

Lukáš Žilka: Dialog Management with Deep Neural Networks

Oponentský posudek

Lukáš Žilka v předkládané tezi prezentuje cíle své disertační práce, shrnuje dosavadní výsledky a stanovuje další směr svého výzkumu na téma řízení dialogových systémů. Práce se zaměřuje na jednu ze zásadních částí dialogových systémů, a to komponentu pro sledování stavu dialogu (Dialog State Tracking). Cílem práce je řešení založené na rekurentních neuronových sítích typu Long Short-Term Memory (LSTM). Takový model umožňuje inkrementální sledování stavu dialogu po každém proneseném slově (word-by-word) a je trénovatelný na základě anotovaných dat (dialogů),

Teze má 20 stran a obsahuje vše potřebné: úvod do problematiky, popis dialogových systémů a jejich komponent včetně přehledu různých řešení, popis hlubokých neuronových sítí, jejich různých typů včetně LSTM, které je pro práci zásadní, dále poměrně zevrubný popis vlastní práce, t.j. návrh modelu sledování stavu dialogů, který je založen na LSTM sítích, a experimentu provedeného na standardních datech z restaurační domény s poměrně povzbudivými výsledky. Následuje analýza výsledků a diskuse nad dalším směřováním výzkumu. Práce je psaná anglicky, přehledně a srozumitelně, bez zásadnějších chyb. Přehled použité literatury je obsáhlý, některé položky by však bylo vhodné doplnit o chybějící informace.

Práci Lukáše Žilky lze hodnotit jako výbornou, svědčící o dobrém směřování a vedení výzkumu. Jedna věc by si ovšem zasloužila hlubší vysvětlení. Autor několikrát zmiňuje, že analýza dialogu word-by-word (tedy predikce stavu po každém proneseném slově) je lepší než varianta turn-by-turn (predikce se děje až potom, co účastník domluví, např. po celé větě), a to zejména kvůli zlepšení schopnosti systému reagovat (responsiveness). V tomto přístupu lze najít dva problémy. Jednak se může stát, že predikce stavu před koncem promluvy je předčasná a může vést k tomu, že systém během promluvy zbytečně "zabloudí". Totiž i poslední slovo ve větě může zcela změnit její význam. Obtížně se také budou zpracovávat jazyky, jejichž slovosled často řadí slovesa, která jsou pro průběh dialogu bezesporu důležitá, až na konec věty (např. němčina). Druhý problém může být ten, že predikce před koncem promluvy není vůbec potřeba. Asi by nebylo vhodné, aby automatický systém člověka přerušoval uprostřed věty, protože došlo ke změně stavu dialogu. "Slušnější" by bylo jej nechat domluvit, provést predikci stavu na základě kompletní promluvy a až následně reagovat. Oba tyto argumenty naznačují, že by přístup turn-by-turn by mohl být vhodnější a přístup word-by-word úlohu zbytečně zesložituje (jeho výhodou samozřejmě je, že není potřeba komponenta SLU a je možné zpracovávat přímo výstup rozpoznávače řeči, kódování celých vět by bylo složitější).