Joint and Pipeline Probabilistic Models for Fine-Grained Sentiment Analysis

Roman Klinger and Philipp Cimiano
Semantic Computing Group, CIT-EC, Bielefeld University, Germany

SENTIRE Workshop at IEEE ICDM, Dallas, TX, USA
What is this talk about?

The battery life of this camera is too short.

Aspect subjective

The battery life of this camera is too short.

Aspect subjective
Outline

1. Introduction
2. Probabilistic model for subjective term and target identification
3. Evaluation of different pipeline orders
4. Joint Model
5. Evaluation of the Pipeline vs. Joint Model
6. Summary and Discussion
Introduction

- Sentiment Analysis/Opinion Mining
  - Often modelled as classification or segmentation task
- Fine-Grained Opinion Mining:
  - Involves prediction of aspect/target, subjective terms, polarity, relations
- Our previous work: Developed model to analyze:
  - Given subjective phrases $\Rightarrow$ impact on target prediction
  - Given targets $\Rightarrow$ impact on subjective phrase prediction
  - Both with perfect and realistic prior knowledge
- Contribution of this paper:
  - Present a flexible model which takes into account inter-dependencies
Previous and Related Work

- **Extracting subjective phrases:**
  B. Yang et al. (2012). "Extracting opinion expressions with semi-Markov conditional random fields". In: EMNLP-CoNLL

- **Given perfect subjective phrases, predict targets:**
  N. Jakob et al. (2010). "Extracting opinion targets in a single- and cross-domain setting with conditional random fields". In: EMNLP

- **ILP approach**
  B. Yang et al. (2013). "Joint Inference for Fine-grained Opinion Extraction". In: ACL

**Our work:**
- Real-world setting, predict all entities
- Relational structure in multiple directions
- Flexible, easy to augment
Outline

1. Introduction

2. Probabilistic model for subjective term and target identification

3. Evaluation of different pipeline orders

4. Joint Model

5. Evaluation of the Pipeline vs. Joint Model

6. Summary and Discussion
A Factor Graph is a bipartite graph over factors and variables

- Factor $\Psi_i$ computes a scalar over all variables
- Let $\vec{x}$ be observed variables, $\vec{y}$ output variables
- Common definition:
  $\Psi_i(\vec{x}_i, \vec{y}_i) = \exp \left( \sum_k \theta_{ki} f_{ki}(\vec{x}_i, \vec{y}_i) \right)$
  (parameters $\theta_{ki}$ and sufficient statistics $f_{ki}(\cdot)$)
- Probability distribution:
  $p(\vec{y}|\vec{x}) = \frac{1}{Z(\vec{x})} \prod_i \Psi_i(\vec{x}_i, \vec{y}_i)$
Templates for Factor Graphs

- Probability distribution

\[
p(\vec{y}|\vec{x}) = \frac{1}{Z(\vec{x})} \prod_i \exp \left( \sum_k \theta_{ki} f_{ki}(\vec{x}_i, \vec{y}_i) \right)
\]

- A Factor Template \( T_j \) consists of
  - parameters \( \theta_{jk} \) and statistic functions \( f_{jk} \)
  - some description of variables yielding tuples \((\vec{x}_j, \vec{y}_j)\)
Introduction

Probabilistic Model

Evaluations 1

Joint Model

Evaluations 2

Summary

Templates for Factor Graphs

Parameters $\theta_{jk}$, feature functions $f_{jk}$ are shared across tuples

$$p(\overline{y}|\overline{x}) = \frac{1}{Z(\overline{x})} \prod_{T_j \in \mathcal{T}} \prod_{(\overline{x}_i, \overline{y}_i) \in T_i} \exp \left( \sum_k \theta_{kj} f_{kj}(\overline{x}_i, \overline{y}_i) \right)$$

Examples for descriptions:
Markov Logic Networks (Richardson et al., 2006)
Imperatively defined factor graphs (McCallum et al., 2009)
Variable Definition

- Extraction of **aspects** and **subjective phrases** as segmentation
- Application of a semi-Markov-like model
- Implementation in FACTORIE (McCallum et al., 2009)

The battery life of this camera is **too** short.

- Aspect
- Aspect
- Subjective
Introduction

Probabilistic Model

Evaluations 1

Joint Model

Evaluations 2

Summary

Templates

Single-Span-Template
- lower-case string, POS, and both
- Combined with IOB-like-prefixes
- Sequence of POS tags

The battery life of this camera is too short.
Templates

Inter-Span-Template (partially inspired by Jakob et al., 2010)

- Does the target span contain the noun that is closest to the subjective phrase?
- Are there spans of both types in the sentence?
- Is there a one-edge dependency relation between subjective phrase and target?
- Single-Span features only if one of those holds!
Learning and Inference

- Inference: Metropolis Hastings sampling (a Markov Chain Monte Carlo method)
- Learning: Sample Rank (Wick et al., 2011)

Objective Function

\[
f(t) = \max_{g \in s} \frac{o(t, g)}{|g|} - \alpha \cdot p(t, g),
\]

- \(t\) is a span, \(g\) is a gold span
- \(o(t, g)\) is length of overlap
- \(p(t, g)\) number of 'outside' tokens
Outline

1. Introduction
2. Probabilistic model for subjective term and target identification
3. Evaluation of different pipeline orders
4. Joint Model
5. Evaluation of the Pipeline vs. Joint Model
6. Summary and Discussion
Cameras (Kessler et al., 2010)

- Given subjective terms, how good is target prediction?
- Predicting subjective terms, how good is target prediction?
- Given target terms, how good is subjective prediction?
- Predicting targets terms, how good is subjective prediction?

![Graph showing F1 scores for various evaluation scenarios](image-url)
Cars (Kessler et al., 2010)

<table>
<thead>
<tr>
<th></th>
<th>Target-F(_1)</th>
<th>Subjective-F(_1)</th>
<th>Target-F(_1) Partial</th>
<th>Subjective-F(_1) Partial</th>
</tr>
</thead>
<tbody>
<tr>
<td>pred. S. → T.</td>
<td>0.51</td>
<td>0.56</td>
<td>0.43</td>
<td>0.33</td>
</tr>
<tr>
<td>pred. T. → S.</td>
<td>0.64</td>
<td>0.66</td>
<td>0.62</td>
<td>0.55</td>
</tr>
<tr>
<td>Gold S. → T.</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Gold T. → S.</td>
<td>1.00</td>
<td>0.74</td>
<td>0.70</td>
<td>0.50</td>
</tr>
<tr>
<td>Jakob 2010</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Twitter (Spina et al., 2012)
Idea for Joint Model

- Model relation explicitly

The battery life of this camera is too short.

- Features in three templates
  - Single Span
  - Inter Span
  - Relation
    (new: similar to inter span, but measuring another variable)
Sampler

Pipeline
- Propose spans, span changes
- Propose adding relations for each aspect-subjective pair

Joint
- Propose subjective phrases
- Propose aspects as targets of each subjective phrase
- Propose span changes, removing relations
Objective functions

**Pipeline**

- Spans as before \( f(t) \)
- Relations accuracy-based

**Joint**

Relation:

\[
    h(su, ta) = \max_{(su^*, ta^*) \in rel^*} \begin{cases} 
    -1 & \text{if } o(su, su^*) = 0 \text{ or } o(ta, ta^*) = 0 \\
    \frac{1}{2} (o(su, su^*) + o(ta, ta^*)) & \text{else}
    \end{cases}
\]

Span:

\[
    g(t) = \beta f(t) + \sum_{(su, ta) \in rel(t)} h(su, ta)
\]

Roman Klinger and Philipp Cimiano
## Pipeline vs. joint

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Pipeline Model</th>
<th>Templates</th>
<th>Joint Model</th>
<th>Iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Single span</td>
<td>Training/Prediction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prediction</td>
<td>Relation</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>Relations</td>
<td>Inter span</td>
<td>Subjectives</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Subjectives</td>
<td>Relation</td>
<td>Aspects</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Aspects</td>
<td></td>
<td>Relations</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Relations</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results

Camera

Car

Aspect Partial  Subjective Partial  Relation Partial
Aspect  Subjective  Relation
Summary

- Joint Modelling has positive impact
- Clearly observable for aspects
- Slight to moderate drop for subjective phrase and relation
- Easy do adapt to other characteristics (opinion holder, polarity, dependencies, etc.)
Bibliography I

Jakob, N. et al. (2010). "Extracting opinion targets in a single- and cross-domain setting with conditional random fields". In: EMNLP.


Yang, B. et al. (2012). "Extracting opinion expressions with semi–Markov conditional random fields". In: EMNLP–CoNLL.

Thank you!