

Predikcia zhoršenia zdravotného stavu

Marián Kravec

školiťel': MSc. František Dráček

Obsah

Motivácia a ciele práce

Dáta

Embedding záznamov

Predikcia budúcich nákladov na pacienta

Príprava na otázky s posudkov

Predikcia zhoršenia
zdravotného stavu

Motivácia a ciele práce

Dáta

Embedding záznamov

Predikcia budúcich
nákladov na pacienta

Príprava na otázky s
posudkov

Motivácia a ciele práce

Predikcia zhoršenia
zdravotného stavu

Motivácia a ciele práce

Dáta

Embedding záznamov

Predikcia budúcich
nákladov na pacienta

Príprava na otázky s
posudkov

- ▶ Slovenské zdravotníctvo má nízku mieru využitia dostupných dát.
- ▶ Predikcia budúcich nákladov pacienta môže zlepšiť plánovanie a prerozdelenie zdrojov.
- ▶ Ciele:
 - ▶ Navrhnuť a implementovať spôsob transformácie záznamov pacienta do číselných vektorov (embedding).
 - ▶ Navrhnuť a implementovať systém na predikciu budúcich nákladov pacienta na základe jeho historických záznamov.
- ▶ Výsledky práce:
 - ▶ Embedované atribúty spĺňali vlastnosť blízkosti podobných záznamov.
 - ▶ Systém na predikciu budúcej nákladovej skupiny pacienta mal presnosť 39.9% (84.4% v prípade povolenia chyby o jednu kategóriu).

Obsah

Motivácia a ciele práce

Dáta

Embedding záznamov

Predikcia budúcich nákladov na pacienta

Príprava na otázky s posudkov

Predikcia zhoršenia
zdravotného stavu

Motivácia a ciele práce

Dáta

Embedding záznamov

Predikcia budúcich
nákladov na pacienta

Príprava na otázky s
posudkov

Dáta

- ▶ Umelo generované dáta podľa štruktúry NCZI: 173 355 pacientov, 133 miliónov záznamov.
- ▶ Typ dát: poistné dáta.
- ▶ Dve skupiny záznamov (datasets):
 - ▶ Výkony ambulantnej zdravotnej starostlivosti
 - ▶ Predpísané lieky
- ▶ Jeden záznam: informácia o jednom výkone/lieku konkrétneho pacienta.
- ▶ Použité atribúty:
 - ▶ Dátum záznamu
 - ▶ Vek pacienta
 - ▶ Medicínsky výkon
 - ▶ Liek
 - ▶ Diagnóza
 - ▶ Cena výkonu/lieku

Predikcia zhoršenia
zdravotného stavu

Motivácia a ciele práce

Dáta

Embedding záznamov

Predikcia budúcich
nákladov na pacienta

Príprava na otázky s
posudkov

Obsah

Motivácia a ciele práce

Dáta

Embedding záznamov

Predikcia budúcich nákladov na pacienta

Príprava na otázky s posudkov

Predikcia zhoršenia
zdravotného stavu

Motivácia a ciele práce

Dáta

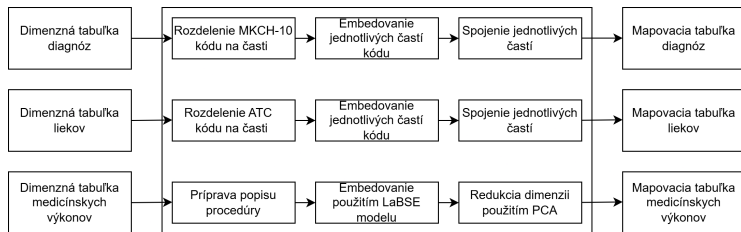
Embedding záznamov

Predikcia budúcich
nákladov na pacienta

Príprava na otázky s
posudkov

Vytváranie embeddingov záznamov

- Každý záznam reprezentovaný vektorom (embedding) z 4 častí:
 - Časová pečiatka (vek pri udalosti)
 - Diagnóza (MKCH-10, hierarchické embeddingy)
 - Liek (ATC, hierarchické embeddingy)
 - Výkon (LaBSE embedding popisu + PCA)
- Cieľ: podobné záznamy majú podobné vektory (Euclidovská vzdialenosť).
- Overenie: klastrovanie, výpočty podobností.



Predikcia zhoršenia
zdravotného stavu

Motivácia a ciele práce

Dáta

Embedding záznamov

Predikcia budúcich
nákladov na pacienta

Príprava na otázky s
posudkov

Časová pečiatka

- ▶ Odhad veku pacienta v dňoch v čase záznamu.
- ▶ Určená podľa veku pacienta v rokoch v čase prvého záznamu a časovým rozdielom (v dňoch) medzi dátumami prvého a daného záznamu.
- ▶ Chyba: v priemere štvrtý rok, najviac polrok.

Predikcia zhoršenia
zdravotného stavu

Motivácia a ciele práce

Dáta

Embedding záznamov

Predikcia budúcich
nákladov na pacienta

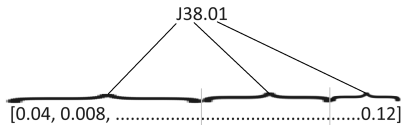
Príprava na otázky s
posudkov

Diagnózy

- ▶ Embedujeme MKCH-10 kódy.

[znak][dvojčísle].[dvojčísle]

- ▶ Hierarchický kód.
- ▶ Každá časť kódovaná samostatne a následne spojené dohromady.
- ▶ Príklad:
 - ▶ J - Choroby dýchacej sústavy
 - ▶ J3 - Iné choroby dýchacích ciest
 - ▶ J38 - Choroba hlasiviek a hrtana
 - ▶ J38.0 - Obrna hlasiviek a hrtana
 - ▶ J38.01 - Jednostranná čiastočná obrna hlasiviek a hrtana

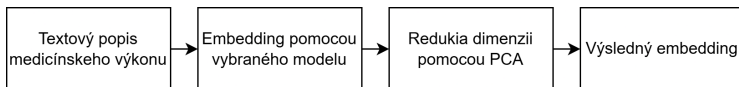


- ▶ Embedujeme ATC kód.

[znak][dvojčísle][dvojica znakov][dvojčísle]

- ▶ Hierarchický kód.
- ▶ Každá časť kódovaná samostatne a následne spojené dohromady.
- ▶ Príklad:
 - ▶ N - Centrálna nervová sústava
 - ▶ N02 - Analgetiká
 - ▶ N02B - Iné analgetiká a antipyretiká
 - ▶ N02BA - Kyselina salicylová a deriváty
 - ▶ N02BA01 - Acylpyrín

- ▶ Embedujeme textový popis výkonu.
- ▶ Vyskúšané modely:
 - ▶ Language-agnostic BERT sentence embedding model (LaBSE)
 - ▶ Lemmatizer + Word2vec model
- ▶ Redukcia dimenzionality pomocou PCA, počet dimenzii vybraný tak aby zachovávali 90% variance.



Validácia embeddingov

Predikcia zhoršenia
zdravotného stavu

Motivácia a ciele práce

Dáta

Embedding záznamov

Predikcia budúcich
nákladov na pacienta

Príprava na otázky s
posudkov

- ▶ Diagnózy a lieky:
 - ▶ Porovnanie podobností (obrátaná hodnota vzdialenosti) dvojíc niekoľkých náhodne vybraných prípadov.
 - ▶ Klastrovanie (K-means) a následné kontrola distribúcií hlavných kategórií prípadov v klastroch.
- ▶ Výkony: klastrovanie a následná vizuálna kontrola obsahu náhodne vybraných klastrov.

Validácia embeddingov

Predikcia zhoršenia
zdravotného stavu

Motivácia a ciele práce

Dáta

Embedding záznamov

Predikcia budúcich
nákladov na pacienta

Príprava na otázky s
posudkov

Diagnózy

Code A	Code B	Similarity
G47.30	G40.09	2.77
G47.30	H40.09	0.53
G47.30	H18.80	0.46
G40.09	H40.09	0.54
G40.09	H18.80	0.45
H40.09	H18.80	0.84

Lieky

Code A	Code B	Similarity
C01EB15	C01CA04	1.24
C01EB15	C10AA07	0.54
C01EB15	J01CA04	0.45
C01CA04	C10AA07	0.64
C01CA04	J01CA04	0.48
C10AA07	J01CA04	0.38

Validácia embeddingov

Medicínske výkony

► Kluster 215:

Výber vhodných príjemcov pre kadaverózne transplantáty z listiny cakatelov

Transplantácia pečene (UH+90301)

Transplantácia obličky (UH+90101)

Transplantácia obličky

Voľná transplantácia šliach

Odobratie chrupkového alebo kostného materiálu na voľnú transplantáciu

Odber pečene na transplantáciu

Transplantácia pečene

Transplantácia pankreasu

Odobratie orgánov alebo časti orgánov na transplantáciu: Pankreas

Odobratie orgánov alebo časti orgánov na transplantáciu: Oblička

Odobratie orgánov alebo časti orgánov na transplantáciu: Srdce

Odobratie orgánov alebo časti orgánov na transplantáciu: Kostná dreň

Odobratie orgánov alebo časti orgánov na transplantáciu: Rohovka

Odobratie orgánov alebo časti orgánov na transplantáciu: Týmus

Odobratie kostného alebo chrupkového materiálu na transplantáciu

Odber kostnej drene na účely transplantácie

Indikácia darcu na odber orgánov na transplantáciu

Celotelové ožarovanie pre transplantáciu kostnej drene

Transplantácia obličiek

Transplantácia srdca

Transplantácia pečene

Transplantácia pankreasu

Transplantácia pľúc

Transplantácia rohovky

Transplantácia skléry

Transplantácia sklery - naklady suvisiace s odberom sklery

Voľný šlachový transplantát

voľný šlachový transplantát

Transplantácia kostnej drene

Predikcia zhoršenia
zdravotného stavu

Motivácia a ciele práce

Údaje

Embedding záznamov

Predikcia budúcich
nákladov na pacienta

Príprava na otázky s
posudkov

Medicínske výkony

► Kluster 45:

Rozbor a plánovanie (komplexná analýza).

Zhodnotenie výsledkov komplexného hemokoagulacného vyšetrenia a klinická interpretácia porúch

Vyhodnotenie KOS a záverečná správa.

Zhodnotenie výsledkov a záver

Resekcia močovodu a reanostomáza

Resekcia a rekonštrukcia žlčových ciest pri nádoroch

Vyhodnotenie KOS a záverečná správa

Vyhodnotenie sociálnej starostlivosti a záverečná správa

vWF antigén - vyšetrenie farmakokinetiky a monitorovanie liečby

vWF Ricof - vyšetrenie farmakokinetiky a monitorovanie liečby

Faktor VII - vyšetrenie farmakokinetiky a monitorovanie liečby

Faktor VIII - vyšetrenie farmakokinetiky a monitorovanie liečby

Faktor IX - vyšetrenie farmakokinetiky a monitorovanie liečby

Počítačové zhodnotenie polysomnografického záznamu a zhodnotenie lekárom

Papilosfinkterotómia a odstránenie konkrementov zo žlčových ciest

Papilosfinkterotómia a odstránenie konkrementov

Papilosfinkterektómia a odstránenie konkrementov zo žlčových ciest alebo pankreatického vývodu

SVLZ - Spoločné vyšetrovacie a liečebné zložky

Obsah

Motivácia a ciele práce

Dáta

Embedding záznamov

Predikcia budúcich nákladov na pacienta

Príprava na otázky s posudkov

Predikcia zhoršenia
zdravotného stavu

Motivácia a ciele práce

Dáta

Embedding záznamov

Predikcia budúcich
nákladov na pacienta

Príprava na otázky s
posudkov

Proces predikcie

Predikcia zhoršenia
zdravotného stavu

Motivácia a ciele práce

Dáta

Embedding záznamov

Predikcia budúcich
nákladov na pacienta

Príprava na otázky s
posudkov

► Kroky pri predikovaní budúcej cenovej kategórie pacienta:

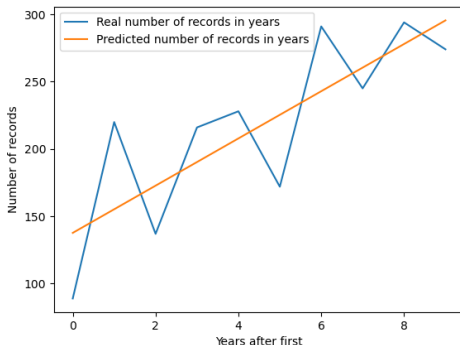
1. Načítanie dát pacienta, embedding a normalizácia.
2. Výpočet počtu záznamov na budúci rok.
3. Predikcia budúcich záznamov.
4. Predikcia cenových kategórii budúcich záznamov.
5. Výpočet cenovej kategórie pacienta.



- Hyperparametre modelov nastavované lokálne na podskupine pacientov.
- Finálne modely trénované na serveri na plnohodnotnom datasete.

Výpočet počtu budúcich záznamov

- ▶ Testované metódy:
 - ▶ Polynomialna regresia počtu záznamov z predchádzajúcich rokov.
 - ▶ Generovanie nových záznamov kým generovaná časová pečiatka nepresiahne rok od posledného skutočného záznamu.
- ▶ Najlepší výsledok: lineárna regresia.



Predikcia zhoršenia
zdravotného stavu

Motivácia a ciele práce

Dáta

Embedding záznamov

Predikcia budúcich
nákladov na pacienta

Príprava na otázky s
posudkov

Predikcia budúcich záznamov

► Testované modely:

- Long short-term memory.
- Decoder-only transformer.

► Vstup: posledných n (veľkosť okna/kontextu) záznamov pacienta.

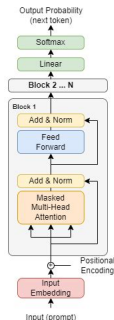
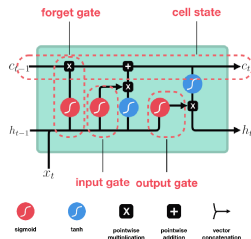
► Výstup: predikcia nasledujúceho záznamu.

► Chybová funkcia: Subpart weighted MSE

$$SubpartWeightedMSE = \frac{1}{p} \sum_{j=1}^p \left(\frac{1}{l_j} \sum_{i=1}^{l_j} (Y_{s_j+i} - \hat{Y}_{s_j+i})^2 \right).$$

► Optimalizované hyperparametre:

- Hĺbka modelu (počet vrstiev).
- Šírka modelu (LSTM).
- Počet hláv (Transformer).
- Dropout rate.



Predikcia zhoršenia zdravotného stavu

Motivácia a ciele práce

Účel

Embedding záznamov

Predikcia budúcich nákladov na pacienta

Príprava na otázky s posudkov

Predikcia cenovej kategórie záznamu

Predikcia zhoršenia
zdravotného stavu

Motivácia a ciele práce

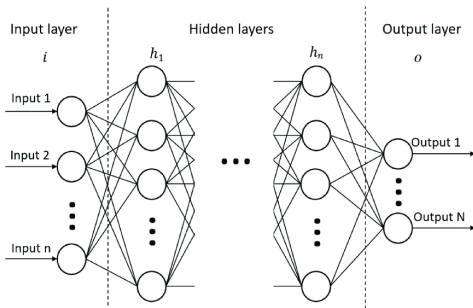
Dáta

Embedding záznamov

Predikcia budúcich
nákladov na pacienta

Príprava na otázky s
posudkov

- Použitý model: Multilayer perceptron.
- Vstup: jeden záznam pacienta.
- Výstup: jedna z 9 cenových kategórii.
- Vyskúšané jednoduchšie modely: Gradient boosting, Ridge regression



Category	Interval
1	$[0,1)$
2	$[1,5)$
3	$[5,10)$
4	$[10,20)$
5	$[20,50)$
6	$[50,100)$
7	$[100,200)$
8	$[200,500)$
9	$[500,\infty)$

Predikcia cenovej kategórie záznamu - trénovanie a validácia

- Optimalizované hyperparametre:
 - Hĺbka modelu.
 - Veľkosti jednotlivých vrstiev.
 - Aktivačné funkcie medzi vrstvami.
 - Chybová funkcia (MSE, Cross Entropy, NLL).
- Miera kvality modelu: presnosť.

Depth	Layer sizes	Activation functions	Mean square error		Cross entropy		Negative log likelihood	
			Test accuracy	Validation accuracy	Test accuracy	Validation accuracy	Test accuracy	Validation accuracy
0			63.2%	63.0%	63.5%	63.3%	62.9%	62.6%
1	[98]	[GELU]	74.5%	74.3%	72.8%	72.6%	73.2%	73.0%
		[Tanh]	71.3%	71.0%	70.9%	70.6%	71.1%	70.9%
	[392]	[GELU]	76.6%	76.3%	74.2%	74.0%	75.5%	75.2%
		[Sigmoid]	70.8%	70.6%	69.9%	69.6%	70.2%	70.0%
	[196]	[GELU]	75.9%	75.6%	74.1%	73.8%	74.1%	73.8%
		[LeakyReLU]	76.0%	75.8%	75.5%	75.3%	74.1%	74.0%
3	[98, 48, 24]	[GELU, GELU, GELU]	74.4%	74.2%	72.0%	71.8%	72.2%	72.0%
		[Sigmoid, ReLU, Tanh]	69.7%	69.5%	69.7%	69.4%	70.3%	70.1%
	[392, 196, 98]	[GELU, GELU, GELU]	77.7%	77.5%	75.2%	75.0%	75.5%	75.3%
		[SiLU, ReLU, GELU]	77.4%	77.2%	74.8%	74.5%	75.2%	75.0%
	[196, 196, 196]	[GELU, GELU, GELU]	77.2%	77.0%	74.7%	74.3%	75.0%	74.7%
		[ReLU, Sigmoid, SiLU]	76.7%	76.5%	73.9%	73.7%	75.0%	74.8%



Predikcia zhoršenia
zdravotného stavu

Motivácia a ciele práce

Dáta

Embedding záznamov

Predikcia budúcich
nákladov na pacienta

Príprava na otázky s
posudkov

Predikcia cenovej kategórie záznamu - trénovanie a validácia

- ▶ Najlepší model pri lokálnom trénovaní:
 - ▶ 8 vrstiev.
 - ▶ Šírky vrstiev: [588, 294, 147, 49, 98, 36, 18, 9].
 - ▶ Aktivačné funkcie medzi vrstvami: [SiLU, GELU, Sigmoid, GELU, SiLU, GELU, LeakyReLU, Softmax].
 - ▶ Chybová funkcia: MSE.
- ▶ Presnosť pri lokálnom trénovaní:
 - ▶ Trénovacia: 78.3%
 - ▶ Validačná: 78.0%
- ▶ Presnosť pri trénovaní na serveri:
 - ▶ Trénovacia: 80.2%
 - ▶ Validačná: 80.1%
- ▶ Validačná presnosť najlepšieho Gradient boosting: 71.4%
- ▶ Validačná presnosť najlepšieho Ridge regression: 66.9%

Predikcia zhoršenia
zdravotného stavu

Motivácia a ciele práce

Dáta

Embedding záznamov

Predikcia budúcich
nákladov na pacienta

Priprava na otázky s
posudkov

Ďakujem za pozornosť

Predikcia zhoršenia
zdravotného stavu

Motivácia a ciele práce

Dáta

Embedding záznamov

Predikcia budúcich
nákladov na pacienta

Príprava na otázky s
posudkov

Obsah

Motivácia a ciele práce

Dáta

Embedding záznamov

Predikcia budúcich nákladov na pacienta

Príprava na otázky s posudkov

Predikcia zhoršenia
zdravotného stavu

Motivácia a ciele práce

Dáta

Embedding záznamov

Predikcia budúcich
nákladov na pacienta

Príprava na otázky s
posudkov



THE BEST THESIS DEFENSE IS A GOOD THESIS OFFENSE.

zdroj: XKCD (<https://xkcd.com/1403/>)