

Testen von Signifikanz

Um ein Modell zu testen braucht man ein Trainings-, Development- und Test-Set. Eine Methode um die Ergebnisse oder die Unterschiede zwischen zwei Systemen zu bewerten ist der Signifikanztest. Der Signifikanzwert kann Aussagen darüber machen, ob sich die Ergebnisse z.B. mindestens um eine bestimmte Wahrscheinlich verbessert haben. Ein Wert wird als signifikant bewertet, wenn das Signifikanzniveau erreicht ist. Übliche Signifikanzlevel sind 0.01, 0.05 und 0.1. Die Gauß'sche Verteilung veranschaulicht die Wahrscheinlichkeitsmasse. Es gibt verschiedene statistische Tests um Ergebnisse von Experimenten auszuwerten, wie z.B. der Sign Test, die Mean Average Precision und der Paired T-Test.