

M-004

異なるソーシャルネットワークサービスにおける統一的なタグを利用した

フレームワークの評価

Evaluation of Framework for Uniformly Using Tags of Different Social Network Services

廣瀬 崇人[†] 野口 裕貴[†] 山口 崇志[†]

Takahito Hirose Yuki Noguchi Takashi Yamaguchi

マッキンケネス ジェームス[†] 永井 保夫[†]

Kenneth James Mackin Yasuo Nagai

1. はじめに

近年ソーシャルネットワークサービス(SNS)を利用した情報推薦により、ユーザにとって有益な情報が得られる仕組み作りがなされている。しかしながら、SNSを利用して情報推薦を行う場合で、タグが少数しか付いていない場合がある。その場合推薦の精度が落ちてしまう。

推薦精度が落ちてしまう問題を解決する為タグデータベースを作成することで推薦精度が上がると期待出来る。

本稿では、複数の異なるタグデータを統一的に利用できるフレームワークを提案しウェブページの推薦を行う。

キーワード:ねこ

タイトル	カテゴリ	タグ	VideoID
特訓するねこ	ペットと動物	ねこ,猫	hPzNi6NKAG0
funny cats	ペットと動物	funny, cats	JZAH15_LTG0
Ronaldinho Touch of God	エンターテインメント	ronaldinho,brasil,futsal	ls06D1nrwKc
Ronaldinho vs Ronaldo	スポーツ	ronaldinho,cristiano,ronaldo,brasil,portugal,barcelona,manchester,united,football,skills	OS5tQvQ0B-Y
Funny Football	コメディ	funny,football,soccer,bloopers,blunders,graffs,gay,referee,referee	YFQ_N4imYVE
David Blaine street Magic: YouTube Edition!	コメディ	DavidBlaine,fake,street,magic,Mikey,Day,Mic hael,Naughton,Wildnout,MTV,Groundlings,ske tch,funny,card,trick,spoof,silpa	AYxu_MQSTTY
Billy Wingrove's Learn Freestyle Football	スポーツ	Nike,Joga,Bonito,Billy,Wingrove,Football,Free style,Soccer,Footy,Tricks,Skills,Adidas,Beck ham,Ronaldinho,Brazil,Futebol	rPcdG8gToPU
Laker Kobe Bryant attempts massive stunt...and succeeds! Real?	自動車と乗り物	laker,kobe,bryant,massive,stunt,aston,martin,jump,over,car,real,special,effect	yURa9T0-Rjk
Funny Cats	ペットと動物	cat,cats,funny,bloopers,animal,kitten,kittens,caturday,sound,effects	nTasT5h0LEg
extreme break dancing	エンターテインメント	break, dancing	Tyxlai6UF0

図1 YouTube の API を使い連続動画再生結果

図1は、「ねこ」というキーワードから動画を連続で動画を再生させるプログラムを書いた時の結果である。この実験では、YouTube の API を利用し「ねこ」というキーワードを送り、検索結果から一つ選び猫に関する動画を自動的にスタートさせるものだ。次の動画は、再生中の動画のタグを利用し、検索しランダムに次再生するようにセットして、動画が終わったら次の動画が始まる

ような仕組みである。

実験結果は、3 個目の動画のタグを見ると解のように「ronaldinho」「brasil」「futsal」と「ねこ」という最初のキーワードからズレが生じ、キーワードにマッチしない動画が再生されている。検索アルゴリズムは公開されていない。しかし、こちらで改善することはできる。

問題点として、検索キーワードのタグが少なかった可能性がある。

2. フレームワーク

YouTube のカテゴリを一つのタグとして扱うことにより livedoor ブックマークと同じ階層となる。

SNS ごとタグを付ける集団の違いがあるため、付ける傾向、目的なども変わってくる。その違いを埋めればタグから Web ページ推薦の補助になると思われる。

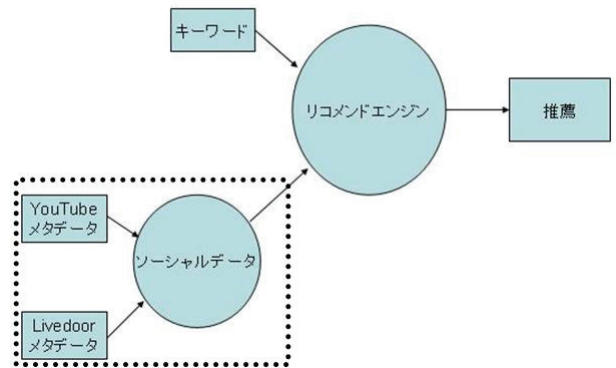


図2 推薦の仕組み

図2の点線で囲まれた部分が本稿の提案部分である。ソーシャルデータとは Livedoor クリップのメタデータ(タイトル、タグ、URL)YouTube のメタデータ(タイトル、カテゴリ、タグ、URL)を持ち、YouTube メタデータのカテゴリはタグと同じように扱う。本稿では、メタデータの集まりをソーシャルデータと呼ぶ。これを使い、アイテム推薦を行う。

3. 主成分分析

今回は既存の方法を主に使用している。まず、各 SNS

[†]東京情報大学 総合情報学部情報システム学科 Department of Information Systems, Tokyo University of Information Sciences

をクラスタリングして同義語などタグの揺らぎをなくす。

今回は主成分分析法を使用する。主成分分析で次元削減すると、共起している意味の近い語を抽象的に表す圧縮された文書ベクトルが与えられるという考えである。

	d1	d2	d3	d4	d5
Linux	4	3	0	0	0
Mac	0	0	2	0	0
Apple	0	0	1	1	0
Ruby	1	0	0	0	0
Ubuntu	2	0	0	0	0
shell	0	1	0	0	0

表1 アイテム(URL)に付随するタグの例

図2の d1~d5 はアイテム(URL)で、Linux などはタグである。つまり、d1 のアイテムには Linux というタグが4つ Ruby というタグが1つ、Ubuntu というタグが2つ付けられている。

アイテムが増えればタグが増えます。タグがまばらなことをうまく利用することで計算量を押しさえ込むことができる。

タグの関係もはっきりする。例えば Apple でもりんごと Apple 社の二つがある。この場合、Mac タグが付いているものが他にあった場合そちらを優先することにより間違っておりんごを推薦することもなくなる。

この二つの SNS のタグを統一的に使い推薦を行う。しかし、YouTube のメタデータと livedoor のメタデータでは付ける集団が違うためタグの揺らぎが存在する。

4. おわりに

本稿では、複数の異なるタグデータを統一的に利用できるフレームワークを提案しウェブページの推薦を行った。今後実データによる実験を行う。

参考文献

- [1] Toby Segaran 著 富山仁健 鴨澤真夫 訳 “集合地プログラミング” オライリー・ジャパン 2008
- [2] 高橋信 著 “マンガでわかる統計学” オーム社 2004
- [3] 嗜好の遷移過程に着目した Web 推薦アルゴリズムの提案 古川忠延 松尾豊 内山幸樹 石塚満 JSAI (2008)
- [4] 独立成分分析を用いた Web 閲覧履歴の解析と Web ページ推薦への応用 鶴原翔夢 高須賀清隆 丸山一樹 寺田実 DEWS (2008)