

局所ヒストグラム法の拡張によるサムネイル画像のタグ付け

Thumbnail Tagging by Extended Bag of Frequent Local Histogram Method

河野邊 早 尾崎知伸
Saki Kawanobe Tomonobu Ozaki

日本大学文理学部
College of Humanities and Sciences, Nihon University

Frequent local histogram method is an image classification technique to effectively utilize frequent patterns on image features. In this paper, we extend the method by introducing new pattern selection criteria for multi-label classification problems of thumbnail images in social media.

1. はじめに

本研究では、動画内のワンシーンであるサムネイル画像に対して適切なタグを付与する問題を画像特徴量を用いたマルチラベル分類問題として捉え、局所ヒストグラム法 [1] をマルチラベル分類へと拡張した手法を提案する。

局所ヒストグラム法とは、各画像を Visual Words (VWs) の局所的な分布 (ヒストグラム) の集合で表現するとともに、識別性や網羅性の観点から特徴的な頻出パターンを抽出し、分類のための属性として利用するものであり、実験的に単純に VWs を利用する場合と比較して高い精度が得られることが知られている。局所ヒストグラム法は (多値の) 分類問題を念頭とした手法であり、特徴的なパターン抽出においてラベル (タグ) 間の関連性とデータのスパース性を考慮していないことから、マルチラベル分類として定式化されるソーシャルメディアの投稿動画へのタグ付けに直接適用することは難しい。本研究ではこの問題に対して、網羅性・識別性の評価をタグの組み合わせに対して行うことでタグの関連性を扱うように局所ヒストグラム法を拡張する。加えて、画像毎に評価の高いパターンを特定することで、データのスパース性に対処する。これらの拡張により、より精度の高い分類が期待できる。

2. 局所ヒストグラム法

局所ヒストグラム法 [1] は、下記に示す手順に従い、画像特徴量を用いて各画像を特徴ベクトルに変換し、SVM 等を用いて分類器を構築する (図 1 参照)。

ステップ 1: N 個の画像集合 $I = \{I_1, \dots, I_N\}$ に対し、SIFT など画像の特徴点及び特徴量を求め、それらを K-平均法等でクラスタリングすることで Visual Words を生成する。生成された VWs の集合を W と表記する。各特徴点に対し、その k 近傍中の各 VWs と出現数の対の集合 (局所ヒストグラム) $\{vw : n \mid vw \in W, n > 0\}$ を割り当てることで、各画像 I_j を局所ヒストグラムの集合とみなす。

ステップ 2: 全局所ヒストグラムの集合から (飽和) 頻出パターンを抽出し、以下の評価関数 S に従い、上位 K 位以内のパターンを選択する。選択されたパターンの集合を P_k と表記する。

クラス集合 $C = \{1, \dots, |C|\}$ に対するパターン t の評価値 $S(t, C)$ は、識別性 $D(t, C)$ と代表性 $O(t, C)$ の積

$$S(t, C) = D(t, C) \times O(t, C)$$

で与えられる。ここで識別性 $D(t, C)$ は、

$$D(t, C) = 1 + \frac{\sum_{c \in C} p(c|t) \cdot \log p(c|t)}{\log |C|}$$

where $p(c|t) = \frac{F(t|I_j) \cdot p(c|I_j)}{\sum_{i=1}^N F(t|I_i)}$
and $F(t|I_j) = |\{s \in I_j \mid t \subseteq s\}|$

と定義される。一方、網羅性 $O(t)$ は、

$$O(t, C) = \max_{c \in C} (\exp - [D_{KL}(P(I|t_c^*) \parallel P(I|t))])$$

と定義される。ここで $D_{KL}(\cdot \parallel \cdot)$ は KL-divergence, $P(I|t)$ は I 上での t の分布 (出現数), $P(I|t_c^*)$ はクラス c に対する最適分布 (1/クラス c に属する画像数) をそれぞれ表す。

ステップ 3: 画像 I_j を、 P_k 中の各頻出パターン p にマッチする局所ヒストグラムを持つか否か、すなわち $\exists s \in I_j$ s.t. $p \subseteq s$ を満たすか否かを基準に、 K 次元の 2 値ベクトルに変換する。

ステップ 4: 変換されたデータセットに対し、SVM 等を用い、分類器を構築する。

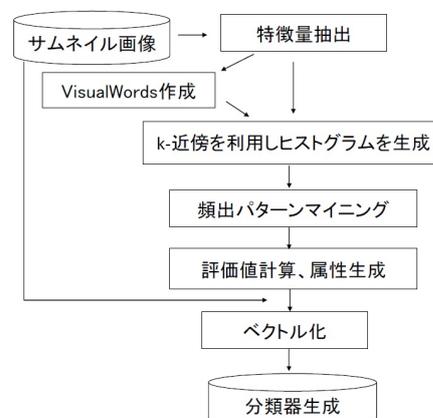


図 1: 局所ヒストグラム法の概要

連絡先: 尾崎知伸, 日本大学文理学部, 〒 156-8550 東京都世田谷区桜上水 3-25-40, ozaki.tomonobu@nihon-u.ac.jp

表 1: 実験結果

		ML-KNN		RAkEL	
		手法 1: $g_1(T)$	手法 2: $g_2(T)$	手法 1: $g_1(T)$	手法 2: $g_2(T)$
train_0	F 値	0.7553 ± 0.0294	0.7559 ± 0.0304	0.5920 ± 0.0360	0.6054 ± 0.0303
	適合率	0.7700 ± 0.0286	0.7705 ± 0.0300	0.6382 ± 0.0542	0.6831 ± 0.0518
	再現率	0.7485 ± 0.0299	0.7495 ± 0.0308	0.5795 ± 0.0396	0.5240 ± 0.0378
test_1	F 値	0.5200 ± 0.0140	0.5267 ± 0.0337	0.4648 ± 0.0336	0.4985 ± 0.0269
	適合率	0.5200 ± 0.0140	0.5267 ± 0.0337	0.4636 ± 0.0347	0.4991 ± 0.0270
	再現率	0.5200 ± 0.0140	0.5267 ± 0.0337	0.4701 ± 0.0323	0.5026 ± 0.0270
test_2	F 値	0.5340 ± 0.0371	0.5367 ± 0.0177	0.4819 ± 0.0460	0.5120 ± 0.0252
	適合率	0.5340 ± 0.0371	0.5367 ± 0.0177	0.4833 ± 0.0472	0.5124 ± 0.0251
	再現率	0.5340 ± 0.0371	0.5367 ± 0.0177	0.4852 ± 0.0449	0.5143 ± 0.0256
test_3	F 値	0.5323 ± 0.0317	0.5420 ± 0.0435	0.4889 ± 0.0471	0.5160 ± 0.0366
	適合率	0.5333 ± 0.0313	0.5420 ± 0.0435	0.4885 ± 0.0483	0.5155 ± 0.0373
	再現率	0.5333 ± 0.0313	0.5420 ± 0.0435	0.4918 ± 0.0452	0.5203 ± 0.0366
test_4	F 値	0.5496 ± 0.0468	0.5760 ± 0.0285	0.5038 ± 0.0703	0.5323 ± 0.0380
	適合率	0.5502 ± 0.0468	0.5760 ± 0.0285	0.5043 ± 0.0695	0.5339 ± 0.0381
	再現率	0.5495 ± 0.0468	0.5760 ± 0.0285	0.5068 ± 0.0696	0.5338 ± 0.0388
test_5	F 値	0.5373 ± 0.0503	0.5435 ± 0.0308	0.4728 ± 0.0444	0.5135 ± 0.0340
	適合率	0.5373 ± 0.0503	0.5440 ± 0.0317	0.4721 ± 0.0460	0.5145 ± 0.0334
	再現率	0.5373 ± 0.0503	0.5434 ± 0.0307	0.4771 ± 0.0423	0.5139 ± 0.0343

3. マルチラベル分類への拡張

ソーシャルメディアに対するタグ付与などのマルチラベル分類に局所ヒストグラム法を適用する際の問題点として、以下の2点があげられる。第一の問題点は、各画像は複数のラベル(タグ)を持ち、それらのラベル間には階層構造も含めた依存関係が存在することである。局所ヒストグラム法におけるパターンの評価関数 S は、各画像に単一のラベル(クラス)が付与されていることを前提に設計されており、マルチラベル分類に適した形での拡張が必要となる。第二の問題点は、データのスパース性である。局所ヒストグラム法では、ヒストグラム全体から属性として利用されるパターンを選択するが、パターンの偏り等から選択されたパターンを一つも満たさない画像が存在することが考えられる。これらの問題を解決するため、本研究では局所ヒストグラム法に対し、次の3点の拡張を施す。

1. 評価関数をマルチラベル分類を意識した形式に変更する。
2. 各画像ごとに評価値の高いパターンを抽出し、属性として利用する。
3. ベクトル化された画像に対して、マルチラベル分類器 [2] を適用し分類モデルを構築する。

評価関数の変更について、その概要を説明する。画像 I_j に付与されているタグの集合を T_j と表記する。 I_j に対する頻出パターン t の評価値 $S_j(t)$ を以下のように定義する。

$$S_j(t) = \begin{cases} \max_{C \in g(T_j)} S(t, C) & (\exists s \in I_j \text{ s.t. } t \subseteq s) \\ -\infty & (\text{otherwise}) \end{cases}$$

ここで $g(T_j)$ は、タグ集合 T_j からクラス集合の集合を返すメタ関数であり、本研究では以下の2つを考える。

$$g_1(T) = \{C(\{l\}) \mid l \in T\}$$

$$g_2(T) = \{C(s) \mid s \subseteq T, s \neq \emptyset\}$$

ここで $C(L)$ はクラス集合 $\{1,0\}$ を表すと同時に、タグ集合 L を持つ画像にクラス値 1 を、持たない画像にクラス値 0 を割り当てるとする。

属性として利用されるパターンの選択は画像ごとに行う。すなわち、画像 I_j と全頻出パターン集合 T に対し $S_j(t)$ の値上位 K 個のパターンの集合を $T_j(K)$ としたとき、集合 $\bigcup_{I_j \in I} T_j(K)$ を属性として利用する。

4. 評価実験

提案手法の有効性を評価するため、Java 言語を用いて拡張した局所ヒストグラム法を実装し、評価実験を行った。なお、ベクトル化されたデータを対象としたマルチラベル分類器の構築には、ライブラリ Mulan[3]^{*1} で提供されている ML-KNN[4] と RAkEL[5] を利用した。

実験では、国立情報学研究所のダウンロードサービスにより株式会社ドワンゴから提供を受けた「ニコニコ動画コメント等データ」^{*2} から、「スポーツ」カテゴリに属する動画約 8 万件を対象とした。なお、タグの種類が多いため頻度上位 100~150 位のタグのみを利用している。まず、5000 件の画像データを利用し VWs を作成し、1000 件のデータ (train_0) を用いて属性として利用するルール (最小支持度 0.05) を生成した。得られた属性数は、手法 1: 評価関数 $g_1(T)$ が 477, 手法 2:

*1 <http://mulan.sourceforge.net/>

*2 <http://www.nii.ac.jp/dsc/idr/nico/nico.html>

評価関数 $g_2(T)$ が 532 であった。これらの属性を用い、ルール生成に利用したデータセット (train_0) 及び (それとは異なる) 各 1000 件の 5 つのテストデータセット (test_1~test_5) を対象に分類器の構築を行い、その性能を交差検定による評価している。実験結果を表 1 に示す。

実験結果より、今回のデータに関しては、RAkEL より ML-KNN を利用する方が良い結果が得られていることが分かる。ルール生成に利用した train_0 では、ML-KNN で F 値が 0.7 程度、RAkEL では 0.6 程度となり、必ずしも高いわけではないがそれなりの性能が得られることが確認できた。一方で test_1~test_5 に対しては、ML-KNN で F 値 0.52~0.57 程度、RAkEL で F 値 0.46~0.53 程度となり、更なる改善が必要となる結果となった。なお全体として、適合率と再現率に大きな差はなく、その意味ではバランスが取れていることが確認できた。また、評価関数 $g_1(T)$ と $g_2(T)$ を比較すると、若干ではあるが $g_2(T)$ の方が良い結果が得られている。これは $g_2(T)$ の方が考慮するタグ (クラス) の組合せが多いことに起因すると考えられる。

5. まとめ

本研究では、ソーシャルメディアに対するマルチラベル分類を目的に、局所ヒストグラム法の拡張を提案した。今後の課題としては、動画に付与されたコメントや概要文、タイトルなど様々な情報を利用した、精度の向上があげられる。また、ルール選択のための評価基準の改良も、大きな課題の一つである。

謝辞 本研究では、国立情報学研究所のダウンロードサービスにより株式会社ドワンゴから提供を受けた「ニコニコ動画コメント等データ」を利用した。

参考文献

- [1] B. Fernando, E. Fromont and T. Tuytelaars : Effective Use of Frequent Itemset Mining for Image Classification, *Proc. of the 12th European Conference on Computer Vision*, Vol.1, pp.214-227, 2012
- [2] G. Madjarov, D. Kocev, D. Gjorgjevikj and S. Džeroski : An extensive experimental comparison of methods for multi-label learning, *Pattern Recognition*, Vol.45, No.9, pp.3084-3104, 2012
- [3] G. Tsoumakas, I. Katakis and I. Vlahavas : Mining Multi-label Data, *Data Mining and Knowledge Discovery Handbook*, O. Maimon, L. Rokach (Ed.), Springer, 2nd edition, 2010.
- [4] M.-L. Zhang and Z.-H. Zhou : ML-KNN: A lazy learning approach to multi-label learning, *Journal of Pattern Recognition*, Vol.40, No.7, pp.2038-2048, 2007.
- [5] G. Tsoumakas, I. Katakis and I. P. Vlahavas : Random k-Labelsets for Multilabel Classification, *IEEE IEEE Transactions on Knowledge and Data Engineering*, Vol.23, No.7, pp.1079-1089, 2011.