Concept to Code:
Aspect sentiment classification with Deep Learning

Muthusamy Chelliah
Flipkart

Asif Ekbal
IIT Patna

Mohit Gupta
Flipkart
Agenda

1. Introduction (10 min.s) (Asif)
2. LSTM/attention (25 min.s) (Asif)
3. Code - attention (30 min.s) (Mohit)
4. Memory networks (25 min.s) (Asif)
5. Code - Memory networks (30 min.s) (Mohit)
6. Aspect extraction (15 min.s) (Chelliah)
7. RNNs/RecursiveNNs (25 min.s) (Chelliah)
8. Convolutional Memory networks (20 min.s) (Chelliah)
Extraction - Chelliah
Aspect-specific ratings

1. Overall rating alone not good enough for product evaluation

2. Individual features gaining importance

3. Automatic attribute mining for summarization
Fine-grained opinion analysis

Detect subjective expression

- E.g., hate

Characterize intensity

- E.g., strong

Identify target/topic

- What the opinion is all about

[Wiebe 05] Annotating expressions of opinions and emotions in language, LREC
Aspect-based sentiment analysis (ABSA)

Extract targets

- Entities and their features

Summarize opinions

- On individual attributes
- Classify sentiment
  - positive/negative

Voice quality of iphone is great, but its battery sucks
Aspect identification

Opinion target

- Sentence topic: *Like this phone*

General

- Entity evaluated as a whole: *Voice quality of phone is great*

Implicit

- Sentiment indication: *car is cheap*
Aspect categorization

Same concept described

- With different words/phrases

Domain dependent synonyms

- Thesaurus dictionary (e.g., wordnet) not enough

- Image is clear

- Picture/photo -> image

- Call/voice quality

- Movie/picture

- Movie/video

- Expensive/cheap -> price
Opinion identification

Extract sentiment expression
- Polarity, intensity

Determine scope
- Aspect covered in sentence

Voice quality is *not that great*
Battery life is *very long*
Sentiment lexicon

Lexical resource

- Hard to maintain universal version
- Words vary per application domain

Double propagation

- sentiment word/product feature
- E.g., camera takes *great pictures*

Extraction rules

Based on dependency trees

[Qiu 09] Expanding domain sentiment lexicon through double propagation, IJCAI
Domain adaptation

Knowledge transfer

- Topic lexicon
- Labeled data abundant
  - in related domain

Seed generation

- Expand target data
- Exploit
  - Labeled source data
  - sentiment/topic relation

<table>
<thead>
<tr>
<th>Domain</th>
<th>Review</th>
</tr>
</thead>
</table>
| camera | The **camera** is great.  
  - it is a very **amazing** **product**. 
  - i highly recommend this **camera**. 
  - takes **excellent** **photos**. 
  - **photos** had some **artifacts and noise**. |
| movie  | This **movie** has **good script**, **great casting**, **excellent acting**.  
  - I **love** this **movie**. 
  - **Godfather** was the most **amazing** **movie**. 
  - The **movie** is **excellent**. |

[Li 12] Cross-Domain Co-Extraction of Sentiment and Topic Lexicons, ACL
Sequence labeling

Token-level tagging

- BIO scheme
- for each word in sentence

Sentence tagged with scheme

- Target (middle row)
- Expression (bottom row)

B beginning, I tokens inside, O tokens outside

[Choi 05] Identifying sources of opinions with CRFs and extraction patterns, HLT-EMNLP
Patterns

- Part-whole/meronymy
  - engine of car
- No (e.g., noise)

Ranking

- frequency
- relevance
  - HITS - from Web mining

[Zhang 10] Extracting and Ranking Product Features in Opinion Documents, COLING
Pattern mining

- Syntactic structures

Common pattern

- `<Feature/NN>` wildcard
  - To be fit in reviews
  - NN: POS tag of wildcard
- **Product** name *mp3* specified
  - *Screen* matching *mp3* is a feature

[Zhang 10] Extracting and Ranking Product Features in Opinion Documents, COLING
Opinion Expression

Jointly detected with orientation/strength

Parameter sharing vs. cascading 2 separate components

[Choi 10] Hierarchical sequential learning for extracting opinions and their attributes, ACL
Opinion expression (contd.)

The committee, as usual, has refused to make any statements.

Direct subjective (DSE)

- Opinion holder’s
- Explicit mentions of private states
  - Or speech events expressing them

Expressive subjective (ESE)

- Writer’s
- Indicate emotion/ sentiment
  - Not convey directly

[Irsoy 14] Opinion mining with deep recurrent neural networks, EMNLP
Phrase-level extraction

Feature set/function expansion

- Task-specific engineering effort

Pre-processing components

- Dependency parse tree, entity
- Manually crafted lexicons

Relax Markovian assumption

- NOT word level

[Yang 12] Extracting opinion expressions with semi-markov CRFs, EMNLP
Expression/target relationship

Pipelined approach

- Ignores interaction among extraction stage

Leverage knowledge instead

- From predictors that optimize subtasks

S1: [The workers][H_{1,2}] were irked[O_{1}] by [the government report][T_{1}] and were worried[O_{2}] as they went about their daily chores.

S2: From the very start it could be predicted[O_{1}] that on the subject of economic globalization, [the developed states][T_{1,2}] were going to come across fierce opposition[O_{2}].

[Yang 13] Joint inference for fine-grained opinion extraction, ACL
Word preference
Target/opinion candidates

- Semantic/opinion relations (solid/dotted)

Estimate candidate confidence

- From preferred collocation

[Kang 14] Extracting opinion targets and opinion words from online reviews with graph co-ranking, ACL
Dependency subtree polarity

Reversing polarity
- Words in subjective sentences

Tree bank with fine-grained label
- For phrases in sentence parse tree

CRFs with hidden variables
- Vs. bag-of-words (BoW)

[Nakagawa 10] Dependency tree-based sentiment classification using CRFs with hidden variables, NAACL
Seed generation

Frequently occurring nouns

- Filter commonly used one (e.g., thing, one)

Domain relevance measure

- Term frequency combined with Likelihood Ratio Test

Labeled example set

- +ve: features, -ve: noise terms

[Collobert ‘11] Natural language processing (almost) from scratch, JMLR.
RNNs/RecNNN - Chelliah
ProdFeatMin: summary [Xu 14]

Mine product function/attributes

Syntax-based methods use only contextual information

May suffer from data sparsity

Extract seeds automatically; measure semantic similarity between terms

CNN trained on each seed occurrence and classifies all for candidate

Label propagation of prior knowledge to product feature distribution

Exploring semantic relations with all seeds/other candidates

[Xu 14] Product feature mining: semantic clues versus syntactic constituents, ACL
Syntactic patterns

One-hot representation to encode context

- partial/discrete features (e.g., keywords)
- shallow information (e.g., POS tags)

Pattern design

Precision vs. generalization

<table>
<thead>
<tr>
<th>Example sentences</th>
<th>LSP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ... work such author as Herrick, Goldsmith, and Shakespeare</td>
<td>such NP as {NP, }* {(or</td>
</tr>
<tr>
<td>2. Even then, we would trail behind other European Community member, such as Germany, France and Italy</td>
<td></td>
</tr>
<tr>
<td>3. Bruises, wounds, broken bones or other injuries</td>
<td>NP{, NP}*{,} or other NP</td>
</tr>
<tr>
<td>4. Temples, treasuries, and other important civic buildings</td>
<td>NP{, NP}*{,} and other NP</td>
</tr>
<tr>
<td>5. All common-law countries, including Canada and England</td>
<td>NP{,} including {NP}* {or</td>
</tr>
<tr>
<td>6. ... most European countries, especially France, England, and Spain</td>
<td>NP{,} especially {NP}* {or</td>
</tr>
</tbody>
</table>

Syntactic patterns (contd.)

Product-have-feature

(a) can’t find \textit{fm-tuner}
   i) Product mentioned with \textit{player} instead of \textit{mp3}

(b) have replaced by \textit{support}

NP-VB-feature

c) irrelevant case not talking about product

[Xu 14] Product feature mining: semantic clues versus syntactic constituents, ACL
Lexical semantic clue

Noise term extracted even with high contextual feature

d) **Flaws** follow **mp3**

Not a product feature

Verify if candidate relates to target product

[Xu 14] Product feature mining: semantic clues versus syntactic constituents, ACL
Similarity graph

Screen is a product feature of mp3

Lcd is equivalent to screen

- hence a feature itself

Not features

- Terms similar to negative seeds

[Xu 14] Product feature mining: semantic clues versus syntactic constituents, ACL
Word embedding
Located closer in embedding space
- Semantically similar words
- Vectors alike

Distance metric
- Cosine similarity between 2 vectors

Mining of infrequent product features
- independence from term frequency

[Collobert ‘11] Natural language processing (almost) from scratch, JMLR.
Contextual semantic clue

Have

● Part-whole relation

Support

● Product-function relation

S.th have/s.th support

● Product features follow
● S.th replaced by terms referring to target product
  ○ (E.g., mp3, player)

(a) This player has an fm tuner.
(b) This mp3 supports wma file.
(c) This review has helped people a lot.
(d) This mp3 has some flaws.

[Xu 14] Product feature mining: semantic clues versus syntactic constituents, ACL
Convolutional neural networks (CNN)

Semi-supervised model

- Context encoding

Soft pattern miner

- Less sensitive to lexicon change

Consecutive subsequence $q_i$ of $s$ with $t$ and length $l$

- Screen is impressive

[Xu 14] Product feature mining: semantic clues versus syntactic constituents, ACL
Conventional neural models

Candidate term $t$ placed in window center

Best window is bracketed text if $l=5$

- $t = \text{screen}$ at boundary
- Should contain $\text{mp3}$
  - Strong evidence for feature finding

(a) *The [screen of this mp3 is] great.*
(b) *This [mp3 has a great screen].*

[Xu 14] Product feature mining: semantic clues versus syntactic constituents, ACL
Label propagation: clue combination

Each term with label distribution
- Probability of candidate being a feature

Classified results of CNN
- Prior knowledge

Explore candidate semantic relations
- To all seeds/other candidates globally

[Xu 14] Product feature mining: semantic clues versus syntactic constituents, ACL
Bootstrapping framework

Examples for training

- Label propagation

Accurate prior distribution

- CNN classification

Seeds generated after many iterations

- Results produced finally

**Input:** The review corpus $\mathcal{R}$, a large corpus $C$

**Output:** The mined product feature list $P$

**Initialization:** Train word embedding set $EB$ first on $C$, and then on $\mathcal{R}$

**Step 1:** Generate product feature seeds $V_s$ (Section 3.1)

**Step 2:** Build semantic similarity graph $G$ (Section 3.2)

while $\text{iter} < \text{MAX.ITER}$ do

**Step 3:** Use $V_s$ to collect occurrence set $T_s$ from $\mathcal{R}$ for training

**Step 4:** Train a CNN $\mathcal{N}$ on $T_s$ (Section 3.3)

Apply mini-batch SGD on Equ. 9;

**Step 5:** Run Label Propagation (Section 3.4)

Classify candidates using $\mathcal{N}$ to setup $I$;

$L^0 = I$;

repeat

$L^{i+1} \leftarrow (1 - \alpha)M^T L^i + \alpha D I$;

until $||L^{i+1} - L^i||^2 < \epsilon$;

**Step 6:** Expand product feature seeds

Move top $T$ terms from $V_c$ to $V_s$;

$\text{iter}++$

**Step 7:** Run Label Propagation for a final result $L_f$

Rank terms by $L^+_f$ to get $P$, where $L^+_f > L^-_f$;

[Xu 14] Product feature mining: semantic clues versus syntactic constituents, ACL
OpExRNN: summary [Irsoy 14]

Extract opinion expression

Token-level sequence labeling

Deep, narrow networks outperform shallow, wide ones
Recurrent Neural Networks (RNN)

ESE - phrases with subjectivity

- Terms with neutral sentiment in many contexts

Models interpreting context better

- Disambiguating subjective uses of phrases
  - With common words (e.g., as usual, in fact)

Embeddings Vs. parse tree & lexicon

Vs. semi-CRF

Input, hidden, output layer

- Black, orange, red node

[Irsoy 14] Opinion mining with deep recurrent neural networks, EMNLP
All

- ESE (i.e., obviously) omitted by annotator

Semi CRF

- Identifies long, subjective phrases
- Entirely misses subjective expression
  - With no clear sentiment
  - But equally, not yet/enough

Subjective expressions with inside of label

[Irsoy 14] Opinion mining with deep recurrent neural networks, EMNLP
Subjective expressions with \textit{begin} label

Deep RNN identifies

- first ESE in entirety
  - E.g., in any case
- More words as Inside 2nd ESE
  - E.g., it's high time

Shallow RNN labels few tokens as inside ESE

E.g., ANY, TIME

[Irsoy 14] Opinion mining with deep recurrent neural networks, EMNLP
AspExRNN: summary [Liu 15]

Extract opinion target

Token-level sequence labeling

RNNs outperform feature-rich CRF
RNN/Embeddings

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

Figure 1: Elman-type, Jordan-type and LSTM RNNs with a lookup-table layer, a hidden layer and an output layer. The concatenated context vector for the word “disk” at time $t$ is $x_t = [x_{hard}, x_{disk}, x_{is}]$ with a context window of size 3. One memory block in the LSTM hidden layer has been enlarged.

RNN: Short-term dependency between sentence words

Elman: dynamic temporal behavior remembering previous hidden layer

[15] Fine-grained opinion mining with RNNs and word embeddings, EMNLP
Bidirectional Elman RNN

Future information as critical as past

- Know *disk* to tag *hard* as **b-targ**

Aspect term in subjective, not objective, sentence

- E.g., crunchy tuna, to die for
- Crunchy tuna, imported from Norway

---

Figure 2: (a) Bidirectional Elman-type RNN and (b) Linguistic features concatenated with the hidden layer output in Elman-type RNN.

[Liu 15] Fine-grained opinion mining with RNNs and word embeddings, EMNLP
[Xu 14] CNN/label propagation improve each other through bootstrapping towards aspects

RNN/word embedding with token-level sequence labeling towards

[Irsoy 14] opinion expression

[Liu 15] aspect
RecNNSemComp: summary [Socher 13]

Semantic word spaces cannot express meaning of longer phrases
Understanding compositionality requires richer training/test data
More powerful model and Sentiment Treebank

[Socher 13] Recursive deep modeling for semantic compositionality over a sentiment tree bank, EMNLP
Characterizing sentiment/intensity

- Aggregate token vector representation

Tree bank with fine-grained label

- For phrases in sentence parse tree

Sentiment class at every tree node

Capturing negation/scope in sentence

[Socher 13] Recursive deep modeling for semantic compositionality over a sentiment tree bank, EMNLP
Sentiment degree

Slider for annotator

- After showing random phrases

Hit phrases sampled from set of all

- To prevent labels being influenced by what follows

[Socher 13] Recursive deep modeling for semantic compositionality over a sentiment tree bank, EMNLP
Recursive Neural Network (RecNN)

Compute parent vectors bottom-up

- With compositionality function $g$

Node classifier

- Vectors as features

Input vector interact only implicitly

- Through non-linearity (squashing function)

$\begin{align*}
p_1 &= f \left( W \begin{bmatrix} b \\ c \end{bmatrix} \right), \\
p_2 &= f \left( W \begin{bmatrix} a \\ p_1 \end{bmatrix} \right)
\end{align*}$

[Socher 13] Recursive deep modeling for semantic compositionality over a sentiment tree bank, EMNLP
Matrix-Vector RecNN

Combine constituents
- Multiplying matrix of one with vector of other

Compositional function
- Parameterized by participating words

Large number of parameters
- Depends on vocabulary size

\[
p_1 = f \left( W \begin{bmatrix} Cb \\ Bc \end{bmatrix} \right), \quad P_1 = f \left( W_M \begin{bmatrix} B \\ C \end{bmatrix} \right)
\]

[Socher 13] Recursive deep modeling for semantic compositionality over a sentiment tree bank, EMNLP
Recursive Neural Tensor Network (RNTN)

Direct multiplicative interaction

- Between input vectors

Compose aggregate meaning

- From smaller constituents in generic fashion

Dashed box represents one of $d$ slices

- Capturing child’s influence on parent

[Socher 13] Recursive deep modeling for semantic compositionality over a sentiment tree bank, EMNLP
Contrastive conjunction

Phrase X but phrase Y

- Conjunction as argument for 2nd conjunct
- 1st functioning concessively

Phrases are of different sentiments

- Classification needs to be correct for both
- Lowest node that dominates but/Y is correct

Bag of features: longer sentences
Recursive network: shorter phrases

[Socher 13] Recursive deep modeling for semantic compositionality over a sentiment tree bank, EMNLP
Negating positive sentences

RNTN able to structurally learn

Negation less obvious with least

[Socher 13] Recursive deep modeling for semantic compositionality over a sentiment tree bank, EMNLP
Negating negative sentences

Less negative not necessarily positive

E.g., not terrible

[Socher 13] Recursive deep modeling for semantic compositionality over a sentiment tree bank, EMNLP
### Positive/negative phrases

<table>
<thead>
<tr>
<th>( n )</th>
<th>Most positive ( n )-grams</th>
<th>Most negative ( n )-grams</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>engaging; best; powerful; love; beautiful</td>
<td>bad; dull; boring; fails; worst; stupid; painfully</td>
</tr>
<tr>
<td>2</td>
<td>excellent performances; A masterpiece; masterful film; wonderful movie; marvelous performances</td>
<td>worst movie; very bad; shapeless mess; worst thing; instantly forgettable; complete failure</td>
</tr>
<tr>
<td>3</td>
<td>an amazing performance; wonderful all-ages triumph; a wonderful movie; most visually stunning</td>
<td>for worst movie; A lousy movie; a complete failure; most painfully marginal; very bad sign</td>
</tr>
<tr>
<td>5</td>
<td>nicely acted and beautifully shot; gorgeous imagery, effective performances; the best of the year; a terrific American sports movie; refreshingly honest and ultimately touching</td>
<td>silliest and most incoherent movie; completely crass and forgettable movie; just another bad movie. A cumbersome and cliche-ridden movie; a humorless, disjointed mess</td>
</tr>
<tr>
<td>8</td>
<td>one of the best films of the year; A love for films shines through each frame; created a masterful piece of artistry right here; A masterful film from a master filmmaker,</td>
<td>A trashy, exploitative, thoroughly unpleasant experience ; this sloppy drama is an empty vessel.; quickly drags on becoming boring and predictable.; be the worst special-effects creation of the year</td>
</tr>
</tbody>
</table>

**More strongly positive phrases**

**At most \( n \)-gram lengths**

[Socher 13] Recursive deep modeling for semantic compositionality over a sentiment tree bank, EMNLP
PhraseRNN: summary [Nguyen 15]

Identify aspect sentiment in a sentence

Propagating semantics through binary dependency tree not enough

Novel hierarchical structure integrating dependency relations/phrases
Hierarchical sentiment classification

Extract basic phrases of sentence
- Constituent tree
- Preposition, noun, verb

Syntactic relations of aspect
- dependency tree

Word (leaf) or phrase (intermediate node)
- D-dimensional vector

RNN merges word representations
- phrases/sentences

AdaRNN with n composition functions
- Selected with linguistic tags

[Nguyen 15] PhraseRNN for ABSA, EMNLP
Phrase recursive neural network

Set of relation edges between
a) Vertices, b) sub-trees

Phrase List
PP[Except the design]
NP[the phone]
VP[is bad for me]

Target-dependent tree integrating constituent/
dependency trees

[Nguyen 15] PhraseRNN for ABSA, EMNLP
Phrase recursive neural network (contd.)

Composition function

- $G$ Inner and $h$ outer phrases

Feature types

- Label (left & right), dependency type
- E.g., (is-bad,COP), (me-for,POBJ), (bad-for, PREP)

Aspect sentiment category

- Root of binary dependency tree
- Logistic regression

Structure integrates

- dependency tree
- phrases

[Nguyen 15] PhraseRNN for ABSA, EMNLP
AspCatHybFeatLearn: summary [Zhou 15]

N-gram based features fail to capture semantic relations between different words

One-hot representation can’t measure association between words/aspects

Semi-supervised word embedding algorithm

Capturing semantic relations between words and sentiment words-aspects

E.g., delicious, tasty: food

[Zhou 15] Representation learning for aspect category detection in online reviews, AAAI
Hybrid feature learning

- Seed words help assign category labels
- Sentiment-aspect pair with dependency patterns

Noun->SBJ->W<-PRB<-ADJ

Aspect-specific word embeddings
- From corpus with noisy labels

Deeper features with neural networks
- Stacked on word vectors

[Zhou 15] Representation learning for aspect category detection in online reviews, AAAI
2-layer feed-forward network
- Trained to fit aspect categories
- Outputs binary variable
- Learns same shared features

Different network/output value per category
- Aspect-specific features in hidden layer

2-class logistic regression classifier
Trained on hybrid features for each aspect

[Zhou 15] Representation learning for aspect category detection in online reviews, AAAI
Composition/categorization - literature

[Socher 13] longer phrases need powerful compositional models and richer training/evaluation resources

[Nguyen 15] enrich aspect representation with constituent/dependency trees towards ASC

[Zhou 15] hybrid features for aspect categorization concatenating shared and aspect-specific features
Convolutional Memory Networks
ConvMemNw: summary [Fan 18]

Memory networks with single slot can’t model complex expressions - multiple words

One-hot representation can’t measure association between words/aspects

Convolutional network with attention instead computes weights

Multiple memory units corresponding to multi-word

<table>
<thead>
<tr>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>not be disappointed</td>
<td>did not enjoy</td>
</tr>
<tr>
<td>is not hard</td>
<td>not good for</td>
</tr>
<tr>
<td>not hard to</td>
<td>do not like</td>
</tr>
<tr>
<td>is never disappointing</td>
<td>can not work</td>
</tr>
</tbody>
</table>

[Fan 18] Convolution-based memory network for ABSA, SIGIR
Convolutional memory network

Extract word sequence features

Contiguous subsequence of memory units - chunks

Store context in a fixed-size window
- Capture long-distance dependency

Neural attention
- retrieves/feeds parts into downstream components

Figure 1: (a): A single hop version of our model. (b): A multiple hop version of our model.

[Fan 18] Convolution-based memory network for ABSA, SIGIR
Figure 2: The changes in each hop of attention

Producing +ve sentiment multiword
- Highest attention weight

Window size
- Commensurate noise
- Store context into different memory slots
- Capture context info into proper seq.

[Fan 18] Convolution-based memory network for ABSA, SIGIR
TNet: summary [ACL 18]

Attention-based approaches keep word-level features static

Aggregate them with weights as final representation

CNNs fail for sentences of different sentiments over multiple targets

Extract active local/n-gram features

Preserve original, contextual information

Favorite dish never tired

Great food but dreadful service

Long battery life vs. startup time

[Li 18] Transformation networks for target-oriented sentiment classification, ACL
Transformation networks

Contextualized word representations
- Bi-directional LSTM with hidden layers

Context-preserving transformation
- Target info. Into word representation
- Learn more abstract word-level features

Position-aware convolutional layer
- Encode positional relevance between word/target
- Extract informative features for classification

[Li 18] Transformation networks for target-oriented sentiment classification, ACL
Context preserving transformation

Consolidate word/target representations
  ● Tailor-made target-specific transformation

Target-specific word representation
  ● Deep neural architecture

Loss-less forwarding
  ● Directly feed features to next layer

Active scaling
  ● Gating function to control passed proportion

[Li 18] Transformation networks for target-oriented sentiment classification, ACL
### TNet: case study

<table>
<thead>
<tr>
<th>Sentence</th>
<th>BILSTM-ATT-G</th>
<th>RAM</th>
<th>TNet-LF</th>
<th>TNet-AS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Air has higher [resolution]$_p$ but the [fonts]$_N$ are small.</td>
<td>(N$^X$, N)</td>
<td>(N$^X$, N)</td>
<td>(P, N)</td>
<td>(P, N)</td>
</tr>
<tr>
<td>3. Sure it’s not light and slim but the [features]$_p$ make up for it 100%</td>
<td>N$^X$</td>
<td>N$^X$</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>5. [startup times]$_N$ are incredibly long: over two minutes.</td>
<td>P$^X$</td>
<td>P$^X$</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>7. The [staff]$_N$ should be a bit more friendly.</td>
<td>P$^X$</td>
<td>P$^X$</td>
<td>P$^X$</td>
<td>P$^X$</td>
</tr>
</tbody>
</table>

**Target/n-gram feature color coded**
- E.g., resolution, air has higher
- $+$ve, $-$ve, neutral: P, N, O

**Input targets wrapped in brackets**
- Labels as subscripts
  - X indicates incorrect prediction

[Li 18] Transformation networks for target-oriented sentiment classification, ACL
Parameterized CNN: summary [ACL 18]

CNNs don’t consider aspect terms
Memory networks can’t handle local patterns
Get aspect-specific features with target term information
Parameterized filters/gates

[Huang 18] Parameterized CNNs for ABSC, EMNLP
Parameterized CNN: architecture

<table>
<thead>
<tr>
<th>PF-CNN</th>
<th>PG-CNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence: great food but the service was dreadful.</td>
<td>Sentence: great food but the service was dreadful.</td>
</tr>
<tr>
<td>Aspect: food</td>
<td>Aspect: service</td>
</tr>
</tbody>
</table>

**Concatenate target vector with general sentiment for classification features**

**Control how much info. is passed to next layer**

[Huang 18] Parameterized CNNs for ABSC, EMNLP
ProxWeiCNN: summary [Zhang 19]

Syntactic dependencies of aspects with context ignored
Aspects thus attend to contextual words descriptive of other aspects

Position/dependency proximity weight

Its size is ideal and weight is acceptable

[Zhang 19] Syntax-aware ASC with proximity-weighted convolution network, SIGIR
Proximity-weighted CNN

Attention weight calculated typically

- With vectors in latent semantic space

Descriptively near but not syntactically correlated

- **Size acceptable**
- Proximity weight instead

Aspect sentiment polarity decided by key phrase

- CNN captures n-gram

[Zhang 19] Syntax-aware ASC with proximity-weighted convolution network, SIGIR
Dependency proximity

Position information approximates syntactical proximity

Distance between words in syntax dependency parsing tree

Shortest path length in tree between context word and food

Sequence of tree-based distance for all sentence words wrt aspect term aluminum

[Zhang 19] Syntax-aware ASC with proximity-weighted convolution network, SIGIR

Food is awesome - definitely try striped bass
## Syntax-aware ASC

<table>
<thead>
<tr>
<th>Method</th>
<th>Visualization</th>
<th>Pred.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Att.</td>
<td>great food but the service was dreadful!</td>
<td>negative</td>
</tr>
<tr>
<td>Pos.</td>
<td>great food but the service was dreadful!</td>
<td>positive</td>
</tr>
<tr>
<td>Dep.</td>
<td>great food but the service was dreadful!</td>
<td>positive</td>
</tr>
</tbody>
</table>

Attention wrongly renders term dependencies and decides on which context word depicts *food*

[Zhang 19] Syntax-aware ASC with proximity-weighted convolution network, SIGIR
Classification with convolution - literature

[Huang 18] parameterized filters/gates for aspect integration into CNN

[Li 18] context-preserving and position relevant transformation

[Fan 18] compute weights of multiple memory units towards multi-words

[Zhang 19] proximity-weighted CNN for syntax-aware context representation
Backup
Characterizing sentiment/intensity

[Socher 13] Recursive deep modeling for semantic compositionality over a sentiment tree bank, EMNLP
Sentiment degree

Characterizing sentiment/intensity

- Aggregate token vector representation

Tree bank with fine-grained label

- For phrases in sentence parse tree

[Socher 13] Recursive deep modeling for semantic compositionality over a sentiment tree bank, EMNLP