Vyhledávání slov v rozsáhlém archivu mluvené řeči

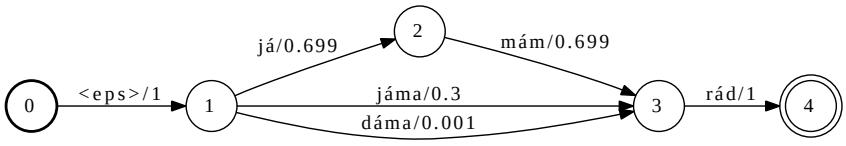
Jan Vavruška[1]

1 Úvod

Úloha vyhledávání slov v rozsáhlém archivu mluvené řeči (angl.: STD - Spoken Term Detection) představuje spojení dvou odvětví oboru Umělé inteligence: vyhledávání informací (IR - Information Retrieval) a automatického rozpoznávání řeči (ASR - Automatic Speech Recognition). Cílem IR systému je zprostředkovat uživateli přístup k datům z řečového archivu na základě nějakého jeho dotazu. Jeho hlavní komponentu představuje *index*, což je nějaká datová struktura, obsahující všechny relevantní informace o takovém archivu. STD systém tedy indexuje výstup systému rozpoznávání řeči, který je reprezentován formou tzv. *slovních* a *fonémových mřížek*. Hlavním cílem práce bylo nastudovat a aplikovat přístup k indexaci a vyhledávání v těchto mřížkách, jak je popsaný v [Can a Saraclar (2010)]. Následně jej pak otestovat na zvolených experimentálních datech a vyhodnotit jeho přínos z hlediska přesnosti, rychlosti vyhledávání a nároků na datový prostor.

2 Indexace řečových archivů s pomocí vážených konečných transducerů

V typických úlohách ASR v reálném čase, je výsledkem jedno nejlepší řešení posloupnosti rozpoznaných slov. V případě offline rozpoznávání řeči z archivu je výsledkem n nejlepších slovních hypotéz. Výstup pak může být reprezentován formou hranově ohodnoceného, orientovaného a topologicky uspořádaného grafu (podle času). Jeho uzly vymezují začátky a konce slov a hrany představují jednotlivá slova. Ohodnocení hran tvoří pravděpodobnosti vyslovení slova v příslušném časovém intervalu. Takovéto grafy se nazývají *slovní mřížky*. Ukázkový příklad vidíme na obrázku 1. Jiným typem může být *fonémová mřížka*, jejímž hranám namísto slov odpovídají jednotlivé fonémy, tj. hlásky. Narozdíl od slovních, nepotřebuje systém ASR při jejich generování znalost jazykového modelu (odhadované posloupnosti slov) a mřížky jsou tak výsledkem pouze akustické analýzy řečového signálu (akustického modelu řeči).



Obrázek 1: Ukázkový příklad slovní mřížky ve formátu WFST.

Mřížka strukturou odpovídá váženému konečnému automatu, přesněji váženému konečnému transduceru (převodníku), dále jen WFST (angl.: Weighted Finite-State Transducer). Transducer je obecně automat, obsahující množinu stavů a přechodů mezi nimi, abecedu vstupních a výstupních symbolů a také ohodnocení jednotlivých přechodů. Automat přejde ze svého počátečního do koncového stavu tehdy, jestliže existuje nějaká cesta mezi těmito stavy, jejíž vstupní symboly jednotlivých hran odpovídají posloupnosti symbolů posílaných na vstup

[1] student navazujícího studijního programu Aplikované vědy a informatika, obor Kybernetika a řídící technika, specializace Umělá inteligence, e-mail: sandokan@students.zcu.cz

automatu. Přitom současně automat zobrazí na výstup posloupnost symbolů, odpovídajících výstupním symbolům této cesty. Tato cesta je ohodnocena váhou, danou součtem vah dílčích hran náležejících této cestě. Potom říkáme, že vstupní posloupnost symbolů je tímto automatem přijata.

2.1 Index slovních mřížek

Index slovních mřížek z celého řečového archivu, popsaný v práci [Can a Saraclar (2010)], velmi zjednodušeně představuje sjednocení všech těchto mřížek ve formě WFST do jednoho společného transduceru, jehož množinu vstupních symbolů tvoří rozpoznaná slova z celého indexu (slova ze slovníku systému ASR). Tento automat zobrazí hledanou posloupnost slov (odpovídá-li jí nějaká cesta) na jeho výstupní symboly, které představují číselné identifikátory jednotlivých mřížek (archivovaných řečových promluv) ve kterých se hledané slovo nachází. Váhu této cesty tvoří trojice {pravděpodobnost slova, počáteční čas, koncový čas} jeho výskytu v odpovídající mřížce.

2.2 Index fonémových mřížek

Výše popsaný index nad slovními mřížkami je použitelný pro vyhledávání slov, která se nacházejí ve slovníku systému ASR. Pokud se ale jedná o slova mimo slovník (angl.: OOV, out-of-vocabulary), systém je v řeči nerozpozná a ve slovních mřížkách se nebudou nacházet. Tento problém se řeší fonetickou transkripcí takového slova a jeho následným vyhledáním v indexu fonémových mřížek. Výsledky jsou analogické s předchozími, pouze namísto posloupnosti slov zde odpovídají posloupnosti fonémů, reprezentujících hledané slovo. Takovýto index na úrovni subslovních jednotek obsahuje řádově daleko více stavů a přechodů, než index celých slov.

3 Zhodnocení systému a budoucí práce

V porovnání se systémem popsaným v práci [Psutka et al. (2011)], který namísto celých mřížek indexuje jen jejich jednotlivé hrany s nejlepším ohodnocením, poskytuje systém s WFST při srovnatelné nebo i lepší přenosti detekce (fonémové mřížky) daleko rychlejší časy vyhledávání a navíc umožňuje vyhledat podřetězce libovolné délky. Problematické jsou pouze jeho nároky na úložný prostor, zejména v případě fonémového indexu. Moje budoucí práce bude tedy pokračovat hlavně ve směru optimalizace těchto nároků a pozdějšího začlenění hotového vyhledávání do systému [Psutka et al. (2011)].

Poděkování

Příspěvek byl podpořen Ministerstvem Kultury České Republiky, grantovým projektem č. DF12P01OVV022.

Literatura

Can, Dogan; Saraclar, Murat; 2010. Timed Indexation of Weighted Automata - Application to Spoken Term Detection. *IEEE Transactions on Audio, Speech and Language processing, vol. 18, No. 8*.

Psutka, J.; Švec, J. Psutka, J. V.; Vaněk, J.; Pražák, A.; Šmídl, L.; Ircing P.; 2001. System for fast lexical and phonetic spoken term detection in a Czech cultural heritage archive. *EURASIP Journal on Audio, Speech, and Music Processing*.