Fine-grained Opinion Mining: Current Trend and Cutting-Edge Dimensions

Wenya Wang\textsuperscript{1}, Jianfei Yu\textsuperscript{2}, Sinno Jialin Pan\textsuperscript{1} and Jing Jiang\textsuperscript{2}

\textsuperscript{1}Nanyang Technological University, Singapore
\textsuperscript{2}Singapore Management University, Singapore
# TABLE OF CONTENTS

- Introduction
- Aspect-Based Extraction
- Target-Oriented Sentiment Classification
- Opinion Summarization
- Cutting-Edge Dimensions
Introduction
What is Fine-Grained Opinion Mining

8.7 Fabulous · 11,276 reviews

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Staff</td>
<td>8.5</td>
</tr>
<tr>
<td>Facilities</td>
<td>8.7</td>
</tr>
<tr>
<td>Cleanliness</td>
<td>8.9</td>
</tr>
<tr>
<td>Comfort</td>
<td>8.9</td>
</tr>
<tr>
<td>Value for money</td>
<td>8.1</td>
</tr>
<tr>
<td>Location</td>
<td>8.8</td>
</tr>
<tr>
<td>Free WiFi</td>
<td>8.0</td>
</tr>
</tbody>
</table>

Antoine

France

Reviewers' choice · Reviewed: 23 October 2018

Luxury, relaxation, great service and food.

Amazing service and luxury. I almost did not leave my suite as it was so comfortable. The dining service was also outstanding. I'm looking forward to going back there.
Subtasks

- **Extraction**: aspect terms (opinion targets), opinion expressions, aspect categories, opinion holders, opinion relations

- **Sentiment prediction**: sentiment scores (polarities) towards aspect terms or aspect categories

- **Summarization**: multi/single-document, aspect/product-centered, phrase/sentence-based

- Transfer learning, multi-task, multimodal learning
Challenges

- **Uncertainty**: low frequencies, same entity with different expressions
  - “UI” vs “user interface”, “macbook pro” vs “mac pro”

- **Variability**: multiple targets, contrastive views for certain aspects
  - “The laptop has very **small size** which is **convenient for mobility**, but **uneasy for reading**.”

- **Flexibility**: target entities are not restricted to specific POS tags, dual-functioning word acting as both aspects and opinions
  - “I **recommend** this restaurant to everyone.”
  - “The laptop is **lightweight**, and its **ease of use** attracts me.”

- **Scarcity**: Limited annotated resources
Objectives

- An overview of existing methods, traditional machine learning or deep learning, for fine-grained opinion mining.
- Categorize existing approaches based on relationship manipulation.
- Present both advantages and limitations.
- Emphasize the correlations across different subtasks.
- Pose future directions with potential research values.
Aspect-Based Extraction
OUTLINE

● Background

● Methodology
  ○ Unsupervised/Semi-supervised Learning
    ■ Pattern mining
    ■ Topic modeling
    ■ Deep learning
  ○ Supervised Learning
    ■ Feature engineering
    ■ Deep learning with syntactic information
    ■ Deep learning without external knowledge

● Summary
OUTLINE

● **Background**

● **Methodology**
  ○ **Unsupervised/Semi-supervised Learning**
    ■ Pattern mining
    ■ Topic modeling
    ■ Deep learning
  ○ **Supervised Learning**
    ■ Feature engineering
    ■ Deep learning with syntactic information
    ■ Deep learning without external knowledge

● **Summary**
What Can be Extracted

- Aspect-oriented
  - Corpus-based
  - Sentence-based

John thinks the hotel service is terrible.

Aspect category: SERVICE

Opinion pair

Opinion holder

Aspect term (opinion target)

Opinion expression
What Can be Extracted

- Aspect-oriented
  - Corpus-based
  - Sentence-based

Aspect category: SERVICE

John thinks the hotel service is terrible.

Aspect (Category) Identification

battery life, sound quality, Ease of use, ...
What Can be Extracted

- Aspect-oriented
  - Corpus-based
  - Sentence-based

Aspect category: SERVICE

John thinks the hotel service is terrible.

Opinion holder: John
Aspect term (opinion target): hotel service
Opinion expression: terrible

Opinion pair

Aspect (Category) Identification
- battery life, sound quality, Ease of use, ...

Opinion Identification and Sentiment Prediction
- great (+), long (+), convenient (+), ...
What Can be Extracted

- Aspect-oriented
  - Corpus-based
  - Sentence-based

Aspect category: SERVICE

John thinks the hotel service is terrible.

Opinion holder: John
Aspect term (opinion target): hotel service
Opinion expression: terrible

Opinion pair

Aspect (Category) Identification
- battery life, sound quality, Ease of use, ...

Opinion Identification and Sentiment Prediction
- great (+), long (+)
  - convenient (+), ...

Summary Presentation
- battery life: ⭐⭐⭐⭐
  - sound quality: ⭐⭐⭐⭐⭐⭐
What Can be Extracted

- Opinion-oriented
  - Opinion-related entities: opinion expressions (O), opinion targets (T), opinion holders (H)
    - Direct subjective expressions (DSEs): explicit mentions of private states or speech events expressing private states.
    - Expressive subjective expressions (ESEs): expressions that indicate sentiment, emotion, etc. without explicitly conveying them.
  - Opinion relations: IS-ABOUT, IS-FROM

(1) The International Committee of the Red Cross, [as usual][ESE], [has refused to make any statements][DSE].

S1: [The workers][H1,2] were irked[O1] by [the government report][T1] and were worried[O2] as they went about their daily chores.
OUTLINE

● Background

● Methodology
  ○ Unsupervised/Semi-supervised Learning
    ■ Pattern mining
    ■ Topic modeling
    ■ Deep learning
  ○ Supervised Learning
    ■ Feature engineering
    ■ Deep learning with syntactic information
    ■ Deep learning without external knowledge

● Summary
Pattern Mining

- Frequency
- Association
  - Statistical tests
  - Alignment model
- Syntax
  - Word-based
  - Phrase-based
Frequency & Association Mining

- POS: select nouns/noun phrases
- Mining: find all frequent itemsets (>=1% support)
- Pruning: prune multi-word itemsets that are meaningless, single-word itemsets that are redundant
- Opinion word: adjacent adjectives around frequent features
- Infrequent feature: nearest noun/noun phrase around opinion expressions

Hu and Liu. Mining Opinion Features in customer reviews. AAAI 2004
Frequency & Association Mining

- Start with a small set of feature seeds
- Iteratively enlarges by mining associations (likelihood ratio tests, latent semantic analysis):
  - Feature-opinion
  - Feature-feature
  - opinion-opinion

<table>
<thead>
<tr>
<th>Reviews:</th>
<th>CF={screen, price, student}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The screen is really big, but the price is too expensive!</td>
<td></td>
</tr>
<tr>
<td>2. The price is expensive, students don’t buy it usually.</td>
<td></td>
</tr>
<tr>
<td>3. The screen is beautiful, but the price is not!</td>
<td></td>
</tr>
<tr>
<td>4. The screen is big and beautiful!</td>
<td></td>
</tr>
<tr>
<td>CO={big, expensive, buy, beautiful}</td>
<td></td>
</tr>
<tr>
<td>S={screen}</td>
<td></td>
</tr>
<tr>
<td>thd=2.0</td>
<td></td>
</tr>
<tr>
<td>F={screen, price}</td>
<td></td>
</tr>
<tr>
<td>O={big, beautiful, expensive}</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>screen</th>
<th>price</th>
<th>student</th>
<th>big</th>
<th>expensive</th>
<th>buy</th>
<th>beautiful</th>
</tr>
</thead>
<tbody>
<tr>
<td>screen</td>
<td>2.5</td>
<td>0.5</td>
<td></td>
<td>3.0</td>
<td>1.5</td>
<td>0.5</td>
<td>3.0</td>
</tr>
<tr>
<td>price</td>
<td>2.5</td>
<td>1.0</td>
<td></td>
<td>1.5</td>
<td>3.0</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>student</td>
<td>0.5</td>
<td>1.0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>1.0</td>
<td>0.5</td>
</tr>
<tr>
<td>big</td>
<td>3.0</td>
<td>1.5</td>
<td>0.5</td>
<td>2.0</td>
<td>0.5</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>expensive</td>
<td>1.5</td>
<td>3.0</td>
<td>0.5</td>
<td>2.0</td>
<td>1.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>buy</td>
<td>0.5</td>
<td>1.5</td>
<td>1.0</td>
<td>0.5</td>
<td>1.0</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>beautiful</td>
<td>3.0</td>
<td>1.5</td>
<td>0.5</td>
<td>2.0</td>
<td>2.0</td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>
Extraction via Word Alignment

- Mine associations between targets and opinions via word translation model: capture long-span relations

\[ \hat{A} = \arg \max_A P(A \mid S) \]

\[ S = \{w_1, w_2, \ldots, w_n\} \]

\[ A = \{(i, a_i) \mid i \in [1, n]\} \]

The phone has a **colorful** and even **amazing** screen

Translation

The phone has a **colorful** and even **amazing** screen

---

Liu et al. Opinion Target Extraction Using Word-Based Translation Model. EMNLP 2012
Extraction via Word Alignment

- Mine associations between targets and opinions via word translation model: capture long-span relations

\[
\hat{A} = \arg \max_A P(A | S)
\]

\[
S = \{w_1, w_2, ..., w_n\}
\]

\[
A = \{(i, a_i) | i \in [1, n]\}
\]

- Graph-based algorithm to extract opinion targets

\[
C^{t+1} = (1 - \lambda) \times M^T \times M \times C^t + \lambda \times S
\]
Word-Based Syntactic Rule Mining

- **Double propagation**: there are dependency relations between opinion words and aspect words → iteratively expand the opinion and target lexicons

<table>
<thead>
<tr>
<th>RuleID</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{11}$</td>
<td>$O \rightarrow O-\text{Dep} \rightarrow T$ s.t. $O \in {O}, O-\text{Dep} \in {MR}$, $\text{POS}(T) \in {NN}$</td>
</tr>
</tbody>
</table>

Qiu et al. Opinion word expansion and target extraction through double propagation. Comput Linguist 2011

Somasundaran and Wiebe. Recognizing Stances in Online Debates. ACL and AFNLP 2009

Popescu and Etzioni. Extracting Product Features and Opinions from Reviews. EMNLP 2005

Zhuang et al. Movie review mining and summarization. CIKM 2006
Phrase-Based Syntactic Rule Mining

Treat a phrase as a single unit

Wu et al. Phrase Dependency Parsing for Opinion Mining. EMNLP 2009
Lifelong Learning

- Borrow the idea of recommendation to extract aspects based on the information in reviews of a large number of other products
  - Similarity-based recommendation
  - Association-based recommendation

---

**Algorithm 1 AER(D_t, R^-, R^+, O)**

**Input:** Target dataset $D_t$, high precision aspect extraction rules $R^-$, high recall aspect extraction rules $R^+$, seed opinion words $O$

**Output:** Extracted aspect set $A$

1. $T^- \leftarrow $ DPextract($D^t$, $R^-$, $O$);
2. $T^+ \leftarrow $ DPextract($D^t$, $R^+$, $O$);
3. $T \leftarrow T^+ \setminus T^-;
4. T^s \leftarrow $ Sim-recom($T^-$, $T$);
5. $T^a \leftarrow $ AR-recom($T^-$, $T$);
6. $A \leftarrow T^- \cup T^s \cup T^a$.  

---

**Base extraction through DP**

**Recommendation**
Limitation

- Mostly only restricts aspect terms to be noun/noun phrases, opinions to be adjectives
- Rules/Patterns are inflexible
- Easy to produce meaningless features
OUTLINE

● Background

● Methodology
  ○ Unsupervised/Semi-supervised Learning
    ■ Pattern mining
    ■ Topic modeling
    ■ Deep learning
  ○ Supervised Learning
    ■ Feature engineering
    ■ Deep learning with syntactic information
    ■ Deep learning without external knowledge

● Summary
Topic Modeling

- Treat aspect categories as clustered topics
- Beneficial when aspects are implicit

- Knowledge-Free
  - Topic without association
  - Topic with association
  - Context
  - Sentiment
  - Hierarchy

- Knowledge-Infused
  - Prior
  - Labeled data
LDA With Local/Global Context

- Normal LDA: tend to produce global topics (product brand)

I. Choose distribution of topics $\theta_d \sim \text{Dir}(\alpha)$

II. For each word $i$
   A. Choose topic $z_{d,i} \sim \theta_d$
   B. Choose word $w_{d,i} \sim \varphi_{z_{d,i}}$
LDA With Local/Global Context

- Multi-grain LDA: models two distinct types of topics: global topics (properties of reviews) and local topics (ratable aspects)

I. Choose global topic $\theta_{d}^{gl} \sim Dir(\alpha^{gl})$

II. For each sentence $s$, choose $\psi_{d,s}(v) \sim Dir(\gamma)$

III. For each sliding window $v$
   A. Choose $\theta_{d,v}(loc) \sim Dir(\alpha^{loc})$
   B. Choose $\pi_{d,v} \sim Beta(\alpha^{mix})$

IV. For each word $i$
   A. Choose window $\nu_{d,i} \sim \psi_{d,s}$
   B. Choose $\tau_{d,i} \sim \pi_{d,v_{d,i}}$
   C. If global, choose $z_{d,i} \sim \theta_{d}^{gl}$
   D. If local, choose $z_{d,i} \sim \theta_{d,v_{d,i}}^{loc}$
   E. Choose word $w$ from $\phi_{z_{d,i}}^{r_{d,i}}$

---

Brody and Elhadad. An Unsupervised Aspect-Sentiment Model for Online Reviews. NAACL 2010
Titov and MacDonald. Modeling Online Reviews with Multi-grain Topic Models. WWW 2008
## LDA With Local/Global Context

<table>
<thead>
<tr>
<th>label</th>
<th>top words</th>
</tr>
</thead>
<tbody>
<tr>
<td>MG-LDA local (all topics)</td>
<td>sound quality headphones volume bass earphones good settings ear rock excellent games features clock contacts calendar alarm notes game quiz feature extras solitaire usb pc windows port transfer computer mac software cable xp connection plug firewire reset noise backlight slow freeze turn remove playing icon creates hot cause disconnect case pocket silver screen plastic clip easily small blue black light white belt cover button play track menu song buttons volume album tracks artist screen press select battery hours life batteries charge aaa rechargeable time power lasts hour charged usb cable headphones adapter remote plug power charger included case firewire files software music computer transfer windows media cd pc drag drop file using radio fm voice recording record recorder audio mp3 microphone wma formats</td>
</tr>
<tr>
<td>MG-LDA global</td>
<td>ipod music apple songs use mini very just itunes like easy great time new buy really zen creative micro touch xtra pad nomad waiting deleted labs nx sensitive 5gb eax sony walkman memory stick sonicstage players atrac3 mb atrac far software format video screen videos device photos tv archos pictures camera movies dvd files view player product did just bought unit got buy work $ problem support time months</td>
</tr>
<tr>
<td>LDA (out of 40)</td>
<td>ipod music songs itunes mini apple battery use very computer easy time just song creative nomad zen xtra jukebox eax labs concert effects nx 60gb experience lyrics card memory cards sd flash batteries lyra battery aa slot compact extra mmc 32mb radio fm recording record device audio voice unit battery features usb recorder button menu track play volume buttons player song tracks press mode screen settings points reviews review negative bad general none comments good please content aware player very use mp3 good sound battery great easy songs quality like just music</td>
</tr>
</tbody>
</table>
LDA with Sentiment

- Decide if the word is a common English word
- If not, decide the subtopics
- Decide if the word is neutral, positive or negative
- Generate the word

\[
\log(C) = \sum_{d \in C} \sum_{w \in V} c(w : d) \log [\lambda_B p(w | B)] \\
+ (1 - \lambda_B) \sum_{j=1}^{k} \pi_{d_j} \times (\delta_{j,d,F} p(w | \theta_j) \\
+ \delta_{j,d,P} p(w | \theta_P) + \delta_{j,d,N} p(w | \theta_N))
\]

Mei et al. Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs. WWW 2007
LDA With Hierarchy

- Aspects usually form hierarchies: Aspect-Sentiment tree

Kim et al. A Hierarchical Aspect-Sentiment Model for Online Reviews. AAAI 2013
LDA With Hierarchy

- Likelihood generation
  - Draw aspect-sentiment node \( c \sim T \)
  - Draw sentiment \( s \sim \text{Multinomial}(\pi) \)
  - Draw subjectivity \( \theta \sim \text{Beta}(\alpha) \)
  - For each word: draw subjectivity and word \( p \sim \text{Binomial}(1, \theta) \)
    \( w \sim \text{Multinomial}(\phi^s_{c \times p}) \)

Kim et al. A Hierarchical Aspect-Sentiment Model for Online Reviews. AAAI 2013
LDA With Hierarchy

Kim et al. A Hierarchical Aspect-Sentiment Model for Online Reviews. AAAI 2013
LDA with Prior Knowledge

- **Objective:** Generating coherent aspects
- **Incorporate prior knowledge in LDA**
  - **must-link:** 2 noun phrases that shared one or more words are likely to fall into the same topic
  - **cannot-link:** people normally will not repeat the same feature in the same sentence

\[
\begin{align*}
\theta & \sim \text{Dirichlet}(\alpha) \\
z_i | \theta_m & \sim \text{Multinomial}(\theta_m) \\
\varphi & \sim \text{Dirichlet}(\beta) \\
s_i | z_i, \varphi & \sim \text{Multinomial}(\varphi_{z_i}) \\
\eta & \sim \text{Dirichlet}(\gamma) \\
w_i | z_i, s_i, \eta & \sim \text{Multinomial}(\eta_{z_i,s_i})
\end{align*}
\]

S-set representing the must-links among words
{battery, life} {battery, long}

Chen et al. Leveraging Multi-Domain Prior Knowledge in Topic Models. IJCAI 2013
Chen et al. Exploiting Domain Knowledge in Aspect Extraction. EMNLP 2013
Chen and Liu Aspect Extraction with Automated Prior Knowledge Learning. ACL 2014
LDA with Some Labeled Data

- Given some annotated corpus, construct several multinomial word distributions to differentiate aspects from opinions

\[
w_{d,s,n} \sim \begin{cases} 
\text{Multi}(\phi^B) & \text{if } y_{d,s,n} = 0 \\
\text{Multi}(\phi^A_{z_d,s}) & \text{if } y_{d,s,n} = 1, u_{d,s,n} = 0 \\
\text{Multi}(\phi^A_{g}) & \text{if } y_{d,s,n} = 1, u_{d,s,n} = 1 \\
\text{Multi}(\phi^O_{z_d,s}) & \text{if } y_{d,s,n} = 2, u_{d,s,n} = 0 \\
\text{Multi}(\phi^O_{g}) & \text{if } y_{d,s,n} = 2, u_{d,s,n} = 1 
\end{cases}
\]

- Use information (POS tags) to discriminate between aspect/opinion words with maximum entropy (MaxEnt) model

\[
p(y_{d,s,n} = l | x_{d,s,n}) = \pi_{l}^{d,s,n} = \frac{\exp \left( \lambda_l \cdot x_{d,s,n} \right)}{\sum_{l'=0}^{2} \exp \left( \lambda_{l'} \cdot x_{d,s,n} \right)}
\]

Zhao et al. Jointly Modeling Aspects and Opinions with a MaxEnt-LDA Hybrid. EMNLP 2010
Limitation

- The objective function of topic models does not always correlate well with human judgments
- Hard to extract low-frequency aspects
- Hard to deal with multi-word aspect phrases
- Hard to differentiate and associate between aspect and opinion expressions
OUTLINE

● Background

● Methodology
  ○ **Unsupervised/Semi-supervised Learning**
    ■ Pattern mining
    ■ Topic modeling
    ■ **Deep learning**
  ○ **Supervised Learning**
    ■ Feature engineering
    ■ Deep learning with syntactic information
    ■ Deep learning without external knowledge

● Summary
Aspect Extraction with RBM

- Reflects the generation process of reviews by introducing a heterogeneous structure into the hidden layer and incorporating informative priors.

Wang et al. Sentiment-Aspect Extraction based on Restricted Boltzmann Machines. ACL-IJCNLP 2015
Aspect Extraction with RBM

- Reflects the generation process of reviews by introducing a heterogeneous structure into the hidden layer and incorporating informative priors.

- Construct optimal weight matrix

\[
E(v, h) = -\sum_{j=1}^{F} \sum_{k=1}^{K} W_j^k h_j \hat{v}^k - \sum_{k=1}^{K} \hat{v}^k \hat{v}^k - \sum_{j=1}^{F} h_j a_j,
\]

\[
P(h_j = 1 | \hat{v}^k) = P(h_j = 1 | h_{-j}, \hat{v}^k) = \sigma(a_j + W_j^k \hat{v}^k).
\]

- Priors as regularizers

Wang et al. Sentiment-Aspect Extraction based on Restricted Boltzmann Machines. ACL-IJCNLP 2015
Aspect Extraction with Attention

- Improves coherence by exploiting the distribution of word co-occurrences through the use of neural word embeddings
- Use attention to de-emphasize irrelevant words during training

\[ a_i = \frac{\exp(d_i)}{\sum_{j=1}^{n} \exp(d_j)} \]

\[ d_i = e_{w_i}^\top \cdot M \cdot y_s \]

\[ y_s = \frac{1}{n} \sum_{i=1}^{n} e_{w_i} \]

- Sentence reconstruction to enhance coherence

\[ J(\theta) = \sum_{s \in D} \sum_{i=1}^{m} \max(0, 1 - r_s z_s + r_s n_i) \]

He et al. An Unsupervised Neural Attention Model for Aspect Extraction. ACL 2017
Aspect Extraction with Attention

- Multi-seed aspect extractor: every aspect is represented as a matrix consisting of seed embeddings
- Multi-task objective: aspect-relevant words are good indicators of the product’s domain

\[ p_{s}^{\text{dom}} = \text{softmax}(W_{C}v_{s} + b_{C}) \]

- Objective

\[ J_{\text{MT}}(\theta) = J_{r}(\theta) - \lambda \sum_{s \in C_{\text{all}}} \log p(d_{s}) \]

Angelidis and Lapata. Summarizing Opinions: Aspect Extraction Meets Sentiment Prediction and They Are Both Weakly Supervised. EMNLP 2018
OUTLINE

- Background
- **Methodology**
  - Unsupervised/Semi-supervised Learning
    - Pattern mining
    - Topic modeling
    - Deep learning
  - **Supervised Learning**
    - Feature engineering
    - Deep learning with syntactic information
    - Deep learning without external knowledge
- Summary
Feature Engineering for Sequence Labeling

- In sequence labeling, a word can take different roles
  - The word is a beginning component of an entity: B
  - The word is within an entity: I
  - The word is not an entity: O

- Incorporate word/label dependencies

![Diagram showing labels and features for sequence labeling.](image-url)
Graphical Models - HMM

- Integrate linguistic features (part-of-speech) and lexical patterns into HMMs
- Define an observable state as a pair (word, POS(word))
- Objective: given a sequence of words (W) and POS (S), find most probable tag sequence (T).

\[
\hat{T} = \arg \max_T P(T | W, S) = \arg \max_T \frac{P(W, S | T)P(T)}{P(W, S)} \\
\hat{T} = \arg \max_T P(W, S | T)P(T) = \arg \max_T P(S | T)P(W | T, S)p(T) \\
\hat{T} = \arg \max_T \prod_{i=1}^{n} \left( \frac{P(s_i | w_{i-1}, t_i) \times P(w_i | w_{i-1}, s_i, t_i) \times P(t_i | w_{i-1}, t_{i-1})}{P(w_i | w_{i-1}, s_i, t_i) \times P(t_i | w_{i-1}, t_{i-1})} \right)
\]

Graphical Models - CRF

- HMM is a generative model and is hard to integrate rich features
- CRF is a discriminative model and is flexible in terms of structures

\[
P(Y \mid X) = \frac{1}{Z(X)} \exp \left( \sum_{e \in E, i} \gamma_i t_i(e, Y \mid e, X) + \sum_{v \in V, i} \mu_i s_i(v, Y \mid v, X) \right)
\]

---

Li et al. Structure-Aware Review Mining and Summarization. COLING 2010
Choi et al. Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns. HLT/EMNLP 2005
Breck et al. Identifying expressions of opinion in context. IJCAI 2007
Graphical Models - CRF

Word Feature:
- Word token
- Word lemma
- Word part of speech
- Previous word token, lemma, part of speech
- Next word token, lemma, part of speech
- Negation word appears in previous 4 words
- Is superlative degree
- Is comparative degree

Dictionary Feature
- WordNet Synonym
- WordNet Antonym
- SentiWordNet Prior Polarity

Sentence Feature
- Num of positive words in SentiWordNet
- Num of negative words in SentiWordNet
- Num of Negation word

Syntactic Features:
- Parent word
- Parent SentiWordnet Prior Polarity
- In subject
- In copular
- In object

Edge Feature
- Conjunction word
- Syntactic relationship
Graphical Models - CRF

- CRF fails to model segment-level information: syntactic constituent
- Semi-Markov CRF performs sequence labeling at segment level
  \[ s = < s_1, \ldots, s_n > \quad s_i = (t_i, u_i, y_i) \]
- Identify opinion expressions (DSE/ESE)

\[
p(s|x) = \frac{1}{Z(x)} \exp \left\{ \sum_i \sum_k \lambda_k g_k(i, x, s) \right\} \quad Z(x) = \sum_{s' \in S} \exp \left\{ \sum_i \sum_k \lambda_k g_k(i, x, s') \right\}
\]
Joint Inference with Component Classifiers

- Global optimization to optimize subtasks in one goal

Opinion Entity Classifier (CRF)
- Word
- POS
- Lexicon
- Grammatical role

Opinion Relation Classifier (CRF)
- Word
- Phrase type
- Grammatical role
- Distance
- Dependency path

Joint Inference with constraints (ILP)
\[
\arg \max_{x,u,v} \lambda \sum_{i \in S} \sum_{z} f_{iz} x_{iz} \\
+ (1 - \lambda) \sum_{k} \sum_{i \in O} \left( \sum_{j \in A_k} r_{ij} u_{ij} + r_{i0} v_{ik} \right) \\
\sum_{z} x_{iz} = 1 \\
\sum_{z \neq N} x_{iz} + \sum_{z \neq N} x_{jz} \leq 1
\]

Choi et al. Joint Extraction of Entities and Relations for Opinion Recognition. EMNLP 2006
OUTLINE

● Background

● **Methodology**
  ○ **Unsupervised/Semi-supervised Learning**
    ■ Pattern mining
    ■ Topic modeling
    ■ Deep learning
  ○ **Supervised Learning**
    ■ Feature engineering
    ■ **Deep learning with syntactic information**
    ■ Deep learning without external knowledge

● Summary
General Architecture

- **Input**: distributed word embeddings (encode semantic regularities)
- **Hidden**: high-level features encoding input interactions
- **Output**: segmentation labels

<table>
<thead>
<tr>
<th>Labels</th>
<th>N</th>
<th>N</th>
<th>N</th>
<th>N</th>
<th>BO</th>
<th>BA</th>
<th>IA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Features</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$h_1$</td>
<td>$h_2$</td>
<td>$h_3$</td>
<td>$h_4$</td>
<td>$h_5$</td>
<td>$h_6$</td>
<td>$h_7$</td>
<td></td>
</tr>
</tbody>
</table>

Input $x$:
The phone has a **good** screen size
Syntax-Encoded Embedding

- Focus on learning meaningful word embeddings
- Encode dependency path into distributed representations

\[
\sum_{(w_1, w_2, r) \in C_1} \sum_{r' \sim p(r)} \max\{0, 1 - (w_2 - w_1)^T r + (w_2 - w_1)^T r'\} \sum_{(c, w) \in C_2} \sum_{c' \sim p(w)} \max\{0, 1 - w^T c + w^T c'\}
\]
Syntax-Encoded Deep Neural Networks

- Encode dependency tree into the deep learning structure
  - Recursive neural networks to encode syntactic interactions
  - CRF for final sequence tagging to encode sequential interactions

\[
\begin{align*}
    h_n &= f(W_v \cdot x_w + b + \sum_{k \in \mathcal{K}_n} W_{rnk} \cdot h_k) \\
    p(y|h) &= \frac{1}{Z(h)} \prod_{c \in C} \psi_c(h, y_c) \\
    \psi_c(h, y_c) &= \exp \langle W_c, F(h, y_c) \rangle
\end{align*}
\]
Explicit Syntax Incorporation with ILP

- Obtain probability predictions for aspect label sequence \( \{a_1, \ldots, a_n\} \) and opinion label sequence \( \{o_1, \ldots, o_n\} \) using multi-task NN

- Adopt ILP global inference with 3 constraints. Binary prediction sequence of ILP for aspect and opinion \( \{p_1, \ldots, p_n\} \{q_1, \ldots, q_n\} \)
  - Intra-task constraint: e.g., I-AT (I-OT) should not follow O-AT (O-OT)
  - Inter-task constraint:
  
  \[
  q_{\text{parent}(i)1} + q_{\text{parent}(i)2} \geq p_{i1} + p_{i2}, \quad \forall z_i \in \{\text{subj}\} \cap r_i \in \text{NN} \cap r_{\text{parent}(i)} \in \text{JJ}
  \]
  - Lexicon constraint

- Inference

\[
\max \frac{1}{n} \sum_{i=0}^{n-1} \left( \sum_{j=0}^{2} p_{ij} \log a_{ij} + \sum_{k=0}^{2} q_{ik} \log o_{ik} \right)
\]

s.t. \( p_{ij} \in \{0, 1\}, \quad q_{ik} \in \{0, 1\}, \)

\[
\sum_{j=0}^{2} p_{ij} = 1, \quad \sum_{k=0}^{2} q_{ik} = 1,
\]

Yu et al. Global Inference for Aspect and Opinion Terms Co-Extraction Based on Multi-Task Neural Networks. IEEE/ACM TASLP 2019
OUTLINE

- Background
- **Methodology**
  - Unsupervised/Semi-supervised Learning
    - Pattern mining
    - Topic modeling
    - Deep learning
  - **Supervised Learning**
    - Feature engineering
    - Deep learning with syntactic information
    - **Deep learning without external knowledge**
- Summary
Recurrent Neural Networks

- Effect of different variants of RNN on aspect extraction
- Input with different pre-trained word embeddings and features
- Deep RNN improves upon shallow RNN for phrases that implicitly convey subjectivity (ESE)

Irsoy and Cardie. Opinion Mining with Deep Recurrent Neural Networks. EMNLP 2014
Liu et al. Fine-grained Opinion Mining with Recurrent Neural Networks and Word Embeddings. EMNLP 2015
Katiyar and Cardie. Investigating LSTMs for Joint Extraction of Opinion Entities and Relations. ACL 2016
CNN with Rich Embedding

- A simple CNN with well-pretrained word embeddings
- Leverage both general embeddings and domain embeddings
  - General embedding:
    - Glove
  - Domain embedding:
    - Amazon
    - Yelp

Xu et al. Double Embeddings and CNN-based Sequence Labeling for Aspect Extraction. ACL 2018
Interaction-based Attention

- Coupled attentions: aspect attention & opinion attention
- Memory network: multiple layers of attentions

Wang et al. Coupled Multi-layer Attentions for Co-extraction of Aspect and Opinion Terms. AAAI 2017
Interaction-based Attention

- Coupled attentions: aspect attention & opinion attention
- Memory network: multiple layers of attentions

Wang et al. Coupled Multi-layer Attentions for Co-extraction of Aspect and Opinion Terms. AAAI 2017
Interaction-based Attention

\[ r_j^a = \tanh([h_j^\top G^a u^a : h_j^\top D^a u^p]) \]

\[ u_{t+1}^a = \tanh(Q^a u_t^a) + o_t^a \]

\[ o_t^a = \sum_j \alpha_t^a h_j \]
Multi-task attentions for Joint Extraction

- Extraction of both aspect/opinion terms together with aspect categories

- Tensor decomposition:
  \[ G_c = f(G, Z_c) \]

- Context-aware multi-task feature learning:
  \[ \hat{r}_c = g(r_c, S) \]

- Auxiliary task
Memory Interaction

- The aspect-opinion relationship is established based on neural memory interactions

- Multi-task framework
  - Aspect extraction task:
    - A-LSTM
  - Opinion extraction task:
    - O-LSTM
  - Sentimental sentence classification:
    - S-LSTM

Memory operations:
- **READ**: select aspect (opinion) hidden states
- **DIGEST**: distill an aspect (opinion) -specific summary
- **INTERACT**

Li and Lam. Deep Multi-Task Learning for Aspect Term Extraction with Memory Interaction. EMNLP 2017
OUTLINE

- Background
- Methodology
  - Unsupervised/Semi-supervised Learning
    - Pattern mining
    - Topic modeling
    - Deep learning
  - Supervised Learning
    - Feature engineering
    - Deep learning with syntactic information
    - Deep learning without external knowledge
- Summary
Dataset

- **SemEval Challenge**

<table>
<thead>
<tr>
<th>Description</th>
<th>Training</th>
<th></th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>text</td>
<td>tuple</td>
<td>text</td>
</tr>
<tr>
<td>SemEval-15 Restaurant</td>
<td>1,315</td>
<td>1,654</td>
<td>685</td>
</tr>
<tr>
<td>SemEval-16 Restaurant</td>
<td>2,000</td>
<td>2,507</td>
<td>676</td>
</tr>
<tr>
<td>SemEval-14 Laptop</td>
<td>3,045</td>
<td>1,974</td>
<td>800</td>
</tr>
<tr>
<td>SemEval-14 Restaurant</td>
<td>3,041</td>
<td>–</td>
<td>800</td>
</tr>
</tbody>
</table>

- **Digital Device**

<table>
<thead>
<tr>
<th>Data set</th>
<th>Number of reviews</th>
<th>Number of sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>45</td>
<td>597</td>
</tr>
<tr>
<td>D2</td>
<td>34</td>
<td>346</td>
</tr>
<tr>
<td>D3</td>
<td>41</td>
<td>546</td>
</tr>
<tr>
<td>D4</td>
<td>95</td>
<td>1,716</td>
</tr>
<tr>
<td>D5</td>
<td>99</td>
<td>740</td>
</tr>
<tr>
<td>Avg</td>
<td>62.8</td>
<td>789</td>
</tr>
</tbody>
</table>

- **MPQA**

<table>
<thead>
<tr>
<th></th>
<th>Opinion</th>
<th>Target</th>
<th>Holder</th>
</tr>
</thead>
<tbody>
<tr>
<td>TotalNum</td>
<td>5849</td>
<td>4676</td>
<td>4244</td>
</tr>
<tr>
<td>Opinion-arg Relations</td>
<td>4823</td>
<td></td>
<td>1302</td>
</tr>
<tr>
<td>Implicit Relations</td>
<td>4662</td>
<td></td>
<td>1187</td>
</tr>
</tbody>
</table>

- **Citysearch, BeerAdvocate**

<table>
<thead>
<tr>
<th>Domain</th>
<th>#Reviews</th>
<th>#Labeled sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restaurant</td>
<td>52,574</td>
<td>3,400</td>
</tr>
<tr>
<td>Beer</td>
<td>1,586,259</td>
<td>9,245</td>
</tr>
</tbody>
</table>
Experimental Results

- SemEval

<table>
<thead>
<tr>
<th>Model</th>
<th>Laptop</th>
<th>Restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRF</td>
<td>74.01</td>
<td>69.56</td>
</tr>
<tr>
<td>IHS_RD</td>
<td>74.55</td>
<td>-</td>
</tr>
<tr>
<td>NLANGP</td>
<td>-</td>
<td>72.34</td>
</tr>
<tr>
<td>WDEmb</td>
<td>75.16</td>
<td>-</td>
</tr>
<tr>
<td>LSTM</td>
<td>75.25</td>
<td>71.26</td>
</tr>
<tr>
<td>BiLSTM-CNN-CRF</td>
<td>77.8</td>
<td>72.5</td>
</tr>
<tr>
<td>RNCRF</td>
<td>78.42</td>
<td>-</td>
</tr>
<tr>
<td>CMLA</td>
<td>77.80</td>
<td>-</td>
</tr>
<tr>
<td>MIN</td>
<td>77.58</td>
<td>73.44</td>
</tr>
<tr>
<td>GloVe-CNN</td>
<td>77.67</td>
<td>72.08</td>
</tr>
<tr>
<td>Domain-CNN</td>
<td>78.12</td>
<td>71.75</td>
</tr>
<tr>
<td>MaxPool-DE-CNN</td>
<td>77.45</td>
<td>71.12</td>
</tr>
<tr>
<td>DE-LSTM</td>
<td>78.73</td>
<td>72.94</td>
</tr>
<tr>
<td>DE-OOD-CNN</td>
<td>80.21</td>
<td>74.2</td>
</tr>
<tr>
<td>DE-Google-CNN</td>
<td>78.8</td>
<td>72.1</td>
</tr>
<tr>
<td>DE-CNN-CRF</td>
<td>80.8</td>
<td>74.1</td>
</tr>
<tr>
<td>DE-CNN</td>
<td>*<em>81.59</em></td>
<td>*<em>74.37</em></td>
</tr>
</tbody>
</table>
Experimental Results

- **MPQA**

<table>
<thead>
<tr>
<th>Method</th>
<th>Opinion Expression</th>
<th>Opinion Target</th>
<th>Opinion Holder</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F1</td>
</tr>
<tr>
<td>CRF</td>
<td>84.42</td>
<td>61.61</td>
<td>71.17</td>
</tr>
<tr>
<td>CRF+ILP</td>
<td>73.53</td>
<td>74.89</td>
<td>74.11</td>
</tr>
<tr>
<td>LSTM+WLL</td>
<td>67.88</td>
<td>66.13</td>
<td>66.87</td>
</tr>
<tr>
<td>LSTM+SLL</td>
<td>70.45</td>
<td>66.65</td>
<td>68.37</td>
</tr>
<tr>
<td>LSTM+SLL+RLL</td>
<td>71.73</td>
<td>70.92</td>
<td>71.11</td>
</tr>
<tr>
<td>CRF</td>
<td>80.78</td>
<td>57.62</td>
<td>67.19</td>
</tr>
<tr>
<td>CRF+ILP</td>
<td>71.03</td>
<td>69.72</td>
<td>70.22</td>
</tr>
<tr>
<td>LSTM+WLL</td>
<td>64.47</td>
<td>59.45</td>
<td>61.65</td>
</tr>
<tr>
<td>LSTM+SLL</td>
<td>65.97</td>
<td>61.76</td>
<td>63.60</td>
</tr>
<tr>
<td>LSTM+SLL+RLL</td>
<td>65.48</td>
<td>65.54</td>
<td>65.56</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>IS-ABOUT</th>
<th>IS-FROM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>CRF+ILP</td>
<td>61.57</td>
<td>47.65</td>
</tr>
<tr>
<td>LSTM+SLL+Softmax</td>
<td>36.23</td>
<td>36.12</td>
</tr>
<tr>
<td>LSTM+SLL+RLL</td>
<td>62.48</td>
<td>49.80</td>
</tr>
</tbody>
</table>
## Experimental Results

- **Citysearch, BeerAdvocate**

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>LocLDA</td>
<td>0.898</td>
<td>0.648</td>
<td>0.753</td>
</tr>
<tr>
<td></td>
<td>ME-LDA</td>
<td>0.874</td>
<td>0.787</td>
<td>0.828</td>
</tr>
<tr>
<td></td>
<td>SAS</td>
<td>0.867</td>
<td>0.772</td>
<td>0.817</td>
</tr>
<tr>
<td></td>
<td>BTM</td>
<td>0.933</td>
<td>0.745</td>
<td>0.816</td>
</tr>
<tr>
<td></td>
<td>SERBM</td>
<td>0.891</td>
<td>0.854</td>
<td>0.872</td>
</tr>
<tr>
<td></td>
<td>$k$-means$^3$</td>
<td>0.931</td>
<td>0.647</td>
<td>0.755</td>
</tr>
<tr>
<td></td>
<td>ABAE</td>
<td><strong>0.953</strong></td>
<td>0.741</td>
<td>0.828</td>
</tr>
<tr>
<td>Staff</td>
<td>LocLDA</td>
<td>0.804</td>
<td>0.585</td>
<td>0.677</td>
</tr>
<tr>
<td></td>
<td>ME-LDA</td>
<td>0.779</td>
<td>0.540</td>
<td>0.638</td>
</tr>
<tr>
<td></td>
<td>SAS</td>
<td>0.774</td>
<td>0.556</td>
<td>0.647</td>
</tr>
<tr>
<td></td>
<td>BTM</td>
<td><strong>0.828</strong></td>
<td>0.579</td>
<td>0.677</td>
</tr>
<tr>
<td></td>
<td>SERBM</td>
<td>0.819</td>
<td>0.582</td>
<td>0.680</td>
</tr>
<tr>
<td></td>
<td>$k$-means</td>
<td>0.789</td>
<td>0.685</td>
<td>0.659</td>
</tr>
<tr>
<td></td>
<td>ABAE</td>
<td>0.802</td>
<td><strong>0.728</strong></td>
<td><strong>0.757</strong></td>
</tr>
<tr>
<td>Ambience</td>
<td>LocLDA</td>
<td>0.603</td>
<td>0.677</td>
<td>0.638</td>
</tr>
<tr>
<td></td>
<td>ME-LDA</td>
<td>0.773</td>
<td>0.558</td>
<td>0.648</td>
</tr>
<tr>
<td></td>
<td>SAS</td>
<td>0.780</td>
<td>0.542</td>
<td>0.640</td>
</tr>
<tr>
<td></td>
<td>BTM</td>
<td>0.813</td>
<td>0.599</td>
<td>0.685</td>
</tr>
<tr>
<td></td>
<td>SERBM</td>
<td>0.805</td>
<td>0.592</td>
<td>0.682</td>
</tr>
<tr>
<td></td>
<td>$k$-means</td>
<td>0.730</td>
<td>0.637</td>
<td>0.677</td>
</tr>
<tr>
<td></td>
<td>ABAE</td>
<td><strong>0.815</strong></td>
<td><strong>0.698</strong></td>
<td><strong>0.740</strong></td>
</tr>
<tr>
<td>Feel</td>
<td>$k$-means</td>
<td>0.533</td>
<td>0.413</td>
<td>0.456</td>
</tr>
<tr>
<td></td>
<td>LocLDA</td>
<td>0.938</td>
<td>0.537</td>
<td>0.675</td>
</tr>
<tr>
<td></td>
<td>SAS</td>
<td>0.783</td>
<td>0.695</td>
<td>0.730</td>
</tr>
<tr>
<td></td>
<td>BTM</td>
<td>0.892</td>
<td>0.687</td>
<td>0.772</td>
</tr>
<tr>
<td></td>
<td>ABAE</td>
<td>0.815</td>
<td><strong>0.824</strong></td>
<td><strong>0.816</strong></td>
</tr>
<tr>
<td>Taste</td>
<td>$k$-means</td>
<td>0.543</td>
<td>0.496</td>
<td>0.505</td>
</tr>
<tr>
<td></td>
<td>LocLDA</td>
<td>0.399</td>
<td><strong>0.655</strong></td>
<td>0.487</td>
</tr>
<tr>
<td></td>
<td>SAS</td>
<td>0.543</td>
<td>0.496</td>
<td>0.505</td>
</tr>
<tr>
<td></td>
<td>BTM</td>
<td>0.616</td>
<td>0.467</td>
<td>0.527</td>
</tr>
<tr>
<td></td>
<td>ABAE</td>
<td>0.637</td>
<td>0.358</td>
<td>0.456</td>
</tr>
<tr>
<td>Smell</td>
<td>$k$-means</td>
<td><strong>0.844</strong></td>
<td>0.295</td>
<td>0.422</td>
</tr>
<tr>
<td></td>
<td>LocLDA</td>
<td>0.560</td>
<td>0.488</td>
<td>0.489</td>
</tr>
<tr>
<td></td>
<td>SAS</td>
<td>0.336</td>
<td>0.673</td>
<td>0.404</td>
</tr>
<tr>
<td></td>
<td>BTM</td>
<td>0.541</td>
<td>0.549</td>
<td>0.527</td>
</tr>
<tr>
<td></td>
<td>ABAE</td>
<td>0.483</td>
<td><strong>0.744</strong></td>
<td><strong>0.575</strong></td>
</tr>
<tr>
<td>Taste+Smell</td>
<td>$k$-means</td>
<td>0.697</td>
<td>0.828</td>
<td>0.740</td>
</tr>
<tr>
<td></td>
<td>LocLDA</td>
<td><strong>0.897</strong></td>
<td><strong>0.853</strong></td>
<td><strong>0.866</strong></td>
</tr>
<tr>
<td></td>
<td>SAS</td>
<td>0.804</td>
<td>0.759</td>
<td>0.769</td>
</tr>
<tr>
<td></td>
<td>BTM</td>
<td>0.885</td>
<td>0.760</td>
<td>0.815</td>
</tr>
<tr>
<td></td>
<td>ABAE</td>
<td>0.651</td>
<td><strong>0.873</strong></td>
<td><strong>0.735</strong></td>
</tr>
</tbody>
</table>

*He et al. An Unsupervised Neural Attention Model for Aspect Extraction. ACL 2017*
Opinion Summarization
OUTLINE

- Aspect-based opinion summarization
- Extractive summarization
- Abstractive summarization
- Summary
OUTLINE

● Aspect-based opinion summarization
● Extractive summarization
● Abstractive summarization
● Summary
Different Forms of Summaries

- Statistical summary

**Product: Camera**

- Feature: picture

- Positive: 12

  - Overall this is a good camera with a really good picture clarity.
  - The pictures are absolutely amazing - the camera captures the minutest of details.
  - After nearly 800 pictures I have found that this camera takes incredible pictures.
  
  ...

- Negative: 2

  - The pictures come out hazy if your hands shake even for a moment during the entire process of taking a picture.
  - Focusing on a display rack about 20 feet away in a brightly lit room during day time, pictures produced by this camera were blurry and in a shade of orange.
  
  ...

Different Forms of Summaries

- Structured summary

Lu et al. Rated Aspect Summarization of Short Comments. WWW 2009
Different Forms of Summaries

- Visualized summary
Pipelined Approach

- Identify object features or aspect categories (aspect extraction)
- Identify opinion expressions (opinion extraction)
- Associate opinion expressions with target objects (relation detection)
- Determine sentiment polarities of opinion expressions

Gone With The Wind:

**Movie:**
- *Positive:* great, good, amazing, ..., breathtaking
- *Negative:* bad, boring, waste time, ..., mistake

**Actor:**
- *Positive:* charming, brilliant, great, ..., smart
- *Negative:* poor, fail, dirty, ..., lame

**Music:**
- *Positive:* great, beautiful, very good, ..., top
- *Negative:* annoying, noise, too long, ..., unnecessary
... ...
Pipelined Approach

- Identify object features or aspect categories (aspect extraction)
- Associate each sentence to specific aspects
- Group sentences based on sentiment polarities

Hu M and Liu B. Mining and summarizing customer reviews. KDD 2004
Zhuang et al.. Movie review mining and summarization. CIKM 2006
Titov and MacDonald. A Joint Model of Text and Aspect Ratings for Sentiment Summarization. ACL-HLT 2008
Blair-Goldensohn et al. Building a Sentiment Summarizer for Local Service Reviews. NLPIX 2008
Pipelined Approach

- Identify object features or aspect categories (aspect extraction)
  \[ f = (w_m, w_h) \]

- Predict rating for each aspect from the overall ratings
  - Global \( r(f \in t) = r(t) \)
  - Local
    \[ p(w_m | A_i, r) = \frac{c(w_m, S(A_i, r))}{\sum_{w_m' \in V_m} c(w_m', S(A_i, r))} \]
    \[ r(f) = \arg\max_r \{ p(w_m | A_i, r) | A(f) = i \} \]

- Extract representative phrases
  \[ (A_i, R(A_i), RF(A_i))_{i=1}^k \]

Lu et al. Rated Aspect Summarization of Short Comments. WWW 2009
OUTLINE

- Aspect-based opinion summarization
- **Extractive summarization**
- Abstractive summarization
- Summary
Ranking-based Summarization: Contrastiveness

- Given: a topic, a set of positive sentences and negative sentences
- Goal: generate contrastive sentence pairs
- Criteria: representativeness & contrastiveness

  - Content similarity function \( \phi(s_1, s_2) \)
  - Contrastive similarity function \( \psi(s_1, s_2) \)

\[
\begin{align*}
r(S) &= \frac{1}{|X|} \sum_{x \in X} \max_{i \in [1,k]} \phi(x, u_i) + \frac{1}{|Y|} \sum_{y \in Y} \max_{i \in [1,k]} \phi(y, v_i) \\
c(S) &= \frac{1}{k} \sum_{i=1}^{k} \psi(u_i, v_i)
\end{align*}
\]

\[
S^* = \arg \max_S (\lambda \ r(S) + (1 - \lambda) \ c(S))
\]

<table>
<thead>
<tr>
<th>Contradictory Aspect</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contradictory 1</td>
<td>( u_1 )</td>
<td>( v_1 )</td>
</tr>
<tr>
<td>Contradictory 2</td>
<td>( u_2 )</td>
<td>( v_2 )</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Contradictory k</td>
<td>( u_k )</td>
<td>( v_k )</td>
</tr>
</tbody>
</table>

Kim and Zhai. Generating Comparative Summaries of Contradictory Opinions in Text. CIKM 2009
Ranking-based Summarization: Contrastiveness

- PageRank-like algorithm to select sentences
- Preprocessing: use topic modeling to jointly model topic and viewpoint (A word may depend on the topic, the viewpoint, both, or neither)
- Compute jumping probability from ith excerpt to jth excerpt

\[
P(x_j|x_i, z) = \frac{\text{sim}_z(x_i, x_j)}{\sum_{j' \in X} \text{sim}_z(x_i, x_{j'})}
\]

\[
\text{sim}_0(x_i, x_j) = \text{sim}(x_i, x_j) \sum_{m=1}^{k} P(v = m|x_i)P(v = m|x_j)
\]

\[
\text{sim}_1(x_i, x_j) = \text{sim}(x_i, x_j) \times \sum_{m_1, m_2 \in [1, k], m_1 \neq m_2} P(v = m_1|x_i)P(v = m_2|x_j)
\]

\[
P(x_i)P(x_j|x_i, z = 1) + P(x_j)P(x_i|x_j, z = 1)
\]

Paul et al. Summarizing Contrastive Viewpoints in Opinionated Text. EMNLP 2010
Ranking-based Summarization: Concept & Coherence

- Rank sentences based on concept and coherence (besides extraction)
  - Concept: Define opinion as a concept \( e = <t, a, p> \) (has weight \( w_i \))
  - Coherence: local coherence score between adjacent sentences \( c_{i,j} = w \cdot \phi(x, y) \)
- Decoding as Integer Linear Programming (fix \( w \) and \( c \))

\[
\max \left\{ \lambda \sum_{e_i \in E} w_i e_i + (1 - \lambda) \sum_{a_{i,j} \in A} c_{i,j} a_{i,j} \right\} \\
\text{s.t. } s_i, a_{i,j}, e_i \in \{0, 1\} \; \forall i, j \\
\sum_i m_{ij} s_i \geq e_j \; \forall j \\
\sum_i a_{i,j} + \sum_i a_{j,i} = 2s_j \; \forall j \\
\sum_i a_{i,j} = \sum_i a_{j,i} \; \forall j
\]

Nishikawa et al. Opinion Summarization with Integer Linear Programming Formulation for Sentence Extraction and Ordering. COLING 2010
OUTLINE

- Aspect-based opinion summarization
- Extractive summarization
- **Abstractive summarization**
- Summary
Abstractive vs Extractive Summarization

- Extractive summarization is not coherent
- Extractive summarization does not provide an aggregate view

**Extractive:** Bottom line, well made camera, easy to use, very flexible and powerful features to include the ability to use external flash and lens/filter choices. It has a beautiful design, lots of features, very easy to use, very configurable and customizable, and the battery duration is amazing! Great colors, pictures and white balance.

**Abstractive:** Almost all users loved the Canon G3 possibly because some users thought the physical appearance was very good. Furthermore, several users found the manual features and the special features to be very good. Also, some users liked the convenience because some users thought the battery was excellent. Finally, some users found the editing/viewing interface to be good despite the fact that several customers really disliked the viewfinder.
Abstractive Summarization

- Which features of the evaluated entity were most ‘important’ to the users
- Some aggregate of the user opinions for the important features
- The distribution of those opinions
- The reasons behind each user opinion
- Different from general text summarization
Natural Language Generation System

Reviews, meta-info

Text planner: select/organize

Microplanning: lexical selection, sentence planning

Surface realizer: output text

Quantifiers:

if (relative-number == 1) : [“All users (x people) who commented about the aspect”, “All costumers (x people) that reviewed the aspect”, ...]
if (relative-number >= 0.8) : [“Almost all users commented about the aspect and they”, “Almost all costumers mentioned the aspect and they”, ...]
if (relative-number >= 0.6) : [“Most users commented about the aspect and they mainly”, “Most shoppers mentioned aspect and they”, ...]
if (relative-number >= 0.45) : [“Almost half of the users commented about the aspect and they”, “Almost 50% of the shoppers mentioned the aspect and they”, ...]
if (relative-number >= 0.2) : [“About y% of the reviewers commented about the aspect and they”, “Around y% of the shoppers mentioned the aspect and they”, ...]
if (relative-number >= 0.0) : [“z reviewers commented about the aspect and in overall they”, “z shoppers mentioned about the aspect and they”, ...]

Polarity verbs:

if (controversial(aspect)) : [“had controversial opinions about it”, “expressed controversial opinions about this feature”, ...]
else: if (average <= -2) : [“hated it”, “felt that it was very poor”, “thought that it was very poor”, ...]
if (average <= -1) : [“disliked it”, “felt that it was poor”, “thought that it was poor”, ...]
if (average < 0) : [“did not like it”, “felt that it was weak”, “thought that it was weak”, ...]
if (average == 0) : [“did not express any strong positive or negative opinion about it”, ...]
if (average <= +1) : [“liked it”, “felt that it was fine”, “thought that it was satisfactory”, ...]
if (average <= +2) : [“absolutely liked it”, “really liked this feature”, “felt that it was a really good feature”, “thought that it was really good”, ...]
if (average <= +3) : [“loved it”, “felt that it was great”, “thought that it was great”, ...]

Connectives

[“Also, related to the aspect”, “Accordingly, ”, “Moreover, regarding the aspect, ”, “In relation to the aspect, ”, “Talking about the aspect,”, ...]
Discourse Structure

- An Aspect Hierarchy Tree (AHT) that reflects the importance of aspects as well as the relationships between them
- Assume aspect and opinions are given (or preprocessing)
- Process
  - Obtain a discourse tree representation for every review and modify the tree such that every leaf node only contains the aspect words (ADT)
Discourse Structure

● An Aspect Hierarchy Tree (AHT) that reflects the importance of aspects as well as the relationships between them

● Assume aspect and opinions are given (or preprocessing)

● Process
  ○ Aggregate the trees and generate an Aggregated Rhetorical Relation Graph (ARRG)
Discourse Structure

- An Aspect Hierarchy Tree (AHT) that reflects the importance of aspects as well as the relationships between them.
- Assume aspect and opinions are given (or preprocessing).
- Process
  - Use Weighted PageRank to select a subgraph representing the most important aspects (extract aspects that not only have high weight, but that are also linked with heavy edges to other heavy aspects.)
Discourse Structure

- An Aspect Hierarchy Tree (AHT) that reflects the importance of aspects as well as the relationships between them
- Assume aspect and opinions are given (or preprocessing)
- Process
  - Transforms the subgraph into a tree and provides an AHT as output
Key Phrase Composing

- Aim to generate a set of very concise phrases, each phrase is a summary of a key opinion
- Start with a set of high frequency seed words, gradually form meaningful higher order n-grams based on
  - representativeness (based on a modified mutual information function)
  - readability (based on an n-gram language model)
- Formulated as an optimization problem

\[
M^* = \arg \max_M \sum_{i=1}^{k} [S_{\text{rep}}(m_i) + S_{\text{read}}(m_i)]
\]

\[
s_{\text{rep}}(m_1 \ldots m_n) = \frac{1}{n} \sum_{i=1}^{n} \text{pmi}_{\text{local}}(w_i)
\]

\[
s_{\text{read}}(m_1 \ldots m_n) = \frac{1}{2C} \sum_{i=j-C}^{i+C} \text{pmi}'(w_i, w_j), i \neq j
\]

<table>
<thead>
<tr>
<th>Mp3 Player Y</th>
<th>Restaurant X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very short battery life.</td>
<td>Good service.</td>
</tr>
<tr>
<td>Big and clear screen.</td>
<td>Delicious soup dishes.</td>
</tr>
<tr>
<td>(8 words)</td>
<td>Very noisy at nights.</td>
</tr>
<tr>
<td></td>
<td>(9 words)</td>
</tr>
</tbody>
</table>

Deep-Learning-Based Generation (Attention)

- Input: a set of text units corresponding to the same topic
- Output: a one-sentence abstractive summary

Movie: The Martian
Reviews:
- One the **smartest**, sweetest, and most satisfyingly suspenseful sci-fi films in years.
- ...an intimate sci-fi epic that is **smart**, spectacular and stirring.
- The Martian is a **thrilling**, human and moving sci-fi picture that is easily the most emotionally engaging film Ridley Scott has made...
- It's pretty sunny and often **funny**, a space oddity for a director not known for pictures with a **sense of humor**.
- The Martian **highlights the book’s best qualities**, tones down its worst, and adds its own style...

**Opinion Consensus (Summary):** **Smart, thrilling**, and surprisingly **funny**, The Martian offers a **faithful adaptation of the bestselling book** that brings out the best in leading man Matt Damon and director Ridley Scott.

Wang and Ling. Neural Network-Based Abstract Generation for Opinions and Arguments. NAACL 2016
Deep-Learning-Based Generation (Attention)

- Attention: select relevant input information for generating summarization (similar to general text summarization)

\[
\log P(y|x) = \sum_{j=1, \ldots, |y|} \log P(y_j|y_1, \ldots, y_{j-1}, x)
\]
\[
p(y_j|y_1, \ldots, y_{j-1}, x) = \text{softmax}(h_j) \quad h_j = g(y_{j-1}, h_{j-1}, s)
\]

Wang and Ling. Neural Network-Based Abstract Generation for Opinions and Arguments. NAACL 2016
Deep-Learning-Based Generation (Attention)

- **Attention**: select relevant input information for generating summarization (similar to general text summarization)

\[
\log P(y|x) = \sum_{j=1,...,|y|} \log P(y_j|y_1,...,y_{j-1},x) \\
p(y_j|y_1,...,y_{j-1},x) = \text{softmax}(h_j) \quad h_j = g(y_{j-1}, h_{j-1}, s)
\]

- **Attention over multiple inputs**: importance sampling through ridge regression (gold-standard importance label generated through similarity score)

\[
J(w) = \| \tilde{R}w - \tilde{L} \|^2 + \lambda \cdot \| \tilde{R}^T \tilde{R} - \tilde{L}' \|^2 + \beta \cdot \|w\|^2 \\
\dot{w} = (\tilde{R}^T \tilde{R} + \tilde{R}'^T \tilde{L}' + \beta)^{-1}(\tilde{R}^T \tilde{L} + \tilde{R}'^T \tilde{L}') \\
f(x^k) = r_k \cdot w
\]

*Wang and Ling. Neural Network-Based Abstract Generation for Opinions and Arguments. NAACL 2016*
Deep-Learning-Based Generation (Attention)

- Captures aspect and sentiment information: mutual attention mechanism (aspect/sentiment-aware review representation)

**Context-related**
- Hidden states
  \[ H^c = [h_1^c, \ldots, h_k^c] \]
  \[ v^c = \sum_{i=1}^{k} h_i^c / k \]
- Mutual attention
  \[ emb_x^c = \sum_{i=1}^{k} C_i h_i^c \]

**Sentiment-related**
- Hidden states
  \[ H^s = [h_1^s, \ldots, h_m^s] \]
  \[ v^s = \sum_{i=1}^{m} h_i^s / m \]
- Mutual attention
  \[ emb_x^s = \sum_{i=1}^{m} S_i h_i^s \]

**Aspect-related**
- Hidden states
  \[ H^t = [h_1^t, \ldots, h_n^t] \]
  \[ v^t = \sum_{i=1}^{n} h_i^t / n \]
- Mutual attention
  \[ emb_x^t = \sum_{i=1}^{n} T_i h_i^t \]

Yang et al. Aspect and Sentiment Aware Abstractive Review Summarization. COLING 2018
Deep-Learning-Based Generation (Attention)

- Different styles and words for different aspect category: leverage text categorization task

$$emb_x = [emb^c_x, emb^s_x, emb^t_x]$$

$$\hat{y} = \text{softmax}(V_2 \cdot F_x), \quad F_x = \text{tanh}(V_1 \cdot emb_x)$$

- Generation with 3 attentions and pointer-generator framework

$$s_0 = cemb_x = \text{tanh}(W_\mu \times emb_x)$$

$$s_t = LSTM(s_{t-1}, c_t, w_{t-1})$$

$$a_{t,i} = \text{softmax}(\lambda_1 a_{t,i}^{\text{semantic}} + \lambda_2 a_{t,i}^{\text{sentiment}} + \lambda_3 a_{t,i}^{\text{aspect}}), \quad c_t = \sum_{i=1}^{k} a_{t,i} h^c_i$$

$$P(w_t) = p_{\text{gen}} P_{\text{vocab}}(w_t) + (1 - p_{\text{gen}}) \sum_{i:w_i=w_t} a_{t,i}$$
Deep-Learning-Based Generation (Reconstruction)

- Multi-document summarization in an unsupervised manner
- An auto-encoder module that learns representations for each review and constrains the summaries to be in the language domain
- A summarization module that learns to generate summaries that are semantically similar to each of the input documents

Deep-Learning-Based Generation (Structural)

- Learn a discourse tree in an unsupervised manner, generate a summary from surrounding sentences of the root
- Learn a language model through reconstruction
- DiscourseRank ranks each sentence to focus on the main claims

---

Summary:

Good quality floor puzzle

(1) This floor puzzle is a nice size not huge but larger than normal kid puzzles
(2) The pieces are thick and lock together well even on carpet
(4) I bought this puzzle for my son for his first birthday at the store

Body:

(3) The pieces are cardboard but are very dense almost like wood but not quite that solid
(5) My son put it together on berber carpet without having any issues with pieces not staying together
Deep-Learning-Based Generation (Structural)

- Learn both semantics and structure

  ![Diagram]

- Separate sentence embedding into 2 parts: $s_i = [s_i^e, s_i^f]$

- Encoding: $\hat{s}_i = \tanh(\mathbf{W}_s (\sum_{j=1}^{n} a_{ij} s_j^e) + \mathbf{b}_s)$

- Decoding: $\sum_{i=1}^{n} \sum_{t=1}^{l} \log P(w_t^i | w_{<t}^i, \hat{s}_i, \theta)$

Combine Summarization with Sentiment Prediction

- Jointly performs abstractive text summarization and sentiment classification within a hierarchical end-to-end neural framework
Combine Summarization with Sentiment Prediction

- A self-attention layer as a bridge that connects the summarization layer and the sentiment classification layer

Wang and Ren. A Self-Attentive Hierarchical Model for Jointly Improving Text Summarization and Sentiment Classification. ACML 2018
OUTLINE

- Aspect-based opinion summarization
- Extractive summarization
- Abstractive summarization
- Summary
Dataset

- Amazon product review

<table>
<thead>
<tr>
<th>Domains</th>
<th>Train</th>
<th>Valid</th>
<th>Eval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toys &amp; Games</td>
<td>27,037</td>
<td>498</td>
<td>512</td>
</tr>
<tr>
<td>Sports &amp; Outdoors</td>
<td>37,445</td>
<td>511</td>
<td>466</td>
</tr>
<tr>
<td>Movies &amp; TV</td>
<td>408,827</td>
<td>564</td>
<td>512</td>
</tr>
</tbody>
</table>
Experimental Results

- Amazon product review

<table>
<thead>
<tr>
<th>Domain</th>
<th>Toys &amp; Games</th>
<th>Sports &amp; Outdoors</th>
<th>Movies &amp; TV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R-1</td>
<td>R-2</td>
<td>R-L</td>
</tr>
<tr>
<td><strong>Unsupervised approaches</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TextRank</td>
<td>8.63</td>
<td>1.24</td>
<td>7.26</td>
</tr>
<tr>
<td>Opinosis</td>
<td>8.25</td>
<td>1.51</td>
<td>7.52</td>
</tr>
<tr>
<td>MeanSum-single</td>
<td>8.12</td>
<td>0.58</td>
<td>7.30</td>
</tr>
<tr>
<td>StrSum</td>
<td>11.61</td>
<td>1.56</td>
<td>11.04</td>
</tr>
<tr>
<td>StrSum+DiscourseRank</td>
<td><strong>11.87</strong></td>
<td><strong>1.63</strong></td>
<td><strong>11.40</strong></td>
</tr>
<tr>
<td><strong>Supervised baselines</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seq-Seq</td>
<td>13.50</td>
<td>2.10</td>
<td>13.31</td>
</tr>
</tbody>
</table>
Experimental Results


- Reference: love this game
- Seq-Seq-att: fun game
- Our Model (Full): i love this game

(a)

- Reference: good value
- Seq-Seq-att: good for the price
- Our Model (Full): this is a great product for the price

(b)

- Reference: disappointing
- Seq-Seq-att: great dvd
- Our Model (Full): this is a great movie

(c)
Thank You!