Homework 10

Fabian Otto – Matrikelnummer: 2792549
Foundations of Language Technology

14. Januar 2018

Task 10.1 a)

Da die Features zur Erkennung der Spams bekannt sind, ist die Reduzierung des Recalls einfacher als in einem echte Szenario in dem ein Angreifer/Spammer nur mutmaßen kann. Aufgrund dessen können die wesentlichen zwei Strategien auf folgende Ansätze reduziert werden:

1. Hinzufügen von „guten“ Features, d.h. Features, die für „noSpam“sprechen ODER
2. Reduzieren von „schlechten“ Features, d.h. Features, die für „Spam“sprechen

Für den ersten Fall könnten häufige Worte, wie „linguistics“, „languages“, „discussion“, „language“, etc. mehrmals an die Mail angefügt werden. Dies reduziert parallel die relative Anzahl der out-of-Vocab Wörter (im Vergleich zur Gesamtlänge), welche in den jetzigen Features allerdings nicht enthalten ist. Als konkretes Beispiel für Fall 2 können Zeichen, wie !, $, Sterne und Links bzw. nur der „http“ Teil entfernt werden(siehe dazu auch Code).

Task 10.1 b)

In der Aufgabe befindet sich folgender Text „*Please take care, that you only change the spam messages (the messages from the corpus in the training and the test data that belong to the SPAM class), not the valid emails, as (hopefully) the spammer has no way of changing the valid emails on the user's computer*“. Meiner Meinung nach ist das Verändern von Mails bereits zur Trainingszeit nicht sinnvoll, da ansonsten die nun neu entstehenden Besonderheiten wieder neu gelernt werden würden und der Effekt der Recall Reduzierung verloren gehen würde. Aufgrund dessen habe ich nur die Mails der Test Menge geändert und konnte damit einen Recall von ca. 58% erzielen. Mit Hilfe der alter_mail und category Parameter in der get_feature_set Methode ist allerdings auch ein Training mit den veränderten Mails möglich. Mein implementierter Ansatz erreicht damit allerdings keinen schlechten Recall. Hierfür müssten die meisten Wörter, die in den most_informative_features aufgeführt sind, den Spams hinzugefügt werden (Fall 1). Dies muss so lange wiederholt werden bis kein Unterschied zwischen Spam und NoSpam mehr vorhanden ist. Ohne die veränderten Mails funktioniert das hinzufügen dieser Wörter auch und verschlechtert den Recall immer weiter. Für die wichtigsten 8 ist dies auch im Code zu finden. Zusätzlich zum Fall 1, sind auch einige Änderungen für das Entfernen von Features (Fall 2) zu finden.