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

Wenya Wang[1], Jianfei Yu[2], Sinno Jialin Pan[1] and Jing Jiang[2]

[1]Nanyang Technological University, Singapore

[2]Singapore Management University, Singapore

TABLE OF CONTENTS

- Introduction
- Aspect-Based Extraction
- Target-Oriented Sentiment Classification
- Opinion Summarization
- Cutting-Edge Dimensions

Introduction

NANYANG TECHNOLOGICAL UNIVERSITY

What is Fine-Grained Opinion Mining

8.7 **Fabulous** · 11,276 reviews

Staff	8.5	Value for money	8.1
Facilities	8.7	Location	8.8
Cleanliness	8.9	Free WiFi	8.0
Comfort	8.9		

Antoine
France

10

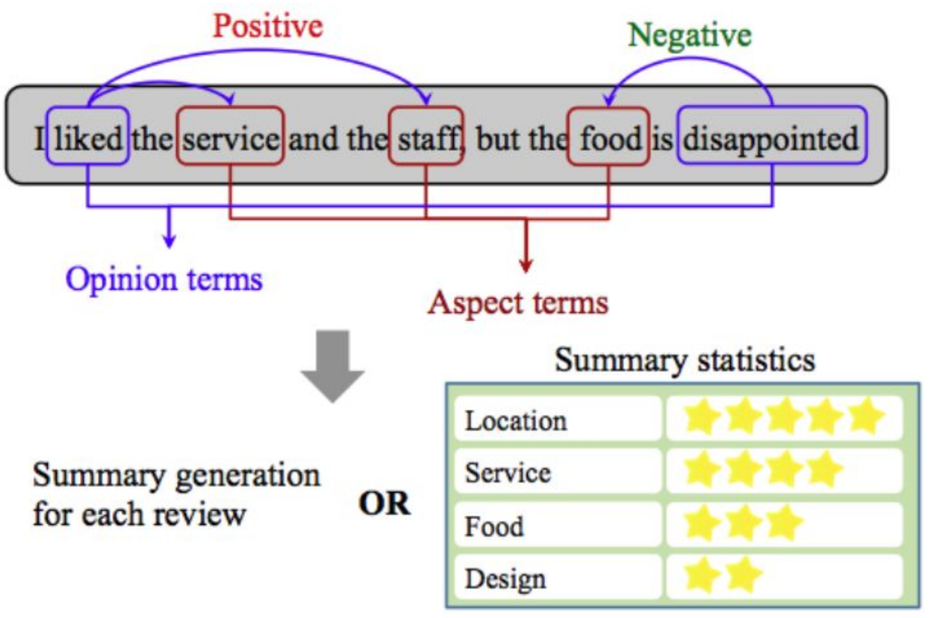
Reviewers' choice · Reviewed: 23 October 2018

Luxury, relaxation, great service and food.

Amazing service and luxury, I almost did not leave my suite as it was so comfortable. The dinning service was also outstanding. I'm looking forward 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

NANYANG TECHNOLOGICAL UNIVERSITY

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.

NANYANG TECHNOLOGICAL UNIVERSITY

Aspect-Based Extraction

NANYANG
TECHNOLOGICAL
UNIVERSITY

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

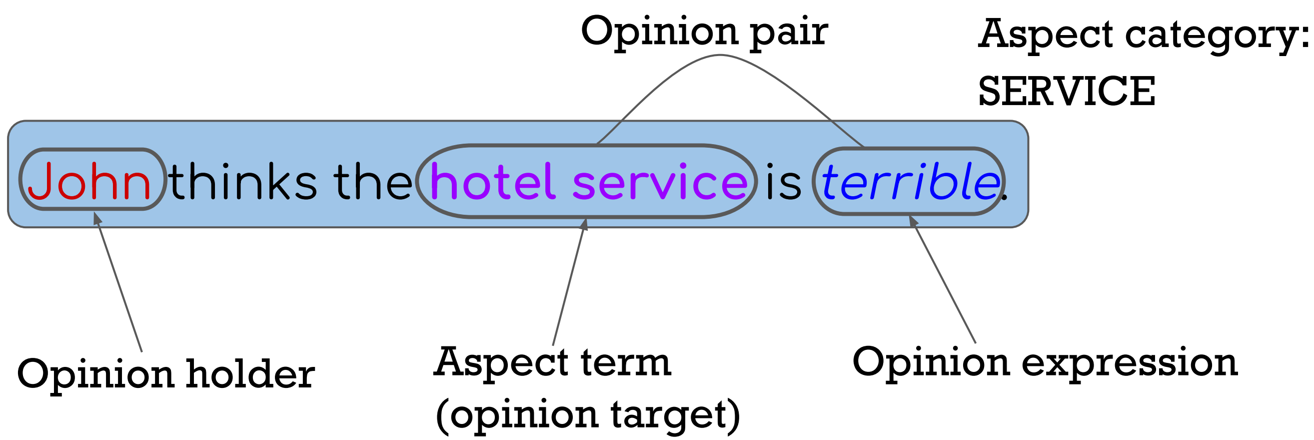
OUTLINE

- **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

NANYANG TECHNOLOGICAL UNIVERSITY

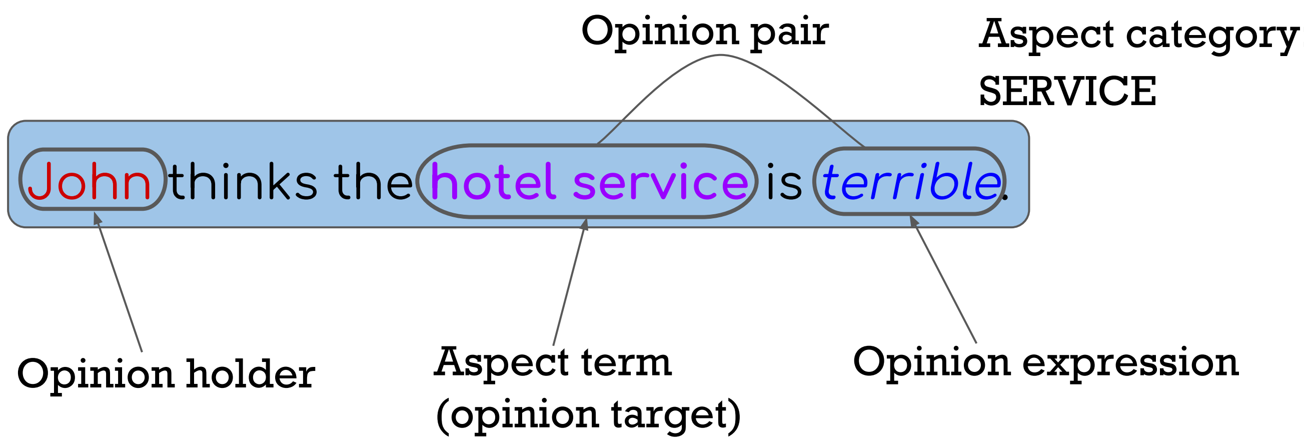
What Can be Extracted

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



What Can be Extracted

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

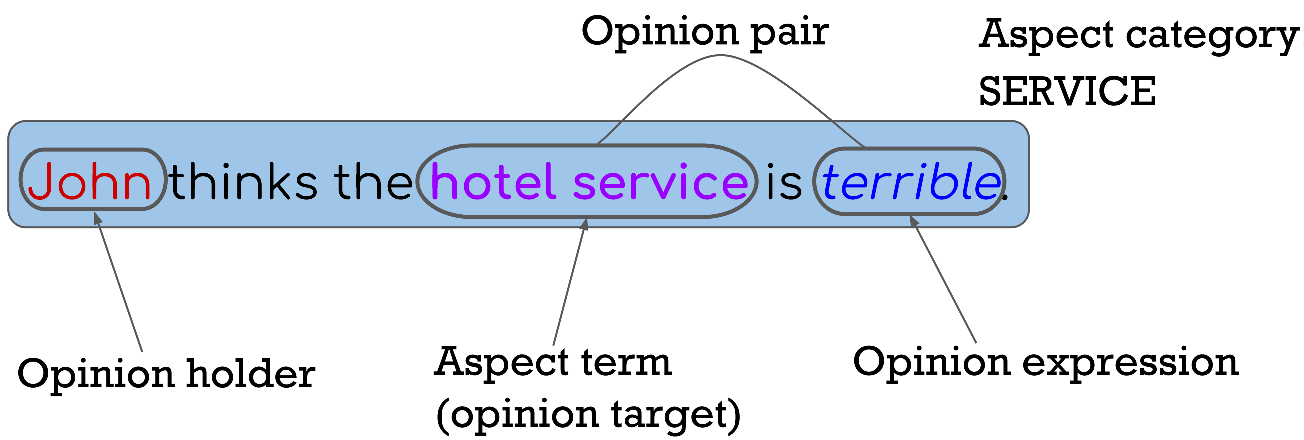


Aspect (Category) Identification

battery life, sound quality,
Ease of use, ...

What Can be Extracted

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



Aspect (Category) Identification

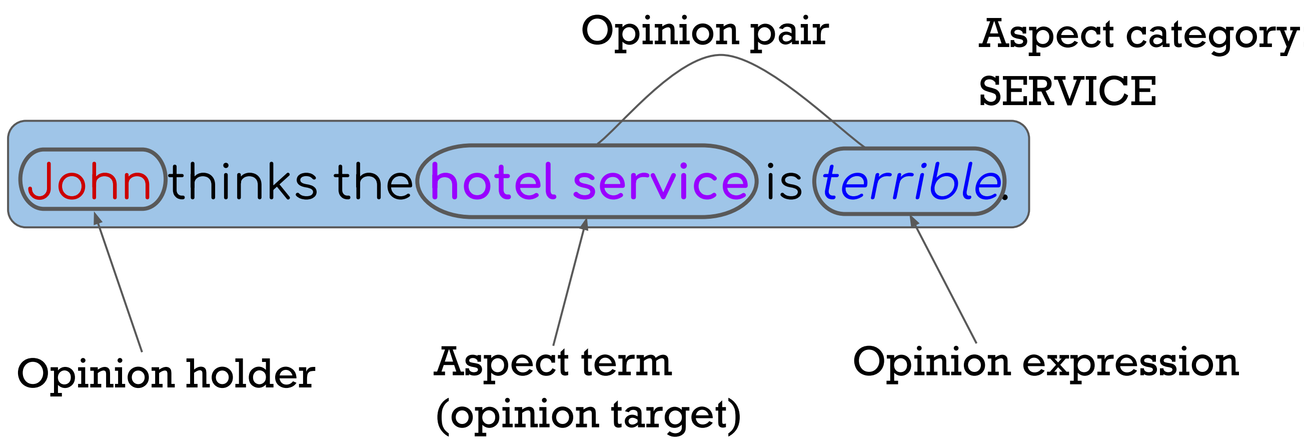
battery life, sound quality,
Ease of use, ...

Opinion Identification and Sentiment Prediction

great (+), long (+)
convenient (+), ...

What Can be Extracted

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



Aspect (Category) 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

S1: [The workers]$_{[H_{1,2}]}$ <u>were irked</u>$_{[O_1]}$ by [the government report]$_{[T_1]}$ and <u>were worried</u>$_{[O_2]}$ as they went about their daily chores.

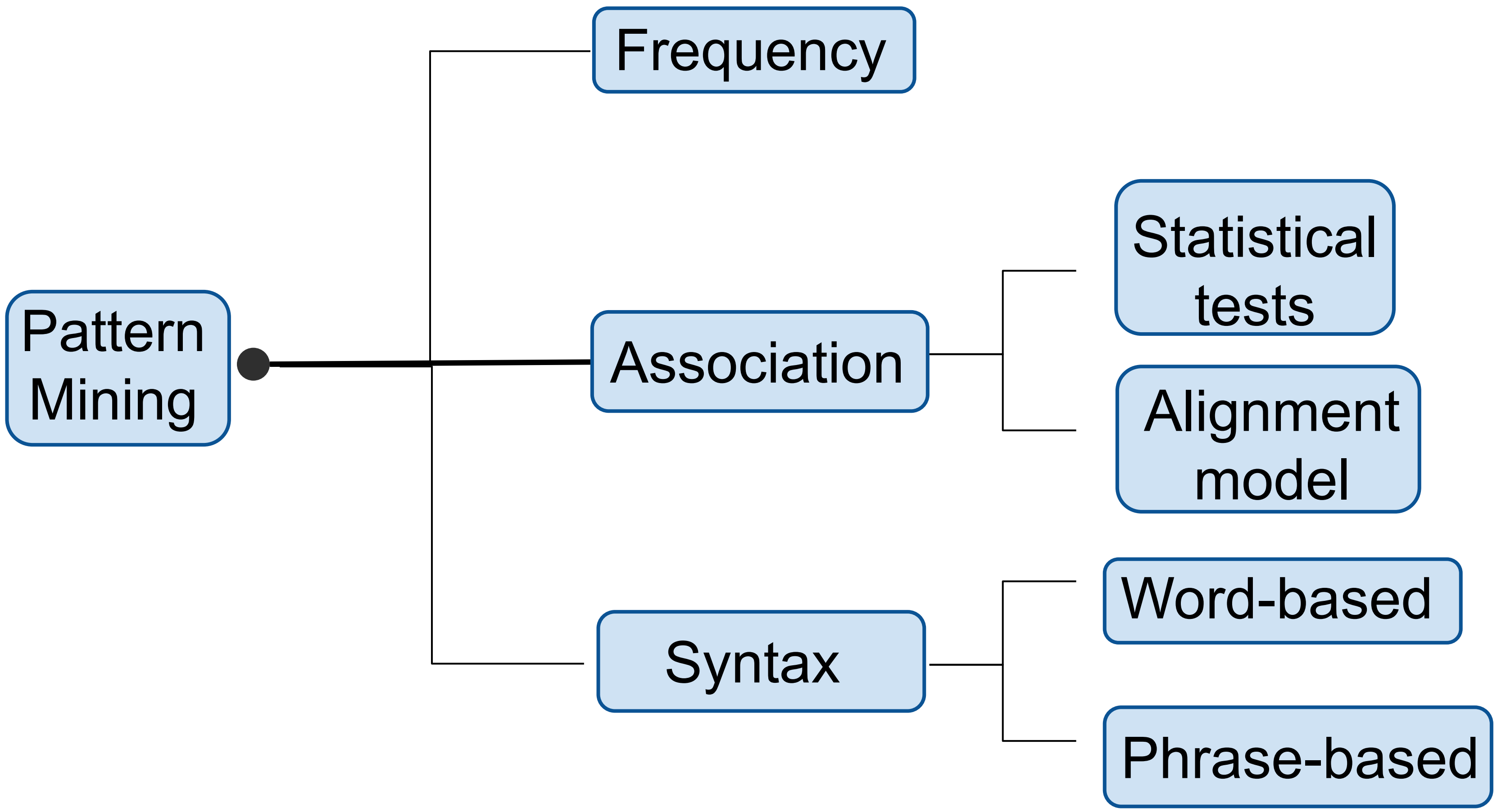
(1) The International Committee of the Red Cross, [as usual]$_{[ESE]}$, [has refused to make any statements]$_{[DSE]}$.

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

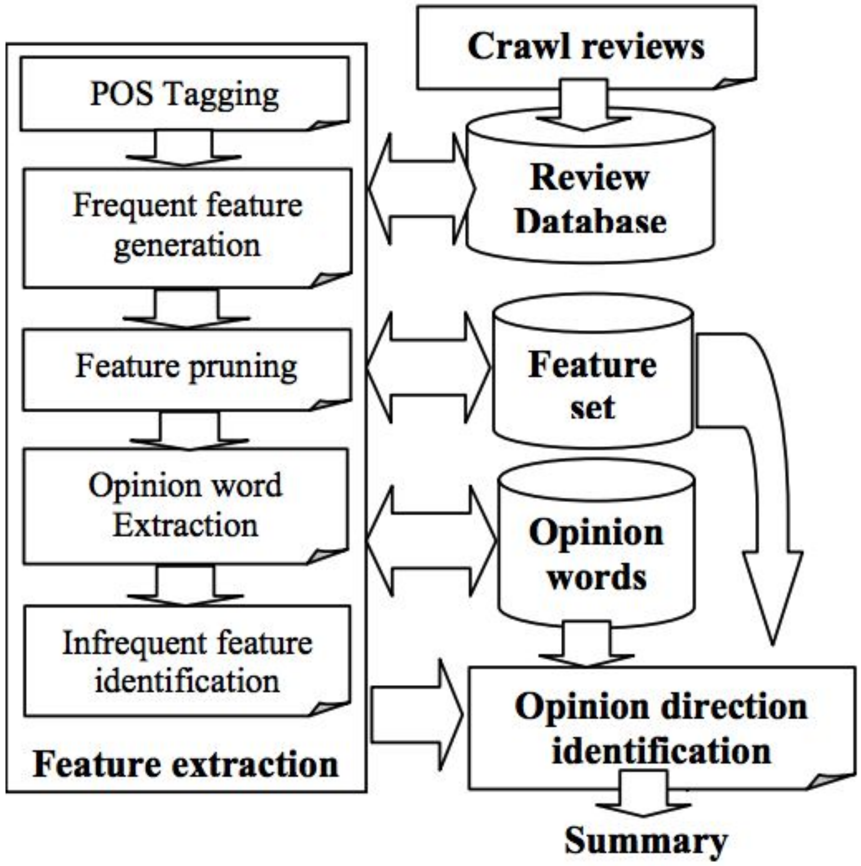
NANYANG TECHNOLOGICAL UNIVERSITY

Pattern Mining



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

Reviews:

1. The screen is really big, but the price is too expensive!
2. The price is expensive, students don't buy it usually.
3. The screen is beautiful, but the price is not!
4. The screen is big and beautiful!

CF={screen, price, student}
CO={big, expensive, buy, beautiful}

S={screen}
thd=2.0
F={screen, price}
O={big, beautiful, expensive}

A	screen	price	student	big	expensive	buy	beautiful
screen		2.5	0.5	3.0	1.5	0.5	3.0
price	2.5		1.0	1.5	3.0	1.5	1.5
student	0.5	1.0		0.5	0.5	1.0	0.5
big	3.0	1.5	0.5		2.0	0.5	2.0
expensive	1.5	3.0	0.5	2.0		1.0	2.0
buy	0.5	1.5	1.0	0.5	1.0		0.5
beautiful	3.0	1.5	0.5	2.0	2.0	0.5	

NANYANG TECHNOLOGICAL UNIVERSITY

Hai et al. One Seed to Find Them All: Mining Opinion Features via Association. CIKM 2012

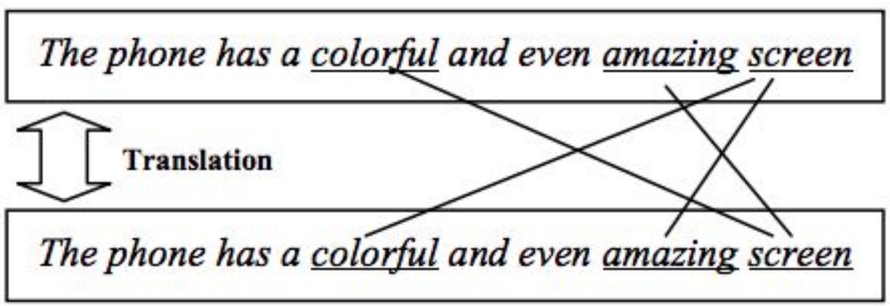
Extraction via Word Alignment

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

$$\hat{A} = \arg\max_{A} P(A \mid S)$$

$$S = \{w_1, w_2, \ldots, w_n\}$$

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



Liu et al. Opinion Target Extraction Using Word-Based Translation Model. EMNLP 2012
Liu et al. Opinion Target Extraction Using Partially-Supervised Word Alignment Model. IJCAI 2013

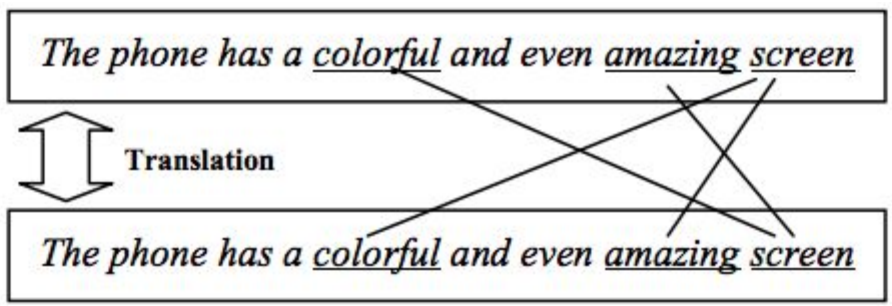
Extraction via Word Alignment

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

$$S = \{w_1, w_2, \ldots, w_n\}$$

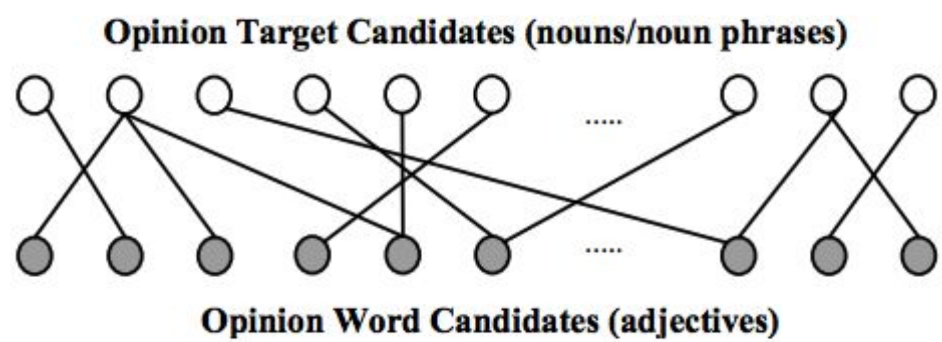
$$\hat{A} = \arg\max_{A} P(A \mid S)$$

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



- Graph-based algorithm to extract opinion targets

$$C^{t+1} = (1-\lambda) \times M^T \times M \times C^t + \lambda \times S$$

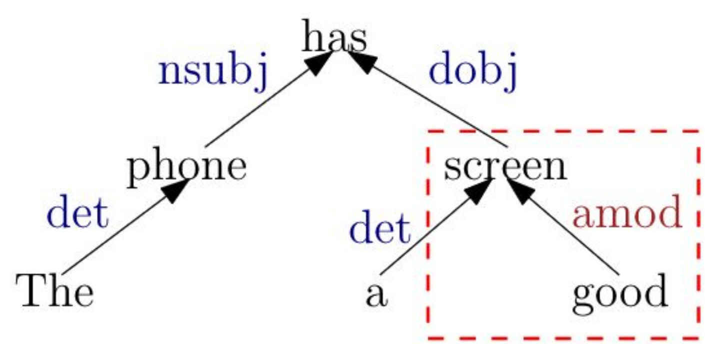


Liu et al. Opinion Target Extraction Using Word-Based Translation Model. EMNLP 2012
Liu et al. Opinion Target Extraction Using Partially-Supervised Word Alignment Model. IJCAI 2013

Word-Based Syntactic Rule Mining

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

RuleID	Observations
$R1_1$	$O \rightarrow$*O-Dep*$\rightarrow T$ **s.t.** $O \in \{O\}$,*O-Dep*$\in \{MR\}, POS(T) \in \{NN\}$



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

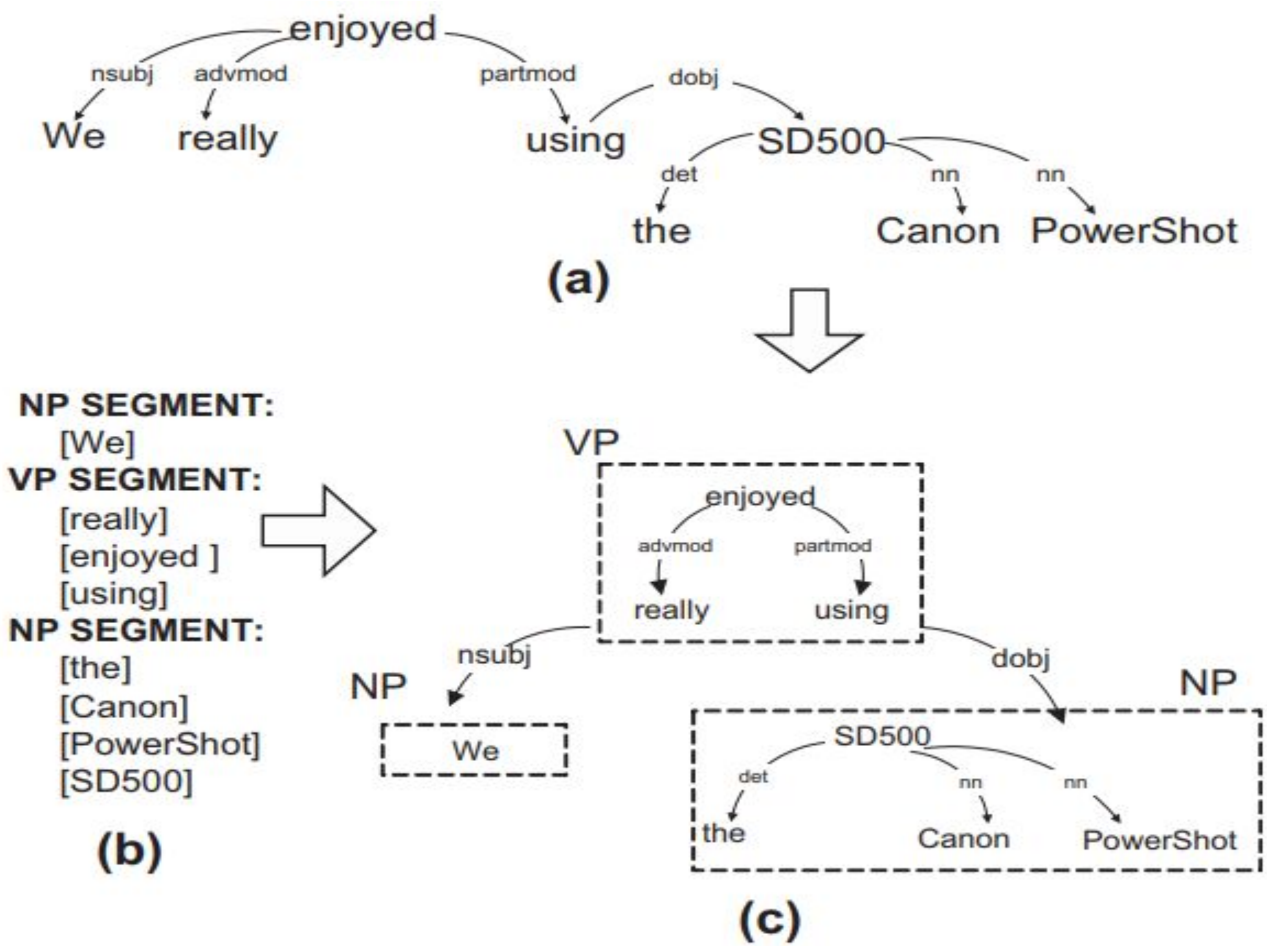
Popescu and Etzioni. Extracting Product Features and Opinions from Reviews. EMNLP 2005

Zhuang et al.. Movie review mining and summarization. CIKM 2006

NANYANG TECHNOLOGICAL UNIVERSITY

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($\mathcal{D}^t, \mathcal{R}^-, \mathcal{R}^+, \mathcal{O}$)

Input: Target dataset $\mathcal{D}^t$, high precision aspect extraction rules $\mathcal{R}^-$, high recall aspect extraction rules $\mathcal{R}^+$, seed opinion words $\mathcal{O}$

Output: Extracted aspect set $\mathcal{A}$

$\mathcal{T}^- \leftarrow \text{DPextract}(\mathcal{D}^t, \mathcal{R}^-, \mathcal{O})$;

$\mathcal{T}^+ \leftarrow \text{DPextract}(\mathcal{D}^t, \mathcal{R}^+, \mathcal{O})$;

$\mathcal{T} \leftarrow \mathcal{T}^+ - \mathcal{T}^-$;

$\mathcal{T}^s \leftarrow \text{Sim-recom}(\mathcal{T}^-, \mathcal{T})$;

$\mathcal{T}^a \leftarrow \text{AR-recom}(\mathcal{T}^-, \mathcal{T})$;

$\mathcal{A} \leftarrow \mathcal{T}^- \cup \mathcal{T}^s \cup \mathcal{T}^a$.

Base extraction through DP

Recommendation

Liu et al.. Improving Opinion Aspect Extraction Using Semantic Similarity and Aspect Associations. AAAI 2016

Limitation

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

NANYANG TECHNOLOGICAL UNIVERSITY

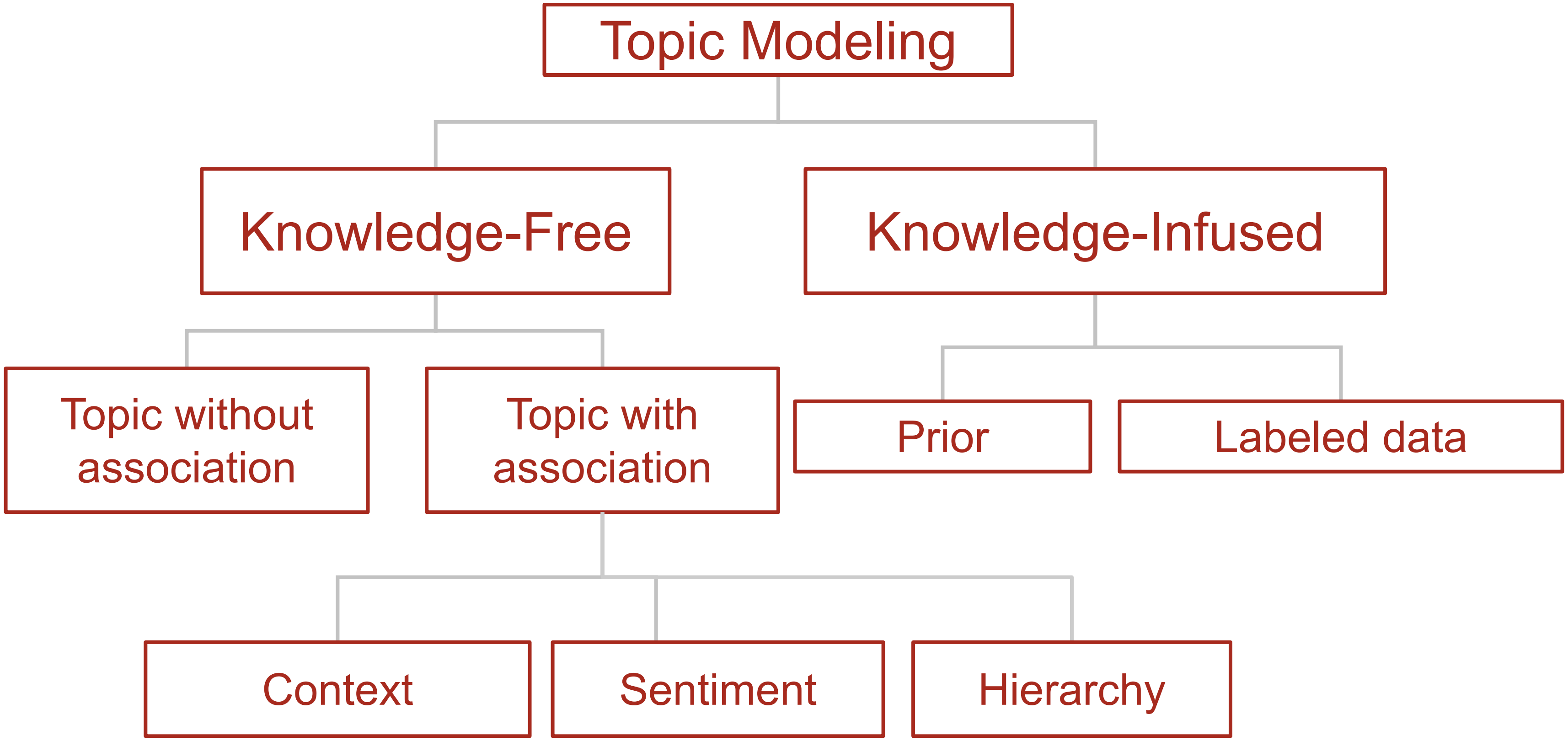
OUTLINE

- 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

NANYANG TECHNOLOGICAL UNIVERSITY

Topic Modeling

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



NANYANG TECHNOLOGICAL UNIVERSITY

LDA With Local/Global Context

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

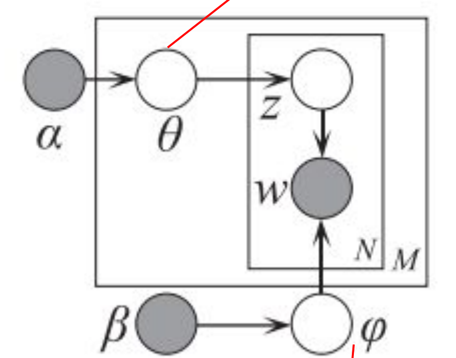
I. Choose distribution of topics $\theta_d \sim Dir(\alpha)$

II. For each word i

  A. Choose topic $z_{d,i} \sim \theta_d$

  B. Choose word $w_{d,i} \sim \varphi_{z_{d,i}}$

Topic distribution per document



(a)

Word distribution per topic

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

- 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

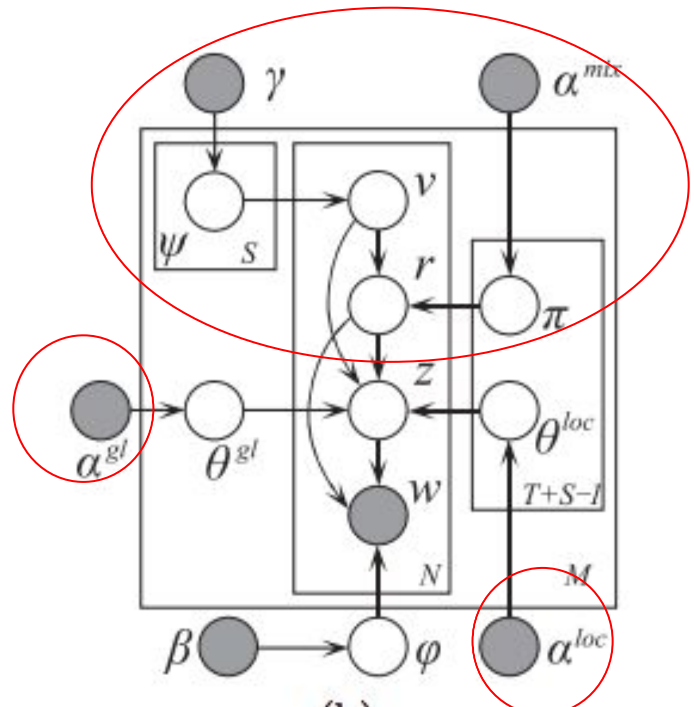
A. Choose window $v_{d,i} \sim \psi_{d,s}$

B. Choose $r_{d,i} \sim \pi_{d,v_{d,i}}$

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 $\phi_{z_{d,i}}^{r_{d,i}}$



(b)

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

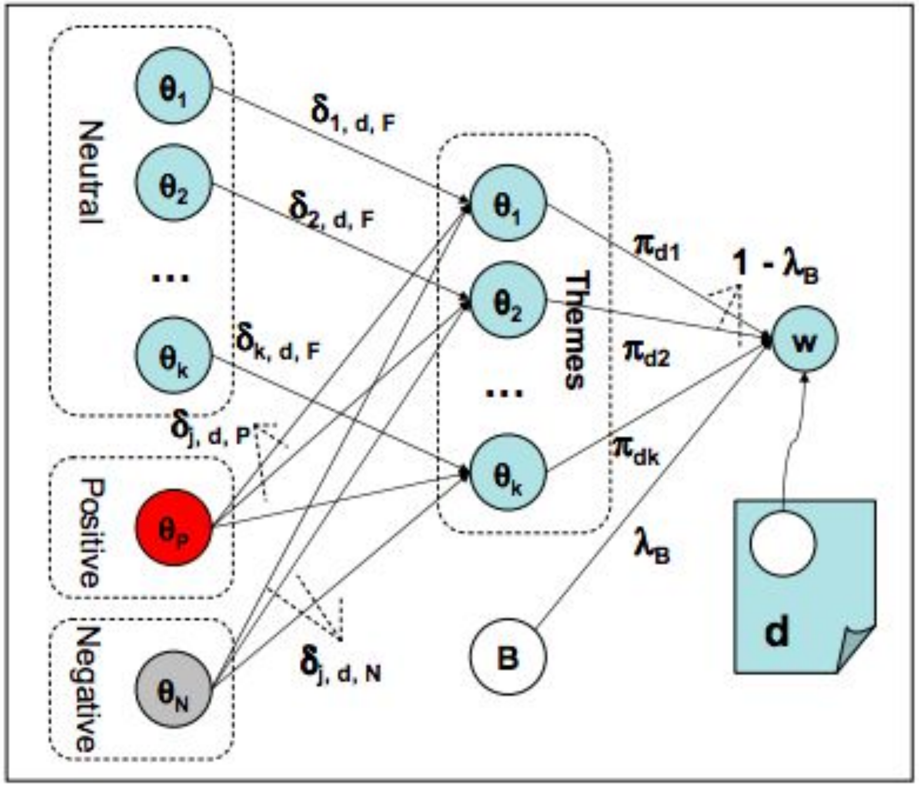
LDA With Local/Global Context

	label	top words
MG-LDA local (all topics)	sound quality	sound quality headphones volume bass earphones good settings ear rock excellent
	features	games features clock contacts calendar alarm notes game quiz feature extras solitaire
	connection with PC	usb pc windows port transfer computer mac software cable xp connection plug firewire
	tech. problems	reset noise backlight slow freeze turn remove playing icon creates hot cause disconnect
	appearance	case pocket silver screen plastic clip easily small blue black light white belt cover
	controls	button play track menu song buttons volume album tracks artist screen press select
	battery	battery hours life batteries charge aaa rechargeable time power lasts hour charged
	accessories	usb cable headphones adapter remote plug power charger included case firewire
	managing files	files software music computer transfer windows media cd pc drag drop file using
	radio/recording	radio fm voice recording record recorder audio mp3 microphone wma formats
MG-LDA global	iPod	ipod music apple songs use mini very just itunes like easy great time new buy really
	Creative Zen	zen creative micro touch xtra pad nomad waiting deleted labs nx sensitive 5gb eax
	Sony Walkman	sony walkman memory stick sonicstage players atrac3 mb atrac far software format
	video players	video screen videos device photos tv archos pictures camera movies dvd files view
	support	player product did just bought unit got buy work $ problem support time months
LDA (out of 40)	iPod	ipod music songs itunes mini apple battery use very computer easy time just song
	Creative	creative nomad zen xtra jukebox eax labs concert effects nx 60gb experience lyrics
	memory/battery	card memory cards sd flash batteries lyra battery aa slot compact extra mmc 32mb
	radio/recording	radio fm recording record device audio voice unit battery features usb recorder
	controls	button menu track play volume buttons player song tracks press mode screen settings
	opinion	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

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 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

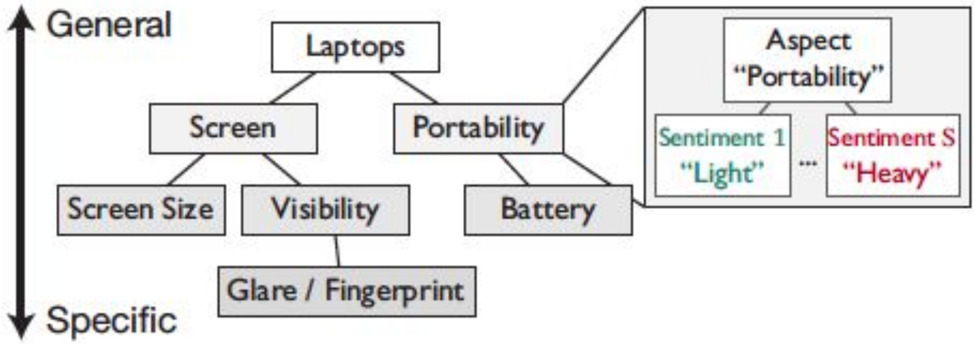


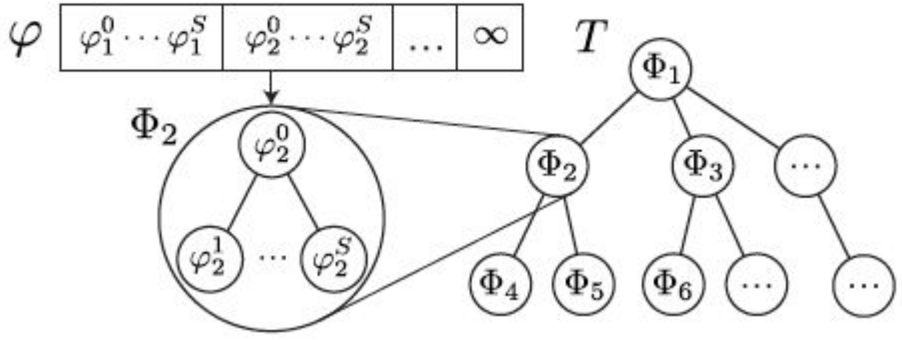
$$log(\mathcal{C}) = \sum_{d \in \mathcal{C}} \sum_{w \in V} c(w:d) log[\lambda_B p(w|B)$$
$$+ (1-\lambda_B) \sum_{j=1}^{k} \pi_{d_j} \times (\delta_{j,d,F} p(w|\theta_j)$$
$$+ \delta_{j,d,P} p(w|\theta_P) + \delta_{j,d,N} p(w|\theta_N))]$$

Sentiment coverage of topic j

LDA With Hierarchy

- Aspects usually form hierarchies: Aspect-Sentiment tree



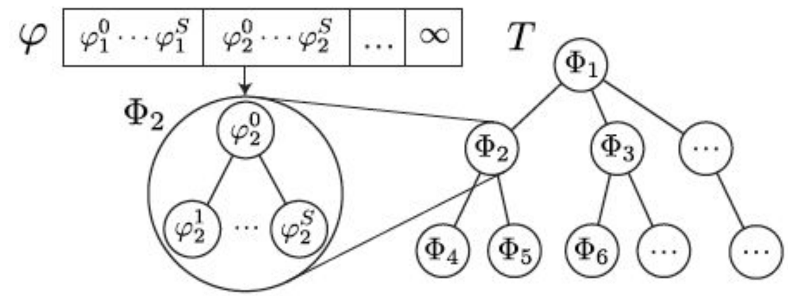


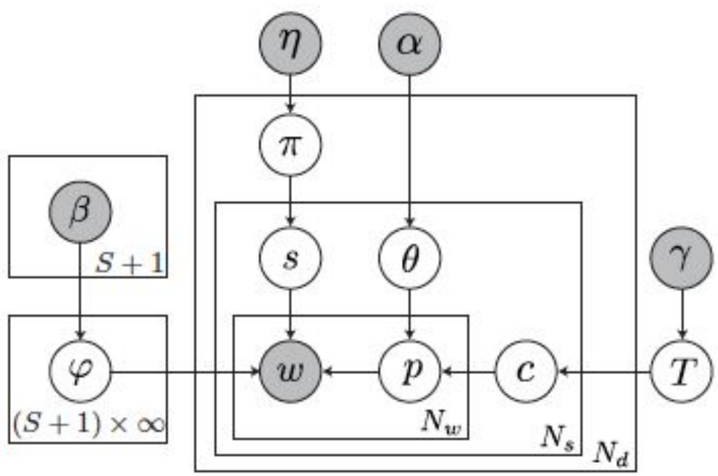
Kim et al. A Hierarchical Aspect-Sentiment Model for Online Reviews. AAAI 2013

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 $p \sim \text{Binomial}(1, \theta)$

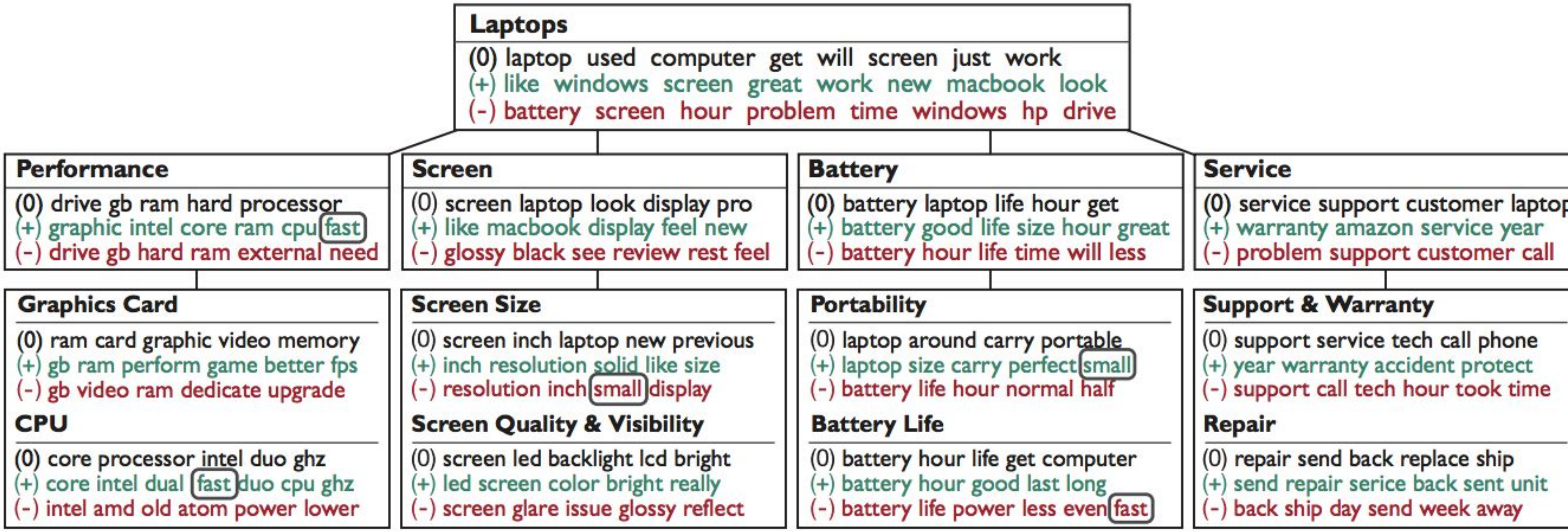
    $w \sim \text{Multinomial}(\varphi_c^{s \times p})$





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

LDA With Hierarchy

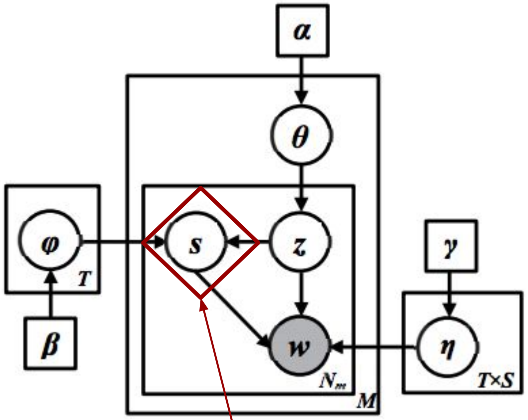


Kim et al. A Hierarchical Aspect-Sentiment Model for Online Reviews. AAAI 2013

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

$$
\begin{aligned}
&\theta \sim Dirichlet(\alpha)\\
&z_i|\theta_m \sim Multinomial(\theta_m)\\
&\varphi \sim Dirichlet(\beta)\\
&s_i|z_i,\varphi \sim Multinomial(\varphi_{z_i})\\
&\eta \sim Dirichlet(\gamma)\\
&w_i|z_i,s_i,\eta \sim Multinomial(\eta_{z_i,s_i})
\end{aligned}
$$



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

$$w_{d,s,n} \sim \begin{cases} \mathrm{Multi}(\phi^{\mathcal{B}}) & \text{if } y_{d,s,n} = 0 \\ \mathrm{Multi}(\phi^{\mathcal{A},z_{d,s}}) & \text{if } y_{d,s,n} = 1, u_{d,s,n} = 0 \\ \mathrm{Multi}(\phi^{\mathcal{A},g}) & \text{if } y_{d,s,n} = 1, u_{d,s,n} = 1 \\ \mathrm{Multi}(\phi^{\mathcal{O},z_{d,s}}) & \text{if } y_{d,s,n} = 2, u_{d,s,n} = 0 \\ \mathrm{Multi}(\phi^{\mathcal{O},g}) & \text{if } y_{d,s,n} = 2, u_{d,s,n} = 1 \end{cases}$$

Background model
T Specific aspect model
General aspect model
T specific opinion model
General opinion model

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

$$p(y_{d,s,n} = l | \boldsymbol{x}_{d,s,n}) = \pi_l^{d,s,n} = \frac{\exp\left(\lambda_l \cdot \boldsymbol{x}_{d,s,n}\right)}{\sum_{l'=0}^{2} \exp\left(\lambda_{l'} \cdot \boldsymbol{x}_{d,s,n}\right)}$$

NANYANG TECHNOLOGICAL UNIVERSITY

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

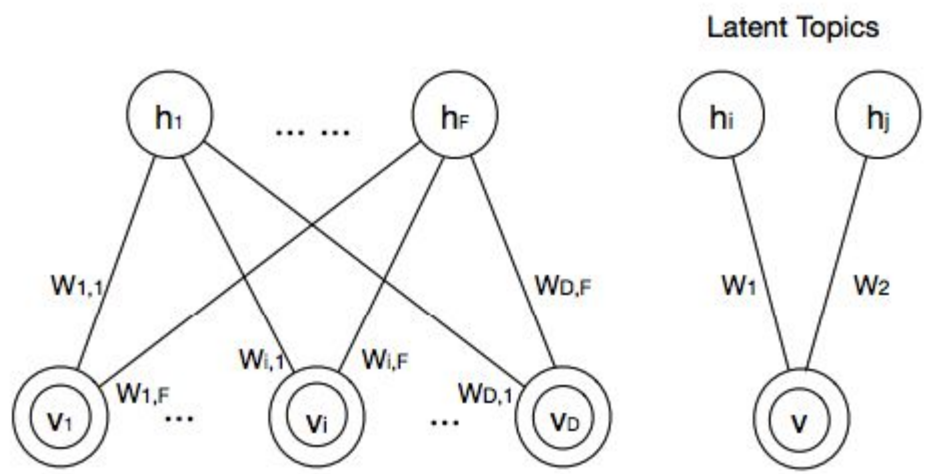
OUTLINE

- 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

NANYANG TECHNOLOGICAL UNIVERSITY

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

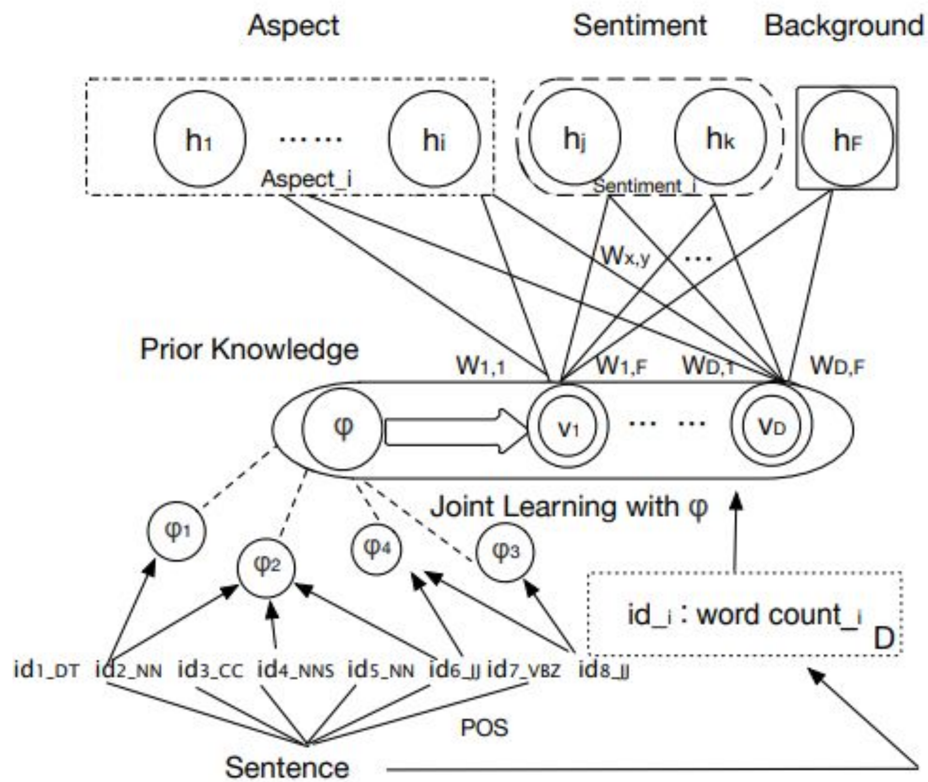
Aspect Extraction with RBM

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

- Construct optimal weight matrix

$$E(\mathbf{v}, h) = -\sum_{j=1}^{F}\sum_{k=1}^{K} W_j^k h_j \widehat{v}^k - \sum_{k=1}^{K} \widehat{v}^k b^k - \sum_{j=1}^{F} h_j a_j,$$

$$\begin{aligned} P(h_j = 1 \mid \widehat{v}^k) =& P(h_j = 1 \mid h_{-j}, \widehat{v}^k) \\ =& \sigma(a_j + W_j^k \widehat{v}^k). \end{aligned}$$



- Priors as regularizers

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

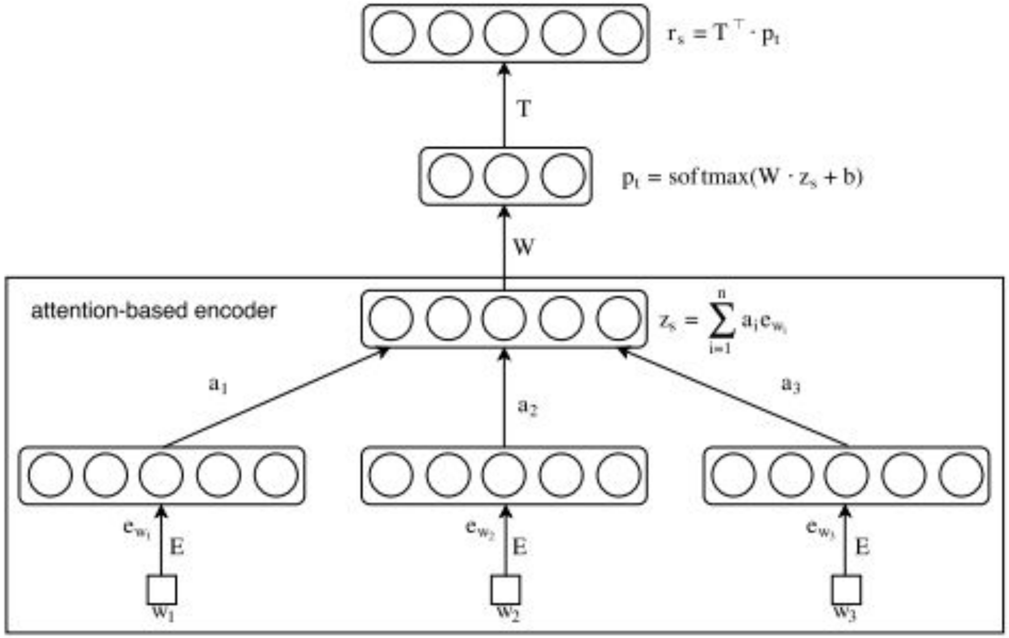
Aspect Extraction with Attention

- 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

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

$$d_i = \mathbf{e}_{w_i}^{\top} \cdot \mathbf{M} \cdot \mathbf{y}_s$$

$$\mathbf{y}_s = \frac{1}{n} \sum_{i=1}^{n} \mathbf{e}_{w_i}$$



- Sentence reconstruction to enhance coherence

$$J(\theta) = \sum_{s \in D} \sum_{i=1}^{m} \max(0, 1 - \mathbf{r}_s \mathbf{z}_s + \mathbf{r}_s \mathbf{n}_i)$$

NANYANG TECHNOLOGICAL UNIVERSITY

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

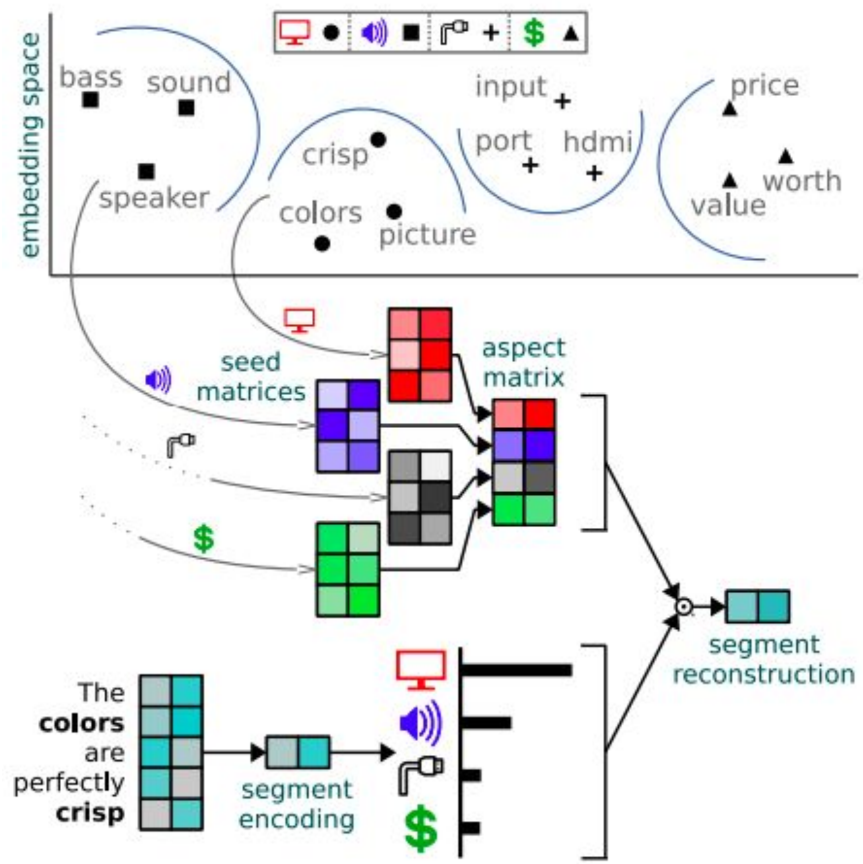
Aspect Extraction with Attention

- 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

$$\mathbf{p}_s^{dom} = \mathrm{softmax}(\mathbf{W}_{\mathrm{C}}\mathbf{v}_s + \mathbf{b}_{\mathrm{C}})$$

- Objective

$$J_{\mathrm{MT}}(\theta) = J_r(\theta) - \lambda \sum_{s \in \mathrm{C}_{all}} \log p^{(d_s)}$$



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

OUTLINE

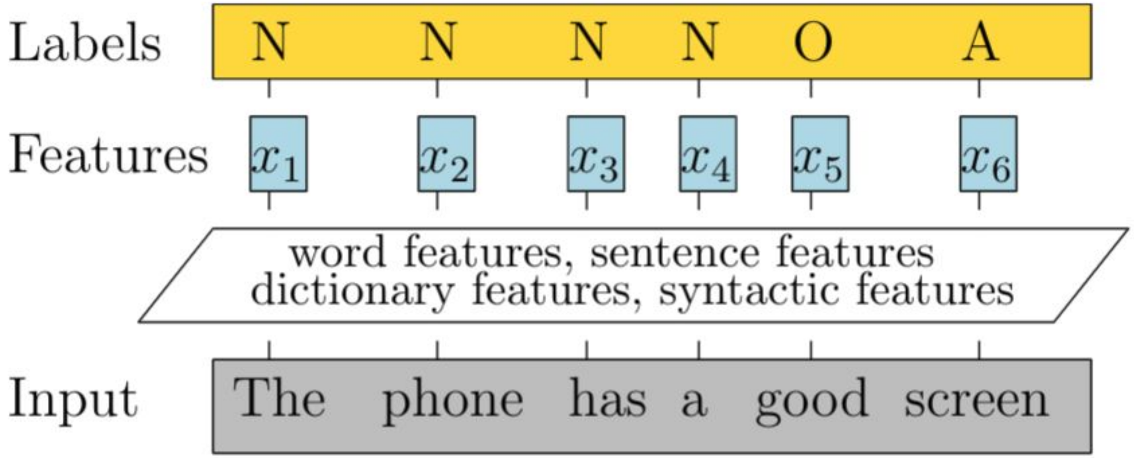
- Background
- **Methodology**
  - Unsupervised/Semi-supervised Learning
    - Pattern mining
    - Topic modeling
    - Deep learning
  - **Supervised Learning**
    - **Feature engineering**
    - Deep learning with syntactic information
    - Deep learning without external knowledge
- Summary

NANYANG TECHNOLOGICAL UNIVERSITY

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



NANYANG TECHNOLOGICAL UNIVERSITY

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).

$$\hat{T} = \arg\max_{T} P(T \mid W, S) = \arg\max_{T} \frac{P(W, S \mid T) P(T)}{P(W, S)}$$

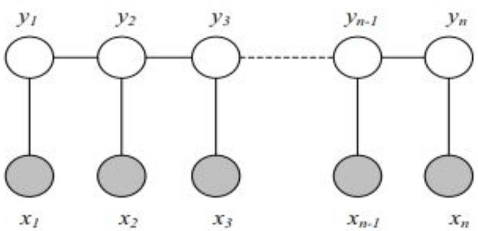
$$\hat{T} = \arg\max_{T} P(W, S \mid T) P(T) = \arg\max_{T} P(S \mid T) P(W \mid T, S) p(T)$$

$$\hat{T} = \arg\max_{T} \prod_{i=1}^{n} \begin{pmatrix} P(s_i \mid w_{i-1}, t_i) \times \\ P(w_i \mid w_{i-1}, s_i, t_i) \times \\ P(t_i \mid w_{i-1}, t_{i-1}) \end{pmatrix}$$

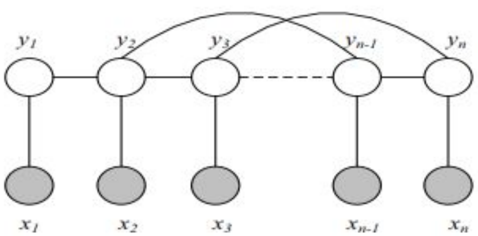
Jin and Ho. A Novel Lexicalized HMM-based Learning Framework for Web Opinion Mining. ICML 2009

Graphical Models - CRF

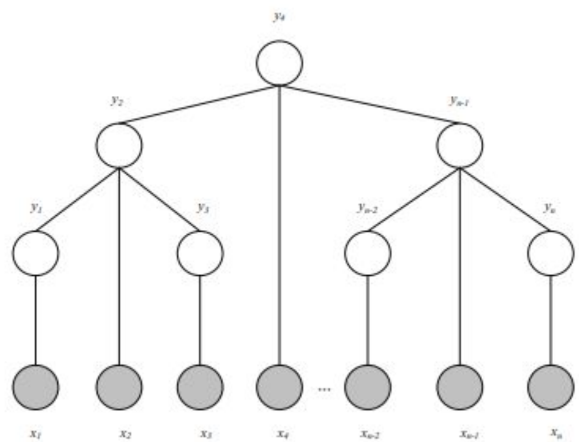
- HMM is a generative model and is hard to integrate rich features
- CRF is a discriminative model and is flexible in terms of structures



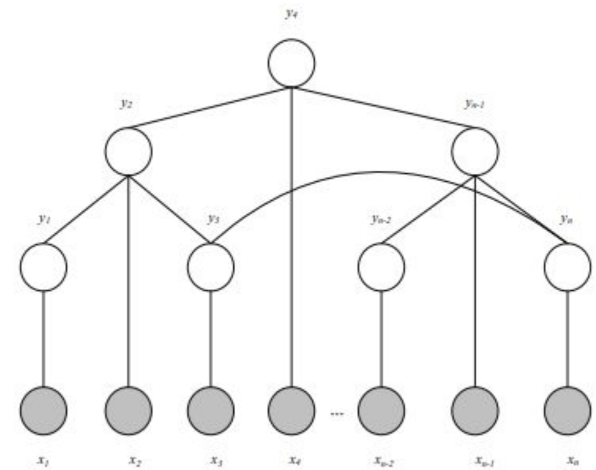
(a) Linear-chain CRFs



(b) Skip-chain CRFs



(c) Tree-CRFs



(d) Skip-Tree CRFs

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

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

Li et al. Structure-Aware Review Mining and Summarization. COLING 2010

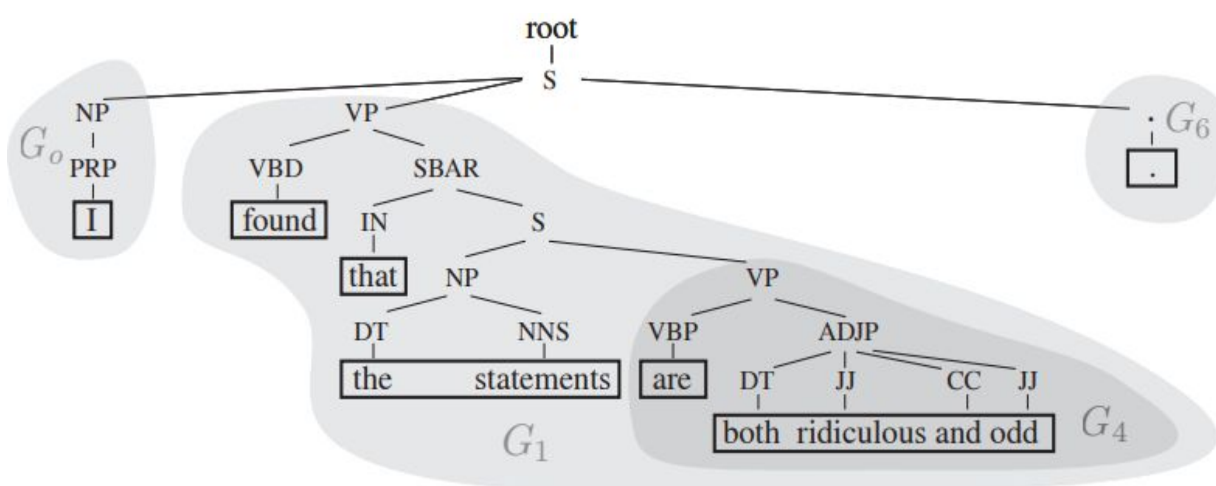
Graphical Models - CRF

- CRF fails to model segment-level information: syntactic constituent
- Semi-Markov CRF performs sequence labeling at segment level

$$s = < s_1, \ldots, s_n > \qquad s_i = (t_i, u_i, y_i)$$

- Identify opinion expressions (DSE/ESE)

$$p(s|x) = \frac{1}{Z(x)} \exp\left\{\sum_i \sum_k \lambda_k g_k(i, x, s)\right\} \qquad Z(x) = \sum_{s' \in S} \exp\left\{\sum_i \sum_k \lambda_k g_k(i, x, s')\right\}$$



Yang and Cardie. Extracting Opinion Expressions with semi-Markov Conditional Random Fields. EMNLP 2012

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)

$$\arg\max_{x,u,v} \lambda \sum_{i \in \mathcal{S}} \sum_{z} f_{iz} x_{iz} + (1-\lambda) \sum_{k} \sum_{i \in \mathcal{O}} \left(\sum_{j \in \mathcal{A}_k} r_{ij} u_{ij} + r_{i\emptyset} v_{ik} \right)$$

$$\sum_z x_{iz} = 1$$

$$\sum_{z \neq N} x_{iz} + \sum_{z \neq N} x_{jz} \leq 1$$

NANYANG TECHNOLOGICAL UNIVERSITY

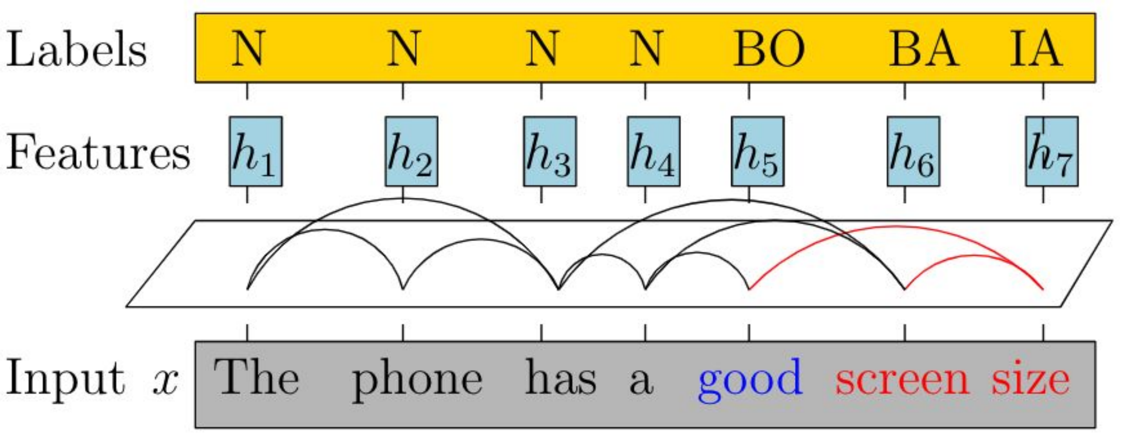
Choi et al. Joint Extraction of Entities and Relations for Opinion Recognition. EMNLP 2006
Yang and Cardie. Joint Inference for Fine-grained Opinion Extraction. ACL 2013

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

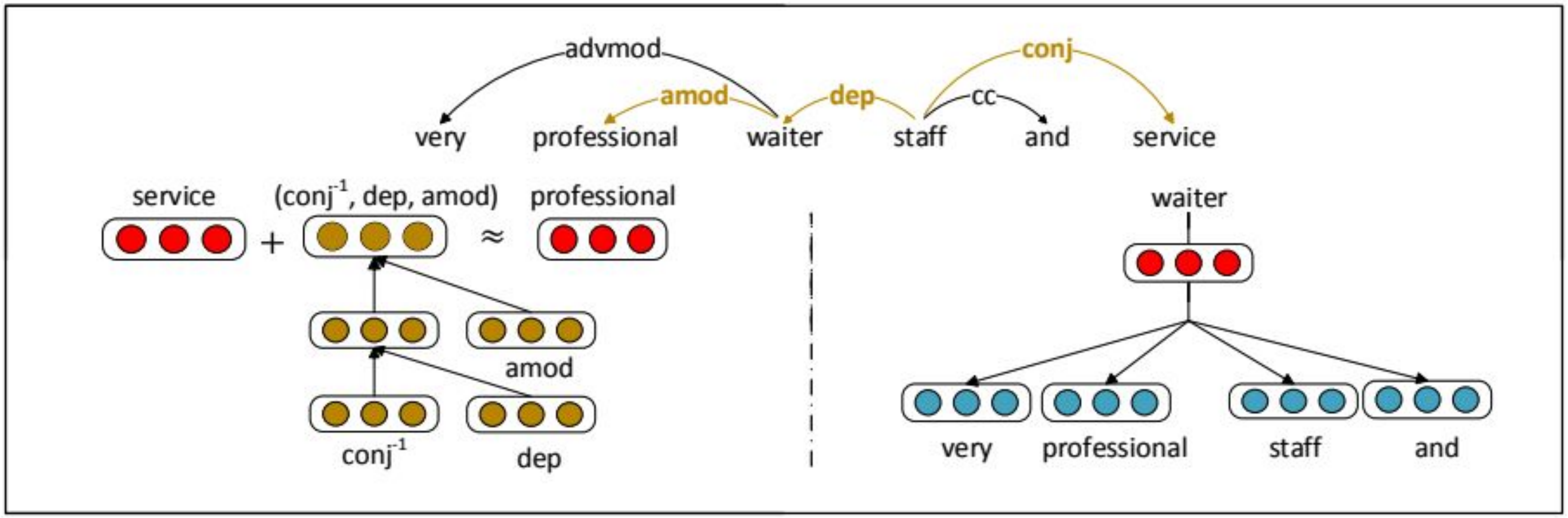
- Input: distributed word embeddings (encode semantic regularities)
- Hidden: high-level features encoding input interactions
- Output: segmentation labels



NANYANG TECHNOLOGICAL UNIVERSITY

Syntax-Encoded Embedding

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



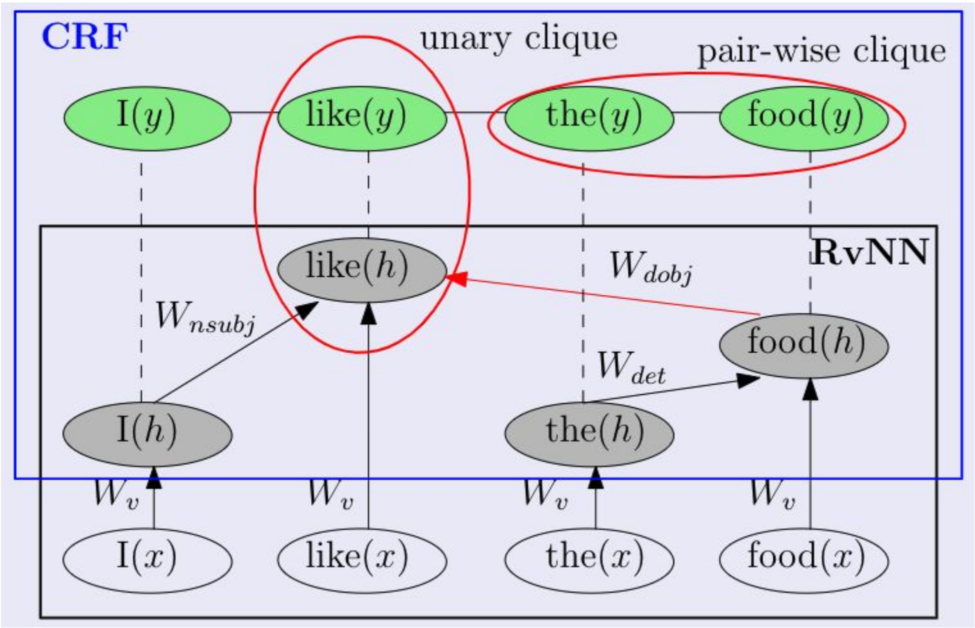
$$\sum_{(w_1,w_2,r)\in C_1} \sum_{r'\sim p(r)} max\{0, 1-(\mathbf{w}_2-\mathbf{w}_1)^\intercal \mathbf{r} + (\mathbf{w}_2-\mathbf{w}_1)^\intercal \mathbf{r}'\} \sum_{(c,w)\in C_2} \sum_{c'\sim p(w)} max\{0, 1-\mathbf{w}^\intercal \mathbf{c} + \mathbf{w}^\intercal \mathbf{c}'\}$$

NANYANG TECHNOLOGICAL UNIVERSITY

Yin et al. Unsupervised Word and Dependency Path Embeddings for Aspect Term Extraction. IJCAI 2016

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



$$\mathbf{h}_n = f(\mathbf{W}_v \cdot \mathbf{x}_w + \mathbf{b} + \sum_{k \in \mathcal{K}_n} \mathbf{W}_{r_{nk}} \cdot \mathbf{h}_k)$$

$$p(\mathbf{y}|\mathbf{h}) = \frac{1}{Z(\mathbf{h})} \prod_{c \in C} \psi_c(\mathbf{h}, \mathbf{y}_c)$$

$$\psi_c(\mathbf{h}, \mathbf{y}_c) = \exp \langle W_c, F(\mathbf{h}, \mathbf{y}_c) \rangle$$

Wang et al. Recursive Neural Conditional Random Fields for Aspect-based Sentiment Analysis. EMNLP 2016

Explicit Syntax Incorporation with ILP

- Obtain probability predictions for aspect label sequence $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ and opinion label sequence $\{\mathbf{o}_1, \ldots, \mathbf{o}_n\}$ using multi-task NN
- Adopt ILP global inference with 3 constraints. Binary prediction sequence of ILP for aspect and opinion $\{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$ $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$
  - Intra-task constraint: e.g., I-AT (I-OT) should not follow O-AT (O-OT)
  - Inter-task constraint: $\boxed{\begin{array}{c}\mathbf{q}_{\text{parent}(i)1} + \mathbf{q}_{\text{parent}(i)2} >= \mathbf{p}_{i1} + \mathbf{p}_{i2}, \\ \forall z_i \in \{\text{subj}\} \;\cap\; r_i \in \mathbf{NN} \;\cap\; r_{\text{parent}(i)} \in \mathbf{JJ}\end{array}}$
  - Lexicon constraint
- Inference

$$\max \frac{1}{n} \sum_{i=0}^{n-1} \left(\sum_{j=0}^{2} \mathbf{p}_{ij} \log \mathbf{a}_{ij} + \sum_{k=0}^{2} \mathbf{q}_{ik} \log \mathbf{o}_{ik} \right)$$

$$\text{s.t.} \quad \mathbf{p}_{ij} \in \{0, 1\}, \quad \mathbf{q}_{ik} \in \{0, 1\},$$

$$\sum_{j=0}^{2} \mathbf{p}_{ij} = 1, \quad \sum_{k=0}^{2} \mathbf{q}_{ik} = 1,$$

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

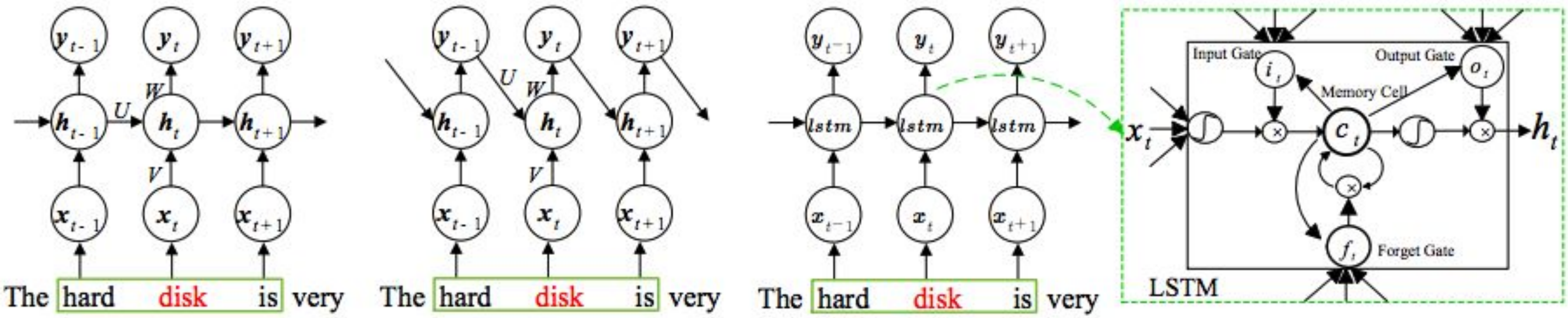
OUTLINE

- 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

NANYANG TECHNOLOGICAL UNIVERSITY

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)

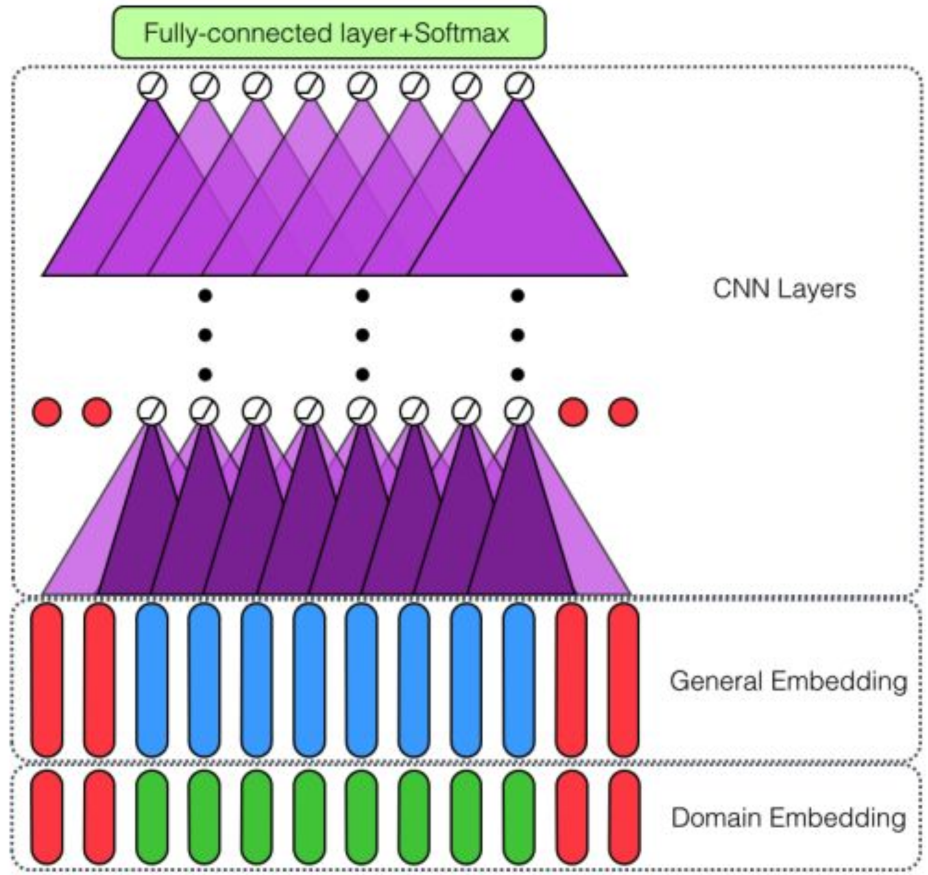


(a) Elman-type RNN (b) Jordan-type RNN (c) Long Short-Term Memory (LSTM) RNN

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

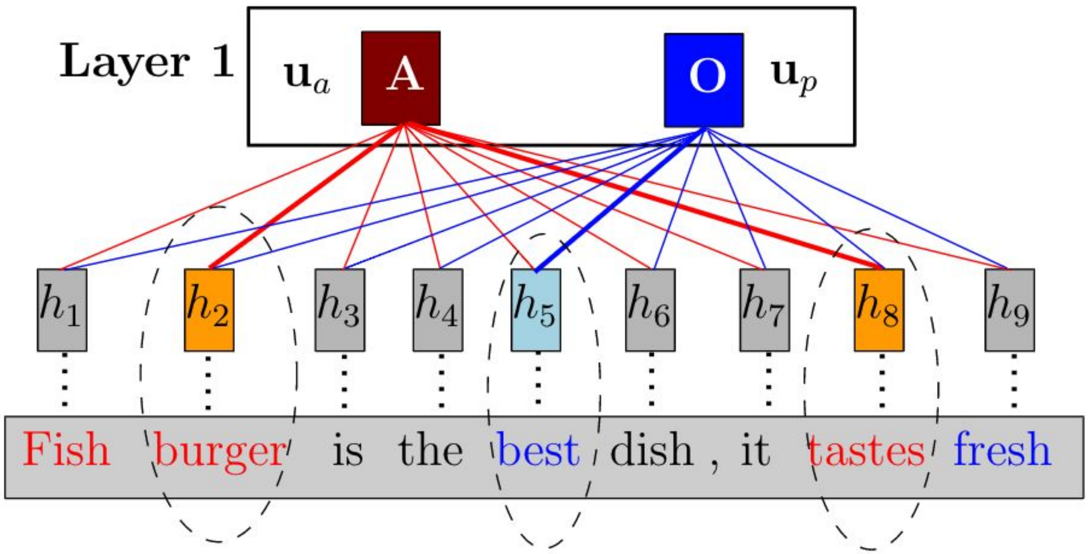
- A simple CNN with well-pretrained word embeddin
- Leverage both general embeddings and domain embeddings
    - General embedding:
        - Glove
    - Domain embedding:
        - Amazon
        - Yelp



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

Interaction-based Attention

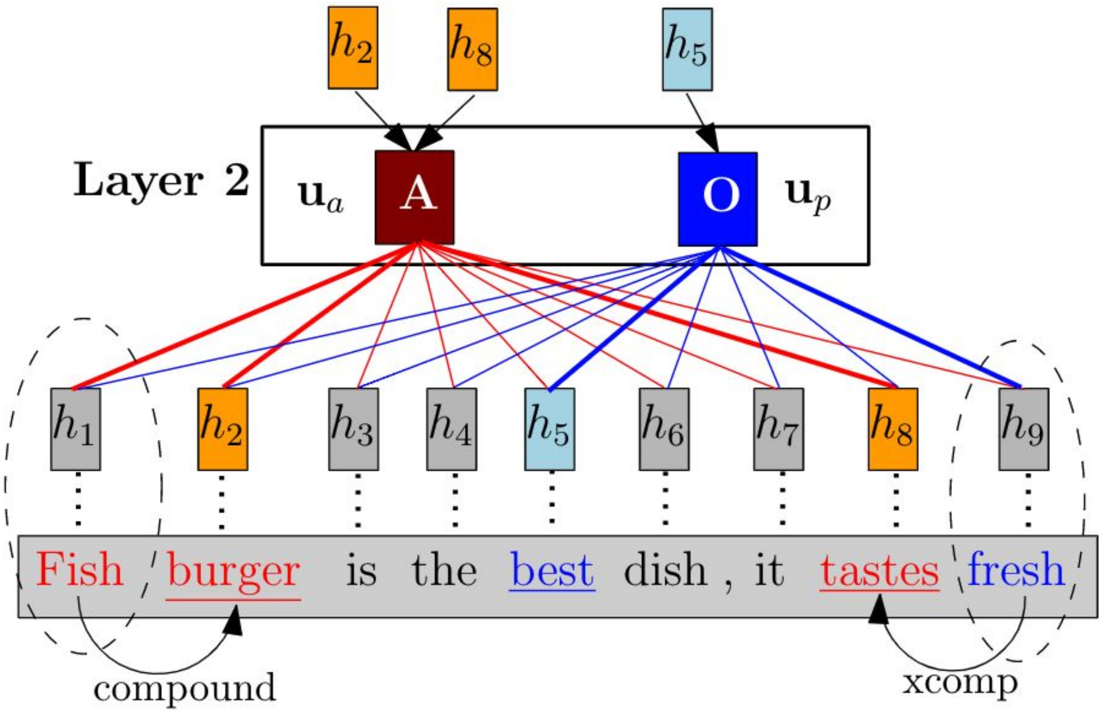
- Coupled attentions: aspect attention & opinion attention
- Memory network: multiple layers of attentions



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

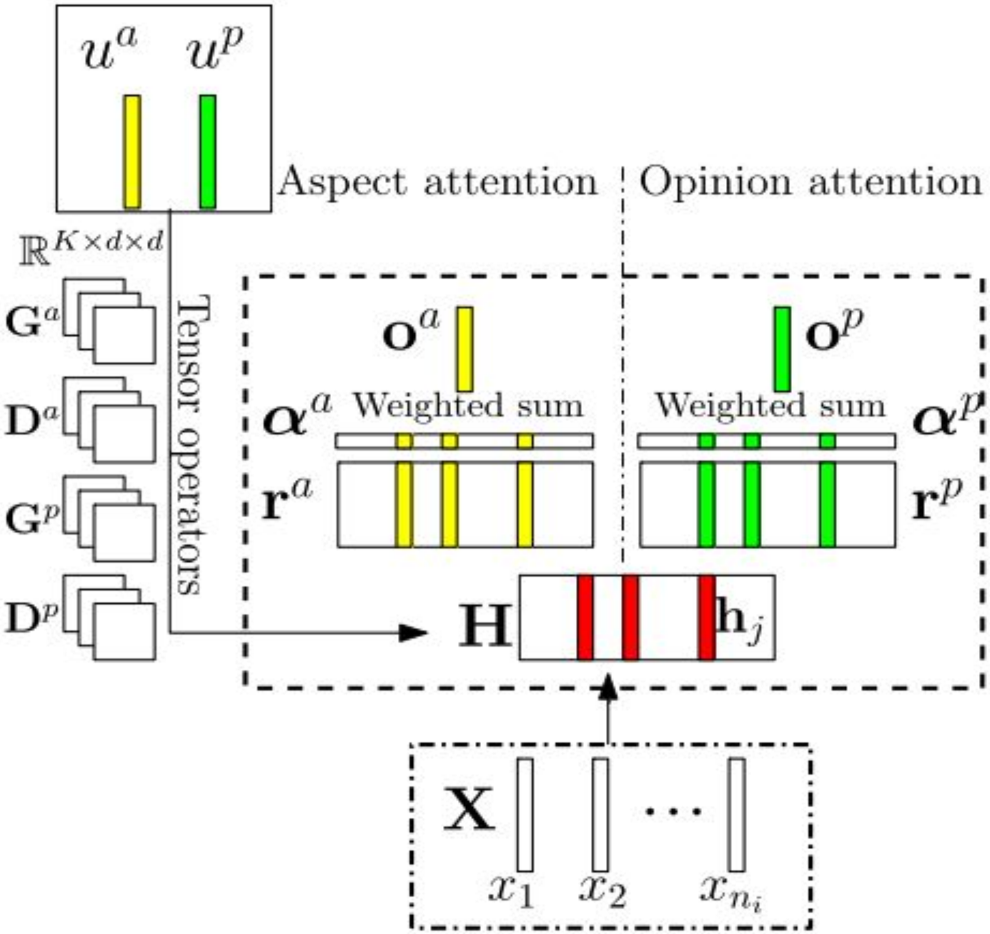
Interaction-based Attention

- 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

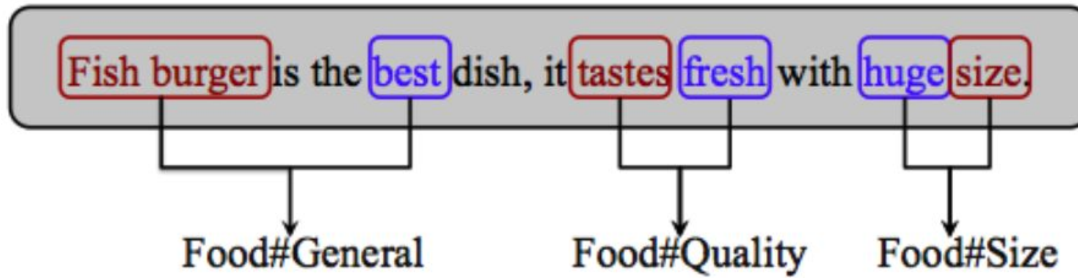


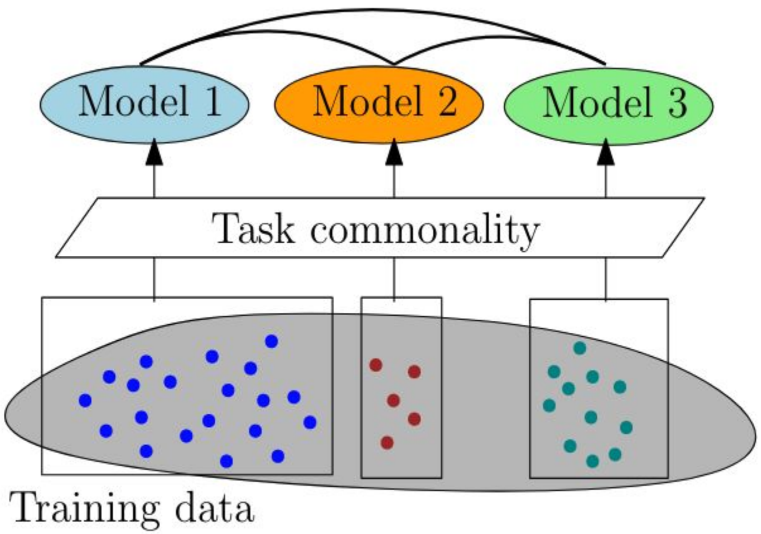
$$
\begin{aligned}
\boldsymbol{r}^a_j &= \tanh([\mathbf{h}_j^\top \mathbf{G}^a \mathbf{u}^a : \mathbf{h}_j^\top \mathbf{D}^a \mathbf{u}^p]) \\
\mathbf{u}^a_{t+1} &= \tanh(\mathbf{Q}^a \mathbf{u}^a_t) + \mathbf{o}^a_t \\
\mathbf{o}^a_t &= \sum_j \alpha^a_t \mathbf{h}_j
\end{aligned}
$$

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

Multi-task attentions for Joint Extraction

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



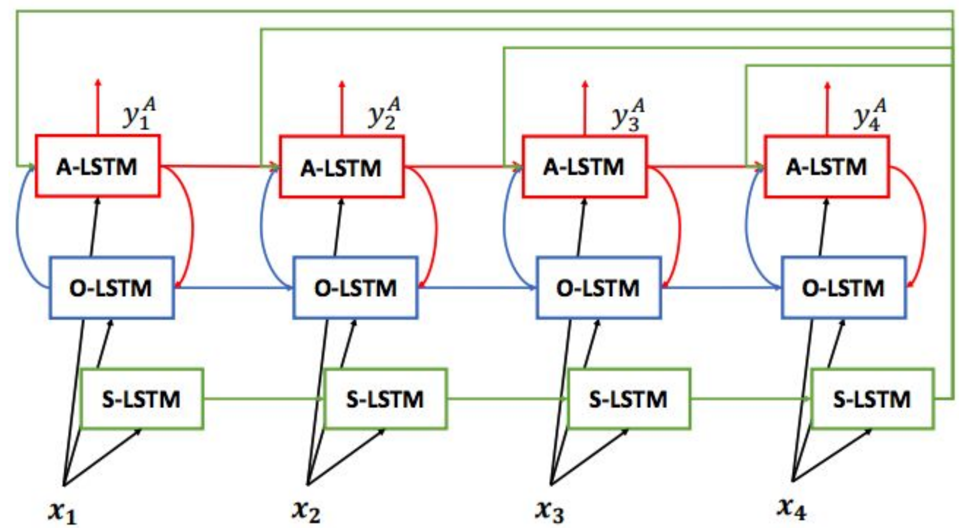


- Tensor decomposition: $\boldsymbol{G}_c = f(\mathcal{G}, \mathbf{Z}_c)$
- Context-aware multi-task feature learning: $\hat{\mathbf{r}}_c = g(\mathbf{r}_c, \mathbf{S})$
- Auxiliary task

Wang et al. Memory Networks for Fine-grained Opinion Mining. AIJ 2018

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**

NANYANG TECHNOLOGICAL UNIVERSITY

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**

NANYANG TECHNOLOGICAL UNIVERSITY

Dataset

- SemEval Challenge

Description	Training		Test	
	text	tuple	text	tuple
SemEval-15 Restaurant	1,315	1,654	685	845
SemEval-16 Restaurant	2,000	2,507	676	859
SemEval-14 Laptop	3,045	1,974	800	545
SemEval-14 Restaurant	3,041	–	800	–

- Digital Device

Data set	Number of reviews	Number of sentences
D1	45	597
D2	34	346
D3	41	546
D4	95	1716
D5	99	740
Avg	62.8	789

- MPQA

	Opinion	Target	Holder
TotalNum	5849	4676	4244

	Opinion-arg Relations	Implicit Relations
IS-ABOUT	4823	1302
IS-FROM	4662	1187

- Citysearch, BeerAdvocate

Domain	#Reviews	#Labeled sentences
Restaurant	52,574	3,400
Beer	1,586,259	9,245

NANYANG TECHNOLOGICAL UNIVERSITY

Experimental Results

- SemEval

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

NANYANG TECHNOLOGICAL UNIVERSITY

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

Experimental Results

- MPQA

	Opinion Expression			Opinion Target			Opinion Holder		
Method	P	R	F1	P	R	F1	P	R	F1
CRF	**84.42**$^{3.24}$	61.61$^{3.20}$	*71.17*$^{2.66}$	**80.38**$^{2.72}$	46.80$^{4.41}$	*59.10*$^{4.06}$	**73.37**$^{4.09}$	49.71$^{3.46}$	59.21$^{3.49}$
CRF+ILP	73.53$^{3.90}$	**74.89**$^{2.51}$	**74.11**$^{2.49}$	*77.27*$^{3.49}$	56.94$^{3.94}$	**65.40**$^{3.07}$	*67.00*$^{3.17}$	**67.22**$^{3.50}$	**67.22**$^{2.54}$
LSTM+WLL	67.88$^{4.49}$	66.13$^{3.20}$	66.87$^{2.66}$	58.71$^{4.87}$	54.92$^{3.23}$	56.50$^{1.51}$	60.33$^{4.54}$	*63.34*$^{2.33}$	61.65$^{2.37}$
LSTM+SLL	70.45$^{5.12}$	66.65$^{3.46}$	68.37$^{3.14}$	63.02$^{4.61}$	56.77$^{3.98}$	*59.65*$^{3.61}$	61.85$^{3.82}$	*63.12*$^{3.59}$	*62.35*$^{2.46}$
LSTM+SLL+RLL	71.73$^{5.35}$	*70.92*$^{3.96}$	*71.11*$^{2.71}$	64.52$^{5.52}$	**65.94**$^{4.74}$	*64.84*$^{1.44}$	62.75$^{3.75}$	*67.17*$^{4.37}$	*64.71*$^{2.23}$
CRF	**80.78**$^{3.27}$	57.62$^{3.24}$	*67.19*$^{2.63}$	**71.81**$^{3.22}$	42.36$^{3.78}$	*53.23*$^{3.69}$	**71.56**$^{3.54}$	48.61$^{3.51}$	57.86$^{3.43}$
CRF+ILP	71.03$^{4.03}$	**69.72**$^{2.37}$	**70.22**$^{2.44}$	**71.94**$^{3.25}$	49.83$^{3.24}$	**58.72**$^{2.80}$	*65.70*$^{3.07}$	**65.91**$^{3.63}$	**65.68**$^{2.61}$
LSTM+WLL	64.47$^{4.79}$	59.45$^{3.52}$	61.67$^{2.26}$	52.72$^{5.01}$	44.21$^{2.54}$	47.85$^{1.41}$	58.41$^{4.72}$	59.72$^{2.52}$	52.45$^{2.23}$
LSTM+SLL	65.97$^{5.46}$	61.76$^{3.69}$	63.60$^{3.05}$	54.46$^{4.49}$	50.16$^{4.38}$	52.01$^{3.05}$	59.80$^{3.29}$	*61.27*$^{3.75}$	60.40$^{2.26}$
LSTM+SLL+RLL	65.48$^{4.92}$	*65.54*$^{3.65}$	*65.56*$^{2.71}$	52.75$^{6.81}$	**60.54**$^{4.78}$	*55.81*$^{1.96}$	59.44$^{3.56}$	*65.51*$^{4.22}$	*62.18*$^{2.50}$

	IS-ABOUT			IS-FROM		
Method	P	R	F1	P	R	F1
CRF+ILP	*61.57*$^{4.56}$	*47.65*$^{3.12}$	*54.39*$^{2.49}$	*64.04*$^{3.08}$	**58.79**$^{4.42}$	**61.17**$^{3.02}$
LSTM+SLL+Softmax	36.23$^{5.10}$	36.12$^{7.75}$	35.40$^{3.35}$	36.44$^{5.26}$	40.19$^{6.13}$	37.60$^{3.42}$
LSTM+SLL+RLL	**62.48**$^{3.87}$	**49.80**$^{2.84}$	**54.98**$^{2.54}$	**64.19**$^{3.81}$	*53.75*$^{6.00}$	*58.22*$^{3.01}$

Katiyar and Cardie. Investigating LSTMs for Joint Extraction of Opinion Entities and Relations. ACL 2016

Experimental Results

- Citysearch, BeerAdvocate

Aspect	Method	Precision	Recall	F_1
Food	LocLDA	0.898	0.648	0.753
	ME-LDA	0.874	0.787	0.828
	SAS	0.867	0.772	0.817
	BTM	0.933	0.745	0.816
	SERBM	0.891	**0.854**	**0.872**
	k-means[3]	0.931	0.647	0.755
	ABAE	**0.953**	0.741	0.828
Staff	LocLDA	0.804	0.585	0.677
	ME-LDA	0.779	0.540	0.638
	SAS	0.774	0.556	0.647
	BTM	**0.828**	0.579	0.677
	SERBM	0.819	0.582	0.680
	k-means	0.789	0.685	0.659
	ABAE	0.802	**0.728**	**0.757**
Ambience	LocLDA	0.603	0.677	0.638
	ME-LDA	0.773	0.558	0.648
	SAS	0.780	0.542	0.640
	BTM	0.813	0.599	0.685
	SERBM	0.805	0.592	0.682
	k-means	0.730	0.637	0.677
	ABAE	**0.815**	**0.698**	**0.740**

Aspect	Method	Precision	Recall	F_1
Feel	k-means	0.720	0.815	0.737
	LocLDA	**0.938**	0.537	0.675
	SAS	0.783	0.695	0.730
	BTM	0.892	0.687	0.772
	ABAE	0.815	**0.824**	**0.816**
Taste	k-means	0.533	0.413	0.456
	LocLDA	0.399	**0.655**	0.487
	SAS	0.543	0.496	0.505
	BTM	0.616	0.467	**0.527**
	ABAE	**0.637**	0.358	0.456
Smell	k-means	**0.844**	0.295	0.422
	LocLDA	0.560	0.488	0.489
	SAS	0.336	0.673	0.404
	BTM	0.541	0.549	0.527
	ABAE	0.483	**0.744**	**0.575**
Taste+Smell	k-means	0.697	0.828	0.740
	LocLDA	0.651	**0.873**	0.735
	SAS	0.804	0.759	0.769
	BTM	0.885	0.760	0.815
	ABAE	**0.897**	0.853	**0.866**

Opinion Summarization

NANYANG TECHNOLOGICAL UNIVERSITY

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 — Feature
      - Positive: 12 — Support
        Sentiment
        - 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.

        ...

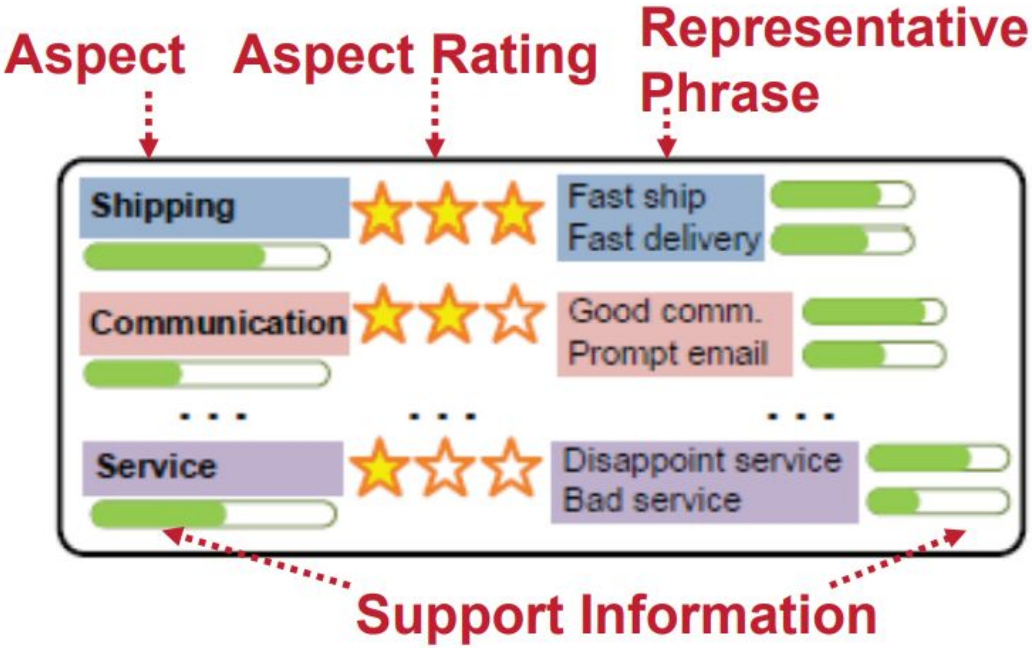
        Raw review Sentences
      - Negative: 2
        - The pictures come out hazy if your hands shake even for a moment during the entire process of taking a picture.
        - Focusing on a display rack about 20 feet away in a brightly lit room during day time, pictures produced by this camera were blurry and in a shade of orange.

        ...

Kim et al. Comprehensive Review of Opinion Summarization.

Different Forms of Summaries

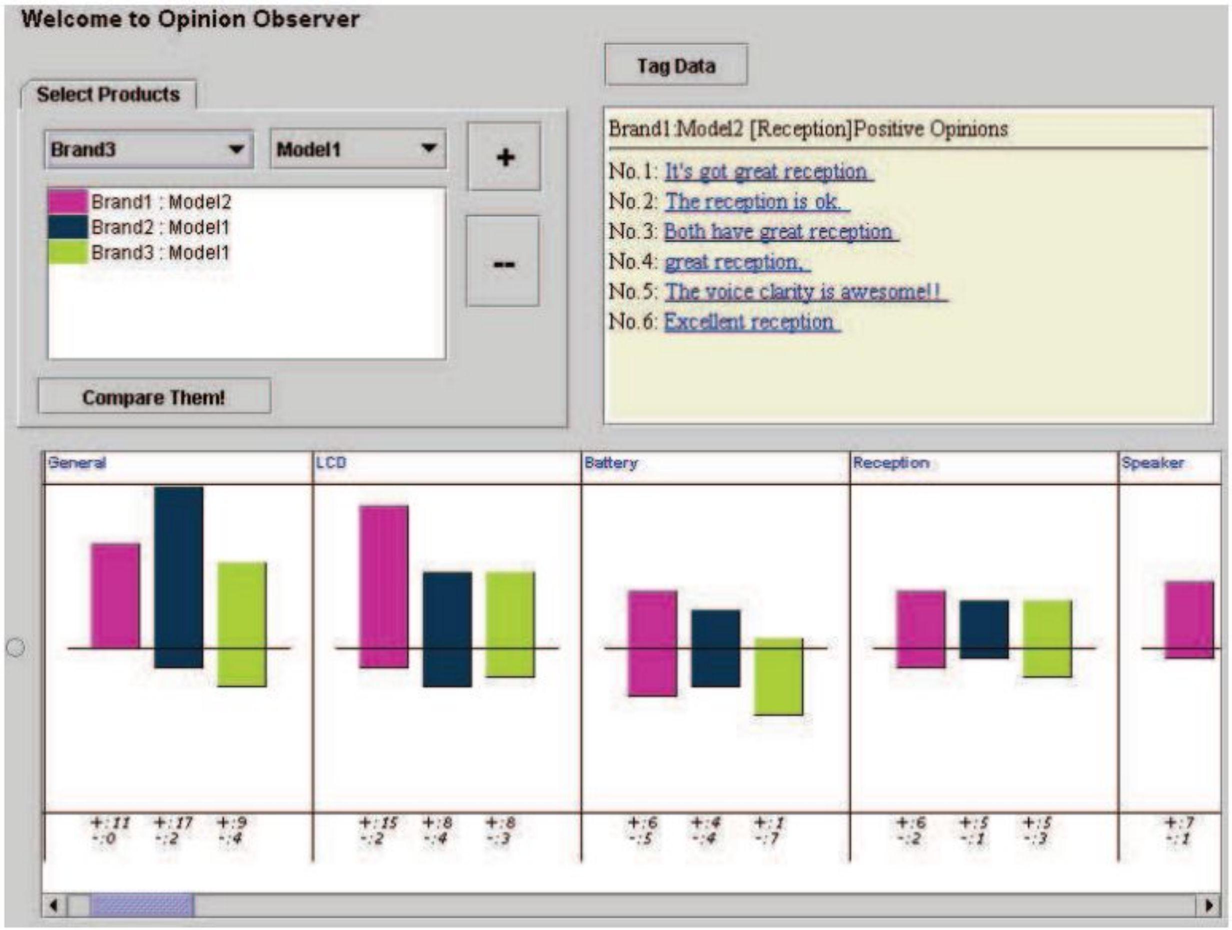
- Structured summary



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

Different Forms of Summaries

- Visualized summary



Liu et al. Opinion Observer: Analyzing and Comparing Opinions on the Web. WWW 2005

NANYANG TECHNOLOGICAL UNIVERSITY

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

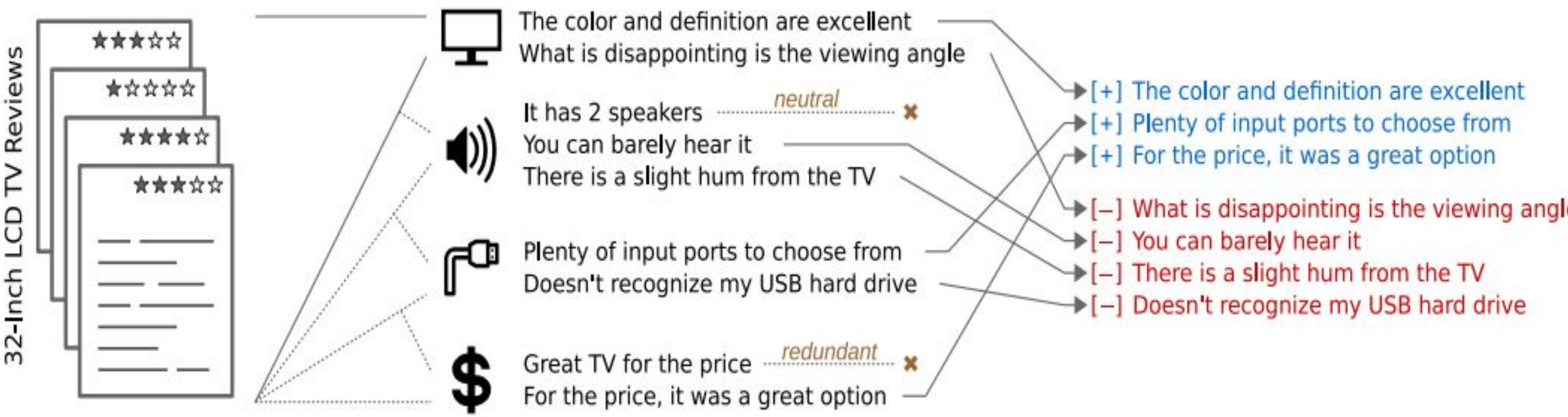
Negative: annoying, noise, too long, … , unnecessary

… …

Li et al. Structure-Aware Review Mining and Summarization. COLING 2010

Pipelined Approach

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



NANYANG TECHNOLOGICAL UNIVERSITY

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)

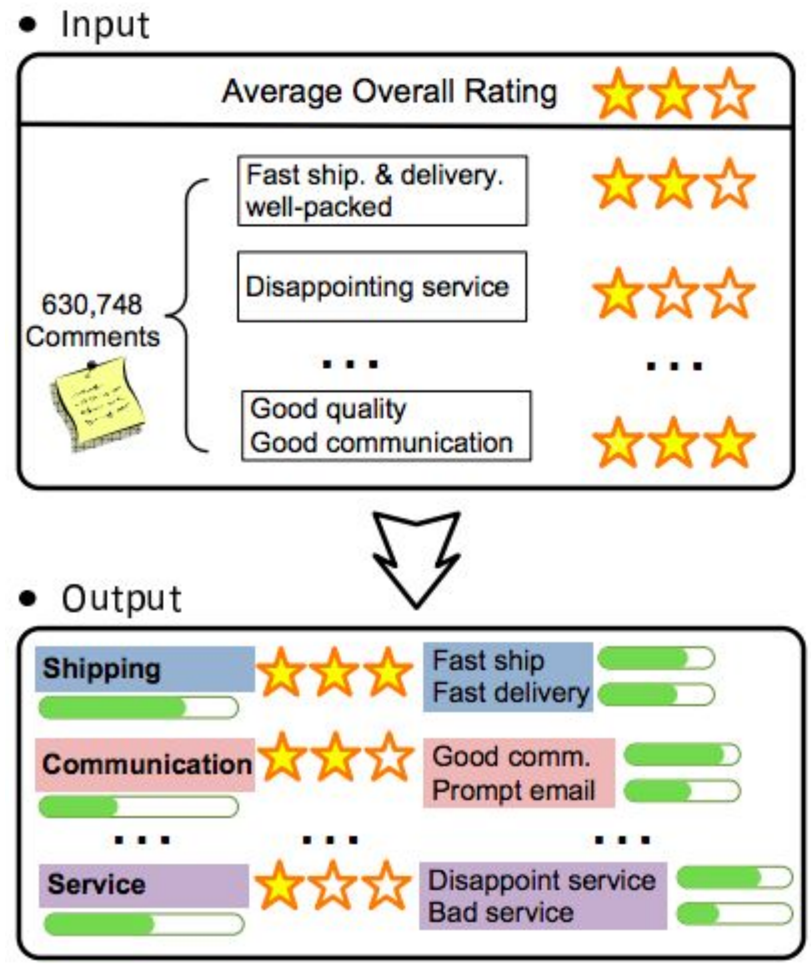
$$f = (w_m, w_h)$$

- Predict rating for each aspect from the overall ratings
  - Global $r(f \in t) = r(t)$
  - Local

$$p(w_m|A_i, r) = \frac{c(w_m, S(A_i, r))}{\sum_{w_{m'} \in Vm} c(w_{m'}, S(A_i, r))}$$

$$r(f) = argmax_r\{p(w_m|A_i, r)|A(f) = i\}$$

- Extract representative phrases



$$(A_i, R(A_i), RF(A_i))_{i=1}^{k}$$

NANYANG TECHNOLOGICAL UNIVERSITY

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

OUTLINE

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

NANYANG TECHNOLOGICAL UNIVERSITY

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 $\psi(s_1, s_2)$

$$r(S) = \frac{1}{|X|} \sum_{x \in X} \max_{i \in [1,k]} \phi(x, u_i) + \frac{1}{|Y|} \sum_{y \in Y} \max_{i \in [1,k]} \phi(y, v_i)$$

$$c(S) = \frac{1}{k} \sum_{i=1}^{k} \psi(u_i, v_i)$$

$$S^* = \arg\max_S \left(\lambda\, r(S) + (1-\lambda)\, c(S)\right)$$

Contradictory Aspect	**Positive**	**Negative**
Contradictory 1	u_1	v_1
Contradictory 2	u_2	v_2
...	...	...
Contradictory k	u_k	v_k

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

$$P(x_j|x_i, z) = \frac{sim_z(x_i, x_j)}{\sum_{j' \in X} sim_z(x_i, x_{j'})}$$

$$sim_0(x_i, x_j) = sim(x_i, x_j) \sum_{m=1}^{k} P(v = m|x_i) P(v = m|x_j)$$

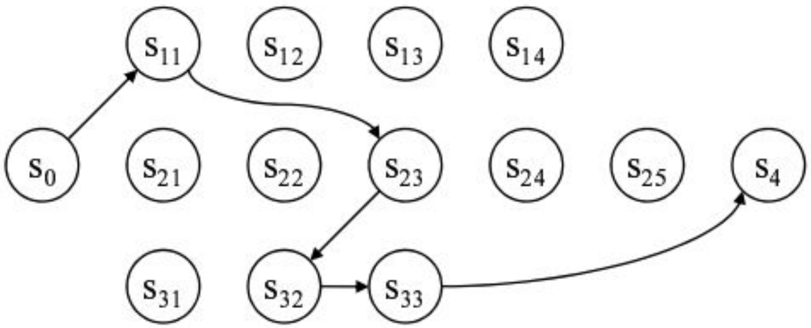
$$sim_1(x_i, x_j) = sim(x_i, x_j) \times \sum_{m_1, m_2 \in [1,k], m_1 \neq m_2} P(v = m_1|x_i) P(v = m_2|x_j)$$

$$P(x_i) P(x_j|x_i, z = 1) + P(x_j) P(x_i|x_j, z = 1)$$

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) $c_{i,j} = \mathbf{w} \cdot \phi(x, y)$



- Decoding as Integer Linear Programming (fix w and c)

$$\max \left\{ \lambda \sum_{e_i \in E} w_i e_i + (1-\lambda) \sum_{a_{i,j} \in A} c_{i,j} a_{i,j} \right\}$$

$$\sum_i a_{i,j} + \sum_i a_{j,i} = 2 s_j \ \forall j$$

$$\sum_i a_{i,j} = \sum_i a_{j,i} \ \forall j$$

$$s_i, a_{i,j}, e_i \in \{0, 1\} \ \forall i, j$$

$$\sum_i m_{ij} s_i \geq e_j \ \forall j$$

NANYANG TECHNOLOGICAL UNIVERSITY

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

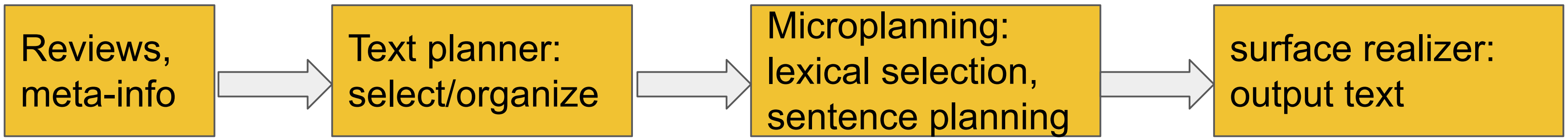
Extractive: Bottom line, well made camera, easy to use, very flexible and powerful features to include the ability to use external flash and lense/filters choices. It has a beautiful design, lots of features, very easy to use, very configurable and customizable, and the battery duration is amazing! Great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



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”, ...]

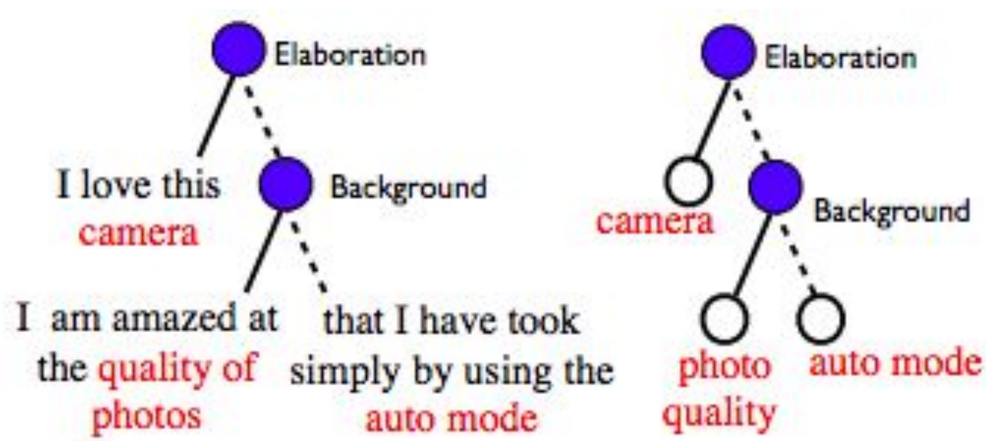
Connectives

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

Carenini and Moore. Generating and evaluating evaluative arguments. AIJ 2006
Gerani et al. Abstractive Summarization of Product Reviews Using Discourse Structure. EMNLP 2014
Di Fabbrizio et al. A hybrid approach to multidocument summarization of opinions in reviews. INLG 2014

Discourse Structure

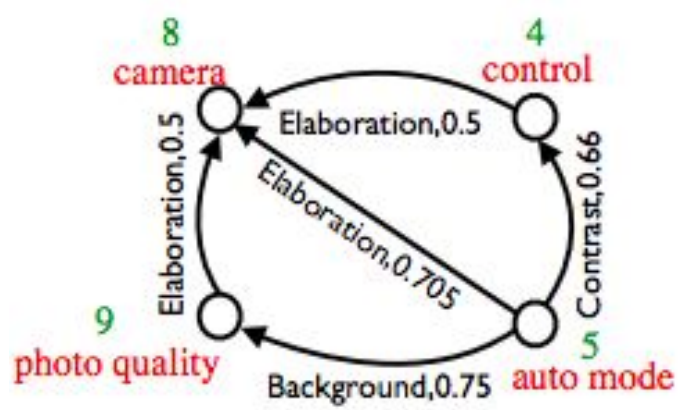
- An Aspect Hierarchy Tree (AHT) that reflects the importance of aspects as well as the relationships between them
- Assume aspect and opinions are given (or preprocessing)
- Process
    - Obtain a discourse tree representation for every review and modify the tree such that every leaf node only contains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



Gerani et al. Abstractive Summarization of Product Reviews Using Discourse Structure. EMNLP 2014

Discourse Structure

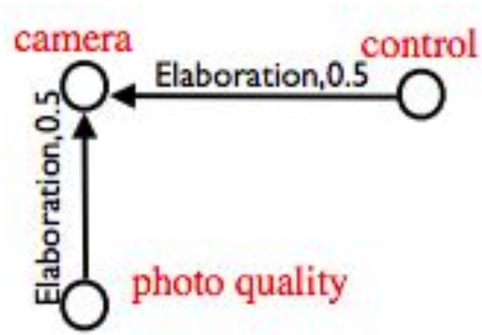
- An Aspect Hierarchy Tree (AHT) that reflects the importance of aspects as well as the relationships between them
- Assume aspect and opinions are given (or preprocessing)
- Process
  - Aggregate the trees and generate an Aggregated Rhetorical Relation Graph (ARRG)



Gerani et al. Abstractive Summarization of Product Reviews Using Discourse Structure. EMNLP 2014

Discourse Structure

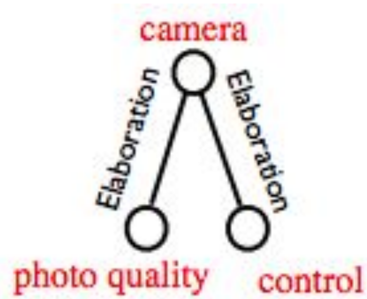
- An Aspect Hierarchy Tree (AHT) that reflects the importance of aspects as well as the relationships between them
- Assume aspect and opinions are given (or preprocessing)
- Process
  - Use Weighted PageRank to select a subgraph representing the most important aspects (extract aspects that not only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



Gerani et al. Abstractive Summarization of Product Reviews Using Discourse Structure. EMNLP 2014

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



Gerani et al. Abstractive Summarization of Product Reviews Using Discourse Structure. EMNLP 2014

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 a modified mutual information function)
  - readability (based on an n-gram language model)
- Formulated as an optimization problem

Mp3 Player Y	Restaurant X
Very short battery life. Big and clear screen. **(8 words)**	Good service. Delicious soup dishes. Very noisy at nights. **(9 words)**

$$M^* = \arg\max_{M=\{m_1,\ldots,m_k\}} \sum_{i=1}^{k} [S_{rep}(m_i) + S_{read}(m_i)]$$

$$\begin{aligned} \textit{subject to} \quad & \textstyle\sum_{i=1}^{k} |m_i| \leq \sigma_{ss} \\ & S_{rep}(M) \geq \sigma_{rep} \\ & S_{read}(M) \geq \sigma_{read} \\ & sim(m_i, m_j) \leq \sigma_{sim} \forall, i, j \in [1, k] \end{aligned}$$

$$S_{rep}(w_1 \ldots w_n) = \frac{1}{n} \sum_{i=1}^{n} pmi_{local}(w_i)$$

$$pmi_{local}(w_i) = \frac{1}{2C} \sum_{j=i-C}^{i+C} pmi'(w_i, w_j), i \neq j$$

NANYANG TECHNOLOGICAL UNIVERSITY

Ganesan et al. Micropinion Generation: An Unsupervised Approach to Generating Ultra-Concise Summaries of Opinions. WWW 2012

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.

NANYANG TECHNOLOGICAL UNIVERSITY

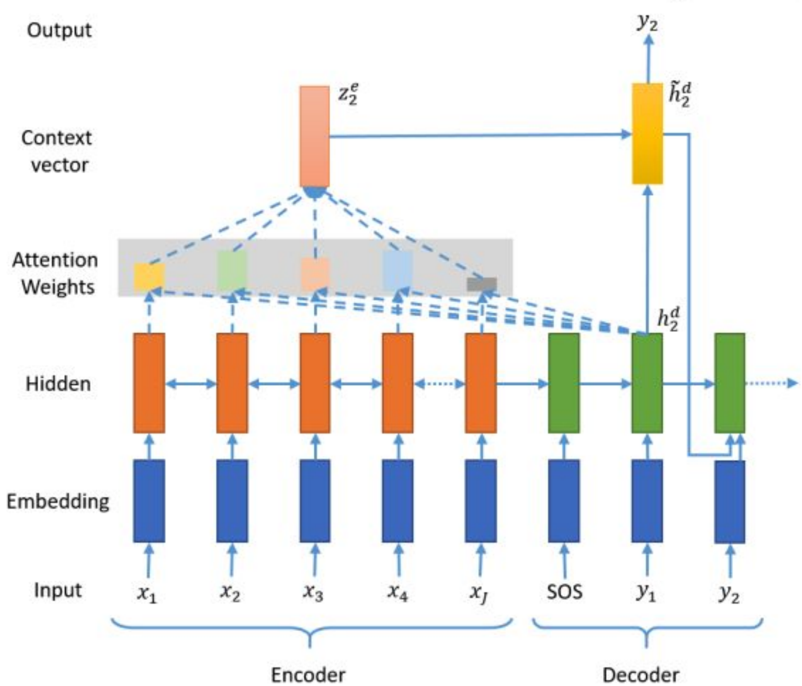
Wang and Ling. Neural Network-Based Abstract Generation for Opinions and Arguments. NAACL 2016

Deep-Learning-Based Generation (Attention)

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

$$\log P(y|x) = \sum_{j=1,\ldots,|y|} \log P(y_j|y_1,\ldots,y_{j-1},x)$$

$$p(y_j|y_1,\ldots,y_{j-1},x) = softmax(\mathbf{h}_j) \qquad \mathbf{h}_j = g(\mathbf{y}_{j-1},\mathbf{h}_{j-1},\mathbf{s})$$



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)

$$\log P(y|x) = \sum_{j=1,\dots,|y|} \log P(y_j|y_1,\dots,y_{j-1},x)$$

$$p(y_j|y_1,\dots,y_{j-1},x) = softmax(\mathbf{h}_j) \qquad \mathbf{h}_j = g(\mathbf{y}_{j-1},\mathbf{h}_{j-1},\mathbf{s})$$

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

Feature vector

Gold label

$$J(\mathbf{w}) = ||\tilde{\mathbf{R}}\mathbf{w} - \tilde{\mathbf{L}}||_2^2 + \lambda \cdot \underline{||\tilde{\mathbf{R}}'\mathbf{w} - \tilde{\mathbf{L}}'||_2^2} + \beta \cdot ||\mathbf{w}||_2^2$$

Enforce text separation

$$\hat{\mathbf{w}} = (\tilde{\mathbf{R}}^{\mathbf{T}}\tilde{\mathbf{R}} + \tilde{\mathbf{R}}'^{\mathbf{T}}\tilde{\lambda}\tilde{\mathbf{R}}' + \tilde{\beta})^{-1}(\tilde{\mathbf{R}}^{\mathbf{T}}\tilde{\mathbf{L}} + \tilde{\mathbf{R}}'^{\mathbf{T}}\tilde{\lambda}\tilde{\mathbf{L}}')$$

$$f(x^k) = \mathbf{r_k} \cdot \mathbf{w}$$

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

$$\mathbf{H}^c = [\mathbf{h}_1^c, \ldots, \mathbf{h}_k^c]$$

$$\mathbf{v}^c = \sum_{i=1}^{k} \mathbf{h}_i^c / k$$

- Mutual attention

$$emb_x^c = \sum_{i=1}^{k} C_i \mathbf{h}_i^c$$

Sentiment-related

- Hidden states

$$\mathbf{H}^s = [\mathbf{h}_1^s, \ldots, \mathbf{h}_m^s]$$

$$\mathbf{v}^s = \sum_{i=1}^{m} \mathbf{h}_i^s / m$$

- Mutual attention

$$emb_x^s = \sum_{i=1}^{m} S_i \mathbf{h}_i^s$$

Aspect-related

- Hidden states

$$\mathbf{H}^t = [\mathbf{h}_1^t, \ldots, \mathbf{h}_n^t]$$

$$\mathbf{v}^t = \sum_{i=1}^{n} \mathbf{h}_i^t / n$$

- Mutual attention

$$emb_x^t = \sum_{i=1}^{n} T_i \mathbf{h}_i^t$$

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

$$emb_x = [emb_x^c, emb_x^s, emb_x^t]$$

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

- Generation with 3 attentions and pointer-generator framework

$$\mathbf{s}_0 = cemb_x = tanh(\mathbf{W}_\mu \times emb_x)$$

$$\mathbf{s}_t = LSTM(\mathbf{s}_{t-1}, \mathbf{c}_t, \mathbf{w}_{t-1})$$

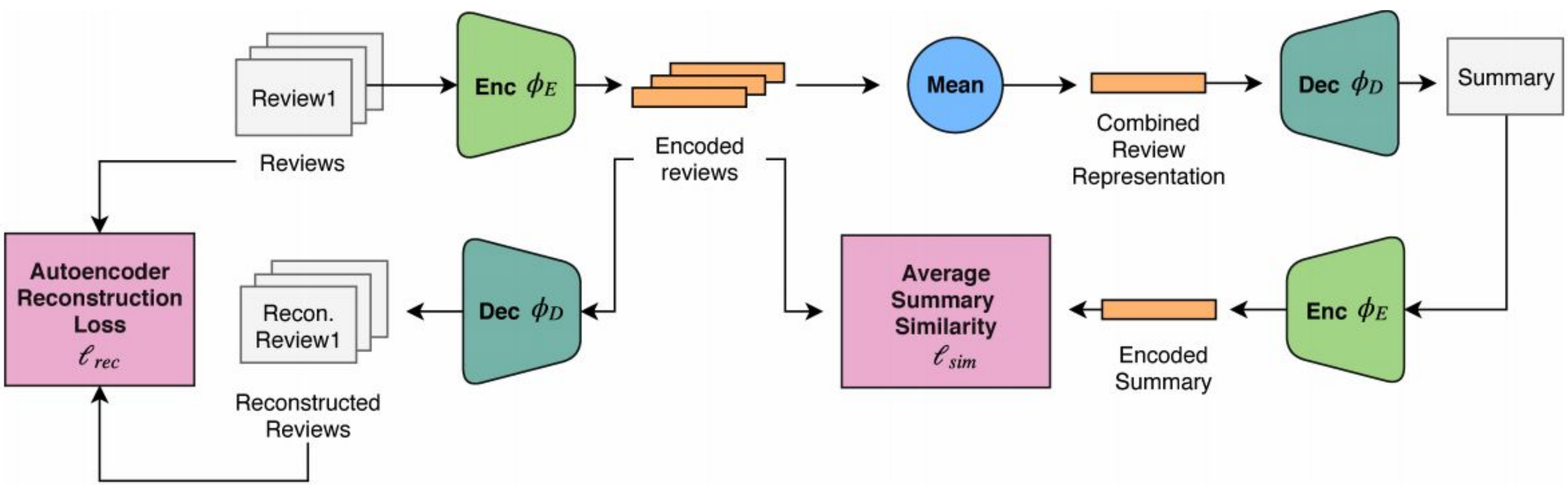
$$a_{t,i} = \text{softmax}(\lambda_1 a_{t,i}^{semantic} + \lambda_2 a_{t,i}^{sentiment} + \lambda_3 a_{t,i}^{aspect}),\quad c_t = \sum_{i=1}^{k} a_{t,i} h_i^c$$

$$P(w_t) = p_{gen} P_{vocab}(w_t) + (1 - p_{gen}) \sum_{i:w_i = w_t} a_{t,i}$$

Yang et al. Aspect and Sentiment Aware Abstractive Review Summarization. COLING 2018

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

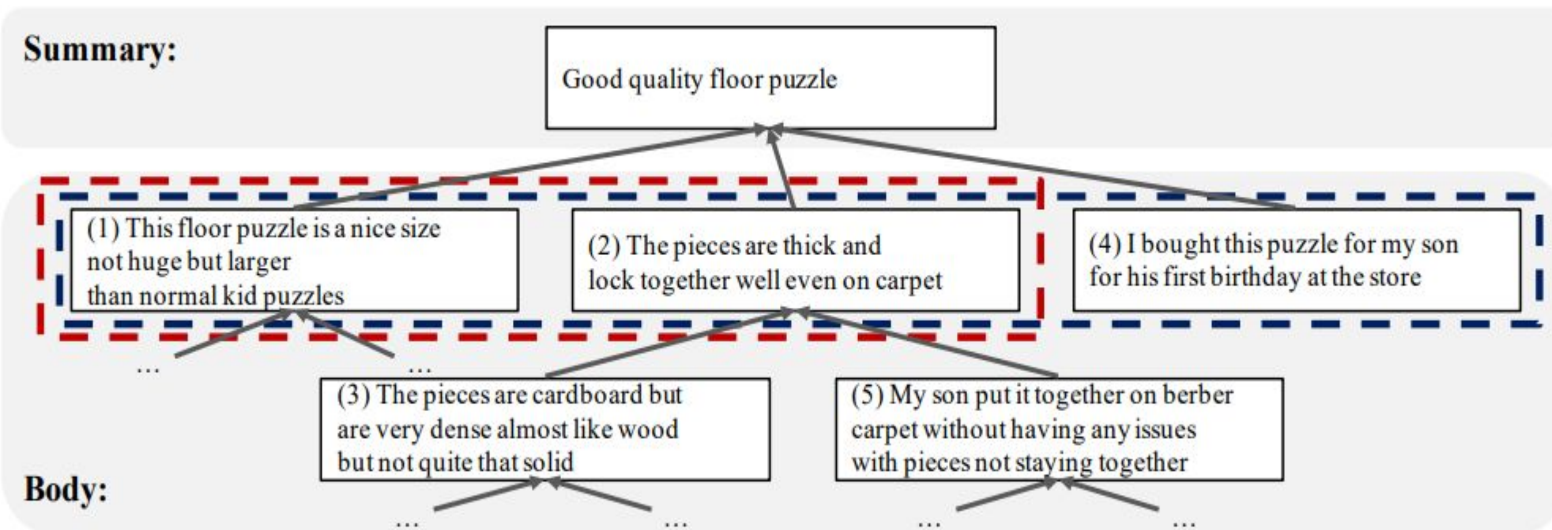


NANYANG TECHNOLOGICAL UNIVERSITY

Chu and Liu. MeanSum : A Neural Model for Unsupervised Multi-Document Abstractive Summarization. ICML 2019

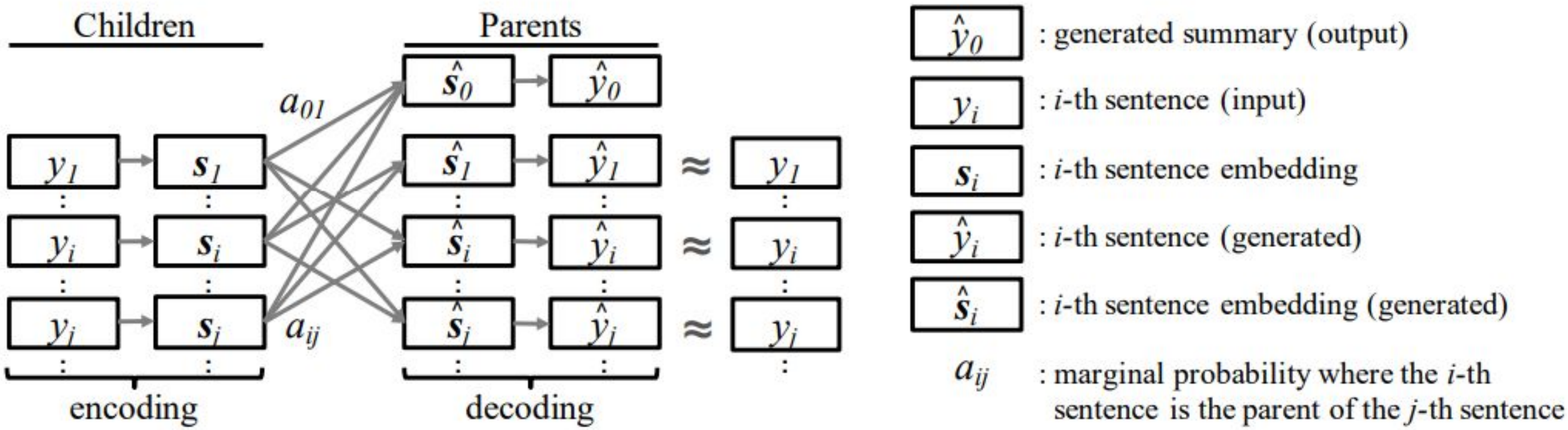
Deep-Learning-Based Generation (Structural)

- 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



Deep-Learning-Based Generation (Structural)

- Learn both semantics and structure

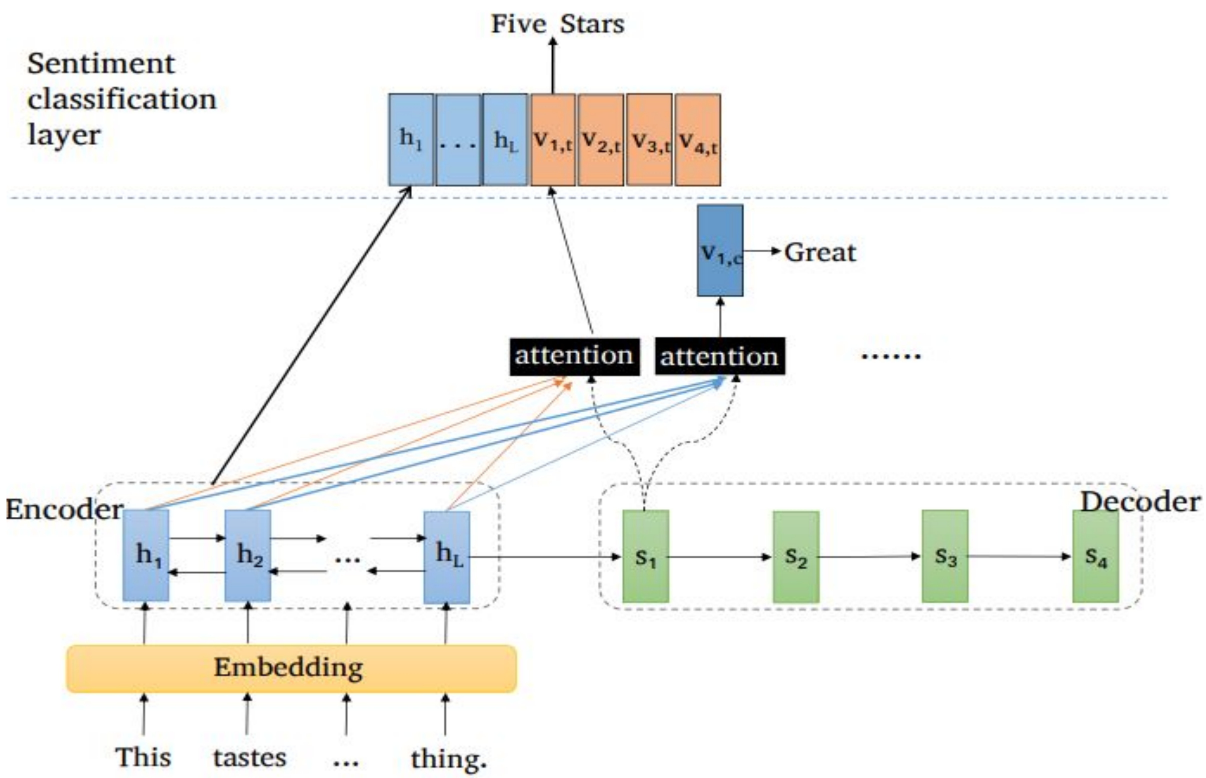


- Separate sentence embedding into 2 parts $\mathbf{s}_i = [\mathbf{s}_i^e, \mathbf{s}_i^f]$
- Encoding: $\hat{\mathbf{s}}_i = tanh\{\mathbf{W}_s(\sum_{j=1}^{n} a_{ij}\mathbf{s}_j^e) + \mathbf{b}_s\}$
- Decoding: $\sum_{i=1}^{n}\sum_{t=1}^{l} \log P(w_i^t | w_i^{<t}, \hat{\mathbf{s}}_i, \theta)$

Isonuma1 et al. Unsupervised Neural Single-Document Summarization of Reviews via Learning Latent Discourse Structure and its Ranking. ACL 2019

Combine Summarization with Sentiment Prediction

- Jointly performs abstractive text summarization and sentiment classification within a hierarchical end-to-end neural framework

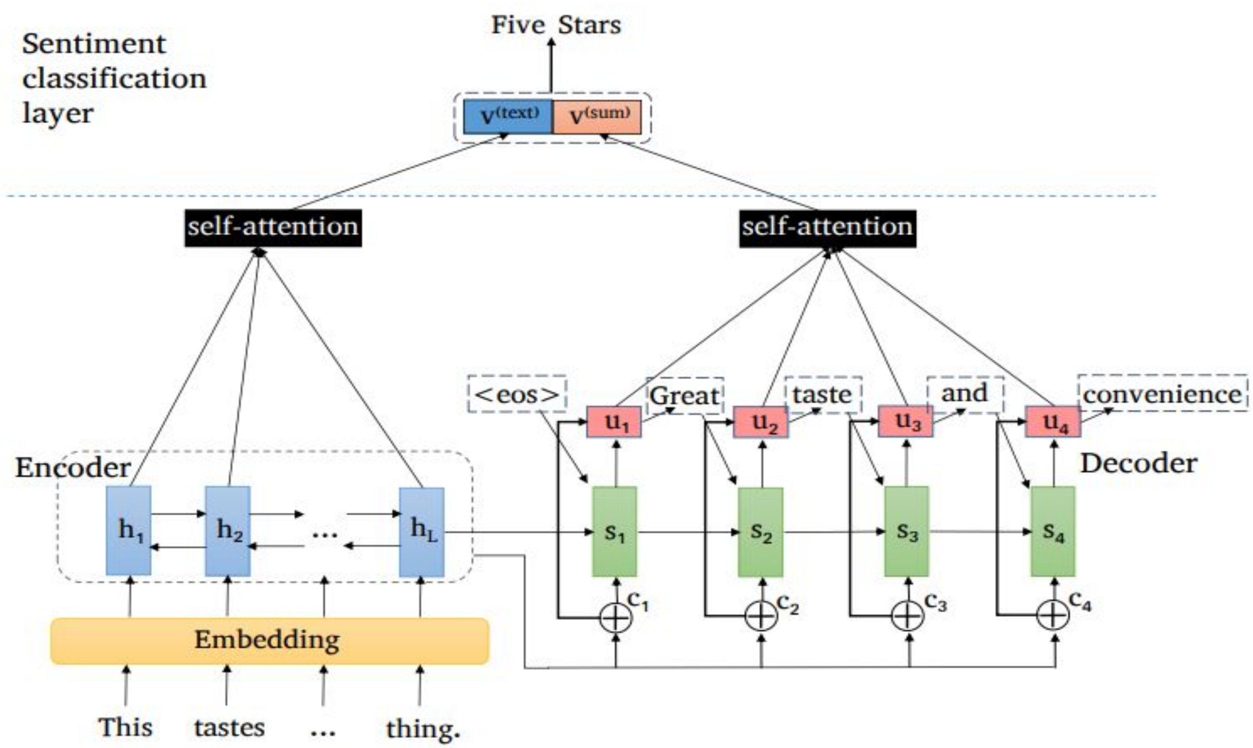


NANYANG TECHNOLOGICAL UNIVERSITY

Ma et al. A Hierarchical End-to-End Model for Jointly Improving Text Summarization and Sentiment Classification. IJCAI 2018

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

Domains	Train	Valid	Eval
Toys & Games	27,037	498	512
Sports & Outdoors	37,445	511	466
Movies & TV	408,827	564	512

NANYANG TECHNOLOGICAL UNIVERSITY

Experimental Results

- Amazon product review

Domain	Toys & Games			Sports & Outdoors			Movies & TV		
Metric	R-1	R-2	R-L	R-1	R-2	R-L	R-1	R-2	R-L
Unsupervised approaches									
TextRank	8.63	1.24	7.26	7.16	0.89	6.39	**8.27**	**1.44**	7.35
Opinosis	8.25	1.51	7.52	7.04	1.42	6.45	7.80	1.20	7.11
MeanSum-single	8.12	0.58	7.30	5.42	0.47	4.97	6.96	0.35	6.08
StrSum	11.61	1.56	11.04	9.15	1.38	8.79	7.38	1.03	6.94
StrSum+DiscourseRank	**11.87**	**1.63**	**11.40**	**9.62**	**1.58**	**9.28**	8.15	1.33	**7.62**
Supervised baselines									
Seq-Seq	13.50	2.10	13.31	10.69	2.02	10.61	7.71	2.18	7.08
Seq-Seq-att	16.28	3.13	16.13	11.49	2.39	11.47	9.05	2.99	8.46

NANYANG TECHNOLOGICAL UNIVERSITY

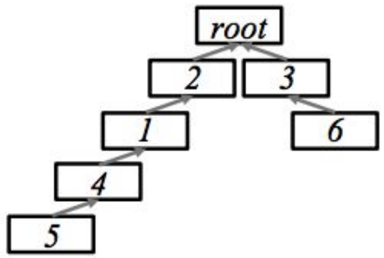
Isonuma1 et al. Unsupervised Neural Single-Document Summarization of Reviews via Learning Latent Discourse Structure and its Ranking. ACL 2019

Experimental Results

Generated Summary | Induced Discourse Tree | Sentences in the Main Body

(a)

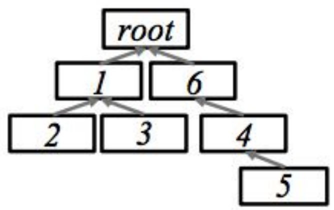
- Reference:
  love this game
- Seq-Seq-att:
  fun game
- **Our Model (Full)**:
  i love this game



1. I love this game
2. It is so much fun
3. I'm all about new and different games
4. I love to play this with my brother because he is very bad at keeping score so I win most of the time and he loves to tell each characters story
5. And he loves to tell each characters story and to tell why each person got what fate
6. It's a must buy if you want a fun and fast card game

(b)

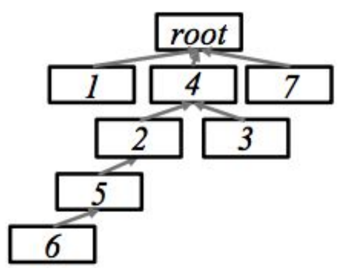
- Reference:
  good value
- Seq-Seq-att:
  good for the price
- **Our Model (Full)** :
  this is a great product for the price



1. have not used it yet at the campground but tested it at home and works fine
2. use a toothpick to hold the valve open so you can deflate it easily
3. if you sit on it and your butt just touches the ground your at the right pressure
4. for the price i would recommend it for occasional use
5. if your a hard core camper you may want a name brand
6. it suits my needs perfectly

(c)

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



1. this had so much potential
2. my favorite 3 guitarist yet the sound is muddied
3. it should have been recorded in 5
4. the video is good
5. the sound is horrible though and that 's what makes this a travesty
6. i am so disappointed as for concert dvds audio is the most important factor
7. not even anamorphic

Isonuma1 et al. Unsupervised Neural Single-Document Summarization of Reviews via Learning Latent Discourse Structure and its Ranking. ACL 2019

NANYANG TECHNOLOGICAL UNIVERSITY

Thank You!