Fine-Grained Sentiment Analysis with Structural Features

Cacilia Zirn, Mathias Niepert,
Heiner Stuckenschmide, Michael Strube

Motivation

Goal:

- Sentiment analysis:
  - determining the polarity of a text with respect to a particular topic.

Existing Solutions:

  - Document level
  - Sentence level

Motivation

Problem:

- Different opinions can even be uttered in the same sentence.

e.g.

Despite the pretty design(Positive) I wound never recommend it(Negative), because the sound quality is unacceptable(Negative).

Motivation

A New Fine-Grained solution:

- Sentiment analysis on subsentence level.
- Motivated by Rhetorical Structure Theory(Mann and Thompson, 1988), every text consists of elementary segments that are connected by relations.

e.g.
s1 = Despite the pretty design
s2 = I wound never recommend it
s3 = because the sound quality is unacceptable

s1 –(concession)– s2 –(cause-explanation-evidence)- s3

Method

Basic idea:

- Integration the heterogeneous features such as polarity scores from sentiment lexicons and neighborhood relations between segments.
- Employ Markov logic networks(MLN) as the framework for combing numerical and structure features.

Method

Markov Networks:

- A Markov network M is an undirected graph whose nodes represent a set of random variables $X = \{X_1, ..., X_n\}$ and whose edges model direct probabilistic interactions between adjacent nodes. More formally, a distribution P is a log-linear model over a Markov network M if it is associated with:
  - a set of features $\{f_1(D_1), ..., f_k(D_k)\}$, where each Di is a clique in M and each f_i is a function from D_i to R.
  - a set of real-valued weights $w_1, \ldots, w_k$, such that

$$P(\mathbf{X} = \mathbf{x}) = \frac{1}{Z} \exp\left(\sum_{i=1}^{k} w_i f_i(D_i)\right),$$

where Z is a normalization constant.

Method

Markov Logic Networks[1,2]:

- A Markov Logic Network (MLN) L is a set of pairs (F_i, w_i) where:
  - F_i is a formula in first-order logic
  - w_i is a real number (the weight of the formula)
- Applied to a finite set of constants $C = \{c_1, \ldots, c_{|C|}\}$ it defines a Markov network $M_{L,C}$:
  - $M_{L,C}$ has one binary node for each possible grounding of each atom in L. The value of the node is 1 if the ground atom is true, 0 otherwise.
  - $M_{L,C}$ has one feature for each possible grounding of each formula F_i in L. The value of the feature is 1 if the ground formula is true, 0 otherwise. The weight of the feature is the weight w_i of the corresponding formula

Method

Markov Logic Networks:

- A ground MLN specifies a joint probability distribution over possible worlds (i.e. truth value assignments to all ground atoms)
- The probability of a possible world x is:

$$p(x) = \frac{1}{Z} \exp\left(\sum_{i=1}^{F} w_i n_i(x)\right)$$

  where:

  - the sum ranges over formulas in the MLN (i.e. clique templates in the Markov Network)
  - $n_i(x)$ is the number of true groundings of formula F_i in x

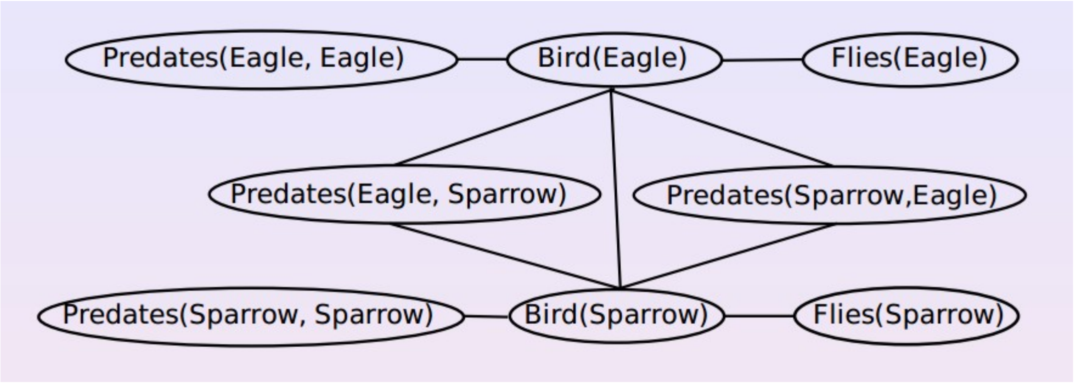
Method

Markov Logic Networks:

- A MLN with two (weighted) formulas:

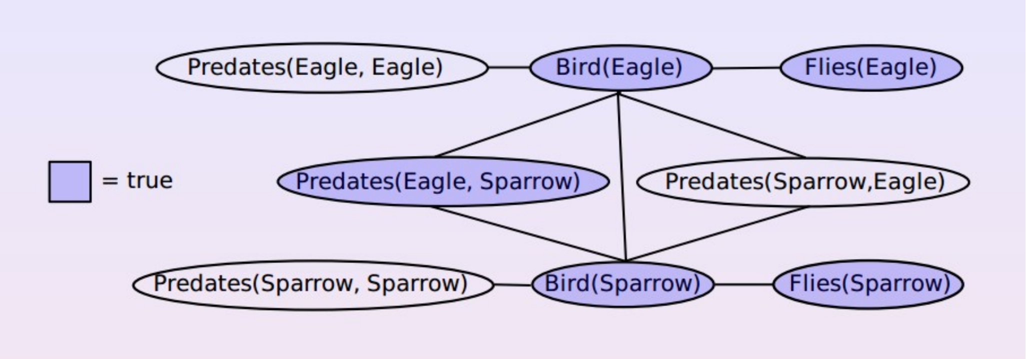
$$w_1 \quad \forall x\ (\texttt{Bird(x)} \Rightarrow \texttt{Flies(x)})$$
$$w_2 \quad \forall x, y\ (\texttt{Predates(x,y)} \wedge \texttt{Bird(y)} \Rightarrow \texttt{Bird(x)})$$

- applied to a set of two constants $\{\texttt{Sparrow}, \texttt{Eagle}\}$
- generates the Markov Network shown in figure

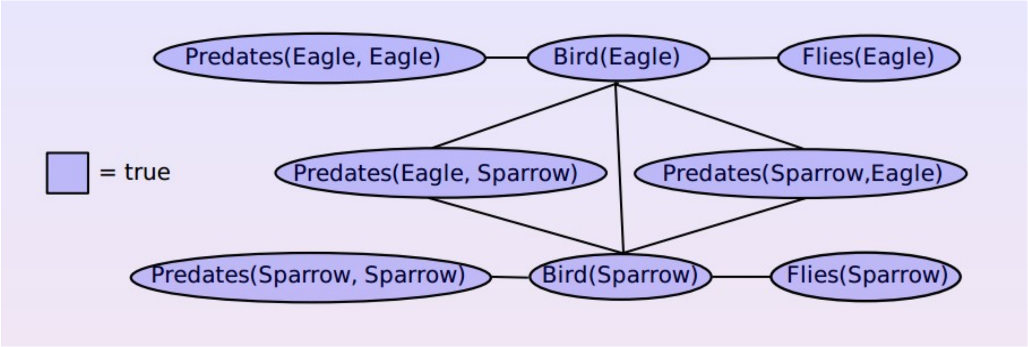


Method

Markov Logic Networks:



$$p(x) = \frac{1}{Z}\exp(2w_1 + w_2)$$



$$p(x) = \frac{1}{Z}\exp(2w_1 + 4w_2)$$

Method

Markov Logic Networks:

- Inference:
  1. (conditional) probability inference
  2. Maximum a-posteriori inference

$$\underset{\mathbf{y}}{\operatorname{argmax}} \sum_{i=1}^{k} w_i f_i(D_i).$$

- Parameter Learning:
  - Employing the Voted perceptron learner for the experiments

Method

Markov Logic Formulation:

- Polarity formulas:
  - A segment is positive or negative but cannot be positive and negative at the same time:

$$\forall x : \neg positive(x) \quad \Rightarrow \quad negative(x)$$
$$\forall x : negative(x) \quad \Rightarrow \quad \neg positive(x)$$

  - Numerical a-priori features such as the polarity scores of individual segments provided by external lexical resources

$$\forall x : positive\_source_{\ell}(x) \quad \Leftrightarrow \quad positive(x)$$
$$\forall x : negative\_source_{\ell}(x) \quad \Leftrightarrow \quad negative(x)$$

Method

Markov Logic Formulation:

- Neighborhood Relations:
  - Intuition: neighboring segments are more likely to have the same polarity.
  - *pre* relation: pre(x,y) if segment x precedes segment y.

$$\forall x, y : pre(x, y) \wedge positive(x) \Rightarrow positive(y)$$
$$\forall x, y : pre(x, y) \wedge negative(x) \Rightarrow negative(y)$$

  - The weights of above formulas are learned during training

Method

Markov Logic Formulation:

- Neighborhood Relations:
  - Intuition: neighboring segments are more likely to have the same polarity.
  - *pre* relation: pre(x,y) if segment x precedes segment y.

$$\forall x, y : pre(x, y) \land positive(x) \Rightarrow positive(y)$$
$$\forall x, y : pre(x, y) \land negative(x) \Rightarrow negative(y)$$

  - The weights of above formulas are learned during training

Method

Markov Logic Formulation:

- Discourse Relations:
  - Intuition: employ discourse relations between discourse segments
  - *contrast and ncontrast* relations:

$$
\begin{aligned}
\forall x, y : contrast(x, y) \wedge positive(x) &\Rightarrow negative(y) \\
\forall x, y : contrast(x, y) \wedge negative(x) &\Rightarrow positive(y) \\
\forall x, y : ncontrast(x, y) \wedge positive(x) &\Rightarrow positive(y) \\
\forall x, y : ncontrast(x, y) \wedge negative(x) &\Rightarrow negative(y)
\end{aligned}
$$

  - Again, the weights of above formulas are learned during training

Experiment

Data:

- A subset of the Multi-Domain Sentiment Dataset
  - Step 1: Pick three categories:
    - Cell Phones & Service
    - Gourmet Food
    - Kitchen & Housewares

Topic	p	n	total
Cell Phones & Service	1392	1785	3177
Gourmet Food	990	616	1606
Kitchen & Housewares	1188	1405	2593
Sum	3570	3806	7376

Table 1: Amount of positive (p) and negative (n) segments.

    - Each category consists of up to 100 reviews, each review is classified as positive/negative according the rating stars.
  - Step 2: Pick 20 longest positive and 20 longest negative for each (amount 120 reviews).

Experiment

Data:

- Gold Standard
  - Three annotators

  - Label: positive negative or neutral for each passage.
    - Passage: a sequence of words sharing the same opinion
    - Agreement:
      - k = 0.40 to 0.45 for negative reviews (Fail)
      - k = 0.60 to 0.84 for positive reviews (Strong agreement)

  - Polarity label for discourse segment:
    - Majority voting of all tokens belonging to the segment

Experiment

Details:

- Polarity features:
  - SentiWordNet
  - Taboada and Grieve's Turney Adjective List(TGL)
  - Unigram Lexicon (UL)

- Discourse Parsing
  - Employ the discourse parser HILDA (Verle & Predngingerr, 2009)
  - Two tasks:
    - Splitting the review text into discourse segments
    - Determining the discourse relations between segments
      - *contrast* v.s. all others (n*contrast)*

- 10 fold cross-validation

Experiment

Results:

	positive			negative			
	P	R	F	P	R	F	A
majority baseline	0.00	0.00	0.00	51.60	100.00	68.07	51.60
SVM	57.05	43.06	49.08	56.44	69.47	62.28	56.66
MLN_polarity	53.21	69.58	60.31	59.90	42.62	49.80	55.67
MLN_neighborhood	66.38	72.94	**69.50**	72.02	65.34	**68.52**	**69.02**
MLN_contrast	61.39	73.47	66.89	69.48	56.65	62.41	64.79

Table 2: Results (%) for the different systems. P = precision, R = recall, F = F-measure, A = accuracy

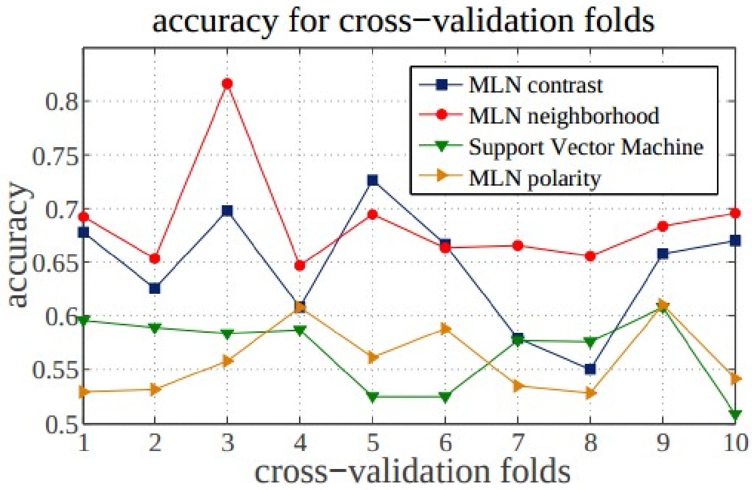


Figure 1: Accuracy values of the various algorithms for the 10 different cross-validation folds

References

[1] Matthew Richardson and Pedro Domingos, Markov Logic Networks, Machine Learning 62 (1-2): 107–136, 2006

[2] Andrea Passerini, Markov Logic Networks, http://disi.unitn.it/~passerini/teaching/complex_systems/slides/MLN.pdf