ハッシュタグ k の共起カウント数上位 50 位までのハッシュタグ $h_{(1...50)}$ それぞれでハッシュタグ検索して得たコメント集合 D_{h_i} の各単語ベクトル $b = (b_1, b_2, \cdots, b_n)$ ごとのコサイン類似度を (1) のように求める．

$$\cos\theta = \frac{\sum_{k=1}^{n} a_k b_k}{\sqrt{\sum_{k=1}^{n} a_k^2}\sqrt{\sum_{k=1}^{n} b_k^2}} \tag{1}$$

以上のプロセスで得たコサイン類似度を，ハッシュタグ k とハッシュタグ k の共起カウント数上位 50 ハッシュタグ $h_{(1...50)}$ それぞれとのコメント間類似度とする．

2.3.5 ハッシュタグ投稿数

それぞれのハッシュタグについて，Instagram でハッシュタグ検索をかけ，その投稿数を取得する．つまりハッシュタグ投稿数が多いほど，Instagram の中でよく使われているハッシュタグと言える．

2.4 ユーザアンケート

本研究では，ハッシュタグの投稿関連度予測モデルの有効性を確認するために，正解データとして，ハッシュタグ k と，ハッシュタグ k の共起カウント数上位 50 ハッシュタグ $h_{(1...50)}$ との関連度が必要である．

そこで，Instagram のユーザー 10 人に対して，Google スプレッドシートを使用したアンケートを実施する．dog, cat, car, horse の 4 ハッシュタグそれぞれについて共起カウント数上位 50 ハッシュタグ，つまり合計 200 個のハッシュタグ i についてキーワードとの関連度を 3 段階（高・中・低）で判断してもらい，関連度「高」の合計値が過半数を超えたものを関連度「高」とし，残りを関連度「低」として扱う．

2.5 予測モデルの作成

作成した変数を用いて，前節の正解値を予測するモデル，すなわち，ハッシュタグの投稿関連度予測モデルを作成する．予測モデル使用する手法には，SVM とランダムフォレストを用いる．訓練データとテストデータは 8:2 の割合で分割する．

3. 実験結果

各変数と正解値との相関分析の結果を表 1 に示す．

表 1 各変数と正解値との相関係数の一覧
Table 1 Correlation coefficient between each variable and the correct answer value.

Variable name	Correlation coef.
Reverse co-occurrence ranking value	0.64
Reverse co-occurrence count	0.63
Exist hashtag count	-0.37
Similarity between comments	0.54

逆共起カウント数（Reverse co-occurrence count），逆共起ランキング値（Reverse co-occurrence ranking value），については，いずれも相関係数 0.6 以上の相関が見られ，コメント間類似度（Similarity between comments）についても 0.54 の相関が見られる．一方，ハッシュタグ投稿数（Exist hashtag count）については，弱い負の相関が見られる．4 変数いずれについても正解値と一定の相関が見られるので，これらを使用して SVM 及びランダムフォレストの 2 手法を用いた予測モデルを作成する．

SVM を用いた予測モデルの正解率は 77.5%である．対して，ランダムフォレストを用いた予測モデルの正解率は 87.5%である．本研究では，ハッシュタグの投稿関連度が高いと思われる変数を作成してモデルを作成する．軸となりうるそれぞれの変数に明快な意味があり，つまるところ「この軸の値がこの範囲なら...」という処理を繰り返して分類を行うことが高い予測精度につながると考えられる．よって，ランダムフォレストが SVM を大きく上回る精度を出したと考えられる．

ランダムフォレストを用いた予測モデルの適合率 (Precision)，再現率 (Recall)，F 値 (F-Measure) を表 2 に示す．また，表中において，Relevance=0 は関連度「低」を，Relevance=1 は関連度「高」を表す．表からいずれも高い値を示しており，本モデルを使ったハッシュタグの投稿関連度予測が可能であると判断できる．

表 2 ランダムフォレストを使用したモデルの予測結果
Table 2 Prediction results of a model using a random forest.

	precision	recall	f-measure	support
Relevance=0	0.91	0.71	0.80	14
Relevance=1	0.86	0.96	0.91	26

4. 結論

本研究では，マイクロインフルエンサーが，得意とする分野・ジャンルを好むターゲットにアプローチする際に有効となる，画像への関連度が高いハッシュタグを適切に選択するために，投稿を表すキーワードとのハッシュタグ関連度を予測するモデルについて提案した．従来研究されてきたハッシュタグを自動で推薦するシステムでは，投稿と直接関係のないハッシュタグを複数推薦してしまうという問題があった．本研究では，逆共起カウント数，逆共起ランキング値，及びコメント間類似度を使ったモデルを提案した．これにより，キーワードとして選択したハッシュタグごとに Instagram から取得した共起ハッシュタグの共起数ランキング上位 50 個に対して，87.5%の正解率を得ることができた．

キーワードとの関連度は，実際の Instagram ユーザに行ったアンケート調査の結果を正解値として使用し，予測モデルにはランダムフォレストを使用した．逆共起カウント数と逆共起ランキング値，及びコメント間類似度は，正解値との相関を示した．これらの結果から，本ハッシュタグ関連度予測モデルは，推薦システムへの搭載が期待できる結果となった．

参考文献

[1] T. Yamasaki, J. Hu, S. Sano, and K. Aizawa. FolkPopularityRank: Tag Recommendation for Enhancing Social Popularity using Text Tags in Content Sharing Services. *International Joint Conferences on Artificial Intelligence*, 3231-3237, 2017.

[2] X. Wang, Y. Zhang, and T. Yamasaki. User-Aware Folk Popularity Rank: User-Popularity-Based Tag Recommendation That Can Enhance Social Popularity, *Proceedings of the 27th ACM International Conference on Multimedia.* 2019. p.1970-1978.

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition* pp. 779-788, 2016

[4] 神野 悦太郎, 小野瀬 良佑, 北 栄輔. SNS 投稿画像の注目度向上のための Human-in-the-Loop を用いたタグ推薦手法の検討. **研究報告.数理モデル化と問題解決**, Vol.2020, 1-2, 2020

[5] 蒲原 智也, 波多野 賢治. 画像に対するハッシュタグ推薦システムの提案. **研究報告データベースシステム** *(DBS)*, 2019(3), 1-4, 2019