December 7, 2013

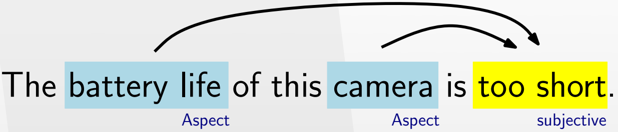
Joint and Pipeline Probabilistic Models for Fine-Grained Sentiment Analysis

Roman Klinger and Philipp Cimiano

Semantic Computing Group, CIT-EC, Bielefeld University, Germany

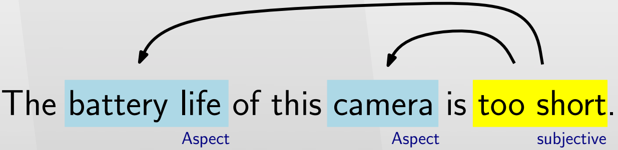
SENTIRE Workshop at IEEE ICDM, Dallas, TX, USA

What is this talk about?



The battery life of this camera is too short.

Aspect Aspect subjective



The battery life of this camera is too short.

Aspect 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 ⇒ impact on target prediction
  - Given targets ⇒ 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

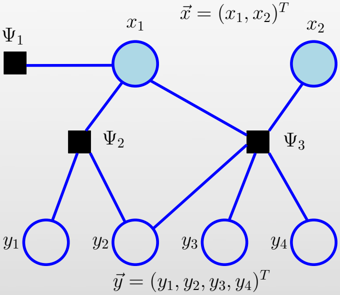
Factor Graphs

A Factor Graph is a bipartite graph over factors and variables

- Factor Ψ_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 θ_{ki} and sufficient statistics $f_{ki}(\cdot)$)

- Probability distribution:
  $$p(\vec{y}|\vec{x}) = \frac{1}{Z(\vec{x})} \prod_i \Psi_i(\vec{x}_i, \vec{y}_i)$$



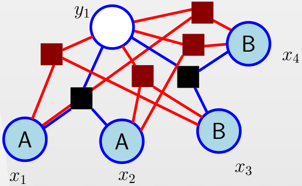
Templates for Factor Graphs

- 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 θ_{jk} and statistic functions f_{jk}
  - some description of variables yielding tupels $(\vec{x}_j, \vec{y}_j)$

Templates for Factor Graphs



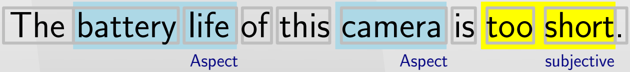
T_1: same value and y_1

T_2: different value and y_1

- Parameters θ_{jk}, feature functions f_{jk} are shared across tupels
- $$p(\vec{y}|\vec{x}) = \frac{1}{Z(\vec{x})} \prod_{T_j \in \mathcal{T}} \prod_{(\vec{x}_i, \vec{y}_i) \in T_i} \exp\left(\sum_k \theta_{kj} f_{kj}(\vec{x}_i, \vec{y}_i)\right)$$
- 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)



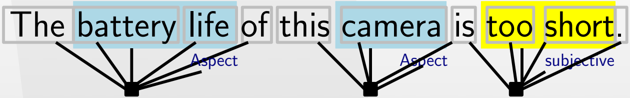
The battery life of this camera is too short.

Aspect

Aspect

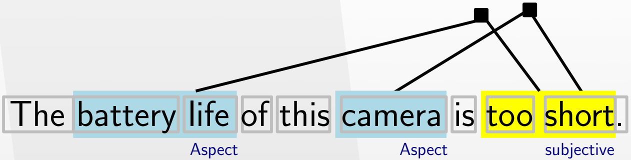
subjective

Templates



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

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

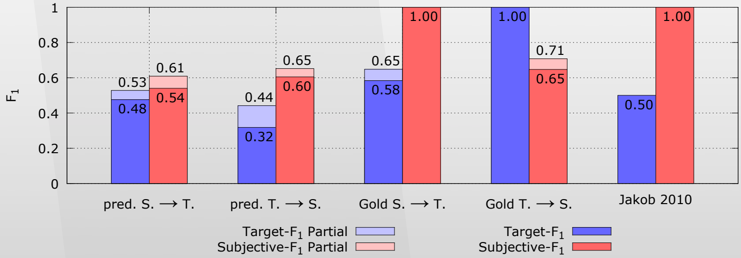
$$f(t) = \max_{g \in s} \frac{o(t, g)}{|g|} - \alpha \cdot p(t, g) \,,$$

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

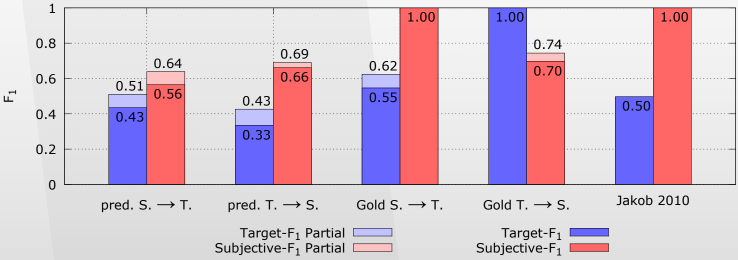
Outline

Cameras (Kessler et al., 2010)

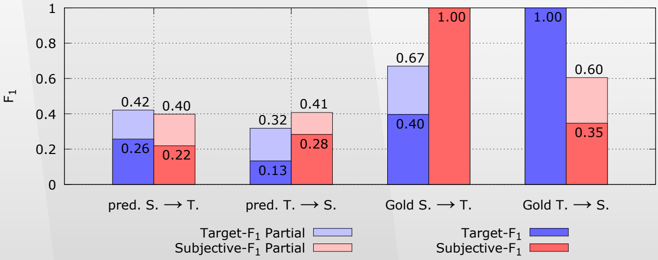
- Given subjective terms, how good is target prediction?
- Predicting subjective terms, how good is target prediction?
- Given target terms, how good is subjective prediction?
- Predicting targets terms, how good is subjective prediction?



Cars (Kessler et al., 2010)



Twitter (Spina et al., 2012)



Outline

1 Introduction

2 Probabilistic model for subjective term and target identification

3 Evaluation of different pipeline orders

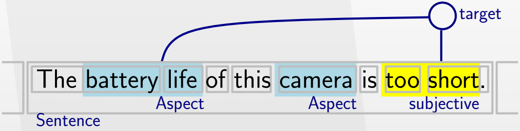
4 **Joint Model**

5 Evaluation of the Pipeline vs. Joint Model

6 Summary and Discussion

Idea for Joint Model

- Model relation explicitly



- 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

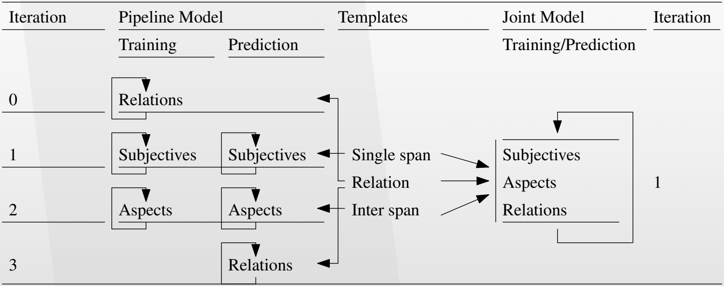
Relation:

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

Span:

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

Pipeline vs. joint



Iteration
Pipeline Model
Training
Prediction
Templates
Joint Model
Training/Prediction
Iteration

0
Relations

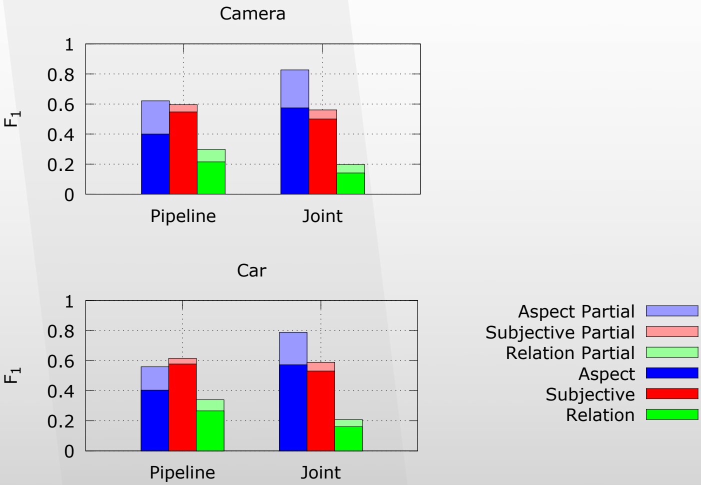
1
Subjectives
Subjectives
Single span
Subjectives

Relation
Aspects
1

2
Aspects
Aspects
Inter span
Relations

3
Relations

Results


Camera
F1
1
0.8
0.6
0.4
0.2
0
Pipeline
Joint
Car
F1
1
0.8
0.6
0.4
0.2
0
Pipeline
Joint
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.

Kessler, J. S. et al. (2010). ``The 2010 ICWSM JDPA Sentment Corpus for the Automotive Domain''. In: ICWSM-DWC 2010.

McCallum, A. et al. (2009). ``FACTORIE: Probabilistic Programming via Imperatively Defined Factor Graphs''. In: NIPS.

Richardson, M. et al. (2006). ``Markov logic networks''. In: Machine Learning 62.1-2, pp. 107–136. ISSN: 0885-6125.

Spina, D. et al. (2012). ``A Corpus for Entity Profiling in Microblog Posts''. In: LREC Workshop on Information Access Technologies for Online Reputation Management.

Bibliography II

Wick, M. et al. (2011). ``SampleRank: Training factor graphs with atomic gradients''. In: ICML.

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

– (2013). ``Joint Inference for Fine-grained Opinion Extraction''. In: ACL.

Thank you!