

Improved Natural Language Learning via Variance-Regularization Support Vector Machines

Shane Bergsma
University of Alberta

Dekang Lin
Google, Inc.

Dale Schuurmans
University of Alberta

1. Introduction

- Setting** Multi-class classifiers with real-valued features (counts, probabilities, scores, etc.)
- Hmm...** Often, good feature weights all have fairly similar values (just like some good unsupervised systems)
- Eureka!** We'll regularize (bias) an SVM so that it prefers weights that are similar (i.e. have low variance). It'll learn faster with fewer examples!

2. Preposition Selection

- Input** I worked in Sweden ____ 1997 to 2001.
- Output** **from** (not **during**, **in**, **on**, etc.)
- Features** \bar{x} : Counts of 34 different prepositions filling slot
- Classifier** $H: \bar{x} \rightarrow y$ N features
K classes

3. Features: N-gram counts

Ask the web! (Bergsma, Lin, Goebel, IJCAI 2009)

- \bar{x}_{from} : Freq(I worked in Sweden **from**),
Freq(worked in Sweden **from** 1997),
Freq(in Sweden **from** 1997 to),
Freq(Sweden **from** 1997 to 2001),
Freq(**from** 1997 to 2001 .), 5-grams
... 4-grams
Google N-gram Corpus
Freq(in Sweden **from**),
Freq(Sweden **from** 1997), 3-grams
Freq(**from** 1997 to)
- \bar{x}_{during} : Freq(I worked in Sweden **during**),
Freq(worked in Sweden **during** 1997),
Freq(in Sweden **during** 1997 to),
...
Freq(**during**1997 to)
- \bar{x}_{in} : Freq(I worked in Sweden **in**),
Freq(worked in Sweden **in** 1997),
Freq(in Sweden **in** 1997 to),
...
Freq(**in**1997 to)

$$\bar{x} = (\bar{x}_{\text{from}}, \bar{x}_{\text{during}}, \bar{x}_{\text{in}}, \dots)$$

E.g. $\bar{x} = (9.8, 4.1, 12.1, 9.3, 2, 7.8, 8.3, 1.3, \dots)$

4. Unsupervised Classification

$$H(\bar{x}) = \operatorname{argmax}_{r=1}^K \{\bar{1} \cdot \bar{x}_r\}$$

4. Multi-Class SVM Models

A) Standard k-SVM:
(Crammer & Singer, JMLR 2001)

$$H_{\mathbf{W}}(\bar{x}) = \operatorname{argmax}_{r=1}^K \{\bar{W}_r \cdot \bar{x}\}$$

K x N matrix → KN weights
preposition-specific weights
all counts

$$\bar{W}_{\text{from}}: (+1, +2, +1, 0, 0, 0.1, -0.2, 0, \dots)$$

$$\bar{x}: (9.8, 4.1, 12.1, 9.3, 2, 7.8, 8.3, 1.3, \dots)$$

Training:
 $\{(\bar{x}^1, y^1), \dots, (\bar{x}^M, y^M)\}$

B) SVM with Class-Specific Features (CS-SVM):

$$H_{\bar{w}}(\bar{x}) = \operatorname{argmax}_{r=1}^K \{\bar{w}_r \cdot \bar{x}_r\}$$

N weights
preposition-specific weights
preposition-specific counts

C) SVM with variance regularization (VR-SVM):

$$\min_{\bar{w}, \xi^1, \dots, \xi^m} \frac{1}{2} \bar{w}^T \mathbf{C} \bar{w} + C \sum_{i=1}^M \xi^i$$

subject to: $\xi^i \geq 0$

$$\forall r \neq y^i, \quad \bar{w}_{y^i} \cdot \bar{x}_{y^i}^i - \bar{w}_r \cdot \bar{x}_r^i \geq 1 - \xi^i$$

$$\mathbf{C} = \operatorname{diag}(\bar{p}) - \bar{p}\bar{p}^T$$

$$\begin{aligned} \operatorname{Var}[W] &= E[W^2] - E[W]^2 \\ &= \bar{w}^T \operatorname{diag}(\bar{p}) \bar{w} - (\bar{w}^T \bar{p})(\bar{p} \bar{w}) \\ &= \bar{w}^T (\operatorname{diag}(\bar{p}) - \bar{p}\bar{p}) \bar{w} \end{aligned}$$

5. Results

Preposition Selection

Unsupervised
Accuracy is 73.7%

Number of Training Examples

System	10	100	1000	10,000	100,000
One-vs.-all SVM	16	50.6	66.1	71.1	73.5
Multi-class SVM	13.7	50.0	65.8	72.0	74.7
CS-SVM	27.1	58.8	69.0	73.5	74.2
VR-SVM	73.8	74.2	74.7	74.9	74.9

Outputs/fillers: {about, above, across, after, against, along, among, around, as, at, before, behind, beneath, beside, between, by, down, during, for, from, in, inside, into, like, of, off, on, onto, over, round, through, to, towards, with}

Context-Sensitive Spelling Correction

Unsupervised
Accuracy is 94.8%

Number of Training Examples

System	10	100	1000	10,000	100,000
CS-SVM	86.0	93.5	95.1	95.7	95.7
VR-SVM	94.9	95.3	95.6	95.7	95.8

Example: "If you (cite/sight/site) me, then I'll (cite/sight/site) you."

Outputs/fillers: {among,between}, {amount,number}, {cite,sight,site}, {peace,piece}, {raise,rise}

Non-Referential Pronoun Detection

Unsupervised
Accuracy is 80.1%

Number of Training Examples

System	10	100	1000
CS-SVM	59.0	71.0	84.3
VR-SVM	70.2	76.2	84.5

Example: "(It) looked as if the ball was in the net."

Outputs = {referential/non-referential}, Fillers={it/its, they/their/them, all pronouns, *, <UNK>}

