Tvorba spam filtru

Report k závěrečnému projektu z předmětu RPH

Filip Krul, Petr Kučera

4. 1. 2021

1 Úvod

Cílem našeho projektu bylo vytvořit spam filtr, který dokáže rozeznat podvodné emaily od emailů, které jsou pro uživatele relevantní.

2 Popis principu algoritmu

Naším cílem bylo vytvořit spam filtr, který by fungoval na dvou rovinách. První rovina statická měla obsahovat algoritmus, který bude klasifikovat emaily na základě dat, která ve většině případů obsahují podvodné emaily. Druhá dynamická část by obsahovala algoritmus, který se dokáže sám učit na základě dat určených k učení. Závěrečná klasifikace, zda je email spam či ne, by byl následně vyhodnocen na základě skóre.

2.1 Načítání dat

Pro lepší práci s daty jsme si vytvořili modul mail, který pomáhá načíst jednotlivé částí mailu, jako je například odesílatel, předmět mailu či samotná emailová zpráva. Email jsme očistili o slova, která nemají žádnou vypovídající hodnotu jako jsou například spojky nebo předložky, unifikovali velikost písmen, odstranili html tagy a složité textové struktury, které mohou znít jinak ale mají stejný význam, nahradili normalizovanými stringy.

2.2 Statická část

Základní filtrování emailů jsme zajistili pomocí již předpřipravených kritérií, která jsme získali z ukázkových dat. Patřilo do nich například filtrace mailů, které nadměrně užívají určité znaky či obsahují slova, která jsou pro spamové maily typická.

2.3 Dynamická část

Dynamickou část filtru nám tvoří algoritmus, který se umí sám učit. Při učení využíváme **Naive Bayes klasifikátor**, který vyhodonocuje data na základě pravděpodobnosti. Aby mohlo učení probíhat efektivně, použili jsme metodu *porter_stemmer* z knihovny *nltk*,[1] která skloňovaná slova konvertuje do základního tvaru. Například z `['love', 'loving', 'lovely']` se stane `['love', 'love', 'love']`.

2.4 Vyhodnocení statusu mailu

Z dat nasbíraných během statické i dynamické části se na závěr vyhodnotilo skóre, které rozhodlo, zda bude filtr klasifikovat email jako spam či ham.

3 Popis týmové práce

3.1 Plánování

Na začátku celého projektu jsme si zavolali. Sdělili si své zkušenosti a rozvrhli, jak budeme postupovat. Vyjasnili jsme si, kdo se pustí, do jakého dílu práce.

3.2 Využívané technologie

Pro organizaci práce jsme vytvořili Trello, kód jsme sdíleli přes GitHub a komunikovali na Discordu.

3.3 Rozdělené práce

Na projektu jsme pracovali dohromady. Filip se spíše více soustředil na psaní kódu, Petr zase vypracoval závěrečný report a prezentaci.

osoba	typ činnosti
Filip Krul	primárně kód, sekundárně report a prezentace
Petr Kučera	primárně report a prezentace, sekundárně kód

4 Závěr

Filtr se nám povedlo dopracovat do stádia, kdy je schopen celkem pokročilé analýzy emailů. Nemůže se ovšem rovnat s algoritmy od softwarových gigantů. Projekt nám umožnil si vyzkoušet postavit algoritmus, který se umí učit za běhu programu. Vyzkoušeli jsme nové metody, které skrývá Python. Zdokonalili jsme se v týmové spolupráci a komunikačních dovednostech.

[1] Použití metody z této knihovny konzultováno a povoleno na cvičení.