Rozpoznávanie reči s využitím HMM*

Lukáš Lechman
Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií
xlechman@stuba.sk

30. september 2021

Abstrakt

V tejto práci by som sa chcel zamerať na modely rozpoznávania reči. Primárne sa budem venovať automatickému rozpoznávaniu reči (ASR). Využitiu Markovho skrytého modelu a jeho implementáciu v ASR. Ďalej sa pokúsim priblížiť, využitie a aplikáciu rozpoznávania reči v bežnom živote. Ako vieme využiť umelú neurónovú sieť, aby počítač komunikoval viac ako človek. A s tým súvisí v neposlednom rade využitie ASR inteligentnými virtuálnymi asistentmi.

1 Úvod

Od dôb turingovho stroja sme ušli dlhú cestu. Počítače a vo všeobecnosti akékoľvek elektronické zariadenia, sa stali kompaktnejšími a kompatibilnejšími. Aj napriek ich väčšej komplexnosti sú jednoduchšie na ovládanie než tomu bolo v minulosti, k ich ovládaniu už nie je potrebný siahodlhý návod ani viacero ľudí. V dnešnej dobe sa už ani nedá nájsť človek ktorý by nevedel pracovať s nejakou elektronikou, alebo by sa to aspoň nevedel naučiť. No ako ľudia máme tendenciu si stále zjednodušovať už aj tak dosť jednoduché veci. Dalo by sa povedať, že to je našou hlavnou evolučnou črtou. A nie je tomu inak ani vo svete počítačov. Jeden taký spôsob, ako si zjednodušiť prácu s elektronikou je využitie rozpoznávania reči, tzv. automatic speech recognition.

Práve rozpoznávanie reči sa stálo populárnym v poslednom desaťročí a stále sa tlačí viac do popredia v oblasti vývoja v IT svete. No nie vždy tomu bolo tak. V minulom storočí, keď išlo len o koncept budúcností sa na to mnoho ľudí pozeralo ako zbytočný krok, keďže mali za to, že komunikácia medzi človekom a počítačom bude vždy nadradená. Veď človek je inteligentnejší ako počítač, sám ho skonštruoval, tak ako by mohol počítač vedieť viac ako on. No dnes už vieme, že počítače sa dokážu učiť, využívať umelú inteligenciu a rôzne algoritmy aby sa mohli v mnohých ohľadoch priblížiť človeku. Síce sme sa ešte nedostali do bodu kedy by boli inteligentnejšími ako my, no už to nie je taká nereálna predstava ako v minulosti. Práve ASR by tomu mohlo dopomôcť. Rozpoznávanie reči má čoraz väčšie využitie, od zjednodušeného vyhľadávania na internete a

*Semestrálny projekt v predmete Metódy inžinierskej práce, ak. rok 2021/22, vedenie: Ján Lučanský

vykonania jednoduchých pokynov až po uľahčenie práce invalidným ľuďom, či dokonca pri práci v armáde. Asi každý z nás sa s ním niekde stretol, no nikdy sme sa nezamýšľali nad tým, ako je možné, že zariadenie vedelo porozumieť našim požiadavkám a následne ich vykonať.

V jednoduchosti by sa dalo povedať, že softvér na rozpoznanie reči funguje na jednoduchom princípe. Po detekcií reči, sa pomocou rôznych zariadení vytvorí súbor s tým čo bolo povedané vo forme vlnových dĺžok. Tento súbor následne prejde rôznymi úpravami, ako je odstránenie hluku v pozadí a pod.. Vyčistený súbor je rozbitý na menšie celky, ktoré sa analyzujú a prechádzajú ASR softvérom, ktorý na základe rôznych algoritmov a modelov predpokladá, aké slovo bolo povedané. No ako to tak už býva nič nie je tak jednoduché ako sa na prvý pohľad môže zdať. Za týmto zdanlivo jednoduchým postupom sa skrýva množstvo komplexných algoritmov a modelov, ktoré musia fungovať s obrovskou presnosťou, aby sme dosiahli požadovaný výsledok. Medzi takéto modely patria Skryté Markove modely (Hidden Markov Models).

2 Hidden Markov models

Pod týmto pojmom rozumieme model pravdepodobnosti pozostávajúci zo série premenných, ktoré reprezentujú isté pozorovania, premenné ktoré sú pre pozorovateľa skryté, z rozloženia počiatočného stavu, prechodovej matice a taktiež z rozloženia parametrov pre všetky pozorovania. [4] Skryté Markove modely pracujú s tzv. štatistickým modelovaním pri spracovaní signálu. Dalo by sa povedať, že skryté markovove modely sa dajú reprezentovať ako štatistické stavové automaty. Tie sú tvorené uzlami a prechodmi medzi nimi, pričom vnútorné vzťahy medzi uzlami sú reprezentované maticou pravdepodobnosti prechodov. Táto matica určuje pravdepodobnosť prechodu z jedného uzla do iného. [2] Prechody medzi stavmi sa riadia súborom pravdepodobností nazývaných pravdepodobnosti prechodu. V konkrétnom stave sa môže generovať výsledok alebo pozorovanie podľa príslušného rozdelenia pravdepodobnosti. Pre vonkajšieho pozorovateľa je viditeľný len výsledok, nie stav, a preto sú stavy pre vonkajšieho pozorovateľa skryté, odtiaľ pochádza názov HMM. [1]

Pre správne definovanie HMM potrebujeme niekoľko prvkov:

- Počet stavov modelu N
- Počet pozorovaných symbolov M
- Množina pravdepodobností prechodu medzi stavmi $A = aij$.
  $a_{ij} = p\{q_{t+1} = j \backslash \mathrm{q}_t = i\}, 1 \leq i, j \leq N, (1)$
  kde qt predstavuje aktuálny stav.
- Rozdelenie pravdepodobnosti v každom stave $B = bj(k)$
  $b_j(k) = p\{o_t = v_k \backslash \mathrm{q}_\mathrm{t} = \mathrm{j}\}, 1 \leq j \leq N, 1 \leq \mathrm{k} \leq \mathrm{M}(2)$
  kde vk označuje k-tý symbol pozorovania a Ot aktuálny vektor parametrov.
- Rozdelenie počiatočného stavu $\pi = pi_i$
  $\pi_i = p\{q_1 = i\}, \quad 1 \leq i \leq N(3)$
- Následne môžeme použiť zápis $\lambda = (A, B, \pi)$ [1]

3 HMM v ASR

HMM majú široké uplatnenie najmä kvôli svoje schopnosti rozpoznať súvislú reč ako tak aj samostatné slová. A taktiež preto, že ich vieme trénovať. Pri rozpoznávaní reči HMM určuje pravdepodobnosť následností 2 foném. Tento proces prebieha v 3 fázach. Prvá kontroluje pravdepodobnosť či daná fonéma je v skutočnosti fonéma, ktorá bola vyslovená, keďže každý človek dokáže povedať jedno slovo rôznymi spôsobmi, a teda sa aj jednotlivé fonémy môžu čiastočne líšiť. Ide o dôsledok prirodzene stochastickej povahy reči. Variabilita vo výslovnosti slova alebo fonémy sa prejavuje dvoma spôsobmi: trvaním a spektrálnym obsahom, ktoré sú známe aj ako akustické pozorovania. Okrem toho sú spektrálne obsahy konkrétnej fonémy ovplyvnené výskytom foném v okolitom kontexte, čo je jav nazývaný koartikulačný efekt. Je preto potrebné, aby HMM zohľadnil tieto koartikulácie. [3] V druhej fáze HMM kontroluje, či sa susedné fonémy môžu vyskytovať vedľa seba. Na základe týchto pravdepodobností HMM určí slovo, ktoré bolo s najväčšou pravdepodobnosťou vyslovené. V tretej fáze HMM kontroluje či slová, ktoré vznikli v druhej fáze dávajú zmysel a teda môžu tvoriť slovné spojenie. Vždy keď HMM narazí na slovo alebo frázu, ktorá nedáva zmysel tak sa vráti do druhej fázy a určí slovo s najväčšou pravdepodobnosťou, tak aby fráza dávala zmysel.

4 Proces učenia HMM

5 Záver

Literatúra

[1] Mbarki Aymen, Ammari Abdelaziz, Sghaier Halim, and Hassen Maaref. Hidden markov models for automatic speech recognition. In *2011 International Conference on Communications, Computing and Control Applications (CCCA)*, pages 1–6, 2011.

[2] Jana Bardovnova, Ivo Provaznik, and Zoltan Szabo. Rozpoznavani izolovanych slov pomoci markovovych modelu. 2000.

[3] K. R. Chowdhary. *Automatic Speech Recognition*, pages 651–668. Springer India, New Delhi, 2020.

[4] Mohd Izhan Mohd Yusoff, Ibrahim Mohamed, and Mohd Rizam Abu Bakar. Hidden markov models: An insight. In *Proceedings of the 6th International Conference on Information Technology and Multimedia*, pages 259–264, 2014.