

PROTOKOLLE REPITITORIUM ZU COMPUTERLINGUISTISCHES ARBEITEN

Ines Röhrer

Centre for Information and Speech Processing, LMU

`I.Roehrer@campus.lmu.de`

1 Signifikanztests, 29.05.2017

Signifikanztests sind dazu da, zu entscheiden ob und wie gut ein System ist. Für diese Tests benötigt man verschiedene Datensets, ein Development-, ein Train- und ein Testset.

Wenn man beispielsweise zwei verschiedene Systeme A und B hat, wobei beide Daten/Beobachtungen liefern, wie z.B. einen F-Score oder Accuracy, wobei System A bessere Ergebnisse als System B erzielt, hilft ein Signifikanztest diese Beobachtung einzuordnen. Man möchte sich sicher sein können, dass die Beobachtung, welches System besser ist, auch auf anderen Daten gleich bleibt, dazu benötigt man einen "Wahrscheinlichkeitsschätzer", der aussagt, ob die Unterschiede in den Ergebnissen verschiedener Tests zufällig sind oder nicht. Das wird mithilfe eines sogenannten P-Wertes bestimmt.