Summarizing Opinions: Aspect Extraction Meets Sentiment Prediction and They Are Both Weakly Supervised

Stefanos Angelidis and Mirella Lapata
Institute for Language, Cognition, and Computation
School of Informatics, University of Edinburgh

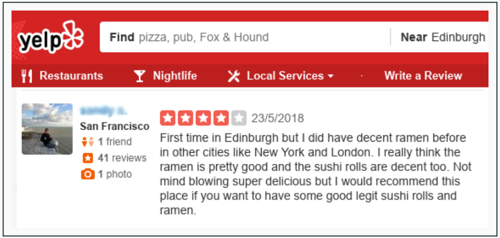
http://stangelid.github.io
✉ s.angelidis@ed.ac.uk

Edinburgh
University of Edinburgh
Natural Language Processing
NLP

Online User Reviews

Heavily influence customer decisions:

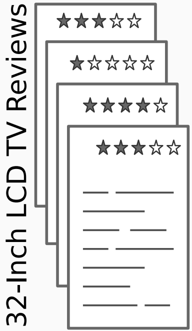
- Travel booking (Ye et al., 2009)
- Box Office success (Duan et al., 2008)
- Shopping (TurnTo.com report, 2018)



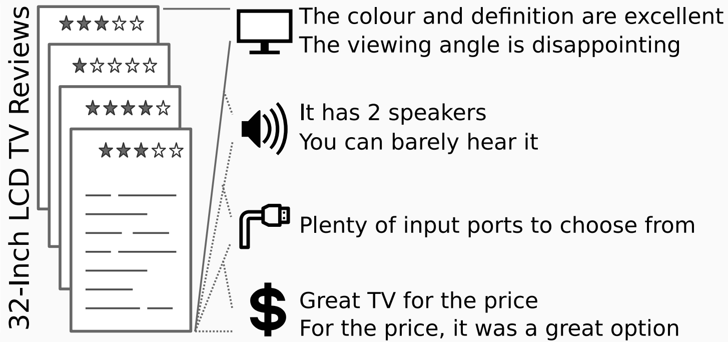
Incredibly rich data source:

- Yelp reviews:
  6.3 million in 2010
  27.3 million in 2017
- Amazon reviews:
  143 million from 1996 to 2014

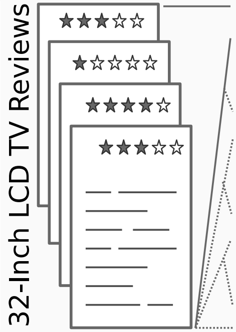
Extractive Opinion Summarization



Extractive Opinion Summarization



Extractive Opinion Summarization



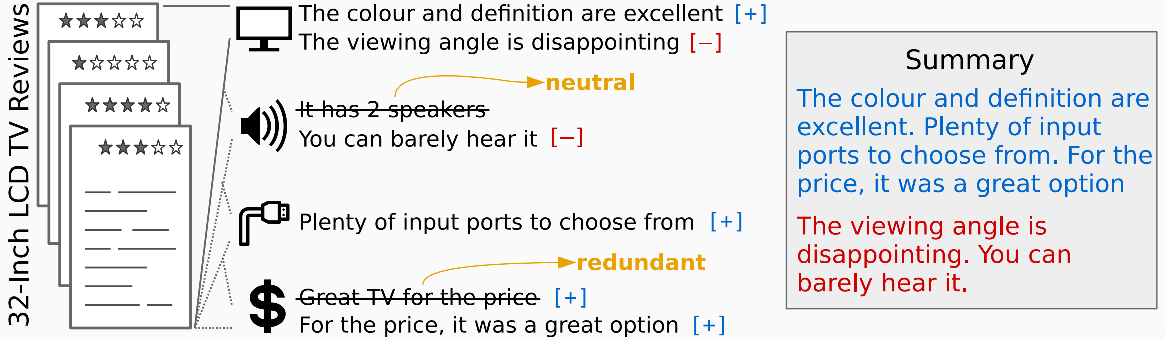
The colour and definition are excellent [+]
The viewing angle is disappointing [–]

It has 2 speakers
You can barely hear it [–]

Plenty of input ports to choose from [+]

Great TV for the price [+]
For the price, it was a great option [+]

Extractive Opinion Summarization



Approach: Neural-based, Weakly Supervised

This Work

Task decomposition

- Aspect Detection
- Fine-grained Sentiment Prediction (Angelidis and Lapata, TACL 2018)
- Summary Construction

No direct supervision for aspects or sentiment
No gold-standard summaries for training

Newly-constructed OpoSum corpus

Aspect Detection

Predict the aspect discussed in each review segment

Television Review	
Overall a good TV!	***General***
Such great picture quality.	**Image**
The colors are perfectly crisp.	**Image**
The sound is the only issue.	**Sound**
If you like good bass,	**Sound**
you must connect it to a Hi-Fi.	**Sound+Connect.**
A good value tv too.	**Price**

Got enough annotated data? Just do classification!

What if we don't?

Aspect Discovery vs. Extraction

Discovery

- No prior knowledge about aspects
- Use topic modeling
- Post-hoc mapping from topics to aspects
- Neural method: Aspect-Based Autoencoder (ABAE; He et al., 2017)

Extraction

- Inject prior knowledge via **seed words**
- No ability to discover new aspects
- Model guided towards aspects of interest
- Neural method: Multi-Seed Aspect Extractor (MATE; this work)

Previous Work: Aspect-Based Autoencoder (ABAE)

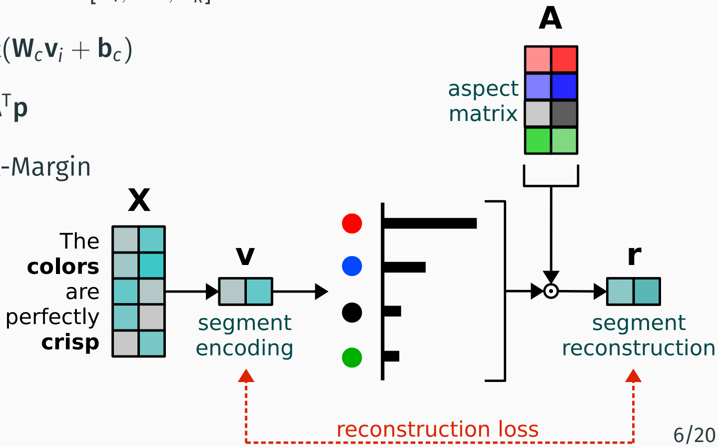
Segment encoding: $\mathbf{v} = \sum_i c_i \mathbf{x}_i$

Aspect dictionary: Aspect Matrix $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_k]$

Aspect prediction: $\mathbf{p} = \mathrm{softmax}(\mathbf{W}_c \mathbf{v}_i + \mathbf{b}_c)$

Segment reconstruction: $\mathbf{r} = \mathbf{A}^\top \mathbf{p}$

Objective: Reconstruction Max-Margin



Previous Work: Aspect-Based Autoencoder (ABAE)

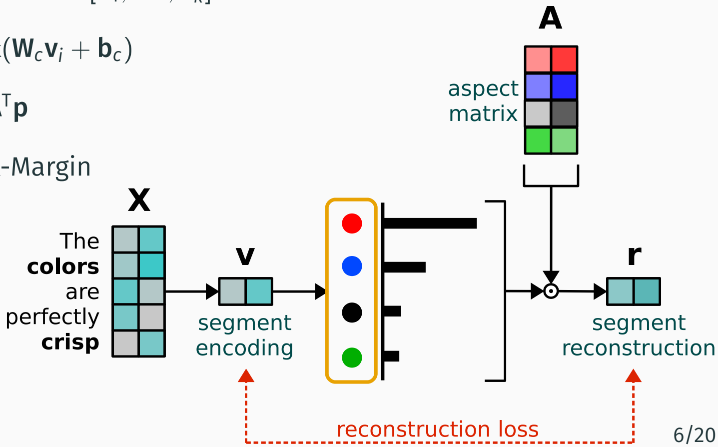
Segment encoding: $\mathbf{v} = \sum_i c_i \mathbf{x}_i$

Aspect dictionary: Aspect Matrix $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_k]$

Aspect prediction: $\mathbf{p} = \text{softmax}(\mathbf{W}_c \mathbf{v}_i + \mathbf{b}_c)$

Segment reconstruction: $\mathbf{r} = \mathbf{A}^\top \mathbf{p}$

Objective: Reconstruction Max-Margin



Multi-seed Aspect Extractor (MATE)

Segment encoding: $\mathbf{v} = \sum_i c_i \mathbf{x}_i$

Aspect dictionary:

Seed Matrices $\{\mathbf{A}_1, \ldots \mathbf{A}_k\}$

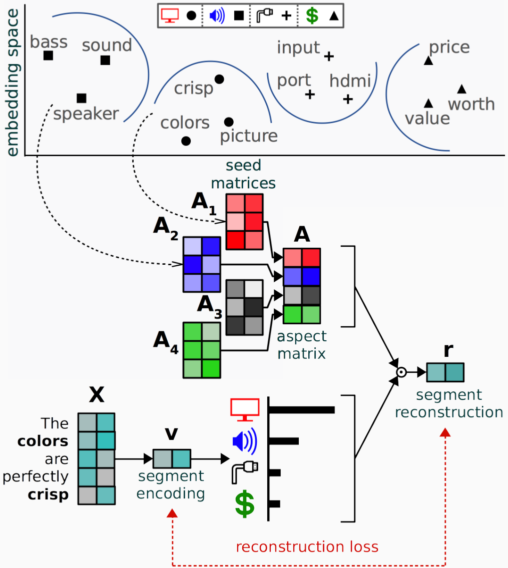
Aspect Vector $\mathbf{a}_i = \mathbf{A}_i^\top \mathbf{z}_i$

Aspect Matrix $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_k]$

Aspect prediction: $\mathbf{p} = \mathrm{softmax}(\mathbf{W}_c \mathbf{v} + \mathbf{b}_c)$

Segment reconstruction: $\mathbf{r} = \mathbf{A}^\top \mathbf{p}$

Objective: Reconstruction Max-Margin



Multi-seed Aspect Extractor (MATE)

Segment encoding: $\mathbf{v} = \sum_i c_i \mathbf{x}_i$

Aspect dictionary:

Seed Matrices $\{\mathbf{A}_1, \ldots \mathbf{A}_k\}$

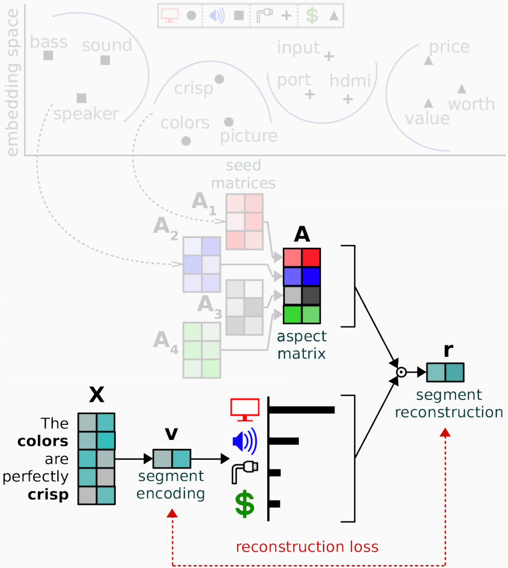
Aspect Vector $\mathbf{a}_i = \mathbf{A}_i^\intercal \mathbf{z}_i$

Aspect Matrix $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_k]$

Aspect prediction: $\mathbf{p} = \text{softmax}(\mathbf{W}_c \mathbf{v} + \mathbf{b}_c)$

Segment reconstruction: $\mathbf{r} = \mathbf{A}^\intercal \mathbf{p}$

Objective: Reconstruction Max-Margin



Multi-seed Aspect Extractor (MATE)

Segment encoding: $\mathbf{v} = \sum_i c_i \mathbf{x}_i$

Aspect dictionary:

Seed Matrices $\{\mathbf{A}_1, \ldots \mathbf{A}_k\}$

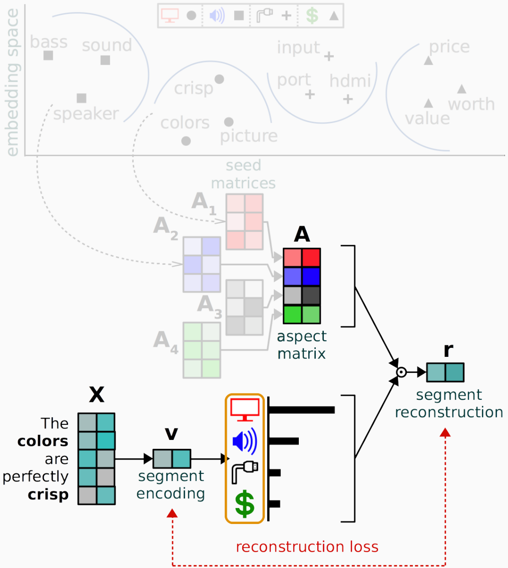
Aspect Vector $\mathbf{a}_i = \mathbf{A}_i^\top \mathbf{z}_i$

Aspect Matrix $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_k]$

Aspect prediction: $\mathbf{p} = \text{softmax}(\mathbf{W}_c \mathbf{v} + \mathbf{b}_c)$

Segment reconstruction: $\mathbf{r} = \mathbf{A}^\top \mathbf{p}$

Objective: Reconstruction Max-Margin



Multi-seed Aspect Extractor (MATE)

Segment encoding: $\mathbf{v} = \sum_i c_i \mathbf{x}_i$

Aspect dictionary:

Seed Matrices $\{\mathbf{A}_1, \ldots \mathbf{A}_k\}$

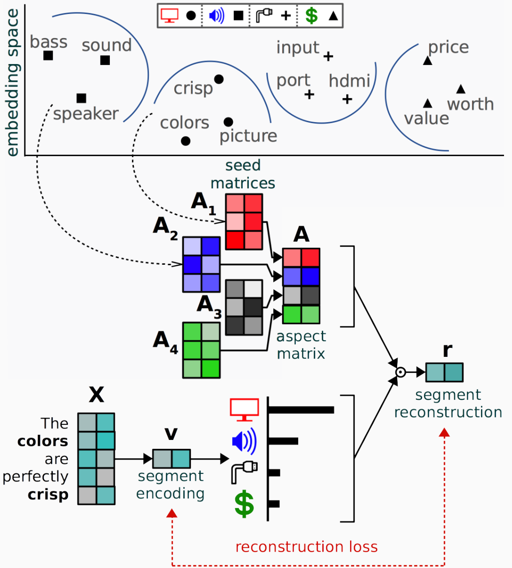
Aspect Vector $\mathbf{a}_i = \mathbf{A}_i^\top \mathbf{z}_i$

Aspect Matrix $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_k]$

Aspect prediction: $\mathbf{p} = \text{softmax}(\mathbf{W}_c \mathbf{v} + \mathbf{b}_c)$

Segment reconstruction: $\mathbf{r} = \mathbf{A}^\top \mathbf{p}$

Objective: Reconstruction Max-Margin



Multi-task Learning (MATE+MT)

Goal: Help encoder focus on aspect words

Idea: Aspect words are domain-specific!

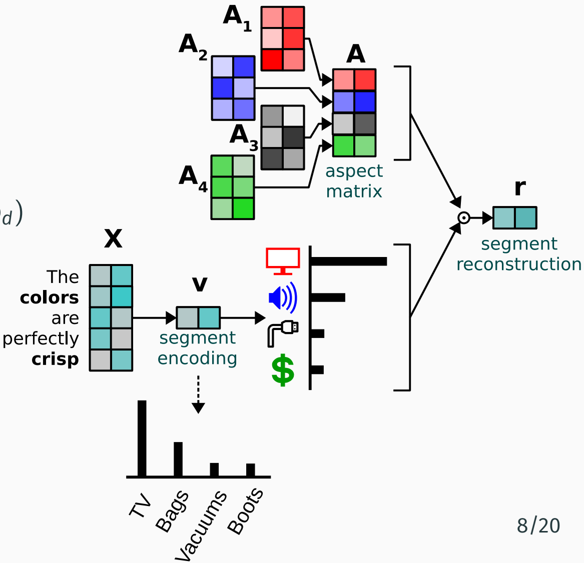
"The colors are perfectly crisp"

"The keys feel great to type on"

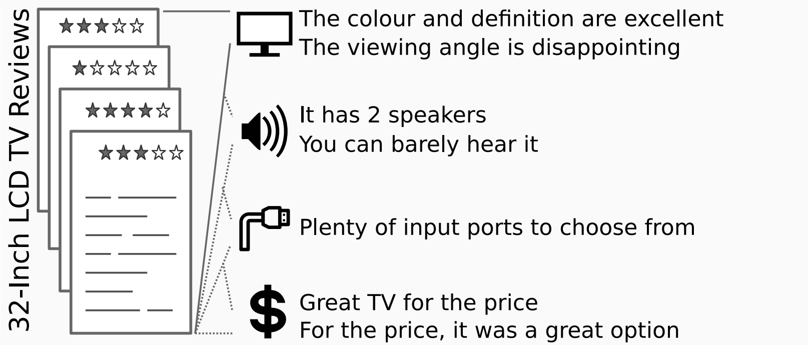
Domain prediction: $\mathbf{p}_{dom} = \text{softmax}(\mathbf{W}_d\mathbf{v} + \mathbf{b}_d)$

Combined objective:

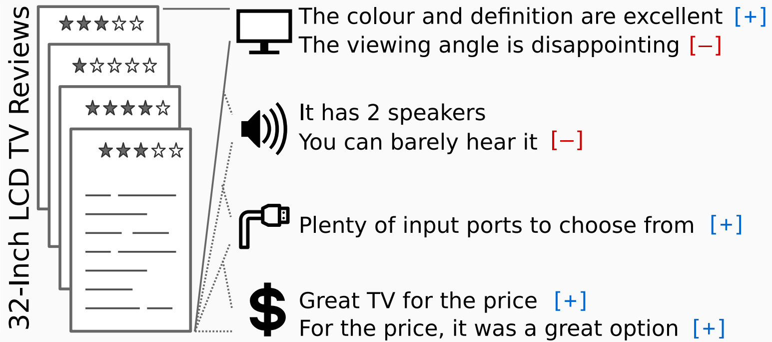
- Reconstruction Max-Margin
- NLL of domain prediction



Extractive Opinion Summarization



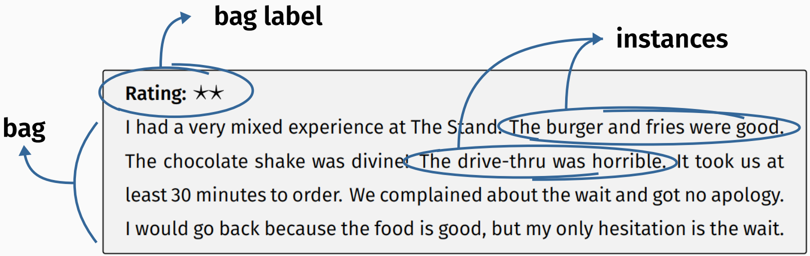
Extractive Opinion Summarization



Fine-grained Sentiment Prediction

Multiple Instance Learning Network (MILNET)

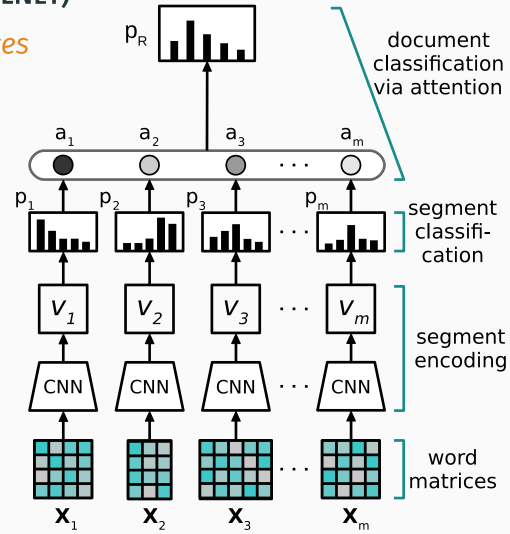
- Training examples → *bags* of *instances*
- Bag labels → supervision
- Instance labels → latent



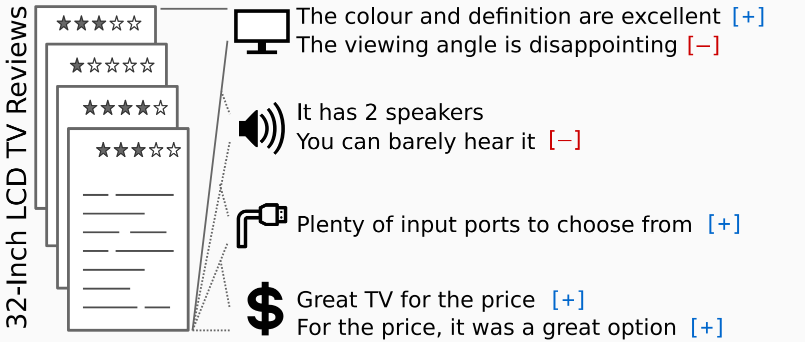
Multiple Instance Learning Network (MILNET)

- Training examples → *bags* of *instances*
- Bag labels → supervision
- Instance labels → latent

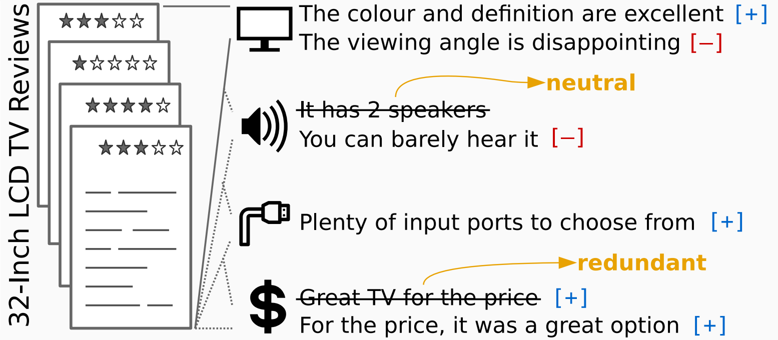
- Model output: $pol_s \in [-1, +1]$



Extractive Opinion Summarization



Extractive Opinion Summarization



Summary

The colour and definition are excellent. Plenty of input ports to choose from. For the price, it was a great option

The viewing angle is disappointing. You can barely hear it.

Opinion Salience

Two sources of information:

- Segment aspect probabilities (via MATE)
- Segment polarities (via MILNET)
- No further training required

What makes a useful opinion?

- Needs to be non-neutral:

  “It has 2 speakers” ✗

  “You can barely hear it” ✓

- Needs to be non-general:

  “Loved it!” ✗

  “Many input ports to choose from” ✓

Opinion Salience

Two sources of information:

- Segment aspect probabilities (via MATE)
- Segment polarities (via MILNET)
- No further training required

What makes a useful opinion? $\longrightarrow$ $|pol_s| \cdot (\max_i p_s^{(a_i)} - p_s^{(\mathrm{GEN})})$

- Needs to be non-neutral:
  "It has 2 speakers" ✗
  "You can barely hear it" ✓
- Needs to be non-general:
  "Loved it!" ✗
  "Many input ports to choose from" ✓

From Rankings to Summaries

Need to avoid redundancy

Greedy opinion selection: (Cao et al., 2015; Yasunaga et al., 2017)

- Start by adding top-ranked opinion to summary
- Keep adding unless:
    a. Cosine similarity with any previous opinion is over 0.5
    b. Summary budget reached (100 words)

The OpoSum Corpus – Training Data

What we need

- Diverse domains
- Thousands of reviews
- No fine-grained labels

Training Data						
Source	Amazon					
Domain	Lapt. Bags	Bluetooth	Boots	Keyboards	Televisions	Vacuums
Products	2040	1471	4723	983	1894	1184
Reviews	43K	80K	78K	34K	57K	68K
Words/Review	98.1	122.5	82.6	127.0	180.4	146.6

The OpoSum Corpus – Evaluation Data

What we need

- Aspect labels
- Opinion salience labels
- Final summaries

The OpoSum Corpus – Evaluation Data

What we need

- Aspect labels
- Opinion salience labels
- Final summaries

Television Review	
Overall a good TV!	***General***
Such great picture quality.	**Image**
The colors are perfectly crisp.	**Image**
The sound is the only issue.	**Sound**
If you like good bass,	**Sound**
you must connect it to a Hi-Fi.	**Sound+Connect.**
A good value tv too.	**Price**

The OpoSum Corpus – Evaluation Data

What we need

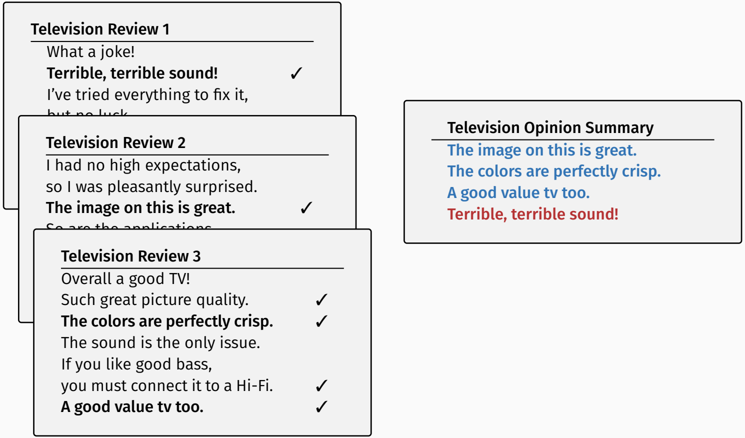
- Aspect labels
- Opinion salience labels
- Final summaries

Television Review	
Overall a good TV!	
Such great picture quality.	✓
The colors are perfectly crisp.	✓
The sound is the only issue.	
If you like good bass,	
you must connect it to a Hi-Fi.	✓
A good value tv too.	✓

The OpoSum Corpus – Evaluation Data

What we need

- Aspect labels
- Opinion salience labels
- Final summaries



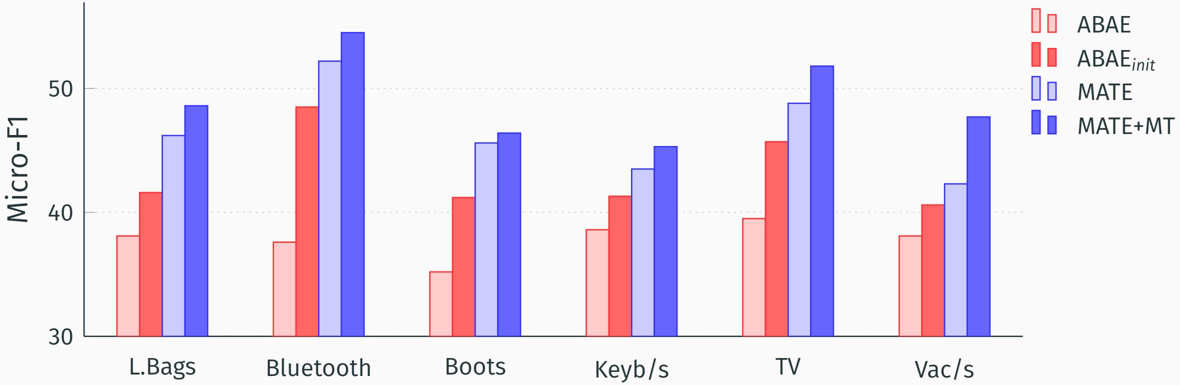
The OpoSum Corpus – Evaluation Data

What we need

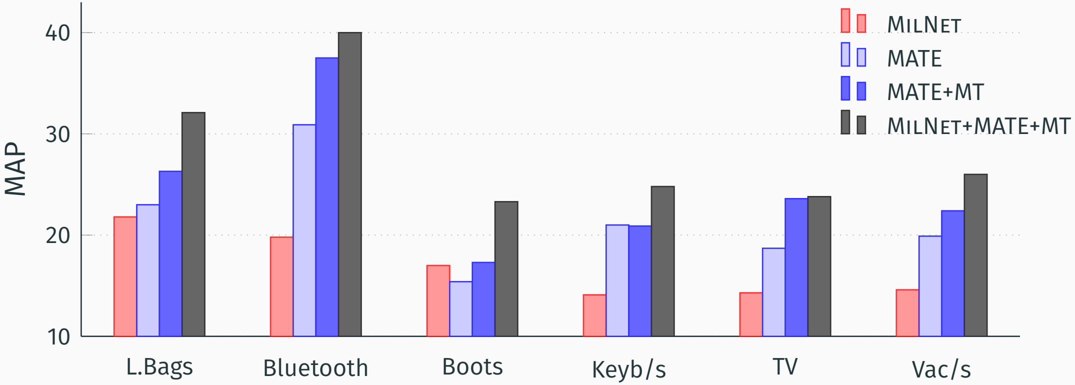
- Aspect labels
- Opinion salience labels
- Final summaries

Evaluation Data						
Domain	Lapt. Bags	Bluetooth	Boots	Keyboards	Televisions	Vacuums
Products	10	10	10	10	10	10
Reviews	100	100	100	100	100	100
Segments	1262	1344	1198	1396	1483	1492

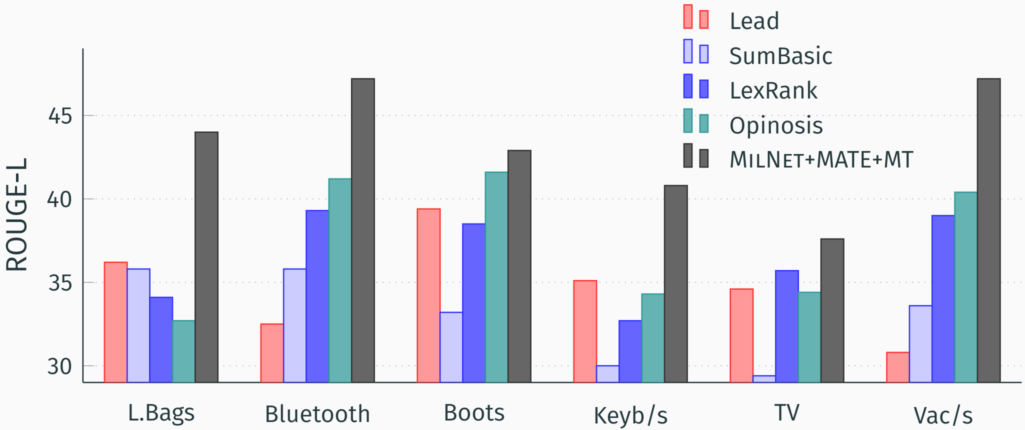
Results – Aspect Extraction



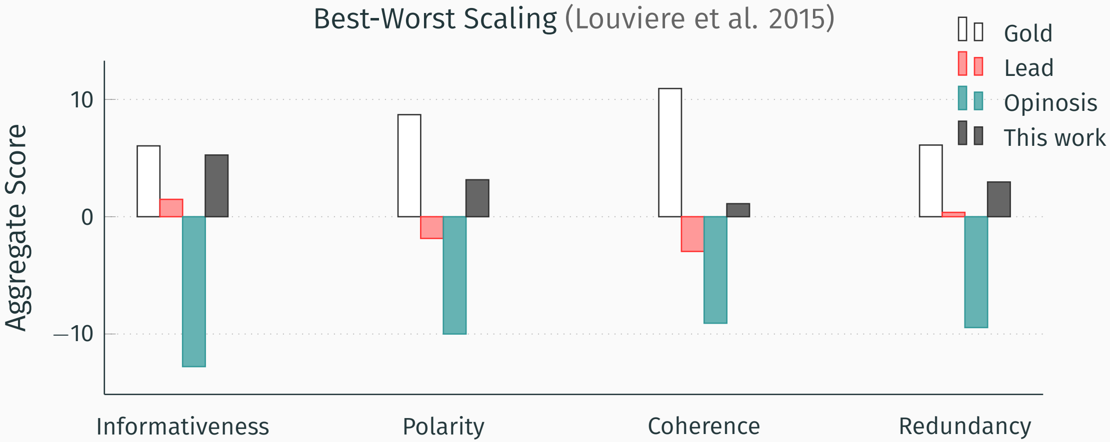
Results – Salient Opinion Retrieval



Results – Opinion Summarization



Results – Human Evaluation



Summary

Opinion summarization without direct supervision

Combines weakly supervised sentiment and aspect detectors

Multi-tasking improves aspect extraction

Automatic and human evaluation show promising results

...and some of our summaries:

Boots	The attention to detail is great! I love the colour plum. Sooooooooo narrow! They have stretched a little, but maybe get a size or half size up. They do not fit well, horribly uncomfortable.
TV	The sound is good and strong. The picture is beautiful. And the price is even better. Unbelieveable picture and the setup is so easy. The Yahoo widgets don't work. And avoid the Sony apps at ALL costs. Communication a bit difficult. :(

Thank you!

Data and Code: `stangelid.github.io`