Detekce klíčových frází ve výstupech fonémového rozpoznávače

Adam Chýlek[1]

1 Úvod

Komunikace člověka s technikou se čím dál více snaží přiblížit přirozené komunikaci mezi lidmi. Jednou z oblastí, ve které je třeba zkoumat nové možnosti a prostředky ke zlepšení kvality této komunikace, je oblast hlasových dialogů.

Práce se zabývá kombinací fonémových mřížek a slovních gramatik, které může být využito pro porozumění v hlasových dialozích prostřednictvím detekce klíčových frází (příp. realizací sémantických entit) nebo též pro vyhledávání v archivech mluvené řeči.

2 Navržený algoritmus

Metody zvolené pro dosažení našeho cíle staví na teorii vážených konečných automatů a v práci jsou uplatněny též metody strojového učení.

Vstupem algoritmu je soubor hledaných lexikálních realizací sémantických entit spolu s přiřazením příslušného označení sémantické entity. Tento soubor je převeden do formy váženého konečného transduceru (WFST) S.

Dalším vstupem je pak buď fonémová mřížka opět v podobně WFST F s pseudo-časovým zarovnáním získaná z fonémového rozpoznávače nebo pouhá první nejlepší hypotéza z této fonémové mřížky. Pseudo-časovým zarovnáním rozumíme uměle vytvořené zarovnání na základě struktury mřížky, kdy jsou přechody očíslovány po topologickém seřazení konečného automatu. Pro námi řešený problém se jedná o dostačující aproximaci reálného časového zarovnání. Mřížka F je výstupem fonémového rozpoznávače a upravená na faktorový automat. Ten umožňuje najít v mřížce i menší části (podřetězce) původních řetězců symbolů.

V algoritmu používáme též WFST L jako slovník pro fonémový přepis slov a transducer E pro určení editační vzdálenosti mezi dvěma řetězci.

Algoritmus vytváří kompozicí (operace nad konečnými automaty, značena $\circ$) transducer $R = S \circ L \circ E \circ F$, jehož hrany mají vstupní symboly reprezentující označení hledaných sémantických entit a výstupní symboly odpovídající fonémové realizaci těchto entit nalezených ve vstupní mřížce F, včetně editační vzdálenosti jako váhy příslušných hran.

Procházením R získáváme příznaky pro klasifikátor. Výstupem klasifikátoru je informace o tom, zda sémantická entita v mřížce je či není. Jako příznaky byly zvoleny: délka nalezeného řetězce, časová značka začátku a konce, editační vzdálenost, počet shodných fonémů oproti hledané fonémové realizaci sémantické entity, nalezené fonémy a fonémy změněné při použití editační vzdálenosti.

Sémantické entity vyskytující se v textu (na základě výstupu klasifikátoru) jsou dále filtrovány tak, aby byly časově se překrývající stejné entity sloučeny do jedné.

Výstupem algoritmu je mřížka sémantických entit vyskytujících se v textu, vč. jejich

[1] student doktorského studijního programu Aplikované vědy a informatika, obor Kybernetika, e-mail: chylek@students.zcu.cz

nalezené fonetické reprezentace a časového zarovnání.

3 Realizace

Pro trénování klasifikátoru (logistická regrese a extrémně znáhodněné stromy) byla použita jednak sada obsahující 5240 mřížek a také mnohem menší sada 570 mřížek z korpusu, který obsahuje nahrávky dotazů na odjezdy a příjezdy vlaků. V rámci experimentů byly detekovány entity nádražních stanic (2806 entit, např. Ústí nad Labem, Praha hlavní nádraží) a typy vlaků (9 entit, jako např. rychlík, osobní vlak, apod.). Výpočet editační vzdálenosti byl na základě experimentů omezen maximálním počtem 2 po sobě následujících vložení/smazání.

Jako baseline byla stanovena detekce klíčových frází hledáním přesné shody fonetické realizace sémantické entity s podřetězcem v mřížce (nulová editační vzdálenost).

V tabulce 1 jsou uvedeny nejlepší dosažené hodnoty AUC v porovnání s baseline a velikostí trénovacího souboru. Sledovaný parametr AUC je definován jako obsah plochy pod křivkou danou závislostí míry detekce DR na míře falešných poplachů FPR. Míru detekce počítáme jako $DR = \frac{\text{(\# entit obsažených v referenci a zároveň ve výstupu)}}{\text{(počet entit v referenci)}}$, míra falešných poplachů je definována $FPR = \frac{\text{(\# entit ve výstupu, které nejsou v referenci)}}{\text{(počet hledaných entit)}}$. Tyto hodnoty se mění v závislosti na prahu Ψ, který používáme při binární klasifikaci pro stanovení příslušné třídy. Formálně tedy $AUC = \int_0^1 DR(FPR)\mathrm{d}FPR$.

Sémantické entity	AUC baseline	AUC slovní	Počet mřížek	AUC
Nádražní stanice	0,458	0,785	5240	$0,732$
			570	$0,663$
Typy vlaků	0,601	0,883	5240	$0,869$
			570	$0,868$
Typy vlaků (klasifikátor nádražních stanic)	0,601	0,883	5240	$0,869$
			570	$0,869$

Tabulka 1: Dosažené hodnoty AUC, testovací sada 1439 mřížek.

4 Závěr

Zajímavých výsledků bylo dosaženo u hledání velmi malé množiny sémantických entit určujících typ vlaků. Zde je vidět, že i přes natrénování klasifikátoru na zcela jiné množině entit je možné nadále s vysokou úspěšností detekovat nové entity.

Navržený algoritmus dosáhl nejlépe AUC 0.732 nad sémantickými entitami vlakových nádraží, resp. 0.869 nad entitami typů vlaků. Při porovnání výsledku detekce nad první nejlepší hypotézu ze *slovních* mřížek, která dosahovala AUC 0.785, resp. AUC 0.883, je však vhodné připomenout, že k tvorbě slovních rozpoznávačů vyžadujeme jazykový model vytvořený z mnohem většího množství dat, než které využíváme u námi navržené metody.

Navržená metoda je tedy vhodná pro detekci klíčových frázích v oblastech, pro které nemáme k dispozici dostatek dat k vytvoření jazykového modelu.