EMNLP 2021
Online & in the Dominican Republic
Conference on Empirical Methods in Natural Language Processing

Recommend for a Reason: Unlocking the Power of Unsupervised Aspect-Sentiment Co-Extraction

Zeyu Li[1], Wei Cheng[2], Reema Kshetramade[1], John Houser[1], Haifeng Chen[2], Wei Wang[1]
[1]University of California, Los Angeles,
[2]NEC Labs America

Paper ID: 792

UCLA Samueli School of Engineering
NEC LABORATORIES AMERICA, INC.
Relentless passion for innovation

Introduction

- Compliments and concerns in reviews are valuable for understanding users' shopping interests and their opinions to specific aspects of items.
- Existing work: Ignores the fact that users may hold different attentions to various properties of the merchandise.
- User may show strong ***attentions*** to certain ***properties*** but indifference to others.

Reviews	Microphone	Comfort	Sound
R1 [5 stars]: *Comfortable. Very high quality sound. . . . Mic is good too. There is an switch to mute your mic. . . I wear glasses and these are comfortable with my glasses on. . . .*	*good* (satisfied)	*comfortable*	*high quality* (praising)
R2 [3 stars]: *I love the comfort, sound, and style but the mic is complete junk!*	*complete junk* (angry)	*love*	*love*
R3 [5 stars]: *. . . But this one feels like a pillow, there's nothing wrong with the audio and it does the job. . . . con is that the included microphone is pretty bad.*	*pretty bad* (unsatisfied)	*like a pillow* (enjoyable)	*nothing wrong*

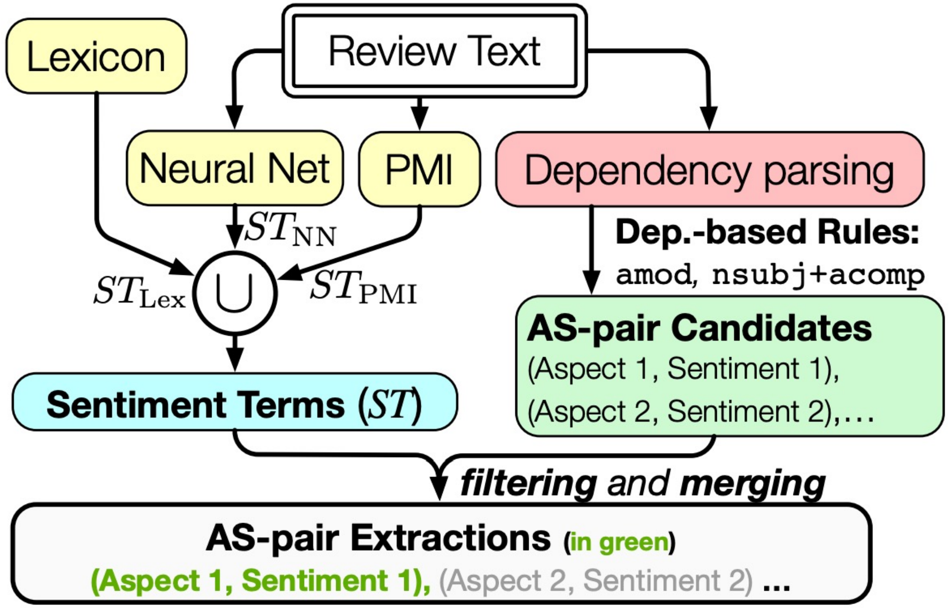
Table 1: Example reviews of a headset with three aspects, namely microphone quality, comfort level, and sound quality, highlighted specifically. The extracted sentiments are on the right. R1 vs. R2: Different users react differently (microphone quality) to the same item due to distinct personal attentions and, consequently, give divergent ratings. R1 vs. R3: A user can still rate highly of an item due to special attention on particular aspects (comfort level) regardless of certain unsatisfactory or indifferent properties (microphone and sound qualities).

Introduction

- Compliments and concerns in reviews are valuable for understanding users' shopping interests and their opinions to specific aspects of items.
- Existing work: Ignores the fact that users may hold different attentions to various properties of the merchandise.
- User may show strong ***attentions*** to certain ***properties*** but indifference to others.
- We propose a tightly coupled two-stage approach: Aspect-Sentiment Pair Extractor (***ASPE***) + Attention-Property-aware Rating Estimator (***APRE***)

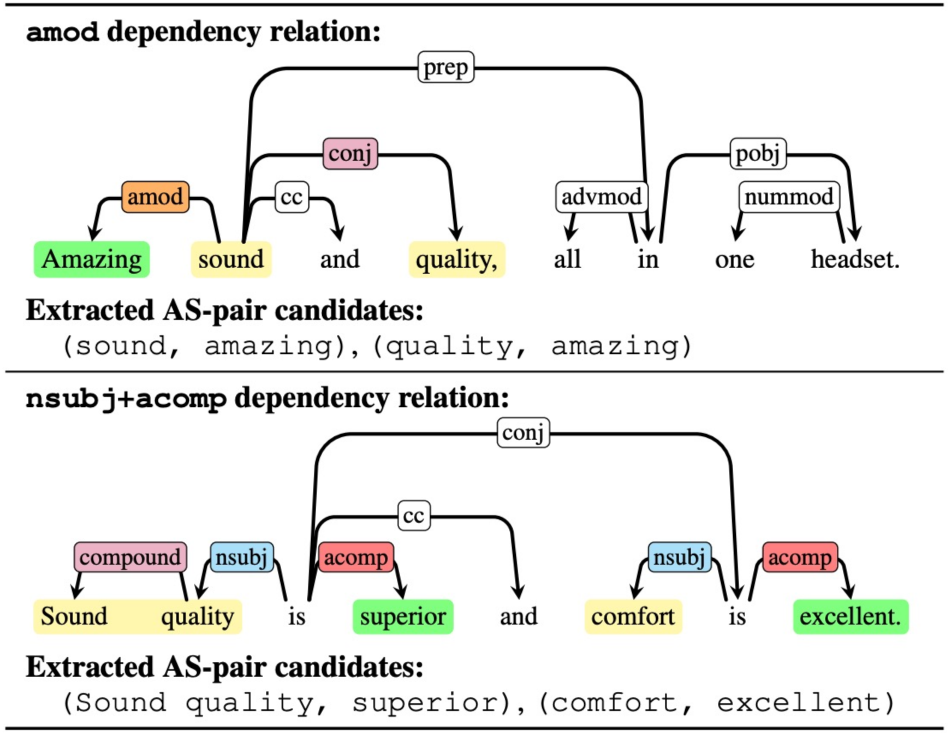
ASPE

- First Step: Sentiment terms extraction
  - Pointwise Mutual Information-based (PMI-based) (Polarity)
  - Neural Network-based (NN-based) (Linguistic patterns)
  - Knowledge/Lexicon-based (Existing knowledge)
  - $ST = ST_{PMI} \cup ST_{NN} \cup ST_{Lex}$



ASPE

- First Step: Sentiment terms extraction
- Second Step: AS-pair extraction
  - First: label AS-pair candidates using dependency parsing
  - Second: filter out non-sentiment-carrying candidates using ST



UCLA Samueli School of Engineering

NEC LABORATORIES AMERICA, INC.
Relentless passion for innovation

APRE

- Language encoding with pre-trained BERT
- ***Explicit*** aspect-level attitude modeling

$$\alpha_{u,r}^{(a)} = \frac{\exp(\tanh(\boldsymbol{w}_{\text{ex}}^T[\boldsymbol{h}_{u,r}^{(a)}; \boldsymbol{a}^{(u)}]))}{\sum_{r' \in R^u} \exp(\tanh(\boldsymbol{w}_{\text{ex}}^T[\boldsymbol{h}_{u,r'}^{(a)}; \boldsymbol{a}^{(u)}]))} \quad \boldsymbol{g}_u^{(a)} = \sum_{r \in R^u} \alpha_{u,r}^{(a)} \boldsymbol{h}_{u,r}^{(a)}$$

- ***Implicit*** review representation

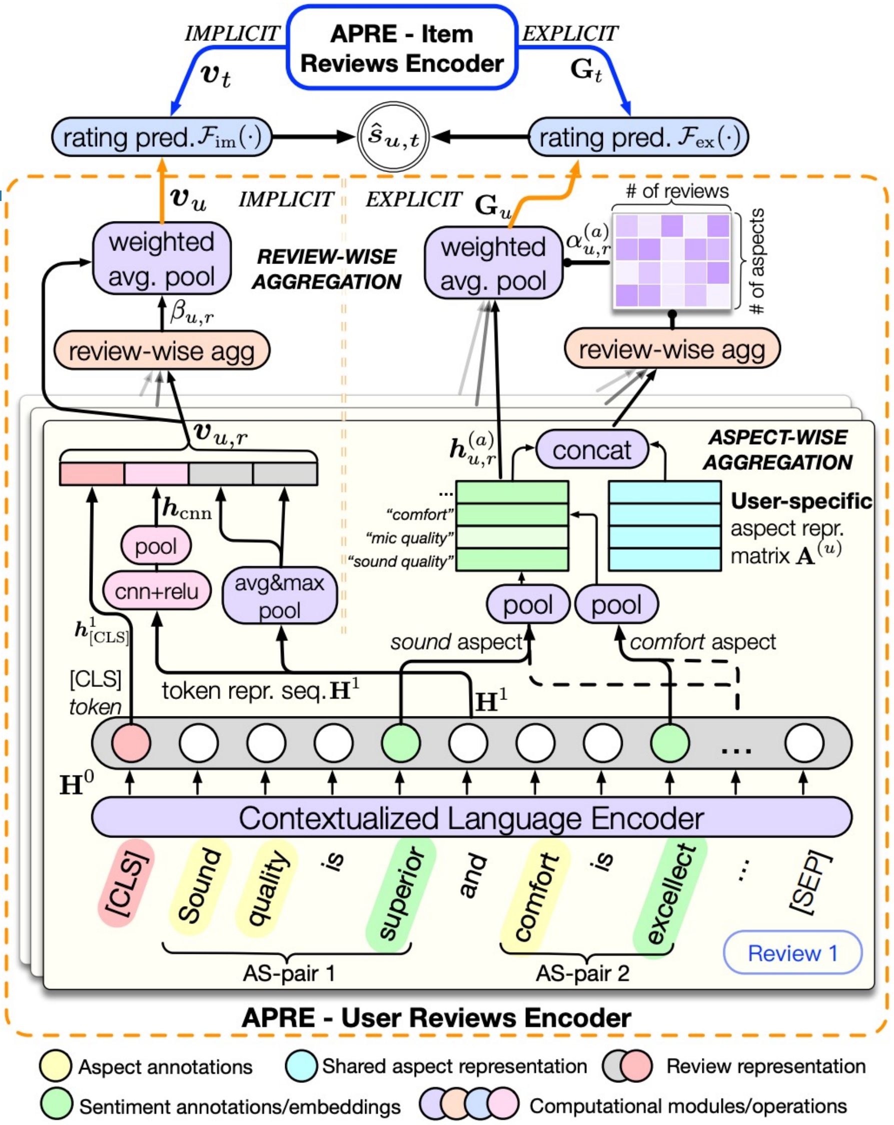
$$\boldsymbol{v}_{u,r} = [\boldsymbol{h}_{[\text{CLS}]}^1; \boldsymbol{h}_{\text{cnn}}; \text{MaxPool}(\mathbf{H}^1); \text{AvgPool}(\mathbf{H}^1)]$$

$$\boldsymbol{h}_{\text{cnn}} = \text{MaxPool}(\text{ReLU}(\text{ConvNN\_1D}(\mathbf{H}^1))).$$

$$\beta_{u,r} = \frac{\exp(\tanh(\boldsymbol{w}_{\text{im}}^T \boldsymbol{v}_{u,r}))}{\sum_{r' \in R^u} \exp(\tanh(\boldsymbol{w}_{\text{im}}^T \boldsymbol{v}_{u,r'}))} \quad \boldsymbol{v}_u = \sum_{r \in R^u} \beta_{u,r} \boldsymbol{v}_{u,r}$$

- Rating regression and optimization

$$\hat{s}_{u,t} = \underbrace{b_u + b_t}_{\text{biases}} + \underbrace{\mathcal{F}_{\text{im}}([\boldsymbol{v}_u; \boldsymbol{v}_t])}_{\text{implicit feature}} + \underbrace{\langle \boldsymbol{\gamma}, \mathcal{F}_{\text{ex}}([\mathbf{G}_u; \mathbf{G}_t]) \rangle}_{\text{explicit feature}}$$



Experiments

- Dataset: 7 Amazon Review Datasets (8:1:1)
- Baseline models: 13 models (traditional + deep learning)
- Evaluation metric: MSE of rating
- Conclusion:
  - Ours vs. the rest: Demonstrates the superior capability of APRE to ***make accurate rating predictions*** in different domains;
  - Ours vs. AHN-B: The performance improvement is ***NOT*** because of the use of BERT

Models	AM	DM	MI	PS	SO	TG	TH
TRADITIONAL MODELS							
MF	1.986	1.715	2.085	2.048	2.084	1.471	1.631
WRMF	1.327	0.537	1.358	1.629	1.371	1.068	1.216
FM	1.082	0.436	1.146	1.458	1.212	0.922	1.050
DEEP LEARNING-BASED MODELS							
ConvMF	1.046	0.407	1.075	1.458	1.026	0.986	1.104
NeuMF	0.901	0.396	0.903	1.294	0.893	0.841	1.072
D-Attn	0.816	0.403	0.835	1.264	0.897	0.887	0.980
D-CNN	0.809	0.390	0.861	1.250	0.894	0.835	0.975
NARRE	0.826	0.374	0.837	1.425	0.990	0.908	<u>0.958</u>
MPCN	0.815	0.447	0.842	1.300	0.929	0.898	0.969
ANR	0.806	0.381	0.845	1.327	0.906	0.844	0.981
DAML	0.829	<u>0.372</u>	0.837	<u>1.247</u>	0.893	<u>0.820</u>	0.962
AHN-B	0.810	0.385	0.840	1.270	0.896	0.829	0.976
AHN	<u>0.802</u>	0.376	<u>0.834</u>	1.252	<u>0.887</u>	0.822	0.967
OUR MODELS AND PERCENTAGE IMPROVEMENTS							
Ours	**0.791**	**0.359**	**0.823**	**1.218**	**0.863**	**0.788**	**0.936**
$\Delta(\%)$	1.390	3.621	1.337	2.381	2.784	4.061	2.350
Val.	0.790	0.362	0.821	1.216	0.860	0.790	0.933
ABLATION STUDIES							
w/o EX	0.814	0.379	0.833	1.244.	0.882	0.796	0.965
w/o IM	0.798	0.374	0.863	1.226	0.873	0.798	0.956

Case Study

- APRE measures the aspect-level contributions of user-attention and item-property interactions via this term:
  - $<\gamma, F_{ex}([G_u;G_t])>$;
- ***Inferred Impact*** row states the interactional effects of user attentions and item properties based on our assumption that *attended aspects bear stronger impacts to the final prediction*.
- This process of decomposition is a great way to ***interpret*** model prediction on an aspect-level granularity

Aspects	material	smell	battery	install	look	price	sound
Attn. of u_*	✓	✓	✓	✓	✓	n/a	n/a
Prop. of t_*	n/a	n/a	✗	✓	✓	✗	✓/✗
Inferred Impact	*Unk.*	*Unk.*	***Neg.***	***Pos.***	***Pos.***	*Unk.*	*Unk.*
$\gamma_i \mathcal{F}_{\text{ex}}(\cdot)_i$ ($\times 10^{-2}$)	1.0	0.8	-0.8	1.9	1.5	0.9	0.6

Thanks!

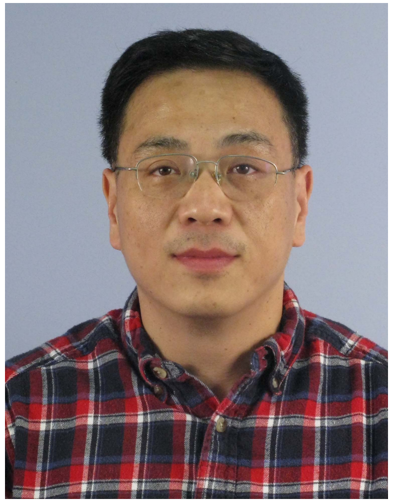
Code on GitHub: `https://github.com/zyli93/ASPE-APRE`

Paper ID: **792**

Looking forward to seeing you in the **POSTER** session!

EMNLP 2021
Online & in the Dominican Republic
Conference on Empirical Methods in Natural Language Processing




UCLA Samueli School of Engineering

NEC LABORATORIES AMERICA, INC.
Relentless passion for innovation