中国科学技术大学

University of Science and Technology of China

PERSONALIZED VIDEO EMOTION TAGGING THROUGH A TOPIC MODEL

Shan Wu, Shangfei Wang and Zhen Gao

- Motivation
- Method
- Experiments
- Conclusion

Motivation

University of Science and Technology of China

Topic: Video emotion tagging has attracted increasing attention, since emotion is the key factor during communication and entertainment.

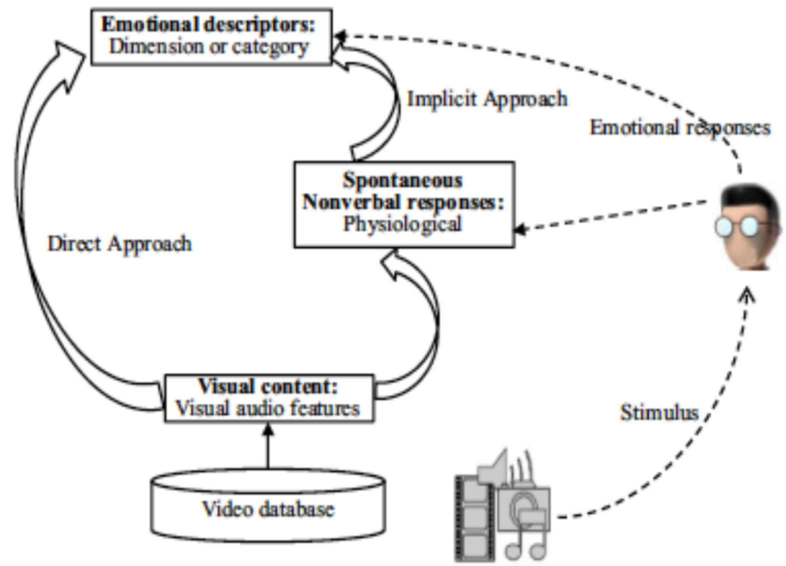


Fig. 1 Components and two major approaches of video affective content analysis [Wang and Ji]

Current approaches:

- Direct emotion tagging — sign emotion labels to videos directly from related audiovisual features
- Implicit emotion tagging — infer emotion labels of videos based on an automatic analysis of a user's spontaneous response when watching the videos

Motivation

Two kinds of emotion:

Direct approaches

- ✧ Expected emotions — common emotions, communicated through visual and aural elements based on film grammar.

Implicit approaches

- ✧ Induced emotions — audiences' emotions elicited during watching videos, personalized emotions, the same video may induce different emotions from different audiences due to their various personalities and culture backgrounds

Focus: Personalized video emotion tagging, involving video content, personal characteristics data and personalized emotion labels.

Difficulties:

- ✧ Complexity of human factors, Personality, Culture, etc.
- ✧ Lack of databases

Motivation

University of Science and Technology of China

- CP-QAE-I database: including videos, users' personality traits, culture traits and perceived affect
- Analysis proves the role of personal characteristics in perceived enjoyment.

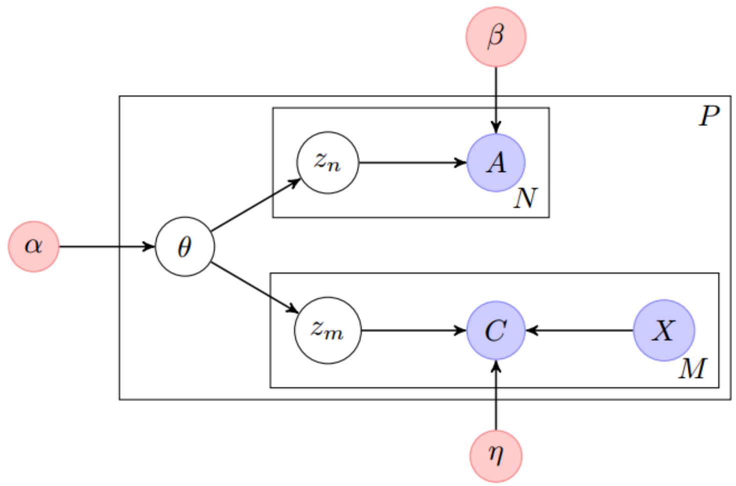
Example: Video clips from Forest Gump

ID	EXTRAVERSION	AGREEABLENESS	CONSCIENTIOUSNESS	NEUROTICISM	OPENNESS	POWER_DISTANCE	INDIVIDUALISM	MASCULINITY	UNCERTAINTY_AVOIDANCE	PRAGMATISM	INDULGENCE	ENJOYMENT
10	5	8	5	5	7	-25	-35	-35	80	65	-5	3
14	5	4	6	3	7	-35	35	0	55	10	-45	5
19	3	4	7	8	7	0	105	-70	5	55	0	1
23	2	6	6	6	6	-95	105	-35	45	130	100	4

[1] The CP-QAE-I: A Video Dataset for Exploring the Effect of Personality and Culture on Perceived Quality and Affect in Multimedia', IEEE Proceedings of the 7th International Workshop on Quality of Multimedia Experience, 2015.
[2] Do Personality and Culture Influence Perceived Video Quality and Enjoyment?, TMM 2016

- Motivation
- Method
- Experiments
- Conclusion

Method



A: personal characteristic
C: emotion label
X: video feature

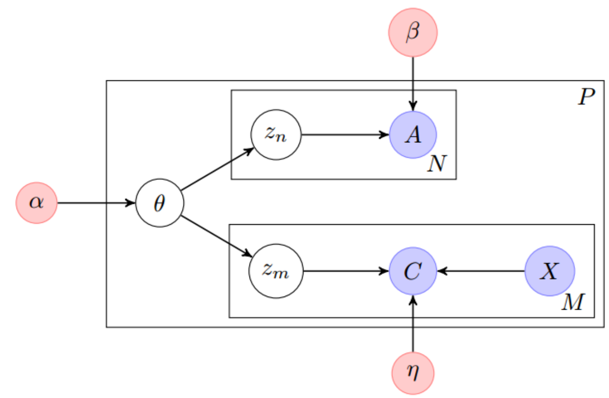
$\alpha \in \Re^{K}, \beta_n \in \Re^{K\times S^n}, \eta \in \Re^{K\times L}$

1. Draw topic proportions $\theta \sim Dir(\alpha)$.
2. For each personal characteristic $a_n, n \in \{1, 2, ..., N\}$:
   (a) Draw the topic assignment $z_n|\theta \sim Mult(\theta)$.
   (b) Draw the personal characteristic $a_n|z_n \sim Mult(\beta^n_{z_n})$.
3. For each movie tagged by this person $c_m, x_m, m \in \{1, 2, ..., M\}$:
   (a) Draw the topic assigment $z_m|\theta \sim Mult(\theta)$.
   (b) Draw the personalized video emotion tags $c_m|z_m \sim Beta(\sigma(\eta^T_{z_m} x_m))$.

Method

-Training

$$\log p(A,C|\Theta) = \log \int_\theta \sum_z \frac{p(A,C,\theta,z|\Theta,X)q(\theta,z)}{q(\theta,z|\gamma,\phi)}d\theta$$
$$\geq \int_\theta \sum_z q(\theta,z)\log\frac{p(A,C,\theta,z|\Theta,X)}{q(\theta,z)}d\theta$$
$$= E_q[\log p(A,C,\theta,z|\Theta,X)] - E_q[\log q(\theta,z)]$$
$$= L(\Theta,\gamma,\phi)$$

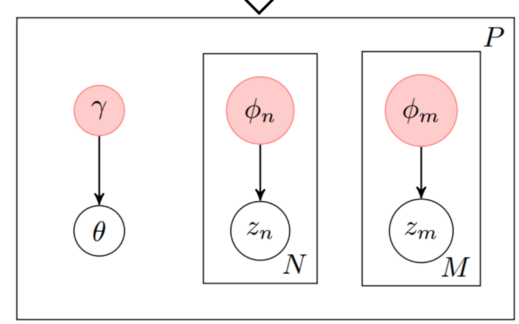


Variantional Inference

$$\log p(A,C\mid\Theta) = L(\Theta,\gamma,\phi) + KL\left[q(\theta,z\mid\gamma,\phi)\,\|\,p(\theta,z\mid A,C,X,\Theta)\right]$$

E-step: fix Θ **,** $\arg\max_{\gamma,\phi} L(\Theta,\gamma,\phi)$

M-step: fix γ£¬ϕ **,** $\arg\max_{\Theta} KL[q(\theta,z\mid\gamma,\phi)\,\|\,p(\theta,z\mid A,C,X,\Theta)]$



-Testing

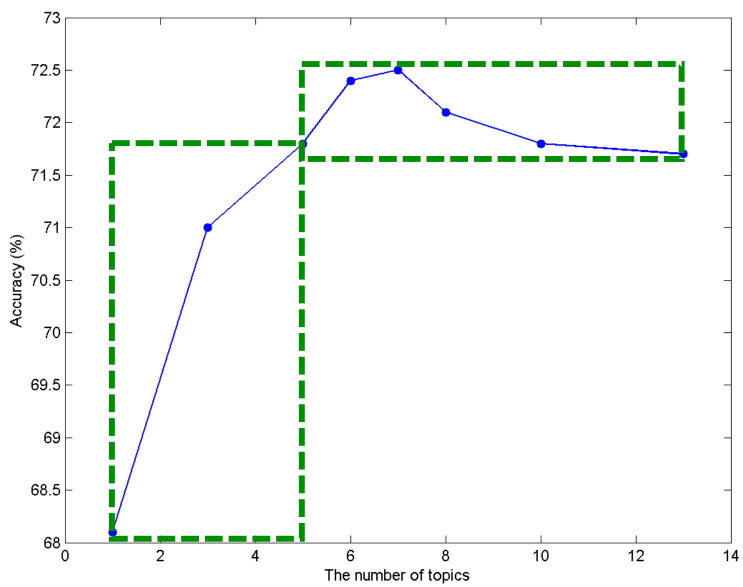
$$p(c=1|P_i,x_{new}) = \sum_{j=1}^{K} p(c=1|\eta_i,x_{new})q_i(\theta_j|\gamma)$$

- Motivation
- Method
- Experiments
- Conclusion

Experiments: Experimental Conditions

- Databases: CP-QAE-I database
  - 144 videos (12 short movie clips varying by frame rate, frame dimension and bit rate)
  - 114 participants with 11 personal characteristics (6 cultural + 5 personality)
  - personalized video tags
- Personalized emotion tags: subjective enjoyment labels
- Personal characteristics: personality traits and cultural traits
- Video features: audio, visual and video quality
- Personalized video emotion tagging:
  - Existing users: Leave one video out cross validation
    - SVM (input: video features and personal characteristics)
    - Our method
  - New users: Leave one subject out cross validation
    - SVM (input: video features and personal characteristics)
    - Method*: using topic model without personal characteristics
    - Our method

Experimental Results: latent topics



- Accuracy quickly increases from 68.1% to 71% when K varies from 1 to 3
- When K changes from 4 to 12, the accuracy of our model varies from 71.5% to 72.5%. The sliding interval is quite small
- 6 topics

John Holland's theory: most personalities of people is a combination of six basic personality types: realistic, investigative, artistic, social, enterprising, and conventional.

topic 1	topic 2	topic 3	topic 4	topic 5	topic 6
inventive careless solitary analytical sensitive	inventive efficient solitary analytical confident	consistent efficient outgoing friendly sensitive	consistent careless outgoing friendly sensitive	consistent efficient solitary analytical sensitive	inventive careless outgoing friendly sensitive
artistic	**investigative**	**social**	**realistic**	**conventional**	**enterprising**

Experimental Results for Personalized Video Emotion Tagging

	leave-one-video-out				leave-one-subject-out					
	SVM		Our method		SVM		Method*		Our method	
	Low	High	Low	High	Low	High	Low	High	Low	High
Low	322	189	308	203	296	215	227	284	314	197
High	205	516	139	582	217	504	163	558	201	520
Acc.	68.0%		72.2%		64.9%		63.7%		67.69%	
F1.	0.620		0.643		0.578		0.504		0.612	
Kappa	0.344		0.418		0.278		0.226		0.335	

Experimental Results for Personalized Video Emotion Tagging

中国科学技术大学
University of Science and Technology of China

	leave-one-video-out				leave-one-subject-out					
	SVM		Our method		SVM		Method*		Our method	
	Low	High	Low	High	Low	High	Low	High	Low	High
Low	322	189	308	203	296	215	227	284	314	197
High	205	516	139	582	217	504	163	558	201	520
Acc.	68.0%		72.2%		64.9%		63.7%		67.69%	
F1.	0.620		0.643		0.578		0.504		0.612	
Kappa	0.344		0.418		0.278		0.226		0.335	

- Our method outperforms SVM. Our method exploits the latent space to capture the structure of potential human factors, and leads to better performance.

Experimental Results for Personalized Video Emotion Tagging

	Existing subjects				New subjects					
	leave-one-video-out				leave-one-subject-out					
	SVM		Our method		SVM		Method*		Our method	
	Low	High	Low	High	Low	High	Low	High	Low	High
Low	322	189	308	203	296	215	227	284	314	197
High	205	516	139	582	217	504	163	558	201	520
Acc.	68.0%		72.2%		64.9%		63.7%		67.69%	
F1.	0.620		0.643		0.578		0.504		0.612	
Kappa	0.344		0.418		0.278		0.226		0.335	

- Our method outperforms SVM. Our method exploits the latent space to capture the structure of potential human factors, and leads to better performance.
- The performances of both SVM and our method decrease for new subjects.

Experimental Results for Personalized Video Emotion Tagging

Existing subjects | New subjects

	leave-one-video-out				leave-one-subject-out					
	SVM		Our method		SVM		Method*		Our method	
	Low	High	Low	High	Low	High	Low	High	Low	High
Low	322	189	308	203	296	215	227	284	314	197
High	205	516	139	582	217	504	163	558	201	520
Acc.	68.0%		72.2%		64.9%		63.7%		67.69%	
F1.	0.620		0.643		0.578		0.504		0.612	
Kappa	0.344		0.418		0.278		0.226		0.335	

- Our method outperforms SVM. Our method exploits the latent space to capture the structure of potential human factors, and leads to better performance.
- The performances of both SVM and our method decrease for new subjects.
- In leave-one-subject-out experiments, Method*, which ignores user difference, performs the worst. It predicts the same video sample as the same label across different viewers, which is unreasonable.

- Motivation
- Method
- Experiments
- Conclusion

Conclusion

- ✧ Propose a topic model to capture the inherent relationships among video content, users' characteristic data and personalized emotional tags for personalized emotion video tagging.

- ✧ Experimental results on the CP-QAE-I database demonstrate that the proposed model can generate meaningful latent topics, and improve the performance of personalized video enjoyment tagging.

Thanks!

Any question?