

Oldřich Krůza: Iterativní zdokonalování přepisu nahrávek s využitím zpětné vazby posluchačů

Oponentský posudek

Oldřich Krůza v předkládané tezi prezentuje dosavadní výsledky a další směr svého výzkumu na téma automatické rozpoznávání mluvené řeči a jeho iterativní zdokonalování s využitím zpětné vazby posluchačů. Cílem práce je automatický přepis audio nahrávek přednášek Karla Makoně, které má autor k dispozici, a umožnění vyhledávání v nich. Vyvinuté metody by ovšem měly být použitelné i na jiné audio záznamy.

Obsah

Teze má 14 stran a je rozdělena do 5 částí. Po stručném úvodu (část 1) a představení díla Karla Makoně (část 2) autor prezentuje cíle práce a dosavadní stav svého výzkumu (část 3). Následuje plán dalšího postupu (část 4) a závěr (část 5).

Dosavadní výsledky práce zahrnují natrénování systému pro automatický přepis nahrávek a implementace prototypu webové aplikace pro editaci automatických přepisů. Většina úsilí byla věnována přípravě rozpoznávače mluvené řeči založeném na systému HVite. Akustický model byl natrénován na části nahrávek Karla Makoně přepsaných autorem (v celkové délce kolem 3 hodin) a jazykový model na datech z Pražského závislostního korpusu. Nejlepší úspěšnost rozpoznávání, které autor svým systémem po značném úsilí dosáhl, je ovšem pouhých 27% *word precision*. Reportovaná *word accuracy* má dokonce záporné hodnoty.

Hodnocení

Práce Oldřicha Krůzy má v podstatě tři milníky: 1) získání dostatečně kvalitního prvotního přepisu nahrávek, 2) vytvoření aplikace pro opravu automatických přepisů posluchači, a 3) návrh a implementace metody pro využití získaných oprav ke zlepšení úspěšnosti automatického přepisu.

Dostatečný výzkumný potenciál je zřejmě jen v bodě 3), částečně i v bodě 2). Bod 1) se jeví čistě jako technický, ani autor v této oblasti nemá žádné výzkumné ambice, ale jeho úspěšné řešení je zcela zásadní pro celou práci. Je možné, že nedostatečná kvalita prvotního přepisu bude posluchače od oprav odrazovat a bylo by jednodušší nahrávky přepisovat rovnou, nikoli opravovat strojový přepis s velkým množstvím chyb. Navíc, pokud by byly nedostatečně kvalitní přepisy použity pro indexování a vyhledávání, výsledek by byl v takovém případě stěží použitelný. Autor sám připouští, že aktuální výsledky jsou „hluboko pod hranicí použitelnosti“, a pravděpodobně nelze očekávat, že se přílišlepší. Zásadní otázkou je také, jaká minimální kvalita přepisu je onou „hranicí použitelnosti“. Autor se tímto problémem zatím nezabýval, ale i hrubý odhad by leccos napověděl a výzkum v této oblasti by byl jistě zajímavou částí disertační práce (analogie s posteditací strojového překladu). V tezi chybí systematičtější snaha zjistit, proč je kvalita automatického přepisu tak nízká, zda existují nějaké rezervy, případně jaké.

Bod 2) je v podstatě vyřešený, i když na uživatelském rozhraní by bylo ještě co zlepšovat. Bodu 3) se autor zatím nevěnoval. Triviální způsob, který se nabízí, je opravená data přidat k těm trénovacím, ale zde by vhodné zjistit, jestli je ještě nějaké zlepšení tímto způsobem možné (např. analýzou *learning curves*). Pokud autor přepsal již 3 hodiny nahrávek pro

účely trénování, je otázkou, jestli se další úsilí vůbec může vyplatit. Rezervy by mohly být i v jazykovém modelu, kde by snad mohlo pomoci zvětšit trénovací data a případně jejich výběr podmínit doménou nahrávek. V tezi jsou zmíněny metody aktivního učení, které by jistě bylo zajímavé využít – cíleně vybírat takové úseky nahrávek, které by bylo žádoucí opravit a použít pro trénování, případně jak taková data použít pro vylepšení modelu, aniž by bylo nutné jej hned celý přetrénovat (i zde existuje analogie s podobnými postupy v oblasti strojového překladu).

V tezi chybí hlubší rešerše a není jasné, zda je autorův přehled o relevantní literatuře dostatečný.

Závěr

Předkládaná teze nasvědčuje tomu, že autor na disertaci pracuje a je schopen prezentovat jisté dílčí výsledky (prvotní přepis nahrávek, webovou aplikaci pro editaci automatických přepisů). Kvalita prvotního přepisu je ale pravděpodobně nedostatečná a doporučoval bych autorovi požádat o spolupráci/konzultaci zkušenější kolegy např. z KKY FAV ZČU. Analýza *learnig curves* potom napoví, zda zvětšování trénovacích dat o opravy posluchačů může vést ke zlepšení výsledků. V další práci bych doporučoval zaměřit se spíše na směry s větším vědeckým potenciálem, např. zmiňované aktivní učení a inkrementální aktualizace modelů.