



# Kontextový XGBoost algoritmus pro detekci hlasivkových pulsů

Michal Vraštil<sup>1</sup>

## 1 Úvod

Strojové učení, konkrétněji hluboké strojové učení, postupně nahrazuje klasické dosud používané analytické metody. Nejinak je tomu i v případě metod pro zpracování řeči. Jedna z takových úloh je i detekce hlasivkových pulsů v řečovém signálu. Hlasivkový puls je úzký okamžik excitace hlasivek, uložených v hrtanu, které se podílejí na tvorbě znělé řeči. Jejich přesná detekce má velký význam pro zlepšení kvality například konkatenační syntézy, detekce řečníka nebo metod pro prozodickou úpravu hlasu.

## 2 Použité klasifikátory, příznaky, předzpracování a data

XGBoost je algoritmus strojového učení s učitelem, který je postaven nad algoritmem Gradient boosting. Pro experimenty byla použita jeho implementace v jazyce Python ve stejnojmenné knihovně. Tento algoritmus využívá jednotlivých rozhodovacích stromů, které jsou trénovány a na základě jejich reziduí po klasifikaci jsou určeny parametry dalšího stromu v pořadí.

Kontextový XGBoost algoritmus byl implementován v (Vraštil, M. (2020)). Jedná se o rozšíření XGBoost algoritmu. V tomto případě jsou zapojeny dva XGBoost algoritmy sériově a první z nich před-klasifikovává kandidáty na hlasivkový puls a připojuje do datasetu informaci o jejich kontextu. Důkladný popis je v (Vraštil, M. (2020)). Trénovací dataset se sestával z 88 řečových nahrávek se vzorkovací frekvencí 16 kHz o celkové době trvání cirká 11,5 minuty a obsahoval řečové nahrávky řečníků různého pohlaví a národnosti, například české, slovenské, německé nebo francouzské. Testovací dataset se pak sestával z 20 nahrávek ze stejné množiny, ale pro klasifikátor dosud neviděné, o celkové velikosti 2,5 minuty. Dále byla k dispozici, jako informace od učitele, pozice hlasivkových pulsů v signálu od experta a signál z elektroglotografu.

Signál byl předzpracován filtrem typu dolní propust' a byla srovnána hlasitost nahrávek. Počítané příznaky a způsob předzpracování jsou více popsány v (Vraštil, M. (2020)). Namátkou lze zmínit například příznaky v časové oblasti, jako amplituda pozitivních a negativních vrcholů, jejich šířka a korelace a poměr dvou o po sobě jdoucích vrcholů se stejnou polaritou. Ve frekvenční oblasti poté byly počítány příznaky jako logaritmus energie nebo melovské kepstrální koeficienty.

## 3 Experimenty

Klasifikátor byl nejprve natrénován na všech trénovacích datech a následně byl ověřen na testovacím datasetu na vybraných metrikách, jako například skóre F1, přesnost nebo Briérov

<sup>1</sup> student navazujícího doktorského studijního programu Aplikované vědy a informatika, obor Kybernetika, email: vrastilm@ntis.zcu.cz

**Tabulka 1:** Porovnání úspěšnosti klasifikátoru XGBoost s klasickými příznaky (XGB) a kontextový klasifikátor s kontextem 7 a všemi přidanými příznaky (K7-XGB).

Metrika	XGB [%]	K7-XGB [%]
F1	98,261	<b>98,416</b>
Precision	99,032	<b>99,053</b>
Avg. prec.	<b>99,923</b>	99,913
Recall	97,502	<b>97,788</b>
Accuracy	98,166	<b>98,328</b>
Bal. accu.	98,210	<b>98,364</b>
ROC AUC	99,885	<b>99,898</b>
Brier score	0,014	0,014

**Tabulka 2:** Kontingenční tabulka pro McNemarův test pro nejlepší kontextový klasifikátor (K7-XGB) a dosavadní nejlepší XGBoost (XGB).

	Úspěch K7-XGB	Selhání K7-XGB
Úspěch XGB	19918	47
Selhání XGB	80	293

skóre a další. Výsledky jsou shrnuty v následující Tabulce 1. Byl také vyhodnocen McNemarův test statistické významnosti výsledků. Optimální velikost kontextu byla určena separátním experimentem v (Vraštil, M. (2020)). Výsledky tohoto experimentu jsou v Tabulce 1.

## 4 Výsledky a závěr

Jak je patrné z Tabulky 1, kontextový klasifikátor dosahuje lepších výsledků, nežli jeho nekontextová verze a to v 6 z 8 měřených metrik. Tento výsledek je statisticky významný na hladině významnosti 5%, viz Tabulku 2, z níž lze spočítat  $\chi^2 = 8,063, p = 0,005$ . Tento kontextový klasifikátor byl posléze použit v dalších experimentech, jako například (Vraštil, M. (2020)), kde byl porovnán při použití jiného druhu předzpracování a ověřen na zašumělých datech. V (Matoušek, J. - Vraštil, M. (2020)) byl tento klasifikátor ověřen proti ostatním klasickým algoritmům a například v (Vraštil, M. - Matoušek, J. (2021)) byl porovnán s konvoluční neuronovou sítí, kde dosahoval lehce lepších výsledků.

## Literatura

- Vraštil M. Detekce hlasivkových pulsů v řečovém signálu pomocí strojového učení. Diplomová práce. Západočeská univerzita v Plzni, 2020
- Matoušek, J. - Vraštil, M. Context-Aware XGBoost for Glottal Closure InstantDetection in Speech Signal. In *TSD 2020, 23rd International Conference on Text, Speech and Dialogue*, 2020.
- Vraštil, M. - Matoušek, J. On Comparison of XGBoost and Convolutional NeuralNetworks for Glottal Closure Instant Detection. In *TSD 2021, 23rd International Conference on Text, Speech and Dialogue*, 2021. [Přijato].