HMM POS Tagging of Twitter Converstion

Muhammad Alif Akbar, Januar Triandy N.E, Surya and XXX
School of Computing, Informatics Departement
Telkom University

***Abstract*—The abstract goes here.**

***Index Terms*—Computer Society, IEEE, IEEEtran, journal, LaTeX, paper, template.**

1. Introduction

This demo file is intended to serve as a "starter file" for IEEE Computer Society conference papers produced under LaTeX using IEEEtran.cls version 1.8b and later. I wish you the best of success.

mds
August 26, 2015

1.1. Subsection Heading Here

Subsection text here.

1.1.1. Subsubsection Heading Here. Subsubsection text here.

2. Processing

Data then processed using a Hidden Markov Model (HMM) classifier with viterbi algorithm. HMM is chosen because it's simplecity and high accuracy for P.o.S Tagging [**?**], [**?**]. HMM is a classifier belong to statistical/bayesian family based on transition and emision probability. An HMM model is folmulated to

$$P(w_1..n|tag_1..n) = \prod_{i=1}^{n} emission(w_i|t_i) \prod_{i=1}^{n} transition(t_i|t_{i-1})$$

Emision probability calculated using formula:

$$e = \frac{count(word_i, tag_i) + 1}{count(tag_i) + \sum_{i=1}^{n} 1}$$

Transition probability calculated using formula based on unigram matrix of the tags:

$$t = \frac{count(tag_i, tag_{i-1}) + 1}{count(tag_{i-1}) + \sum_{i=1}^{n} 1}$$

We do a smoothing to the formula of emission and transition, add 1 (+1) to each instance of word and tags. The smoothing done to eliminate 'zero (0) occurance of an instance'. 'zero occurance' must be eliminate because whenever a word or tag chain never appear on training corpus appear on testing corpus the probaility of whole sentence hosted the word will return 0, and never be selected as possible outcome.

3. Conclusion

The conclusion goes here.

Acknowledgments

The authors would like to thank...

References

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.