

6. Übungsblatt – Abgabe: 02.01.2022, 23:55

Aufgabe 6.1 - Statistische Modellierung

Ein statistisches Modell zur Klassifikation von Wörtern w in Adjektive (*ADJA*) und Nichtadjektive (*NADJA*) wurde auf einem Trainingscorpus mit einigen Hundert Sätzen trainiert. Man erhält dabei die folgende Frequenztabelle:

Wort w Artikel?	Wort $w+1$ großgeschrieben?	w endet auf -er/-es/-e/-en/-em?	ADJA	NADJA
Falsch	Falsch	Falsch	0	1758
Falsch	Falsch	Wahr	44	932
Falsch	Wahr	Falsch	3	473
Falsch	Wahr	Wahr	255	127
Wahr	Falsch	Falsch	0	32
Wahr	Falsch	Wahr	0	128
Wahr	Wahr	Falsch	0	0
Wahr	Wahr	Wahr	0	233

- (a) Wie viele Klassen gibt es, wie heißen sie?
- (b) Wie viele Features gibt es? Wie viele Werte haben die Features jeweils?
- (c) Welches sind die möglichen Ereignisse, die sich aus unterschiedlichen Featurekombinationen ergeben? Wie groß ist also der Ereignisraum? Geben Sie zu jeder Kombination aus einem Ereignis aus dem Ereignisraum und einer Klasse falls möglich einen Beispielsatz/eine Beispielphrase an, der/die zu dieser Kombination passt. Markieren Sie darin jeweils das Wort, auf das sich die Kombination bezieht. Wenn Sie Probleme haben, ein geeignetes Beispiel zu finden, beschreiben Sie, woran das liegt.
- (d) Für welchen Teil des Ereignisraumes (also für wieviele von wievielen potenziell möglichen Ereignissen) hat das Modell Trainingsinstanzen gesehen? Was bedeutet das für die Abdeckung des Modells auf neuen Daten?
- (e) Formulieren Sie Regeln für einen simplen Klassifikator, die jedem Ereignis genau eine Klasse zuordnen.

Hinweis: Das erste Feature wird Ihnen vielleicht unintuitiv erscheinen. Wenn man den POS-Tag des Wortes so gut kennt, dass man weiß, ob es ein Artikel ist oder nicht, könnte man ja auch gleich überprüfen ob es ein ADJA ist. *Wort w Artikel?* soll hier aber als Kurzschreibweise für *w ist eines der Wörter der, die, das, dem, den, ..., ein, eines, ...* verstanden werden, also ein Wort aus einer fest definierten endlichen Menge. Das geht bei Artikeln ohne großen Aufwand, weil es nur einige wenige gibt. Dagegen ist es nicht so einfach möglich alle Adjektive aufzulisten.

Aufgabe 6.2 - Evaluation

„Sentiment analysis“ (zu Deutsch etwa: „Stimmungserkennung“) bezeichnet ein Verfahren, bei dem Daten analysiert werden, um die Stimmung von Personen zu erkennen. Man kann zum Beispiel einen Bot entwickeln, der die Antworten auf einen Twitter-Beitrag analysiert und feststellt, ob der Beitrag gut ankam. Oder man lässt Produktbewertungen in einem Online-Shop analysieren, um automatisch ermitteln zu lassen, wie die Kunden ein bestimmtes Produkt durchschnittlich bewerten.

Nehmen wir an, dass wir eine Rehaklinik für Drogenabhängige betrachten. Die Klinik hat 1.000 Patienten und bietet vorwiegend Gruppentherapien an. Wenn es einem Patienten oder einer Patientin besonders schlecht geht, wird dieser Person allerdings ein privates Therapiegespräch verordnet. Ohne diese intensivere Behandlung würde die Person rückfällig werden. Da es nicht genug Therapeuten gibt, um die Stimmung einzelner Patienten einschätzen zu können, überwacht die Klinik die Gespräche und Chatverläufe aller Patienten mit einem automatischen Stimmungserkennungs-System. Wird bei einem Patienten oder einer Patientin eine sehr schlechte Stimmung erkannt, ruft das System einen Therapeuten für ein Privatgespräch, in dem abgeschätzt wird, ob die Person eine intensivere Betreuung braucht.

Nehmen wir an, dass von den 1.000 Patienten heute 10 eine so schlechte Stimmung haben, dass sie wieder Drogen nehmen wollen und dass 990 heute gut gelaunt sind. Das Stimmungserkennungssystem klassifiziert von den 990 gut gelaunten Patienten 950 als gut gelaunt und 40 als schlecht gelaunt. Von den 10 schlecht gelaunten Patienten klassifiziert es 8 als schlecht gelaunt und 2 als gut gelaunt.

- (a) Berechnen Sie für die Klassen „gut gelaunt“ und „schlecht gelaunt“ (als „gut“ und „schlecht“ abkürzbar) Precision, Recall und F-Score.
- (b) Vergleichen Sie die Ergebnisse für die beiden Klassen. Was fällt Ihnen auf? Beschreiben Sie umgangssprachlich, wie dieses Ergebnis zustande kommt.
- (c) Welche der beiden Evaluationen (Precision und Recall) ist für die tatsächliche Brauchbarkeit des Modells in der Praxis aussagekräftiger? Begründen Sie kurz.

Abgabe via Moodle. Bei Fragen posten Sie im MS Teams Fragenchannel oder besuchen Sie die Helpsession am Freitag von 16:15 bis 17:45 Uhr auf Teams.