Factoring out Background Words

p(w| θ)

d

Text mining paper

➡

the 0.031
a 0.018

…

text 0.04
mining 0.035
association 0.03
clustering 0.005
computer 0.0009
…
food 0.000001
…

How can we get rid of these common words?