

Zusammenfassung von Thema 5: Signifikanztests

Das Thema von Sitzung 5 behandelte das Testen der Signifikanz für Experimente in NLP-Tasks. Hat man bspw. Daten in Training-, Development- und Testset geteilt und man testet hiermit zwei verschiedene Systeme, so benötigt es Signifikanztests um auszuwerten, welches der Systeme besser ist. Es wurden statistische Signifikanz Tests vorgestellt und erklärt, wie diese funktionieren. Dann wurden einige wichtige Maße vorgestellt, welche für die Auswertung von Experimenten üblicherweise verwendet werden, wie z.B. Accuracy und F-Score und. Anbei wurden Beispiele für die einzelnen Maße gezeigt. Zuletzt wurden einige „Fallen“ des Testens aufgezeigt, auf die man aufpassen soll.