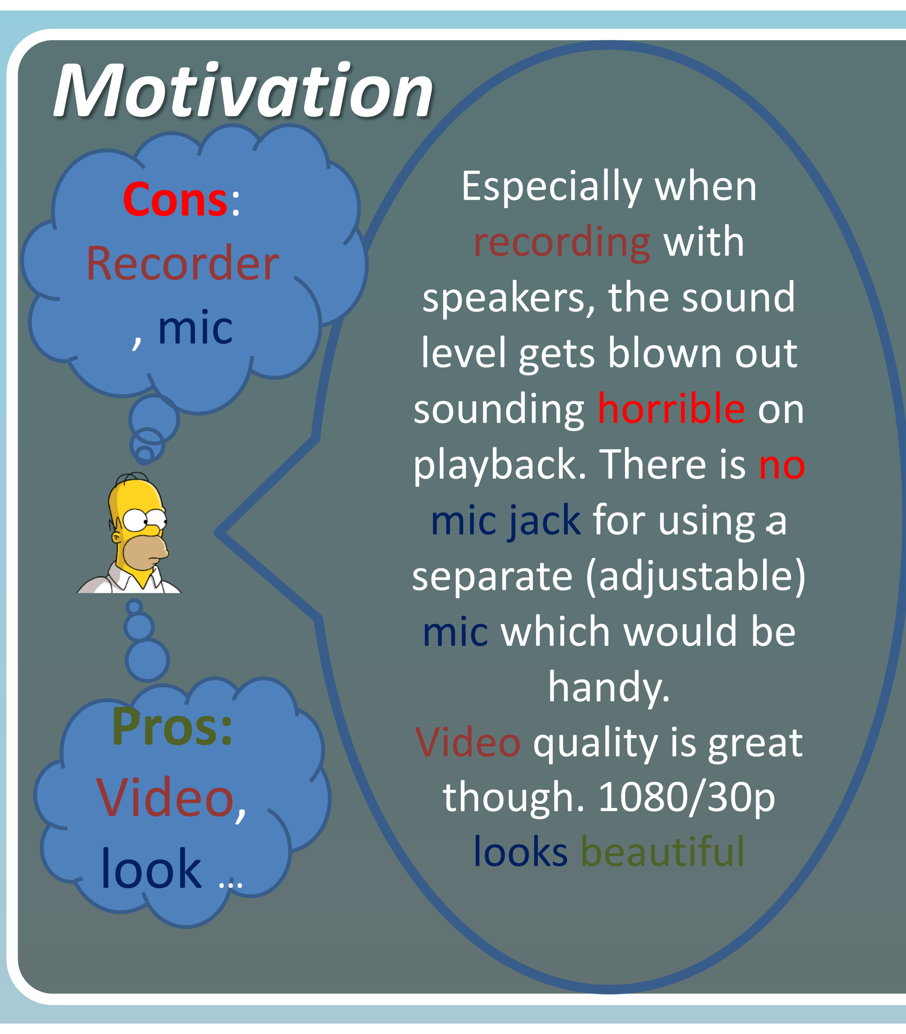
Hidden Topic Sentiment Model

Md Mustafizur Rahman   Hongning Wang
Department of Computer Science, University of Virginia
{mr4xb,hw5x}@virginia.edu

Abstract

User generated reviews contain useful evaluation of various aspects of products with various sentiment, i.e. positive or negative. To better model the sentiment and topics, we propose a novel probabilistic modeling framework called Hidden Topic Sentiment Model (HTSM). HTSM assumes that each sentence in a review document is only associated with one sentiment and one topic; and the sentiment and topic assignment of next sentence will be dependent on the previous sentence. The results show that the topics and sentiment identified by HTSM are much coherent than those topics and sentiment identified by traditional topic models.

Motivation



Observations

1. Sentences used in the reviews either follow the topics of previous sentences or changed into new one.
2. Users have consistent positive or negative attitude towards the same aspects of the entity in one review.

Methods

EM Algorithm

1. E-Step: compute (z, τ, ψ) by alpha- beta recursion

$$\Pr(z_n, \psi_n, \tau_n \mid d, w_1, w_2, \ldots, w_{N_d}, \theta, \beta, \varepsilon, \sigma)$$

$$E(C_{d,z}) = \sum_{n=1}^{N_d} \Pr(z_{d,n} = z, \psi_{d,n} = 1 \mid w_1, w_2, \ldots, w_{N_d})$$

$$E(C_{z,w}) = \sum_{d=1}^{D} \sum_{n=1}^{N_d} \Pr(z_{d,n} = z, w_{d,n} = w \mid w_1, w_2, \ldots, w_{N_d})$$

2. M-Step: update (β, ε ,σ) by maximum likelihood estimation

$$\varepsilon = \frac{\sum_{d=1}^{D} \sum_{n=2}^{N_d} \Pr(\psi_{d,n} = 1 \mid w_1, w_2, \ldots, w_{N_d})}{\sum_{d=1}^{D} (S-1)}$$

$$\sigma = \frac{\sum_{d=1}^{D} \sum_{n=2}^{N_d} \Pr(\tau_{d,n} = 1 \mid w_1, w_2, \ldots, w_{N_d})}{\sum_{d=1}^{D} (S-1)}$$

$$\beta_{z,w} \propto E(C_{z,w}) + \eta - 1$$

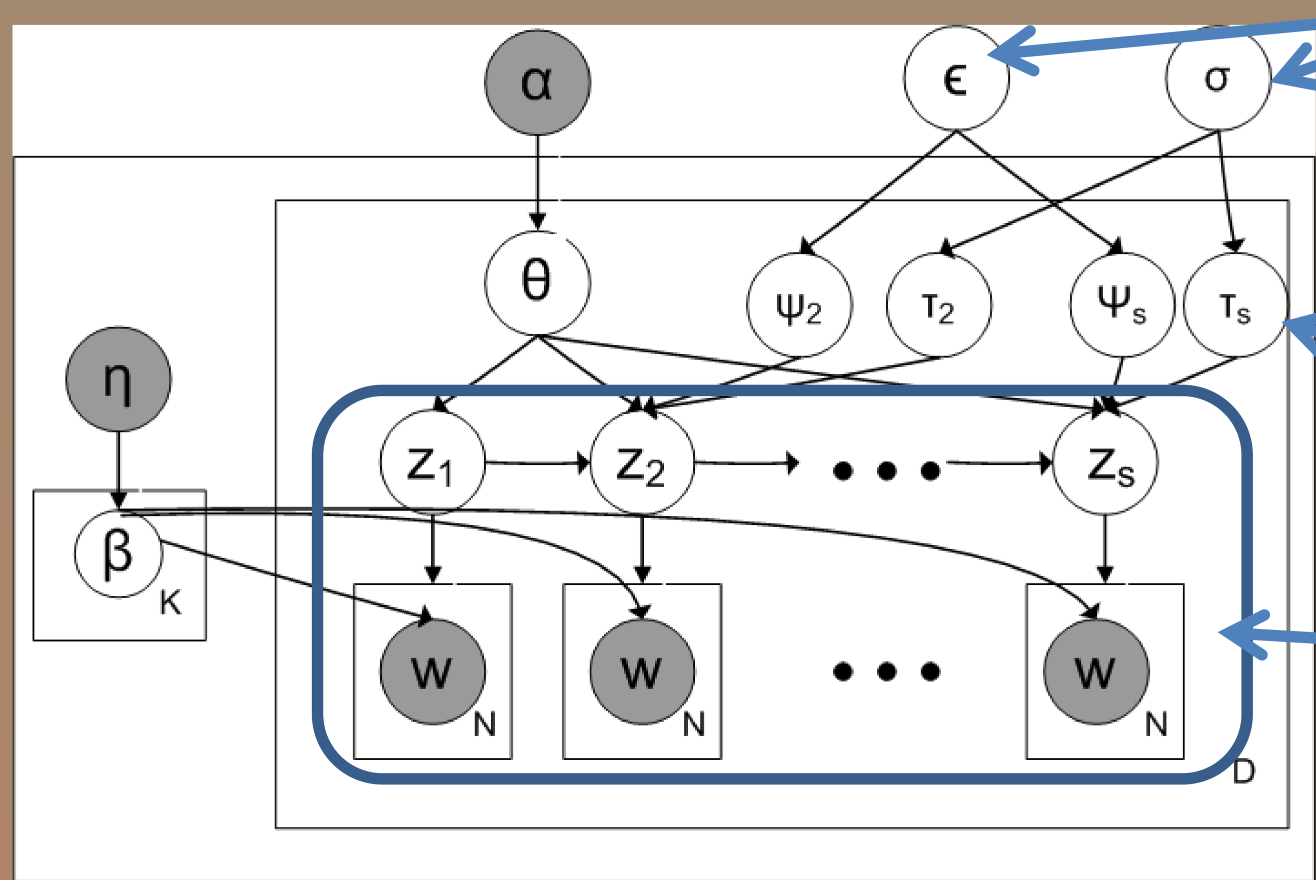


Figure: Graphical Representation of Hidden Topic Sentiment Model

ε is controlling the topic transition and **σ** is for controlling sentiment

Maximum entropy model to predict the **τ** and **ψ**

Hidden Markov Model to model the transition of **topic and sentiment**

Viterbi algorithm for posterior inference

Experimental Results

- **Electronics Reviews from Amazon & NewEgg**

Category	**Amazon**	**NewEgg**
Camera	6919	3020
Tablet	6147	407
Phone	6899	268
Tv	4729	1662

- **Baseline**
  - **Joint Topic Sentiment Model [1]**
  - **Aspect Sentiment Unification Model [2]**

Sentiment seed words

good, nice, excellent, positive, fortunate
bad, nasty, poor, negative, unfortunate

Sentiment Classification

F-1 measure on Positive Sentiment

Category	**JST**	**ASUM**	**HTSM**
Camera	0.693	0.456	**0.779**
Tablet	0.614	0.515	**0.674**
Phone	0.767	0.626	**0.791**
Tv	0.722	0.560	**0.810**

F-1 measure on Negative Sentiment

Category	**JST**	**ASUM**	**HTSM**
Camera	0.484	0.591	**0.708**
Tablet	**0.569**	0.580	0.485
Phone	**0.734**	0.616	0.659
Tv	0.619	0.671	**0.802**

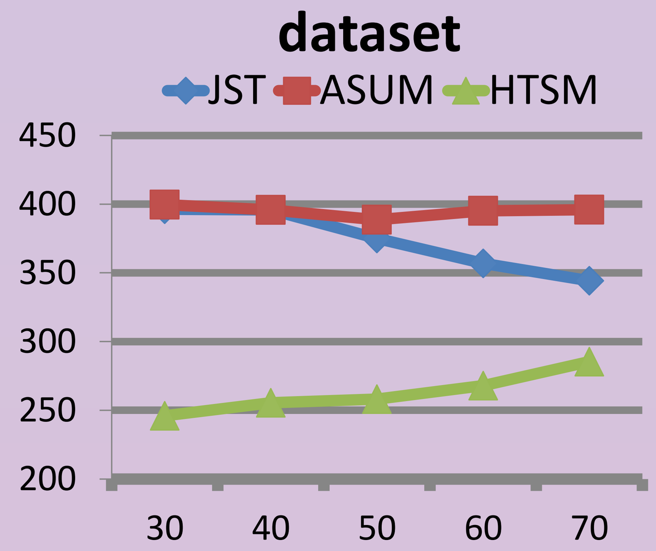
Review summarization on Tablet dataset

Aspects	**Most Probable Sentence**
(-, Apps)	The apps keep freezing up especially yahoo mail
(-, Power)	The power cable is not connecting
(+, Battery)	Battery lasts for week

Top words from Phone dataset

Speaker(p)	**Speaker(n)**	**Message(n)**
good	speaker	Text
speaker	phone	send
sound	bad	receive
great	hear	message
quality	volume	problem

Perplexity on Tablet dataset



Review summarization on Phone dataset

Aspects	**Most Probable Sentence**
(-, Storage)	Storage Capacity very les
(-, Connectivity)	Poor Wi-fi sensitivity
(-, call quality)	Persistent echo during call

Conclusion

HTSM enhances the independence assumption of sentiment and topics in a opinionated text document and thus captures the dependency among them. Through extensive experiment evaluations, HTSM better modeled sentiment and topics in review texts and outperforms existing joint topic sentiment models.

References

1. C. Lin and Y. He. Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 375-384. ACM, 2009.
2. Y. Jo and A. H. Oh. Aspect and sentiment unification model for online review analysis. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 815-824. ACM, 2011