IMAGE SENTIMENT ANALYSIS USING LATENT CORRELATIONS AMONG VISUAL, TEXTUAL, AND SENTIMENT VIEWS

同志社大学 Doshisha University

Marie Katsurai
Doshisha University

Shin' ichi Satoh
National Institute of Informatics

NII 大学共同利用機関法人 情報・システム研究機構 国立情報学研究所 National Institute of Informatics

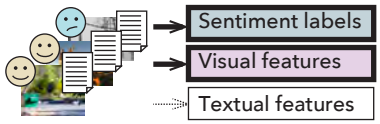
Introduction

Positive or Negative ?

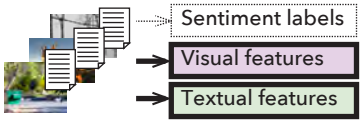
Automatic classification of image sentiment polarity

Related work:

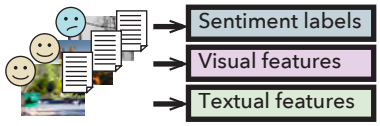
<Image sentiment analysis>



<Image annotation>

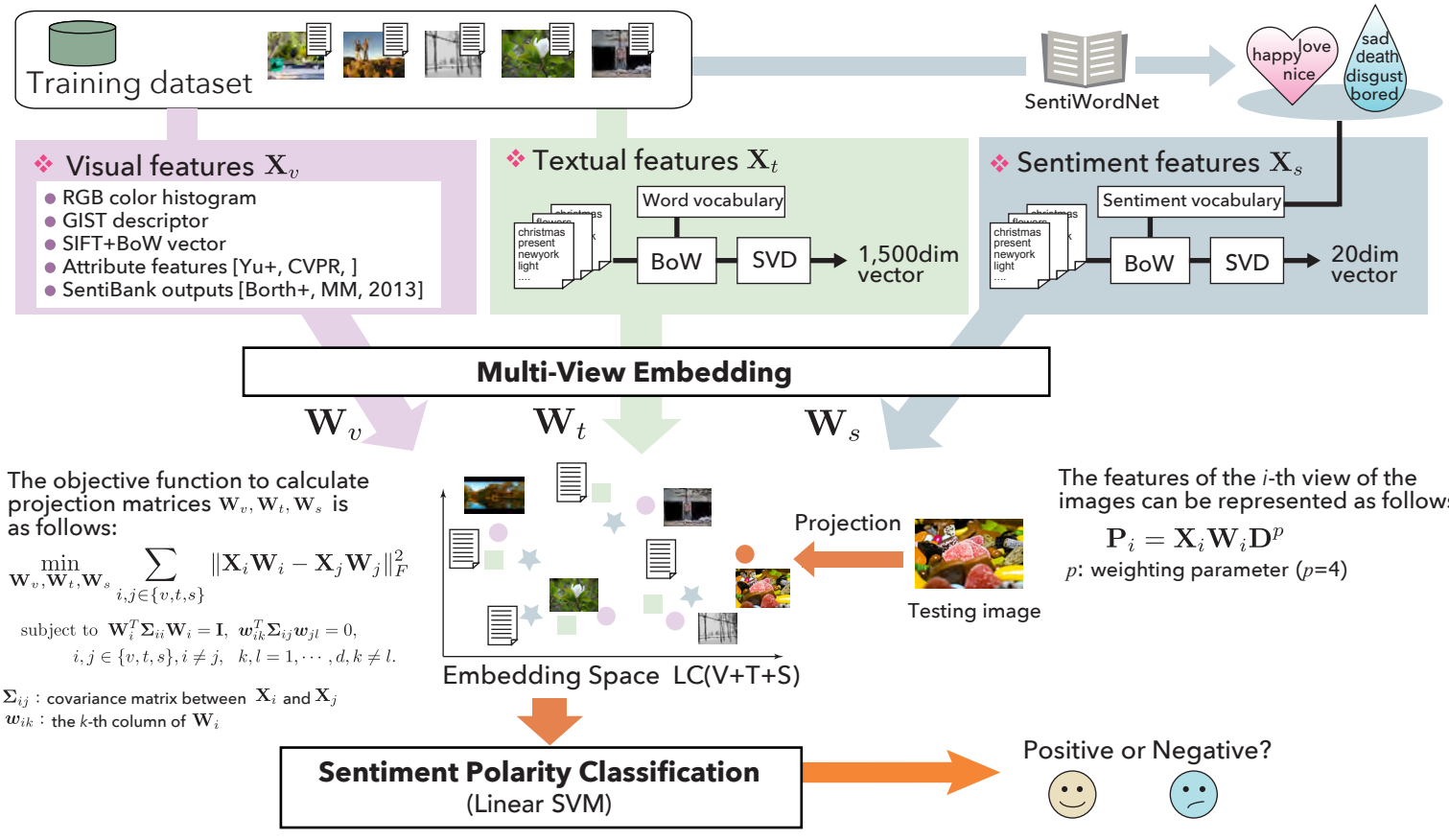


Our method:



We find a latent embedding space in which correlations among the three views are maximized.

Proposed Method



The objective function to calculate projection matrices $\mathbf{W}_v, \mathbf{W}_t, \mathbf{W}_s$ is as follows:

$$\min_{\mathbf{W}_v,\mathbf{W}_t,\mathbf{W}_s} \sum_{i,j\in\{v,t,s\}} \|\mathbf{X}_i\mathbf{W}_i - \mathbf{X}_j\mathbf{W}_j\|_F^2$$

$$\text{subject to } \mathbf{W}_i^T\mathbf{\Sigma}_{ii}\mathbf{W}_i = \mathbf{I},\ \boldsymbol{w}_{ik}^T\mathbf{\Sigma}_{ij}\boldsymbol{w}_{jl} = 0,\quad i,j\in\{v,t,s\}, i\neq j,\quad k,l=1,\cdots,d, k\neq l.$$

$\mathbf{\Sigma}_{ij}$: covariance matrix between $\mathbf{X}_i$ and $\mathbf{X}_j$
$\boldsymbol{w}_{ik}$: the k-th column of $\mathbf{W}_i$

The features of the i-th view of the images can be represented as follows:

$$\mathbf{P}_i = \mathbf{X}_i\mathbf{W}_i\mathbf{D}^p$$

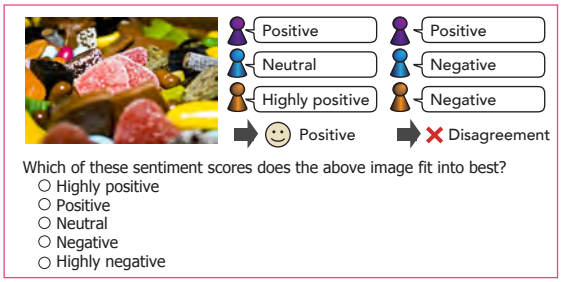
p: weighting parameter (p=4)

Sentiment Polarity Classification
(Linear SVM)

Positive or Negative?

Dataset Construction

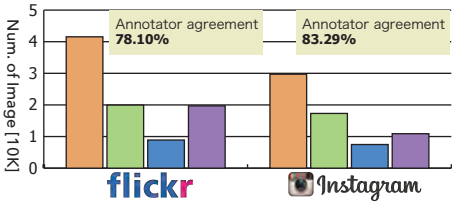
- Collected a set of images and their text from Flickr and Instagram.
- Three users evaluate the polarity of an image on a five point Likert scale.

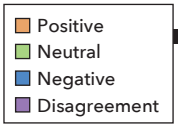


Calculate classification accuracy.

	flickr	Instagram
Training images	11,346	8,802
Test images	1,260	978

Randomly sampled images for training and testing at each run.





The dataset is available on the Web.
http://mm.doshisha.ac.jp/senti/CrossSentiment.html

Performance Evaluation

Averaged and standard deviation of the classification accuracy of 10 runs in each dataset.

	Methods	Dataset: flickr	Dataset: Instagram
Visual only	Random	49.78 ± 1.05%	50.06 ± 1.09%
	[Siersdorfer+, MM, 2010]	69.44 ± 0.85%	67.16 ± 1.28%
	[Borth+, MM, 2013]	70.01 ± 0.63%	67.26 ± 1.12%
	Visual features used in our method	70.54 ± 1.00%	68.03 ± 1.36%
	LC(V+T)+P(V)	70.94 ± 0.67%	68.29 ± 1.42%
	LC(V+S)+P(V)	70.67 ± 0.78%	65.44 ± 1.16%
	LC(V+T+S)+P(V)	72.36 ± 0.41%	68.54 ± 1.14%
Text only	[Thelwall+, JASIST, 2010]	59.30 ± 0.87%	62.78 ± 0.91%
	[Wang+, IJCAI, 2015]	51.87 ± 1.76%	52.61 ± 2.00%
2 View	LC(T+S)+P(T+S)	64.63 ± 0.91%	66.50 ± 0.49%
	LC(V+S)+P(V+S)	68.98 ± 1.01%	69.35 ± 1.08%
	LC(V+T)+P(V+T)	74.42 ± 0.67%	72.43 ± 1.54%
All View	LC(V+T+S)+P(V+T+S)	**74.77 ± 0.82%**	**73.60 ± 0.88%**

Our method using the three views of the images obtained the best average classification accuracy.