

Sitzung am 29.05.2017

Zusammenfassung über das Thema „Practical Significance Testing for Experiments in Natural Language Processing“

Die Verfahren um zu kontrollieren, ob die Methoden eines Experiments auf neuen Daten funktioniert, sind Trainingset (Automatisches Learning von Parametern), Development (nicht schätzwert für die Performance unserer Methoden, denn sein Fluß auf Modell, die man trainiert, sehr unrealistischer ist) und Testset (Noch mal überprüfen, was wir im Experiment bekommen haben. Es ist sehr schätzwert für die Performance unserer Methoden). Wir bekommen dann die Resultate von unseren Methoden, die sich in einem kleinen Wert differenzieren. Wegen diesen Wert kann man nicht unbedingt sagen, dass es eine Methode besser als die andere ist. Deswegen benutzt man das Signifikanz-Test, um diese Schätzung zu machen: Ist diese Schwankung wichtig oder einfach ein Zufall? Die typische Werte für dieses Test sind: 0.05;0,01; 0,001... Je kleiner der Wert ist, desto vertraulicher sind die Resultate. Man muss diese Verteilung finden und kontrollieren, woher diese Unterschiede kommen.