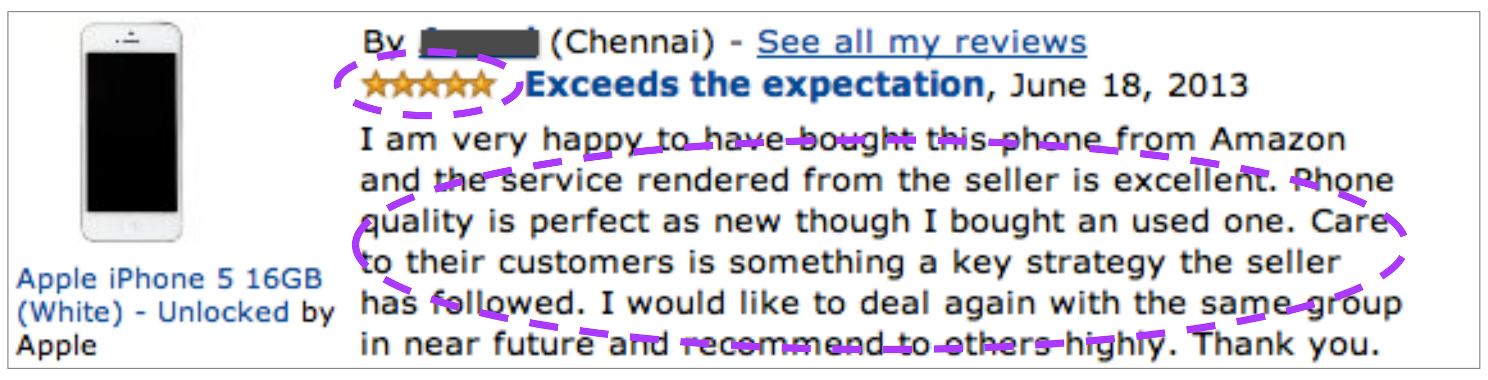
Do Users Rate or Review? Boost Phrase-level Sentiment Labeling with Review-level Sentiment Classification

Yongfeng Zhang, Haochen Zhang, Min Zhang, Yiqun Liu, Shaoping Ma

Information Retrieval Group, Department of Computer Science, Tsinghua University

❖ Motivation and Basic Findings

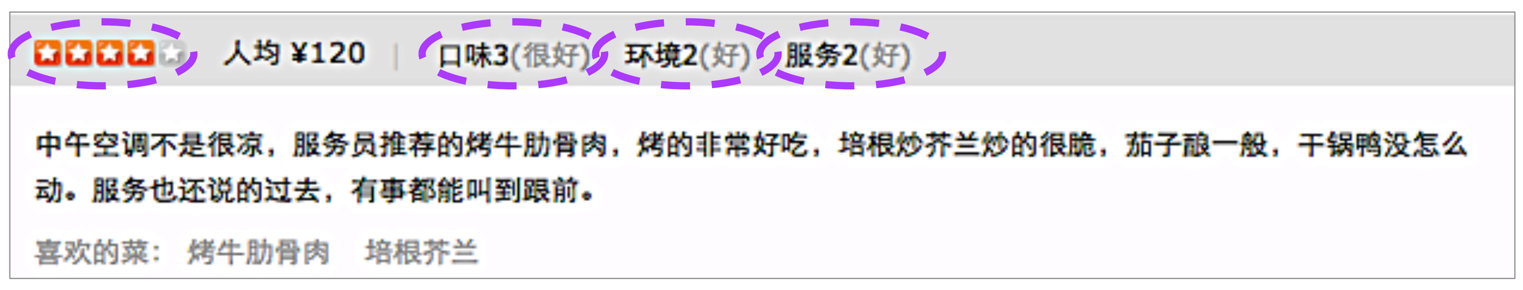
- Many online applications allow users to express his/her opinions by **reviews**. Two basic gradients of a review are the **numerical star rating** and the **review text**.



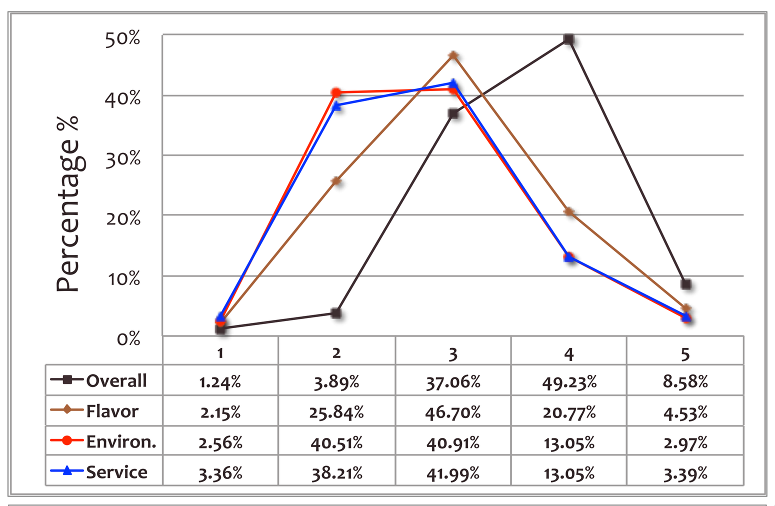
- **Phrase-level Sentiment Analysis** is important in many tasks, e.g. product summarization, keywords extraction.
- A sentiment lexicon is usually constructed
  - (Feature Word, Opinion Word, Sentiment Polarity)
  - e.g. (Phone quality, perfect, positive)
- Current approaches for polarity labeling assume that user's numerical rating represents the overall sentiment of the corresponding review text, however, we find that this assumption is not necessarily true.

❖ User Rating Analysis and Statistics

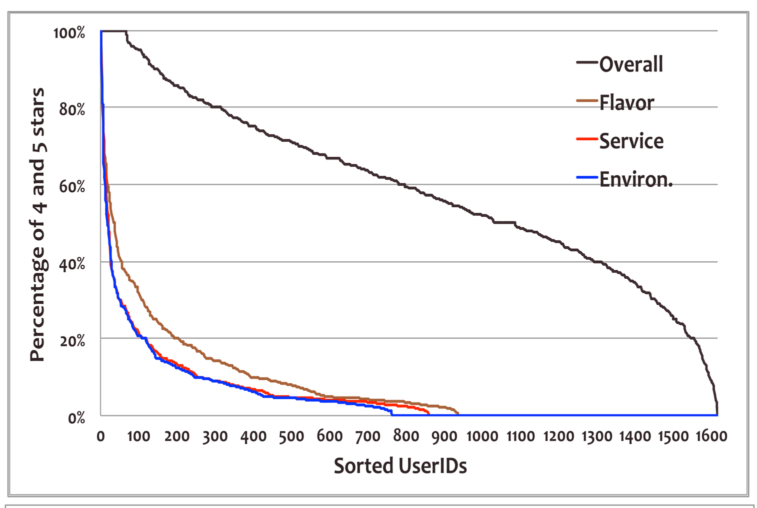
- Adopt the reviews from DianPing.com
  - Each piece of review has an overall rating + three sub-aspect ratings
  - Sub-aspects: Flavor, Environment, Service



- The percentage of each (of the five) stars on Overall rating, Flavor, Environment and Service.
- The percentage of 4+ ratings made by each user.



Users tend to give relatively higher scores on overall rating, while they tend to 'tell the true feelings' and make relatively lower scores on detailed sub-aspect ratings.



Nearly 70% of the users made more than a half 4+ ratings on overall rating, while only less than 5% users did so on the three kinds of sub-aspect ratings.

	Overall	Flavor	Environment	Service
Average Rating	3.6432	3.1547	2.8934	2.8510
Coefficient of Variation	0.1977	0.2522	0.2697	0.2816

- Precisions of review-level sentiment polarity labeling.

	Overall rating	Normalized overall rating	Averaged sub-ratings	Sentiment classification[1]
Label as positive	≥4	≥0	≥4	By algorithm
Label as negative	<4	<0	<4	By algorithm
Positive Review	0.8321	0.5438	0.8009	0.9064
Negative Review	0.7248	0.7859	0.7951	0.8563
Average	0.7970	0.6230	0.7990	0.8900

[1] T. Zagibalov, J. Carrol. Automatic Seed Word Selection for Unsupervised Sentiment Classification of Chinese Text. *Coling* pages 1073-1080, 2008.
[2] M. Hu and B. Liu. Mining and Summarizing Customer Reviews. KDD, 2004.
[3] Y. Lu, M. Castellanos, U. Dayal, and C. Zhai. Automatic Construction of a Context-Aware Sentiment Lexicon: An optimization approach. WWW 2011.

❖ Framework for Sentiment Polarity Labeling

- **Step1.Review-level Sentiment Classification**
  - Classify the sentiment of each review [1][2]
  - Construct review sentiment matrix $\tilde{\mathbf{X}} = [\mathbf{x}_1 \mathbf{x}_2 \cdots \mathbf{x}_m]^T$

    positive: $\mathbf{x} = [1,0]^T$ negative: $\mathbf{x} = [0,1]^T$
- **Step2.Phrase-level Sentiment Polarity Labeling**
  - An optimization framework with four constraints.
  - 1) Review-level Sentiment Orientation.

$$\mathcal{R}_1 = \|\mathbf{A}\mathbf{X} - \tilde{\mathbf{X}}\|_F^2 \quad a_{ij} = I_{ij}^{neg} \cdot \frac{\mathrm{Freq}(i,j)}{\sum_k \mathrm{Freq}(i,k)}$$

  - 2) General Sentiment Lexicon

$\mathcal{R}_2 = \|\mathbf{G}(\mathbf{X} - \mathbf{X}_0)\|_F^2 \quad \mathbf{G}_{ii} = 1$ for fixed-sentiment pairs.

  - 3) Linguistic Heuristics (for 'and' / 'but')

$$\mathcal{R}_3^a = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|\mathbf{X}_{i*} - \mathbf{X}_{j*}\|_F^2 \mathbf{W}_{ij}^a \quad \mathcal{R}_3^b = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|\mathbf{X}_{i*} - \mathbf{X}_{j*}\mathbf{E}\|_F^2 \mathbf{W}_{ij}^b$$

$\mathbf{W}_{ij}^a = \mathbf{W}_{ji}^a = 1$ / $\mathbf{W}_{ij}^b = \mathbf{W}_{ji}^b = 1$ when linked by and / but.

  - 4) Sentential Sentiment Consistency

$$\begin{matrix}\mathbf{W}_{ij}^s \\ \mathbf{W}_{ji}^s\end{matrix} = \begin{cases} 0, \ if\ N_{ij} = 0 \text{ or } \mathbf{W}_{ij}^a \neq 0 \text{ or } \mathbf{W}_{ij}^b \neq 0 \\ \frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}}\left(1 - \frac{dist(i,j)}{length(r_{i_k})}\right), \ else \end{cases} \quad \mathcal{R}_4 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|\mathbf{X}_{i*} - \mathbf{X}_{j*}\|_F^2 \mathbf{W}_{ij}^s$$

- **The Unified Model for Polarity Labeling**

$$\min_{\mathbf{X} \geq 0} \mathcal{R} = \lambda_1\mathcal{R}_1 + \lambda_2\mathcal{R}_2 + \lambda_3(\mathcal{R}_3^a + \mathcal{R}_3^b) + \lambda_4\mathcal{R}_4$$

$$\mathbf{X}_{ij} \leftarrow \mathbf{X}_{ij}\sqrt{\frac{[\lambda_1\mathbf{A}^T\tilde{\mathbf{X}} + \lambda_2\mathbf{G}\mathbf{X}_0 + \lambda_3\mathbf{W}^a\mathbf{X} + \lambda_3\mathbf{W}^b\mathbf{X}\mathbf{E} + \lambda_4\mathbf{W}^s\mathbf{X}]_{ij}}{[\lambda_1\mathbf{A}^T\mathbf{A}\mathbf{X} + \lambda_2\mathbf{G}\mathbf{X} + \lambda_3\mathbf{D}\mathbf{X} + \lambda_4\mathbf{D}^s\mathbf{X}]_{ij}}}$$

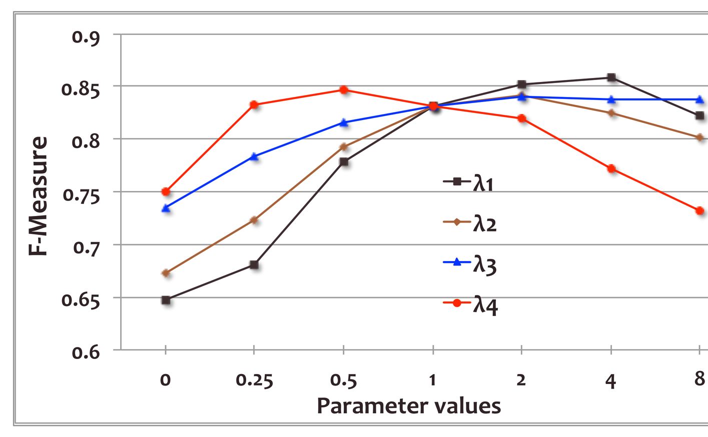
❖ Phrase-level Polarity Labeling Results

- When fixing $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 1$

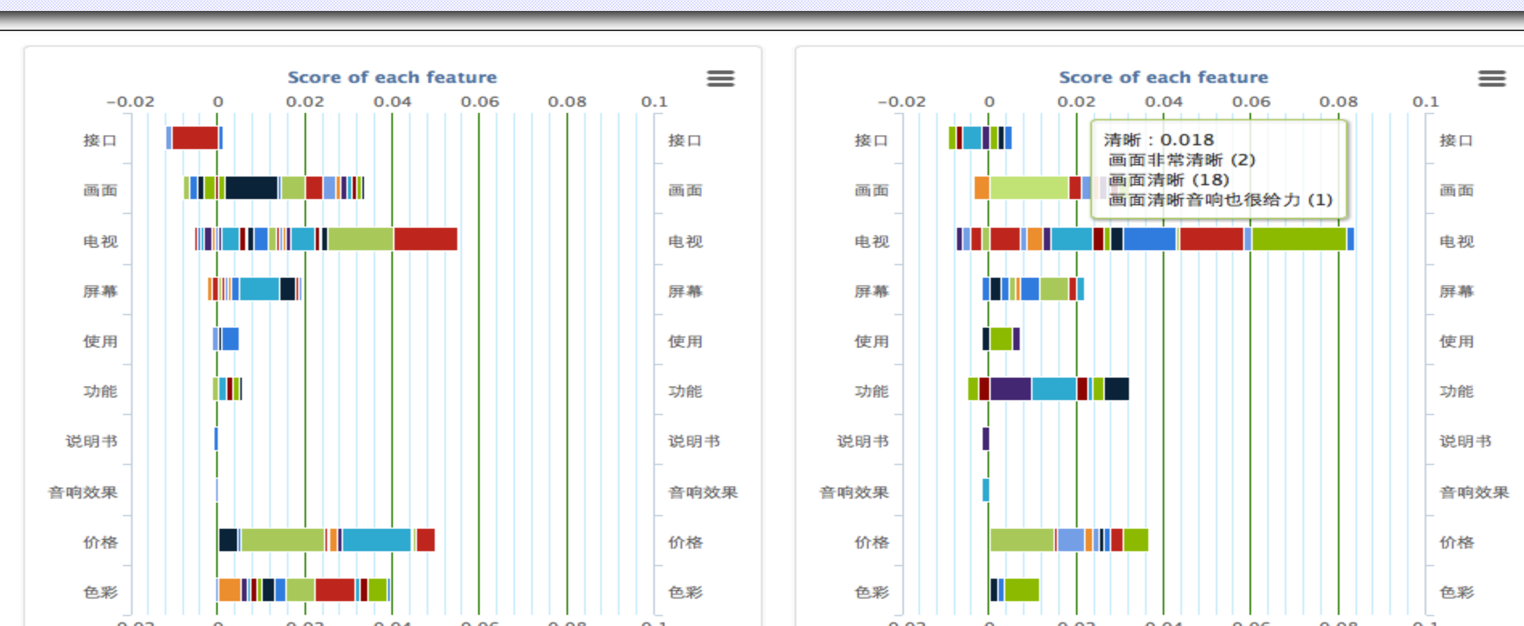
	Precision	Recall	F-measure
MP3 Player Dataset (English)			
By general sentiment lexicon	**0.9238**	0.4201	0.5776
Optimization framework in [3]	0.8269	0.7626	0.7934
Our framework with overall rating	0.8288	0.7525	0.7888
Our full framework	0.8504*	**0.7683**	**0.8073**
Restaurant Review Dataset (Chinese)			
By general sentiment lexicon	**0.9017**	0.3571	0.5115
Optimization framework in [3]	0.8405	0.7760	0.8069
Our framework with overall rating	0.8473	0.7468	0.7938
Our framework with subratings	0.8675	0.7561	0.8079
Our full framework	0.8879*	**0.7818**	**0.8315**

- Parameter Analysis

	λ1	λ2	λ3	λ4	MP3 Player	Restaurant
Default	1	1	1	1	0.8073	0.8315
Knock Out One Term	0	1	1	1	0.6783	0.6476
	1	0	1	1	0.6332	0.6728
	1	1	0	1	0.7461	0.7352
	1	1	1	0	0.7756	0.7504



❖ Demo for Online Product Comparison



Yongfeng Zhang, Dpt. of Computer Science, Tsinghua University, Email: zhangyf07@gmail.com Demo: http://tv.thuir.org