Bayesian Filter (Bayesovský filtr)

Bayesian Filter je statistický algoritmus založený na **Bayesově teorému**, který se používá k třídění dat do kategorií (nejčastěji **Spam** vs. **Ham** – legitimní pošta). Na rozdíl od jednoduchých filtrů, které hledají konkrétní klíčová slova, se bayesovský filtr dokáže učit z příkladů a přizpůsobovat se aktuálním trendům.

Jak to funguje: Princip pravděpodobnosti

Filtr neposuzuje e-mail jako celek, ale rozkládá ho na jednotlivé prvky (slova, znaky, metadata), kterým přiřazuje určitou váhu pravděpodobnosti.

1. Bayesův teorém v praxi

Základní matematický vzorec (zjednodušeně) vypočítává pravděpodobnost, že e-mail je spam, na základě výskytu konkrétních slov:

$$P(Spam|Slovo) = \frac{P(Slovo|Spam) \cdot P(Spam)}{P(Slovo)}$$

- Pokud se slovo „Viagra“ vyskytuje v 90 % spamu, ale jen v 0,1 % legitimní pošty, má toto slovo vysoké „spamové skóre“.
- Pokud se slovo „Projekt“ vyskytuje v 50 % legitimní pošty, ale jen v 1 % spamu, táhne toto slovo výsledné hodnocení směrem k „Ham“.

Proces učení a klasifikace

Bayesovský filtr není po instalaci „vševědoucí“. Musí projít dvěma fázemi:

1. Trénování (Training)

Uživatel nebo správce systému předloží filtru dvě sady dat:

- Velký soubor potvrzených spamů.
- Velký soubor potvrzené legitimní pošty.

Filtr si vytvoří databázi slov a jejich četností v obou skupinách.

2. Bodování (Scoring)

Když dorazí nový e-mail:

```
1. Algoritmus vybere nejvýznamnější slova v textu.
2. Zjistí jejich pravděpodobnost spamu z databáze.
3. Provede kombinovaný výpočet celkového skóre (obvykle od 0.0 do 1.0).
4. Pokud skóre překročí nastavený práh (např. 0.9), e-mail je označen jako spam.
```

Výhody a nevýhody

Výhody

- **Adaptivita:** Pokud začnou spammeři používat nová slova, stačí párkrát označit e-mail jako spam a filtr se sám „přeškolí“.
- **Personalizace:** Filtr pro právníka se naučí, že termíny jako „smlouva“ jsou v pořádku, zatímco pro běžného uživatele mohou být podezřelé.
- **Nízká chybovost:** Správně natrénované filtry mají velmi nízký počet „falešně pozitivních“ výsledků (označení důležité zprávy za spam).

Nevýhody

- **Bayesian Poisoning (Otrávení filtru):** Útočníci do spamu přidávají dlouhé bloky legitimního textu (např. úryvky z románů), aby zmátli statistiku a „naředili“ spamová slova.
- **Nutnost tréninku:** Na začátku vyžaduje součinnost uživatele.
- **Vypočetní náročnost:** Je náročnější na CPU a paměť než jednoduché seznamy zakázaných slov (Blacklisty).

Praktické využití

Dnes je bayesovské filtrování součástí téměř všech velkých antispamových řešení, jako jsou:

- **SpamAssassin:** Populární open-source nástroj pro mailservery.
- **Gmail / Outlook:** Využívají pokročilé varianty tohoto algoritmu v kombinaci s neuronovými sítěmi.
- **Detekce obsahu:** Používá se i mimo e-maily, například pro automatické třídění zpráv v diskusních fórech nebo identifikaci jazyka.

Související pojmy: Spam, Ham, False Positive, Machine Learning, Blacklist, SpamAssassin.