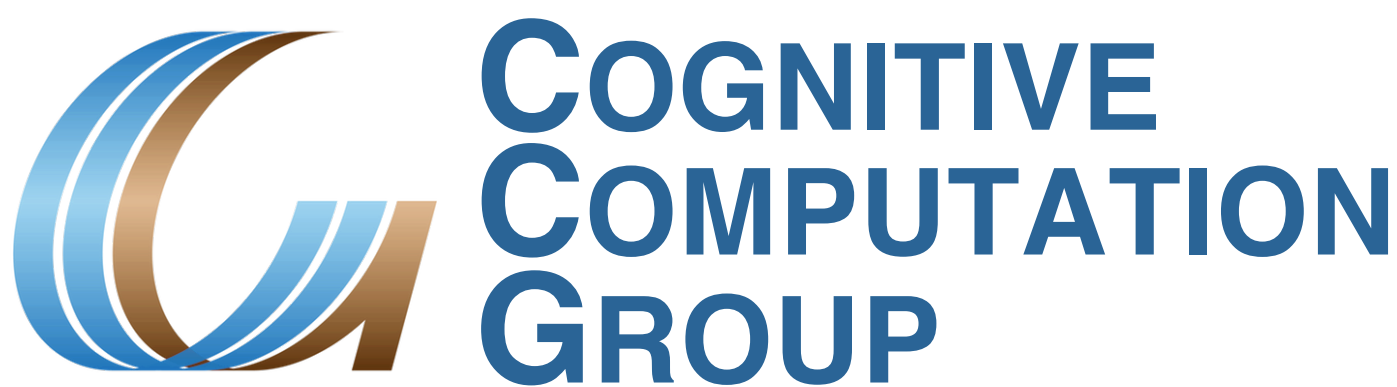# ner and pos when nothing is capitalized

stephen mayhew, tatiana tsygankova, dan roth
university of pennsylvania
{mayhew, ttasya, danroth}@seas.upenn.edu

Problem: NER and POS models aren't robust to casing issues.

Experiments: try 6 different solutions, report scores on cased (C) and uncased (U) data.

Key takeaway: data augmentation is the best option.

Casing Robustness is a REAL ISSUE

NOW HEAR THIS: NAVY ABANDONS ALL CAPS

Official Communications, Long Written Large, Can Use Mixed Case; No Shouting

Modern Tools Have TROUBLE***

her mother, susannah, lived near paris

David's Darkest Moments Came [PERSON] After Goliath

The Internet gives LIMITED HELP

answered Jul 15 '17 at 20:38
Christopher Manning
8,545 ● 27 ● 43

Approach 1: Caseless models. W will work much better on all lowerca

Approach 2: Use the truecaser. W into formally edited capitalization. Y

Experimental Setting: modify only the data. Models are standard BiLSTM-CRF.

	Train Data	Test Data
1.	<Do nothing>	<Do nothing>
2.	Lowercase	Lowercase
3.	Augment with Lowercase	<Do nothing>
3.5	Sample with Lowercase	<Do nothing>
4.	<Do nothing>	Truecase
5.	Truecase	Truecase

NER (CoNLL 2003 English)

Exp.	Test (C)	Test (U)	Avg
1. Cased	**92.45**	34.46	63.46
2. Uncased	89.32	89.32	89.32
3. C+U	91.67	89.31	**90.49**
3.5. Half Mixed	91.68	89.05	90.37
4. Truecase Test	82.93	82.93	82.93
5. Truecase All	90.25	**90.25**	90.25

POS (PTB English)

Exp.	Test (C)	Test (U)	Avg
1. Cased	**97.85**	88.66	93.26
2. Uncased	97.45	**97.45**	97.45
3. C+U	97.79	97.35	97.57
3.5. Half Mixed	**97.85**	97.36	**97.61**
4. Truecase Test	95.21	95.21	95.21
5. Truecase All	97.38	97.38	97.38

Mention Detection (Twitter, English)

Exp.	Mention Detection F1
1. Cased	58.63
2. Uncased	53.13
3. C+U	**66.14**
3.5. Half Mixed	64.69
4. Truecase Test	58.22
5. Truecase All	62.66