Weakly-Supervised Opinion Summarization by Leveraging External Information

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Chao Zhao, Snigdha Chaturvedi

Department of Computer Science, UNC Chapel Hill

The University of North Carolina Department of Computer Science

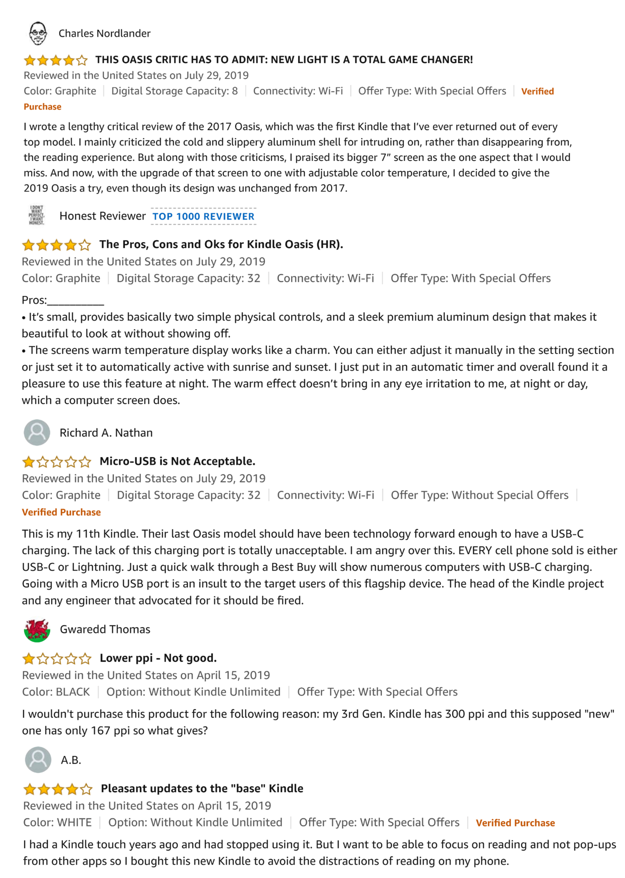
Overview

Opinion summarization from online product reviews is a challenging task involving identifying opinions related to various aspects of the product. In this work,

- We address this task without human supervision – by incorporating domain knowledge from external information;
- We propose AspMem, a generative approach for aspect identification that can use such knowledge effectively;
- We adopt a three-step pipeline to create summaries by identifying aspects, computing salience, and selecting opinions.

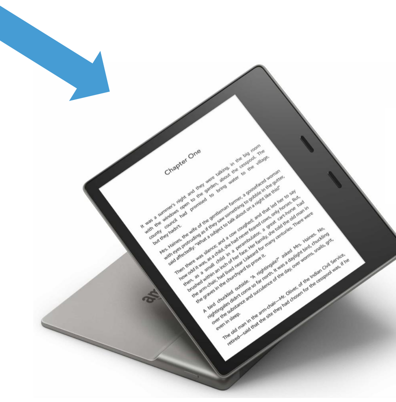
Task Description

Customer Reviews



Feature descriptions

- Our best 7", 300 ppi flush-front Paperwhite display.
- Adjustable warm light to shift screen shade from white to amber.
- Waterproof (IPX8) so you can read in the bath or by the pool.
- Thin and light ergonomic design with page turn buttons.
- Reads like real paper with the latest e-ink technology for fast page
- Instant access to millions of books, newspapers, and audiobooks.
- Works with Audible - pair with Bluetooth headphones or speakers to switch seamlessly between reading and listening.
- A single battery charge lasts weeks, not hours.



Review Summary

It looks and feels much **bigger** than you'd think. The **storage capability** is great. More storage, more books! **Lighting to adaptive** with 2 ambient sensors. One HUGE improvement is that the Oasis is **waterproof**! Overall it's a great e-reader, but abysmal **battery life** negates all the positive features. Shorter than Voyage or Original Oasis. Significantly short. Also, one cannot follow text with **audio**.

Aspect Identification

AspMem contains an array of memory cells $\mathcal{A} = \{a_1, a_2, \cdots, a_k\}$ to store aspect-related information. Similar to topic models, we assume the review segment s is generated from these aspect (topic) memories.

Sentence Representation

$$P(v_i|a_j) \propto \exp(\cos(\boldsymbol{v}_i, \boldsymbol{a}_j)),$$
$$P(v_i) = \sum_j P(v_i|a_j)P(a_j),$$
$$z_i = \frac{P(v_i)}{\sum_j P(v_j)}, \quad \boldsymbol{s} = \sum_i z_i \boldsymbol{v}_i.$$

Sentence Generation

$$P(s|a_i) \propto \exp(\cos(\boldsymbol{s}, \boldsymbol{a}_i)).$$

Likelihood Maximization

$$J(\theta) = -\sum_{s \in \mathcal{X}} \log P(s; \boldsymbol{\theta}) + \lambda \left\| \hat{\mathbf{A}}\hat{\mathbf{A}}^{\mathbf{T}} - \mathbf{I} \right\|_2.$$

Aspect Prediction

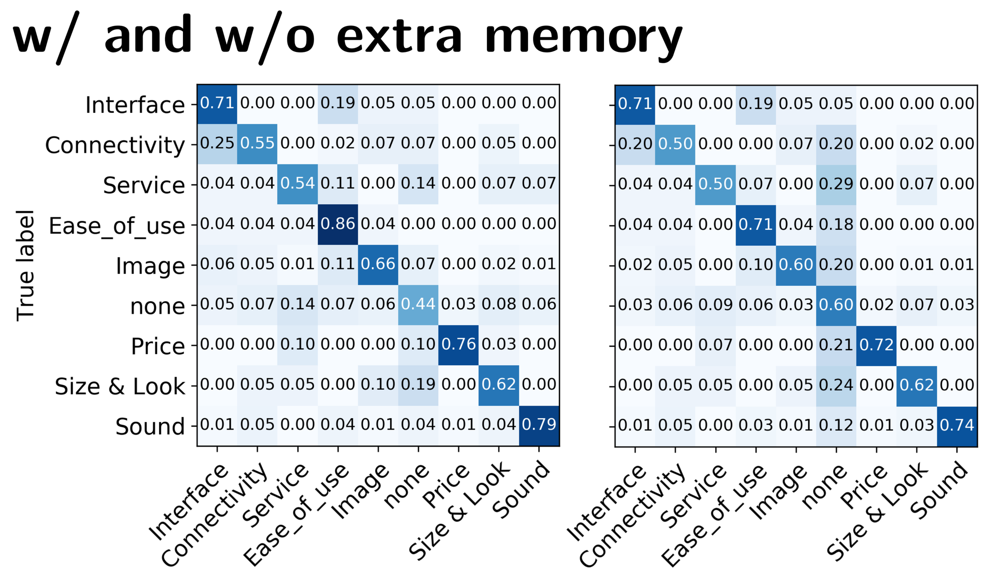
$$P(a_i|s) \propto P(s|a_i)P(a_i).$$

Results

Evaluated by multi-label micro F_1

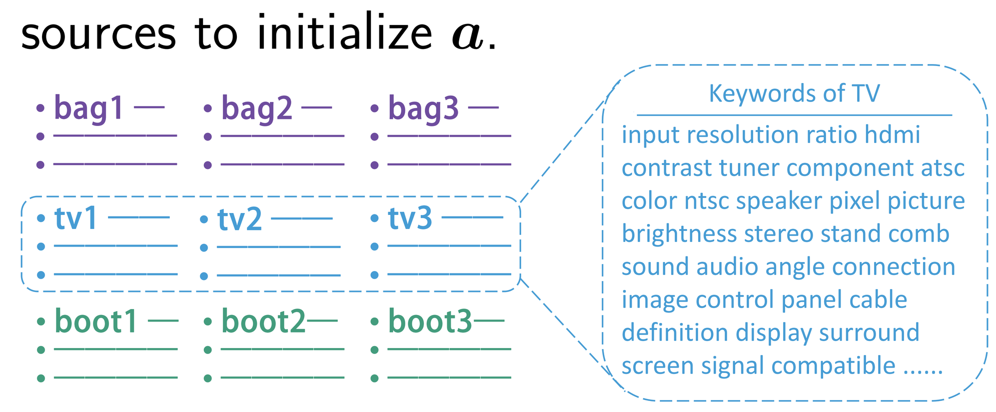
Model	Bags	B/T	Boots	Keyb/s	TVs	Vacs	Ave.
ABAE (2017)	41.6	48.5	41.0	41.3	45.7	40.6	43.2
MATE (2018)	48.6	54.5	46.4	45.3	51.8	47.7	49.1
BERT (2019)	**61.4**	**66.5**	52.0	57.5	**63.0**	60.4	**60.2**
AspMem	52.4	58.1	54.5	51.4	53.9	54.6	54.2
w/ extra mem	60.0	62.0	**55.8**	**61.8**	60.0	**61.8**	**60.2**

w/ and w/o extra memory



External Information

Rather than estimating the aspect memory $\boldsymbol{a}$ by human-annotated data or unsupervised methods, we select the top K keywords with the highest TF-IDF value from external resources to initialize $\boldsymbol{a}$.



Salience Computation

Relevance

✔ "It looks and feels much bigger."

✘ "I love this one!"

$$\mathbb{S}_{rel}(s) = \frac{1}{|s|} \sum_i \max_{j=\{1,\cdots,K\}} g(\cos(\boldsymbol{v}_i, \boldsymbol{a}_j) \cdot w_j).$$

Sentiment

✔ "The storage is great!"

✘ "Bought it for my daughter."

$$\mathbb{S}_{senti}(s) = \mathbf{p}^T \text{softmax}(\text{MLP}(\text{LSTM}(s)))$$

Overall

$$\mathbb{S}_{sal}(s) = \mathbb{S}_{rel}(s) \times \mathbb{S}_{senti}(s).$$

Opinion Selection

An ideal summary would contain high-salience opinions, avoid redundant information, and meet the limit on the length. These goals can be formalized as an ILP problem:

$$\boldsymbol{\alpha} = \arg\max_{\boldsymbol{\alpha}} \sum_i \mathbb{S}_{sal}(s_i)\alpha_i - \sum_{i,j} sim_{ij}\beta_{ij},$$

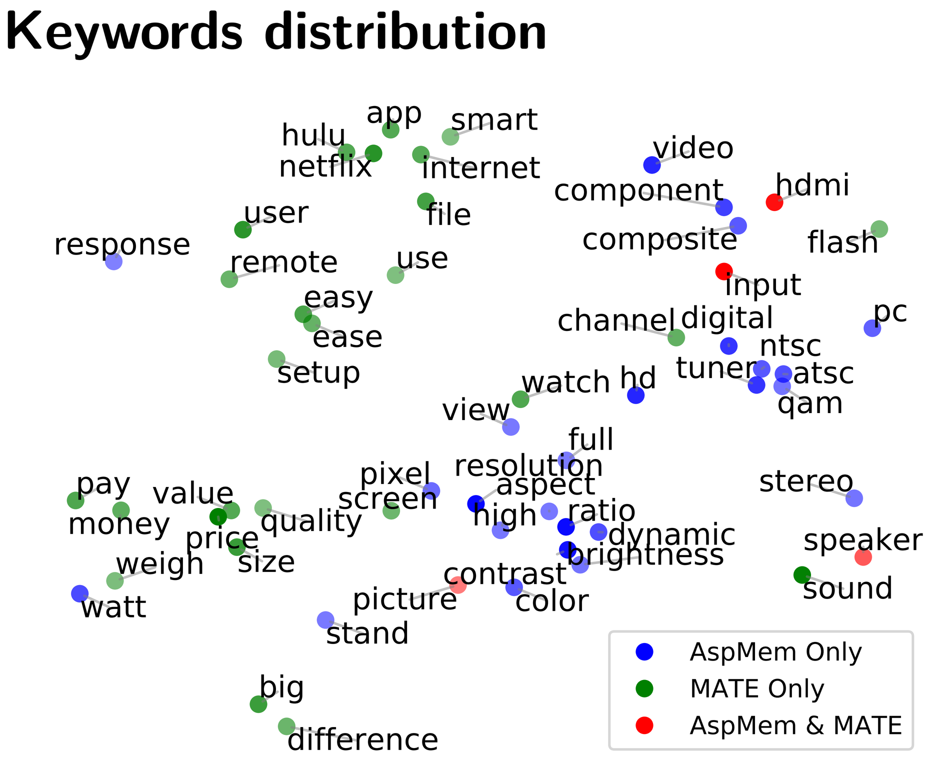
$$\begin{aligned} s.t. \quad & \alpha_i, \beta_{ij} \in \{0,1\} && \forall i,j \\ & \beta_{ij} \geq \alpha_i + '\alpha_j - 1 && \forall i,j \\ & \beta_{ij} \leq \frac{1}{2}(\alpha_i + \alpha_j) && \forall i,j \\ & \textstyle\sum_i \alpha_i l_i \leq L && \forall i \end{aligned}$$

Results

Evaluated by Rouge

Methods	R-1	R-2
Lead	35.5	15.2
LexRank	37.7	14.1
MATE + MILNET [1]	44.1	21.8
AspMemSum	46.6	25.7
w/o filtering	**48.0**	**28.7**
w/o Relevance	41.5	20.5
w/o Sentiment	40.5	18.2
w/o ILP	46.2	25.1
Inter-annotator Agreement	54.7	36.6

Keywords distribution



Summary examples

So this is the **perfect size**. The **padding** is good and it is quite **stylish**. there is just **enough size** for a mouse, power cord, a fre CD's and a cell phone. **Functionality** wise : awesome! The **price** on Amazon is great. Unfortunately over the holidays the **hand strap broke**. As I stated above, getting the **computer** in and out is pain.

Fantastic **sound, great fit** and a simple practicle **set up**. i find this device very good. And with the additional purchase of some acustibud ear pieces the **sounds** is quite good. This is an awesome product. The **design** of the ear pieces **did not fit my ear** anatomy very well. the **sound** isn't as rich as other on ear Bluetooth.

[1] Angelidis, S., Lapata, M. (2018). Summarizing Opinions: Aspect Extraction Meets Sentiment Prediction and They Are Both Weakly Supervised. In EMNLP 2018.

Contact Info

zhaochaocs.github.io/

zhaochaocs@gmail.com

github.com/zhaochaocs/AspMem