

Oceněný student ze soutěže Brno Ph.D. Talent chce překročit hranice neurálních konverzačních systémů

13.5.2024 - | Vysoké učení technické v Brně

Alexandra Poloka k výzkumu motivují omezení aktuálně rozšířených konverzačních systémů, které mají komponentně sestavenou architekturu. Systémy si vedou poměrně dobře v interakcích založených na jednoduchých úkolech, selhávají však v situacích, kdy mají využít delší kontext konverzace a poskytnout přirozený zážitek z dialogu.

Ve výzkumu se proto Polok chce zaměřit na efektivní trénování modelů pro rozpoznání řeči, konverzační jazykové modelování a syntézu řeči. Cílem je integrovat tyto modely do jednoho celku. Sníží se tak informační ztráta při interakci mezi modely, zmenší se velikost výsledného modelu a zkrátí se celková reakční doba. Vzniknout by tak mohl konverzační systém, který najde využití například v projektu využívajícím voicebota pro příjem tísňových volání na linky 150 a 112.

„Snažíme se vyvinout agenta, který dokáže s uživateli komunikovat v případě rozsáhlé katastrofy. Tísňové telefonní linky by v takové krizové situaci byly přetížené, proto chceme vyvinout podpůrný systém k plošnému nasazení. Dokázal by nejen informovat uživatele, ale odpovídat mu také na otázky a současně zpracovávat jeho odpovědi,“ vysvětluje Polok.

Projekt, se kterým doktorand uspěl i v soutěži Brno Ph.D. Talent, je aktuálně rozdělený na 3 fáze. „První je automatické rozpoznávání řeči, kdy potřebujeme mluvenou řeč dostat do podoby psaného textu. Následuje spojení automatického rozpoznávače řeči s jazykovým modelem, abychom dokázali odpovídat a generovat odpovědi na to, co uživatel říká. Třetím krokem je pak integrace těchto dvou fází,“ popisuje Polok.

„Řeč je velmi bohatá, a pokud ji převedeme na text, nutně přijdeme o nějakou informaci – ať už melodii řeči či emoční tonalitu. Já se zaměřuji na to, jak tuto informaci neztratit, ale rovnou ji integrovat do jazykového modelu a syntetizátoru řeči. Výsledkem by tak měla být méně monotónní syntetická řeč,“ dodává.

„Informace v konverzačním systému proplouvají mezi automatickým rozpoznávačem řeči a jazykovým modelem – to je moment, kdy dochází ke ztrátě informace. Přecházíme z informačně bohaté řeči obsahující desítky tisíc bitů za sekundu na text, které obsahuje stovky bitů za sekundu. Já vezmu automatický rozpoznávač řeči, odstraním z něj dekodér, který produkuje text, a ponechám jenom část modelující vstupní signál. Ta je následně transformována tak, aby ležela v prostoru blízko reprezentacím, na kterých byl původně učen jazykový model. Celý systém je následně dotrénován. Informace je tedy předána jazykovému modelu v podobě velké matice čísel a tu model dále zpracuje zase do podoby čísel,“ popisuje práci s neuronovými sítěmi doktorand. Nepochází tak k významné kompresi informace, jako se to děje u standardních komponentně založených systémů. Emoční složka řeči je pak ve výsledku stále zachována.

Výzkum se právě nachází v první fázi. Voicebot připravený pro aplikaci na tísňové linky by měl být hotov do konce roku 2026.

<https://www.zvut.cz/lide/-f38102/oceneny-student-ze-souteze-brno-ph-d-talent-chce-prekrocit-hranice-neuralnich-konverzacnich-systemu-d256947>