Referat Michael Strohmayer (Montag, den 19.06.2017)

Maschine Learning basierte automatische OCR-Korrektur

Ziel dieser Arbeit ist es, Erstellung einer Software zur automatischen Nachkorrektur der eingelesenen OCR-Dokumente, das Trainieren eines Maschine-Learning Systems und der Auswertung der Ergebnisse.
Wir wissen alle, dass sich Schriftbilder, Grammatik und Schreibweisen sich verändern. Auch erkennen die OCR-Systeme manche Wörter nicht zuverlässig.

Wie ist die Vorgehensweise in dieser Arbeit?
Zu erst einmal lesen wir die Dokumente ein. Danach wird die gegebene Featurewerte extrahiert. Anschliessend fügen wir neue Features hinzu. Zudem wird das Machine-Learning Klassifkatoren antrainiert und zuletzt wird das Ganze evaluiert.

Leider enstehen hier bei der Vorgehensweise Probleme. Im Anfangsstadium gibt es Performance Probleme in der Datenverabeitung.
Auch sind der Ausgabe Kofidnezwerte abgeschnitten, deshalb enstehen falsche Trainigswerte.
Durch die Evaluation erzielen wir jedoch deutlich bessere Ergebnisse als zuvor. Und zudem ist die Berechnung von Naive Bayes sehr schnell. Diese bietet hochwertigere Ergebnisse.
Zum Naive Bayes und Libsvm wird ein Vergleich mit einer Tabelle dargestellt, indem man deutlich erkennen kann, dass das Naive Bayes um einiges genauer ist als die Libsvm.

Zuletzt werden noch einige interessante Arbeiten zum Thema als Lesematerial empfohlen, welche man in Michael Strohmayers Handout wiederfinden kann.