Spotlight Id M65

A Bayesian LDA-Based model for semi-supervised part-of-speech tagging

Kristina Toutanova and Mark Johnson

Unlabeled data

Partial supervision

government : {noun}

benefit : {noun,verb}

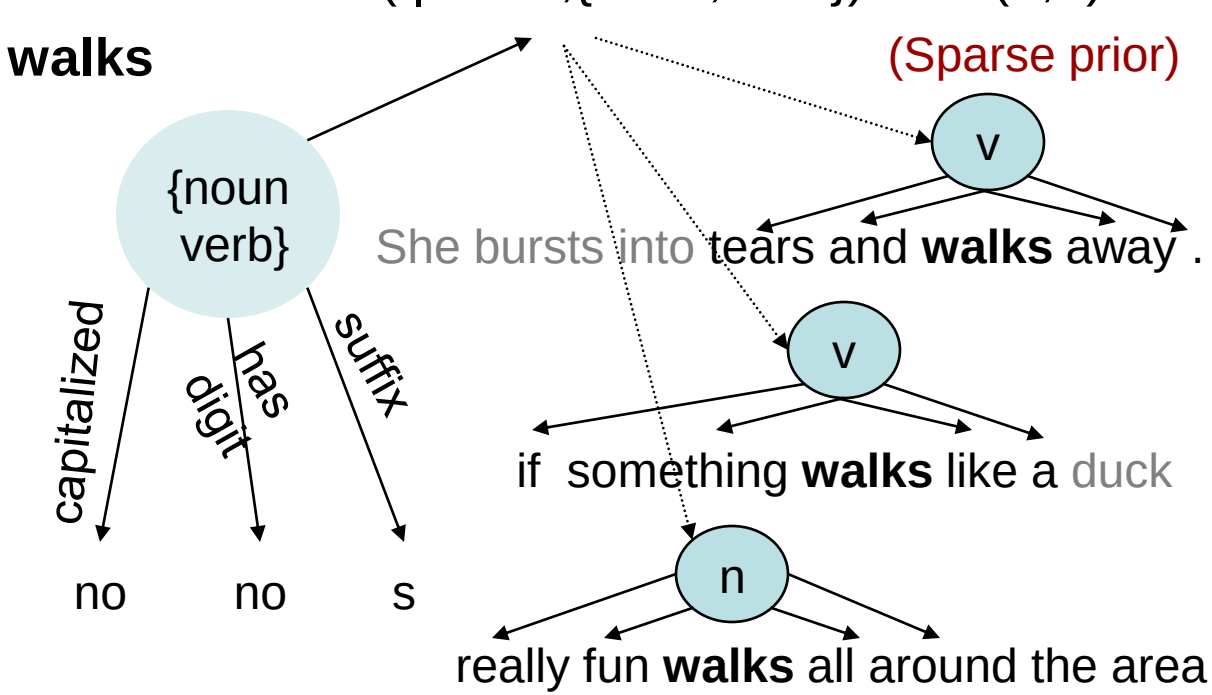
last: {noun, verb, adverb, adjective}

net: {adjective, noun, verb}

at: {noun, prep., part.}

Bayesian generative model

Ambiguity Class Model + Context Model



40 to 50% error reduction over the state of the art