



Strojové učení pro zpracování obrazu a řeči

Jan Švec

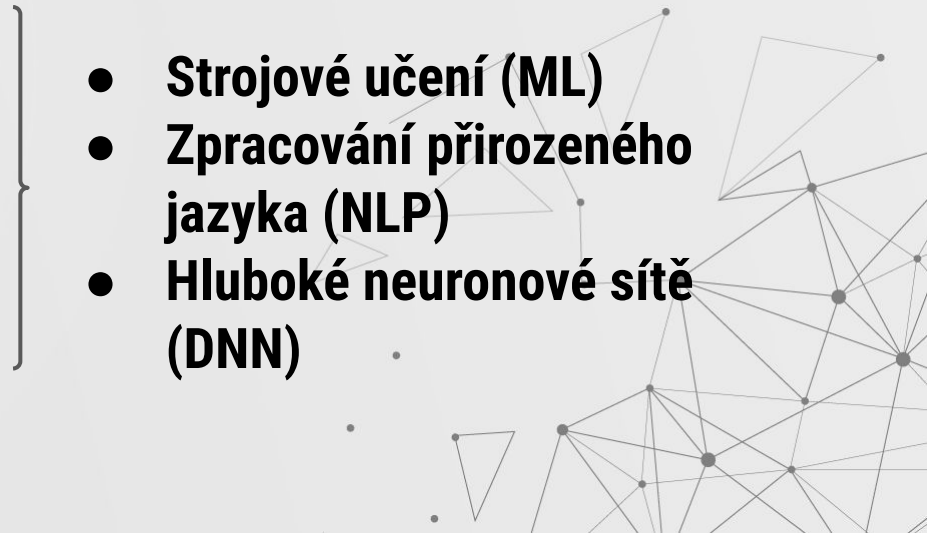
Katedra kybernetiky FAV ZČU / NTIS FAV ZČU

honzas@ntis.zcu.cz

Katedra kybernetiky FAV

Západočeská univerzita v Plzni

- Teorie řízení a řídicí systémy, biokybernetika, **oddělení umělé inteligence**
- **Oblasti umělé inteligence (AI)**
 - Rozpoznávání řeči
 - Syntéza řeči
 - Dialogové systémy a archivy
 - Počítačové vidění
- **Strojové učení (ML)**
- **Zpracování přirozeného jazyka (NLP)**
- **Hluboké neuronové sítě (DNN)**



Strojové učení v kostce

“Téměř každá úloha strojového učení je zobrazení z vektorů na vektory.”

Příznakový vektor (číselná struktura):

$[2, 4, 4, 1.5]^T$

Počet světel
Počet kol
Délka v metrech
Šířka v metrech



Autobus

Osobní auto

Nákladňák

Výstupní
pravděpodobnosti

$[0.01, 0.99, 0.00]^T$

Klasifikace

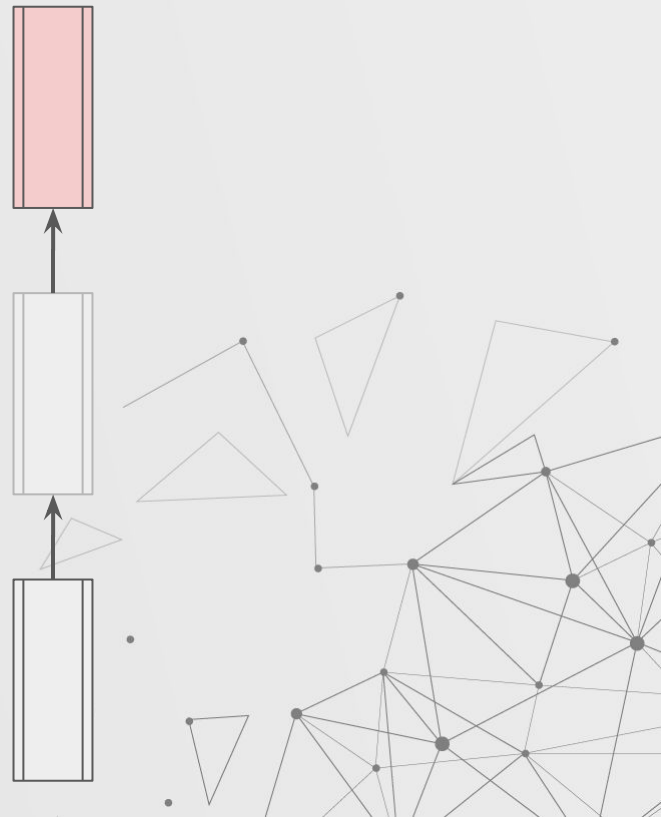
Cílový vektor

- výsledek klasifikace

(volitelně) Mezilehlá
vektorová reprezentace

Vstupní příznakový vektor

- popisuje objekt



Klasifikace posloupností

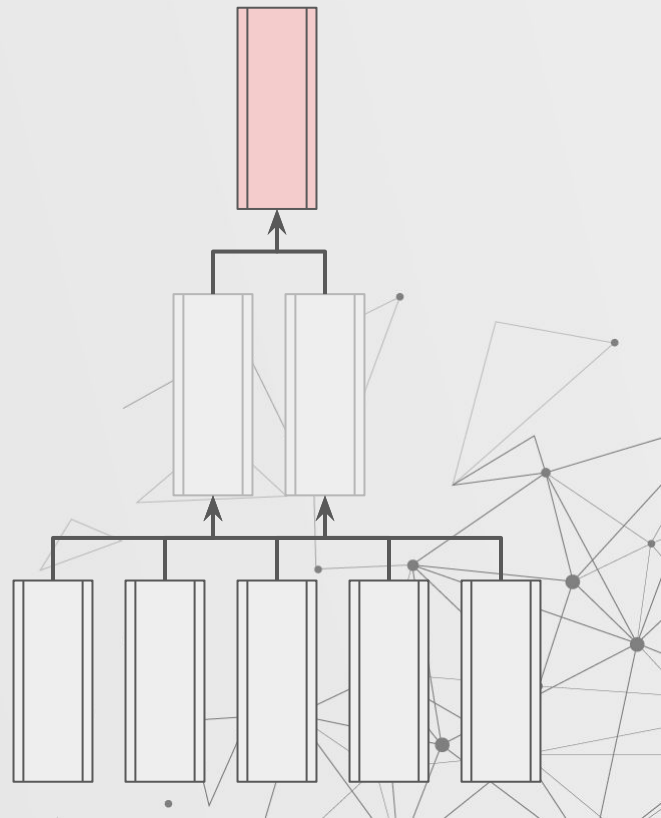
Cílový vektor

- výsledek klasifikace

(volitelně) Mezilehlá vektorová reprezentace
(může být i posloupnost)

Posloupnost příznakových vektorů

- popisuje objekt



Sequence-to-sequence mapování

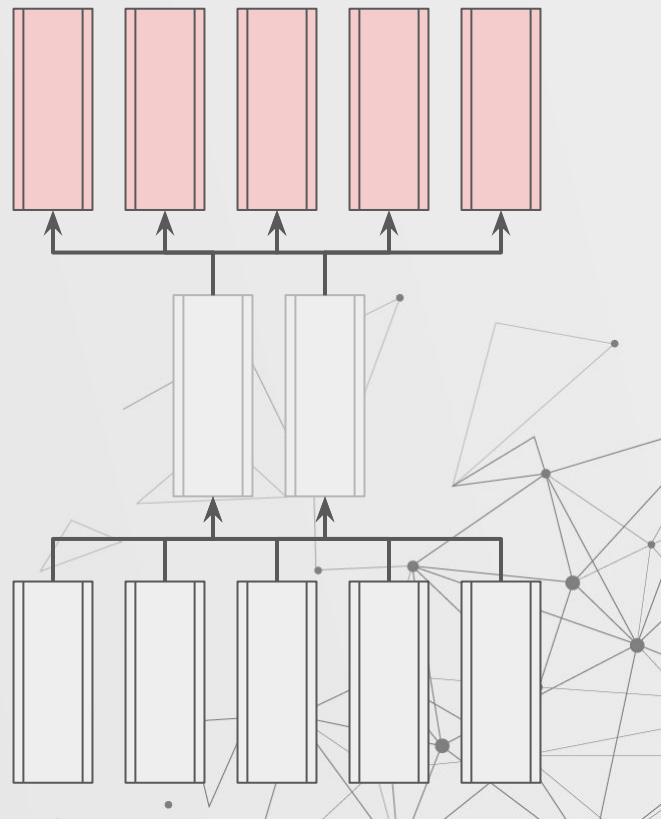
Posloupnost cílových vektorů

- výsledek klasifikace

(volitelně) Mezilehlá vektorová reprezentace
(může být i posloupnost)

Posloupnost příznakových vektorů

- popisuje objekt



Vývoj algoritmů v oblasti zpracování přirozeného jazyka

- **Ručně navržená pravidla**

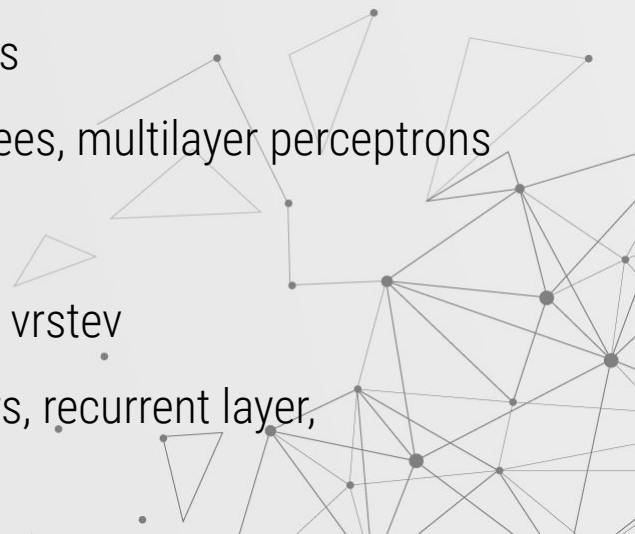
- Pravidla IF-THEN, parametry odhadnuté expertem / z dat

- **Klasické ML algoritmy**

- lineární - logistic regression, Bayes classifier, HMMs
- non-lineární - support vector machines, decision trees, multilayer perceptrons


- **Hluboké neuronové sítě**

- modulární architektury založené na různých typech vrstev
- convolutional layers, residual layers, attention layers, recurrent layer, normalization layers, highway connections



Data pro strojové učení

<i>Velikost trénovacích dat</i>	<i>Počet příkladů</i>	<i>Typ klasifikátoru</i>	<i>Počet parametrů</i>
Malá data <i>Jednotlivé objekty</i>	(stovky)	Ruční pravidla	(jednotky)
Střední data	(tisíce)	Linerární modely	(stovky)
Velká data <i>Posloupnosti</i>	(miliony)	Nelineární modely	(tisíce)
Obrovská data <i>Mnohorozměrné objekty</i>	(miliardy)	Hluboké neuronové sítě	(miliony)
		Hlubší, širší DNN	(miliardy)



“There is no data like more data”



Evolve příznakových vektorů

- **Žádná reprezentace**

- Ruční IF-THEN pravidla. Pracují, ale...

“Every time I fire a linguist, the performance of the speech recognizer goes up.”

Frederick Jelinek (IBM Research), mid. 1970

- **Pečlivě navržené příznakové vektory**

- Ruční IF-THEN pravidla, ale upravená do podoby příznakových vektorů
 - Je vše potřebné součástí příznaků?
 - Mají příznaky stejnou škálu?
 - Mohou nějaké příznaky chybět, co s nimi?



Evolve příznakových vektorů (pravidla)

IF word is **“read”**

THEN pronunciation is **“/ɹiːd/”**

I couldn't read the book I read yesterday.



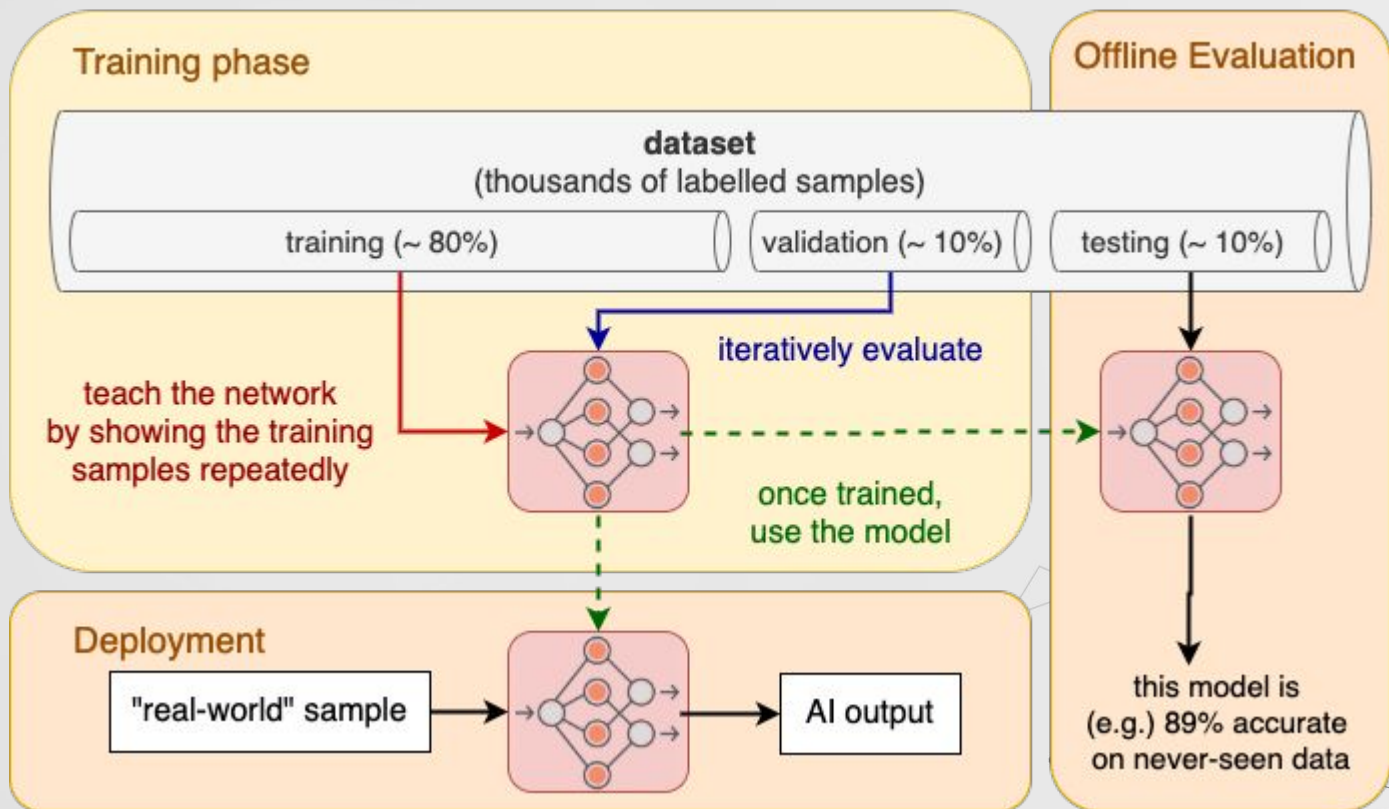
Evolve příznakových vektorů (pravidla, mnohem častější případ)

```
IF (word is "read"  
    and context is "couldn't"  
    and not something else ...)  
THEN pronunciation is "/rɪd/"  
ELSE pronunciation is "/rɛd/"
```

I couldn't read the book I read yesterday.



Trénovací data pro strojové učení



Evoluce příznakových vektorů

- **Natréované příznaky**

- pro data o vysokých dimenzích (mnoho různých slov, velké obrázky), počátky v hlubokých neuronových sítích pro zpracování obrazu

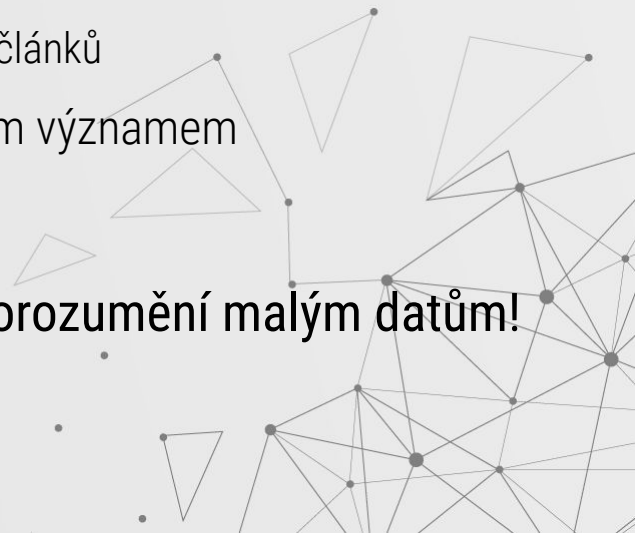
- **Reprezentace textu**

- **Spojité vektory** (word embedding)
 - metody, jak převést izolovaná slova na spojitě vyjádření pomocí příznaků
 - Tomáš Mikolov: **word2vec**
- **Kontextově závislé reprezentace**
 - multi-headed self-attention networks → **Transformer architecture**



Transfer learning pre-training & fine-tuning

- **Před-trénování** modelu na **velkých / obrovských datech** (umělá úloha)
- **Vyladění** modelu na **malých datech** (cílová úloha)
 - Pro definovanou **cílovou úlohu** s využitím **významové reprezentace**
 - Například: strojový překlad, generování novinových článků
 - Schopnost reprezentace synonym a frází se stejným významem
- **Proč?** Můžeme použít znalost z velkých dat pro porozumění malým datům!



T5

(Text-to-text transfer transformer)

Transformer model předtrénovaný na webových datech (Google)

Pro každou větu se část skryje a cílem modelu je větu zrestaurovat:

I have **<1>** and a cat → **<1> a dog**

Předtrénované modely vhodné pro **strojový překlad, porozumění, generování**

- ACL IMDB sentiment analysis

- https://github.com/honzas83/t5s/blob/main/examples/t5s_aclimdb.ipynb

- T5litik

- <https://github.com/honzas83/t5s/blob/main/examples/t5litik.ipynb>

T5 Politik

Koronavirus je zde celých 12 měsíců, situace se nelepší...

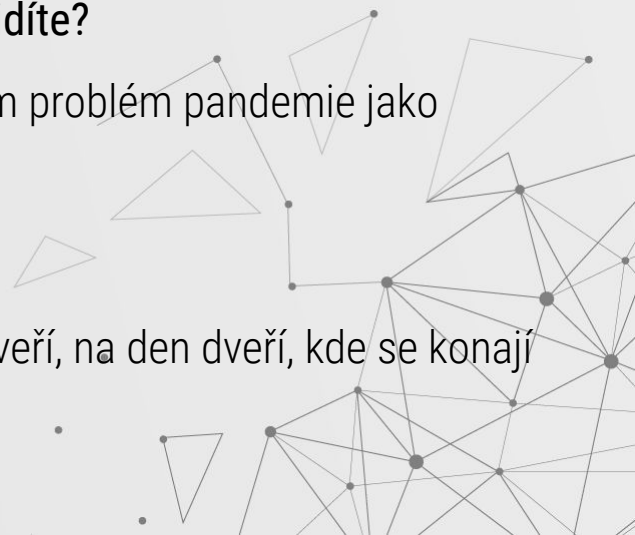
Tak já bych řekl, že situace se zlepšila výrazně, ale já bych to řekl tak, že se situace lepší už jenom z toho důvodu, že je relativně velmi malá pravděpodobnost, že by se ten virus mohl šířit dál.

Chtěl bych se zeptat na váš názor na aktuální pandemii, jak to vidíte?

Tak já si myslím, že ta pandemie je opravdu vážná a není to jenom problém pandemie jako takové, je to i celosvětový problém.

A co byste vzkázal studentům, kteří nás teď poslouchají?

No, abych jim to vysvětlil, tak bych je pozval na den otevřených dveří, na den dveří, kde se konají konference o sexuálním obtěžování a podobně.



Naše AI aplikace Computer vision

Mobilní aplikace

200 druhů hub
Hluboké neuronové
sítě

[Google Play link](#)



Aplikace na houby


Multimodální zpracování

- Zvuk, video, obraz, text
- Indexace a vyhledávání



<https://cemi.zcu.cz/>

Audiovizuální archivy



Naše AI aplikace

Syntéza řeči (KKY & SpeechTech s.ro.)

Dopravní podniky

Praha, Ostrava, Brno (interní služba), Plzeň, Ústí nad Labem

Integrovaná a veřejná doprava

ROPID – Praha, POVED – Plzeňský kraj, KORID – Liberecký kraj

Makro ČR, Olympia ČR, Premium Outlet Prague Airport

Pacienti, kteří mohou přijít o schopnost mluvit

Nahrání a uchování vzorků hlasu

Použití syntézy řeči z textu pro vygenerování libovolné promluvy

[On-line demo SpeechTech TTS](#)

In-store rádia
Veřejná doprava

Konzervace hlasu



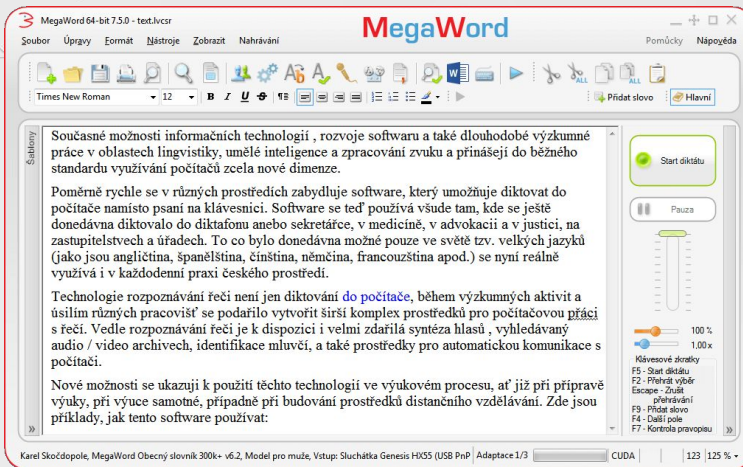
Naše AI aplikace

Syntéza řeči (KKY & SpeechTech s.ro.)



Naše AI aplikace

Rozpoznávání řeči (KKY & SpeechTech s.ro.)



Diktovací software
MegaWord

Přímé využití rozpoznávání řeči

Pořady s čistou audiostopou
Jeden řečník

Technologie stínových řečníků

Diskuzní pořady
Sportovní přenosy
Tvorba high-quality titulků

<https://www.zivetitulky.cz/>

Živé titulky pro televizní
pořady ČT

Naše AI aplikace

Rozpoznávání řeči (KKY & SpeechTech s.ro.)



Naše AI aplikace

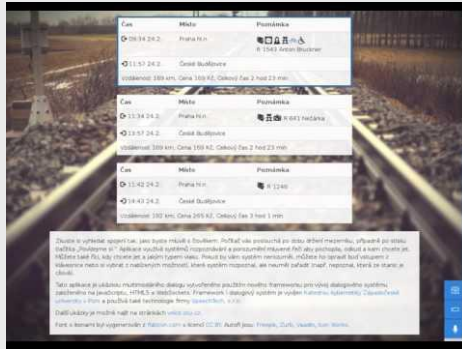
Hlasové dialogy (KKY & SpeechTech s.ro.)

Propojení rozpoznávání a syntézy řeči

Konverzační voice-bot

Python, jednoduchý skript

Propojení do webových aplikací

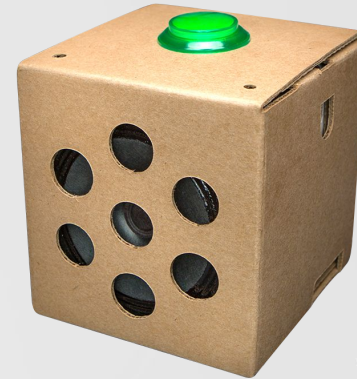


Platforma SpeechCloud

RaspberryPi pro voice-boty

AD/DA převodníky, tlačítko

Propojení na SpeechCloud



VoiceKit

Děkuji za pozornost (a těším se na setkání)

Studium na Katedře kybernetiky FAV ZČU

Studijní program Umělá inteligence a automatizace

#AimtecHackathon Challenge

Sledujte 14.-15.5.2021 <https://hackathon.aimtecglobal.com/challenge>

Plzeňské pyvo

Série workshopů o Pythonu od Plzeňáků pro Plzeňáky

#20 již dnes 12.5.2021 <https://pyvo.cz/plzen-pyvo/>

Přednášky a workshopy





Strojové učení pro zpracování obrazu a řeči

Jan Švec

Katedra kybernetiky FAV ZČU / NTIS FAV ZČU

honzas@ntis.zcu.cz