Appendices

Appendix I

Full List of CRIE features
*CRIE Readability 1.0
*SVM readability prediction 1.0
Paragraphs
Average paragraph length
*Characters
*Words
Adverbs
*Verbs
Type-token ratio
Difficult words
*Low-stroke characters
*Intermediate-stroke characters
*High-stroke characters
*Average strokes
*Two-character words
*Three-character words
*Sentences
*Average sentence length
*Simple sentence ratio
modifiers per NP
Np ratio
*Average propositional phrase
*Sentences with complex structure
Parallelism
Average number of idioms each sentence
*Content words
*Negatives
*Sentences with complex semantic categories
*Number of complex semantic categories
*Intentional words
*Noun word density
*Content word frequency in logarithmic
*Average frequency of content word in domain in Logarithmic
Number of Idioms
*Pronouns
*Personal pronouns
*First personal pronouns
Third personal pronouns
*conjunctions
positive conjunctions
*negative conjunctions
*adversative conjunctions
*causal conjunctions
hypothesis conjunction
condition conjunction
*purpose conjunctions
*figure of speech (simile)
*Content word category

   *feature that is included in the best feature set

Best feature set – LIWC features
WC
WPS
persconc

ppron
we
shehe
they
ipron
quanunit
specart
focuspast
progm
modal pa
general pa
interrog
quant
anx
family
friend
female
differ
see
feel
sexual
drives
achieve
power
risk
motion
work
home
netspeak
assent
Comma
Colon
Exclam
Parenth

Best features set – semantic, sentiment, and follower features
neg sent
char freq
wc over semantic classes
readability
followers

Appendix II

Sample questions used in crowdsourcing experiment

所以在所有东西都还没有调查完之前，就有人能够给人判刑定罪了吗？？？这是中国网民独有的人权？？？

1. 你认为此内容会被微博屏蔽吗？
2. 你认为此微博内容会触发争论或骂战吗？
3. 你认为此微博内容有触发集会、游行示威或暴动的可能吗？

任正非不要再做梦了，孟晚舟被捕，几天前特鲁多就知道，博尔顿在1号以前的晚宴也知道，川普也知道，美加两国情报系统联合行动，不知道罪名，

只因华为是对手，先抓人再说。这样的政治行动，任正非还指望在美加法律框架内"战胜他们"，可能吗?

1. 你认为此内容会被微博屏蔽吗?
2. 你认为此微博内容会触发争论或骂战吗?
3. 你认为此微博内容有触发集会、游行示威或暴动的可能吗?