

Naučili počítač slyšet a rozumět. Ted' už ve dvaceti jazycích

26.5.2025 - Adam Pluhař | Technická univerzita v Liberci

A s docentem Červou jsme si povídali o tom, jak naučit počítač, aby rozuměl lidské řeči nebo o tom, zda se máme bát umělé inteligence.

Pane docente, píšete víc na klávesnici, nebo už jen diktujete?

Píšu na klávesnici, protože sdílím kancelář s více kolegy a diktování by je rušilo.

Ptám se proto, že převod mluveného slova na jedničky a nuly je to, co vaši laboratoř nejvíc zaměstnává. Navíc, rozpoznávání řeči se už používá v řadě u hlasových asistentů, chatbotů, u automatické tvorby titulků nebo hlasovém ovládání auta. Myslíte, že hlas jednou nahradí klasické ťukání do klávesnice?

V některých situacích už se tak do velké míry stalo, například při tvorbě titulků či diktování radiologických nebo jiných medicínských nálezů. Ve většině případů ale bude z důvodů soukromí a rušení stále převládat psaní na klávesnici.

Před 30 lety vyvinula laboratoř SpeechLab na FM první funkční rozpoznávač VoiceCad, jenž uměl na základě 33 povelů ovládat hlasem počítač. Bylo to tehdy něco jako technologický zázrak?

Úplný zázrak to nebyl, protože podobný vývoj probíhal i jinde ve světě, ale v rámci ČR byl tehdy pan profesor Nouza skutečně jedním z prvních, kteří podobnou aplikaci vytvořili.

V roce 2005 jste vyvinuli program MyVoice, díky němuž dokázali handicapovaní ovládat počítač pouhým hlasem. Kolika lidem už pomohl?

Program se stále šíří mezi další a další uživatele, kterých jsou řádově už stovky. Je to nyní zásluha hlavně spolku Silou hlasu. Zakladatelka a předsedkyně tohoto spolku Dita Horochovská za tuto činnost získala Cenu Olgy Havlové 2022 – lidem s omezenou schopností pohybu předává zkušenosti s ovládáním počítače pomocí hlasu. Loni dokonce dostala státní vyznamenání od prezidenta Petra Pavla.

Byli jste tehdy v rámci českých nebo i světových pracovišť unikátní, nebo se podobnou činností zabývalo více pracovišť? A v čem spočívá vaše unikátnost dnes?

V rámci českých pracovišť jsme byli určitě jedni z prvních. Další laboratoře podobného zaměření vznikaly postupně v Plzni, na ČVUT v Praze a pak také na FIT v Brně. Naše unikátnost je nyní právě v počtu jazyků, které zvládneme a v tom, že řada z nich je skutečně prakticky nasazena díky spolupráci s firmou Newton.

Jak dlouho trvá, než se počítač, který rozumí řeči jedniček a nul, naučí porozumět nějakému jazyku? Co je k tomu učení zjednodušeně řečeno zapotřebí?

Nám to trvalo v podstatě 30 let vývoje. Přidat další jazyk je ale dnes pro tým, který má potřebné know-how a hardwarové vybavení, otázka několika měsíců práce. V rámci nich je zejména třeba vytěžit z veřejných zdrojů nebo jinak zpracovat zhruba 1000 hodin zvukových dat s přesným textovým přepisem.

Pomáhá vám v tomto výzkumu živelný rozvoj umělé inteligence a neuronových sítí? A je díky tomu trénování jazykových modelů jednodušší a rychlejší?

Rozvoj algoritmů AI a hlubokých neuronových sítí znamenal v posledních letech do té doby bezprecedentní nárůst přesnosti přepisu. Trénování je dnes ovšem řádově výpočetně náročnější.

V rámci mezinárodního grantu ve spolupráci s firmou Newton a univerzitou NTNU jste nedávno ovládli rozpoznávání velmi složité mluvené norštiny s mnoha dialekty a dvěma psanými formami. Dosáhli jste vyšší přesnosti než obdobné nástroje od Googlu nebo Microsoftu. Jak jste toho dosáhli?

Přesnost našich systémů je založena na použití dobrých, a námi různě optimalizovaných modelů. A pak na poctivé a mravenčí práci, která spočívá v důkladné přípravě trénovacích dat, která nesmí být zatížena žádnými chybami referenčního přepisu. Systém je nyní nabízen firmou Newton ve Skandinávii různým institucím, například pro přepis jednání v parlamentu.

Přepis na diktafon nahraných rozhovorů jako je tento, zabere několik hodin. Volně dostupných nástrojů zápisu řeči jsem už přitom vyzkoušel mnoho a jsou stále nekvalitní a nespolehlivé. Čím to?

I když se přesnost systémů rozpoznávání řeči za poslední období opravdu obecně výrazně zvýšila, tak jsou stále situace, kde, podobně jako lidské ucho a mozek, nefungují spolehlivě. Jde například o přepis řeči v prostředí se silným okolním ruchem a podobně. Navíc open-source modely nebo systémy často trpí tím, že jejich tvůrci si nedali často příliš práci s výběrem a čištěním trénovacích dat, což zhoršuje jejich přesnost.

Na čem nyní pracujete?

Aktuálně pracujeme zejména na vylepšování kvality přepisu v rušném prostředí, typicky například při použití vzdáleného mikrofону.

Pokud byste nyní naším povídáním motivoval budoucí studenty, co je na FM naučíte? A mohou se vůbec přidat k vám do týmu a spolupracovat s vámi třeba na učení počítače novému jazyku?

V rámci studijního programu Informační technologie nabízíme specializaci Inteligentní systémy, kde studenty naučíme porozumět různým metodám strojového učení a naučíme je tyto metody aplikovat při řešení úloh v oblasti zpracování řeči, obrazu i psaného textu. Do našeho týmu se pak mohou přidat již během studia v rámci řešení projektů nebo bakalářských či diplomových prací.

Jaký je váš názor na umělou inteligenci? Bude se její vývoj dál zrychlovat? A bude nám pomáhat, nebo se jí máme bát? "Vypnout" už asi nikdy nepůjde.

Vývoj AI se opravdu neustále zrychluje a je v něm tolik peněz, že vypnout ho již nikdo nedokáže. Už dnes nám pomáhá a třeba v diagnostice různých nemocí nám bude pomáhat ještě více. Bát se musíme spíš sami sebe, neboť lidé dokáží proti sobě zneužít v podstatě cokoli.

doc. Ing. Petr Červa, Ph.D.

(* 1980)

V listopadu 2021 obhájil habilitační práci Adaptation of speech recognition systems to selected real-world deployment conditions a přednesl habilitační přednášku na téma Adaptace systémů rozpoznávání řeči na vybrané reálné podmínky nasazení a byl jmenován docentem ve vědním oboru technická kybernetika.

Působí v Ústavu informačních technologií a elektroniky Fakulty mechatroniky, informatiky a mezioborových studií TUL, kde vede Laboratoř umělé inteligence.

<https://tuni.tul.cz/a/naucili-pocitac-slyset-a-rozumet-ted-uz-ve-dvaceti-jazycich-161373.html>