

Neurónové siete

Projekt

Richard Staňa

1. Príprava dát:

Z daných dát sme vytiahli požadované roky na učenie a testovanie. Pre učenie použijeme hodnoty BZ, RMS_BZ_GSE, SW Proton Density a Dst-index. Keďže v odporúčanom roku 1984 sa nachádza veľa chýbajúcich dát a v odporúčanom roku 2001 sa nachádza 14 búrok, z čoho ale 2 sú zdvojené, čo by mohlo ovplyvniť úspešnosť siete, rozhodli sme sa ako trénovací rok použiť rok 2005.

V odporúčanom roku 2007 sa nachádzajú iba 2 búrky. Preto sme sa rozhodli ako testovací rok použiť rok 2011, aby bolo na čom testovať. Ďalším dôvodom takéhoto výberu testovacieho a trénovacieho roka bolo, že po nahradení chýbajúcich dát neostali žiadne chýbajúce hodnoty.

2. Použité metódy:

- Dáta boli predspracované metódou preprocessing, ktorá z dát odstráni chýbajúce hodnoty, ak je ich za sebou menej ako 24. Odstránenie je lineárne vzhľadom na hodnotu pred prvou a po poslednej chýbajúcej hodnote za sebou.
- Metódou najdenieNormalizacnychKonstant boli nájdené normalizačné konštanty pre oba roky spolu tak, že najmenšie nájdené hodnoty boli zaokrúhlené nadol na desiatky a ešte sa od nich odčítalo 10 a najväčšie nájdené hodnoty boli zaokrúhlené nahor na desiatky a ešte sa k nim pričítalo 10. Pre roky 2001 a 2005 boli normalizačné konštanty nasledovné:

min BZ	-60
max BZ	40
min RMS_BZ_GSE	-10
max RMS_BZ_GSE	40
min W Proton Density	-20
max W Proton Density	100
min Dst-index	-280
max Dst-index	70

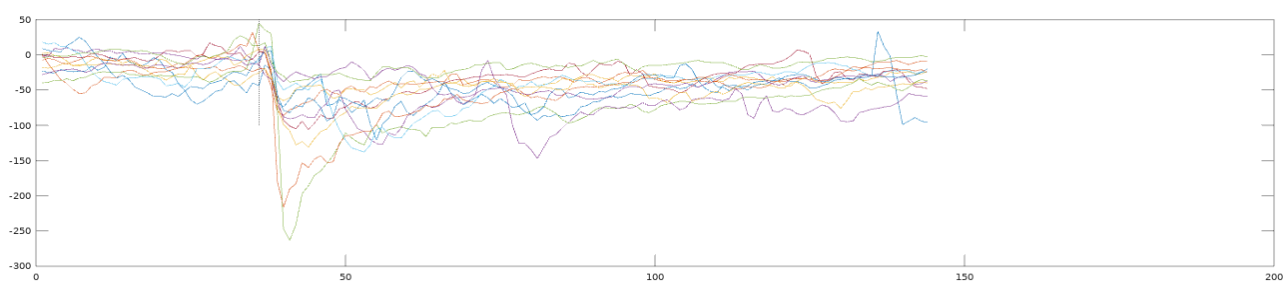
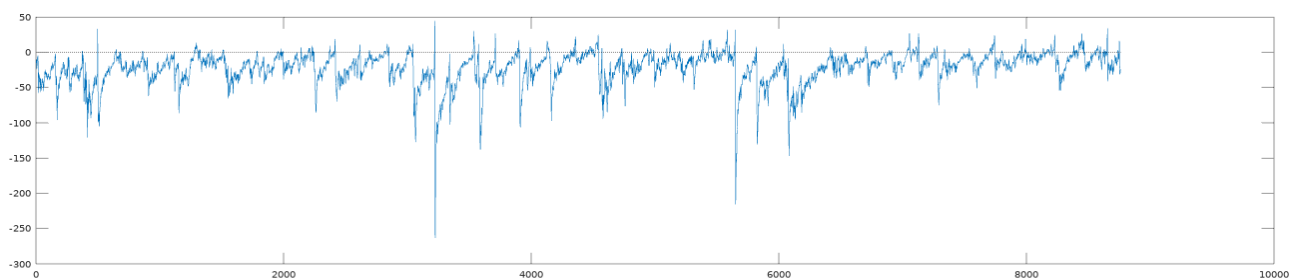
- Metóda hľadanieBurok našla pre jednotlivé roky výskyt búrok v matici v ktorej každých 144 riadkov znamenalo jednu búрку a do konzoly vypísala pozície (hodiny od začiatku roka) začiatkov jednotlivých poklesov Dst-index o viac ako 40.
- Metóda normalizacia prevádzala nájdené výskyty búrok z metódy hľadanieBurok na hodnoty z intervalu (-1, 1) a denormalizacia prevádzala hodnoty, ktoré odhadne sieť z intervalu (-1, 1) naspäť na hodnoty, v ktorých je možné hľadať búrky.

- Metóda hľadanieBurokVOdhadnutychDatach hľadá búrky v denormalizovaných hodnotách, ktoré odhadla sieť.
 - Metódy zadanieA, zadanieB, zadanieC, zadanieD sú samotné implementácie neuronovej siete, ktoré na vstup dostávajú normalizované nájdené búrky v trénovacom roku a normalizovaný testovací rok a vracajú odhadované hodnoty Dst-index v trénovacom roku.
 - Metóda sko vypočíta strednú kvadratickú odchýlku medzi hodnotami Dst-index z testovacieho roka a denormalizovaných hodnôt Dst-index ktoré vypočítala neurónová sieť
3. Po veľa hodinách strávených pri neúspešnom implementovaní projektu pomocou rekurentného Back-Propagation algoritmu sme sa rozhodli implementovať projekt pomocou dobredného Back-Propagation algoritmu, ktorý sme preberali na druhej prednáške.
- Implementovali sme všetky 4 modely v metódach zadanieA, zadanieB, zadanieC, zadanieD.
- V prípade modelu A sa jednalo o sieť s 2 skrytými vrstvami po 10 neurónov. Vstupných neurónov bolo 33 (jeden prahový neurón a 4-krát po 8 vstupných parametrov).
 - V prípade modelu B sa jednalo o sieť s 2 skrytými vrstvami po 10 neurónov. Vstupných neurónov bolo 41 (jeden prahový neurón, 4-krát po 8 vstupných parametrov a 8 neurónov, v ktorých sú uložené predchádzajúce výstupy siete).
 - V prípade modelu C sa jednalo o sieť s 2 skrytými vrstvami po 10 neurónov. Vstupných neurónov bolo 51 (jeden prahový neurón, 4-krát po 8 vstupných parametrov a 18 neurónov, v ktorých je uložený predchádzajúci výstup neurónov skrytých vrstiev okrem prahových).
 - V prípade modelu C sa jednalo o sieť s 2 skrytými vrstvami. V prvej vrstve 10 neurónov, v druhej vrstve 11 neurónov, v jednom z nich je uložený predchádzajúci výstup siete. Vstupných neurónov bolo 51 (jeden prahový neurón, 4-krát po 8 vstupných parametrov a 18 neurónov, v ktorých je uložený predchádzajúci výstup neurónov skrytých vrstiev okrem prahových).
4. Trénovanie:
- Predpovedali sme 2 hodiny dopredu pomocou každého modelu.
 - V každom modeli sme nastavili učiaci pomer na 0,01 a učenie sme opakovali na každej nájdenej búrke 100-krát.

5. Vyhodnotenie:

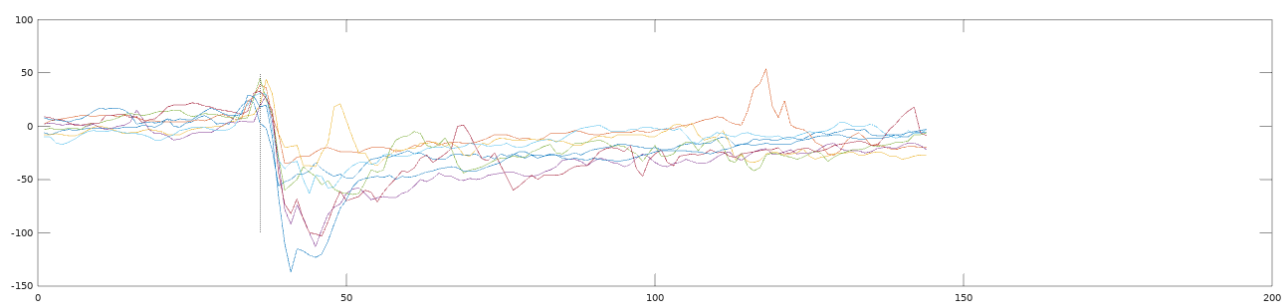
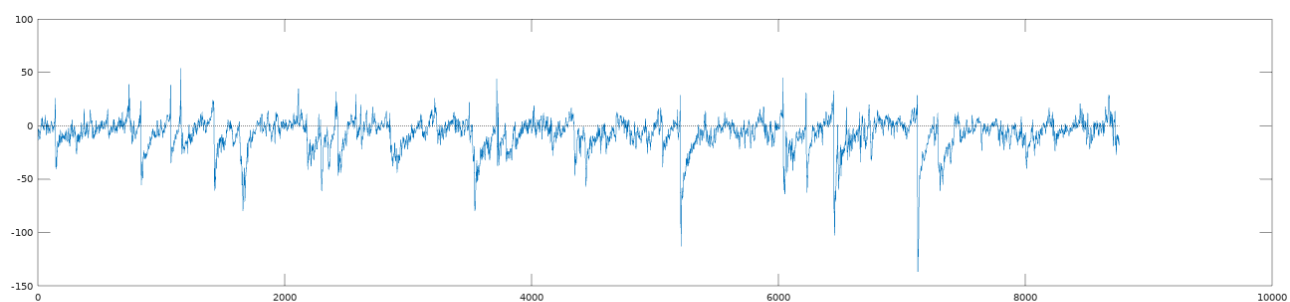
- V tréningovom roku sa nachádzalo 12 búrok na pozíciách:

399, 1152, 1550, 3048, 3221, 3574, 3906, 4571, 5649, 5821, 6039, 7133



- V testovacom roku sa nachádzalo 8 búrok na pozíciách:

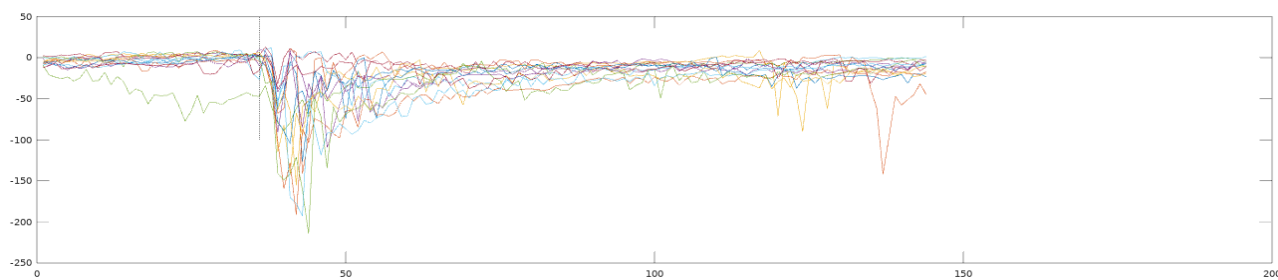
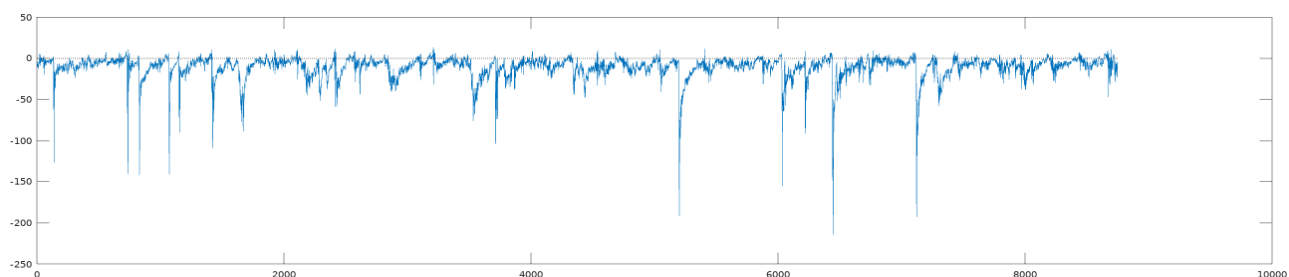
836, 1076, 3718, 5204, 6039, 6225, 6447, 7126



V každom z pokusov učenia vyšli podobné výsledky, uvidíme výsledky jedného z pokusov, pri každom sú grafy odhadov hodnoty Dst-index a nájdených búrok:

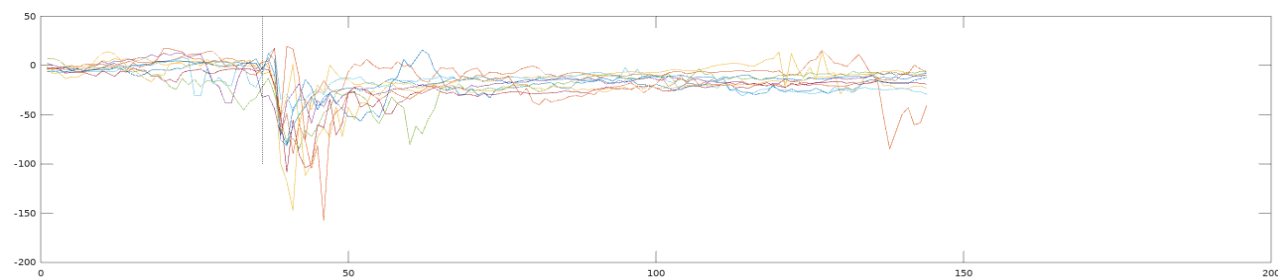
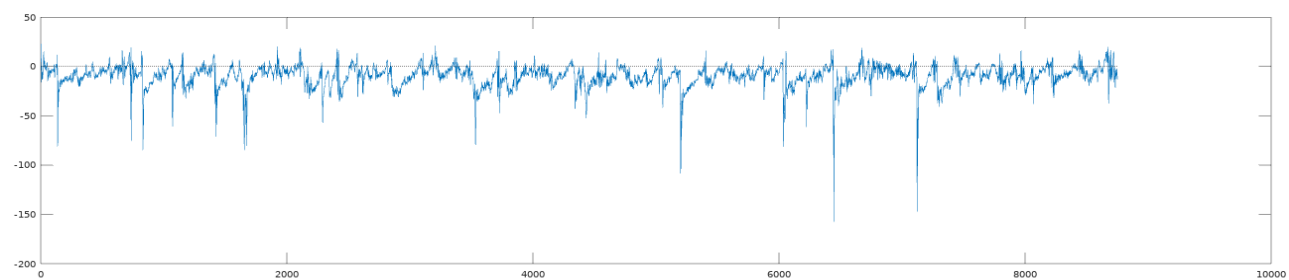
- Model A našiel 14 búrok na pozíciách:

142, 740, 1077, 1423, 1677, 2422, 3218, 3720, 5206, 6040, 6227, 6449, 7126, 7979



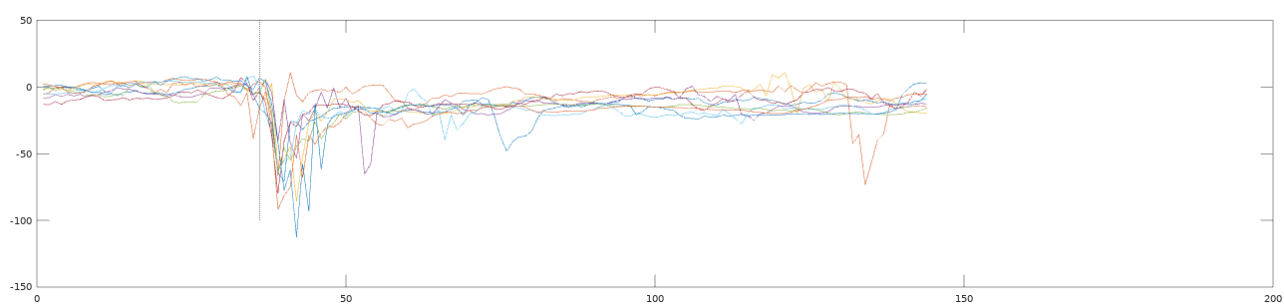
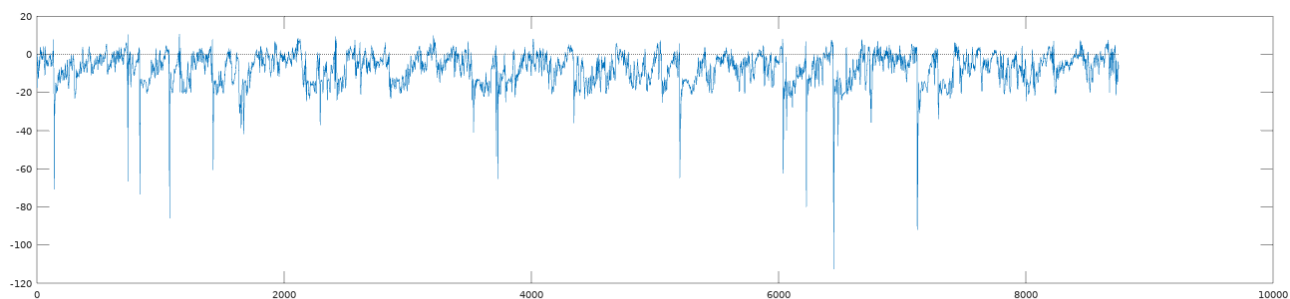
- Model B našiel 10 búrok na pozíciách:

144, 739, 1076, 1431, 1659, 3733, 5206, 6041, 6449, 7128



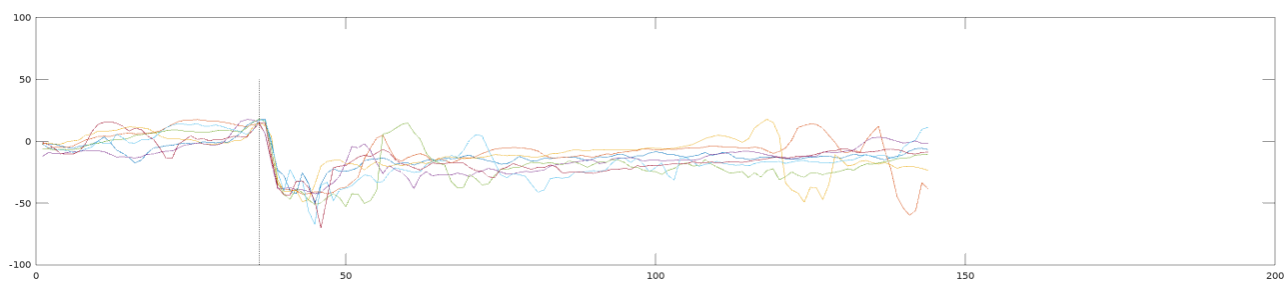
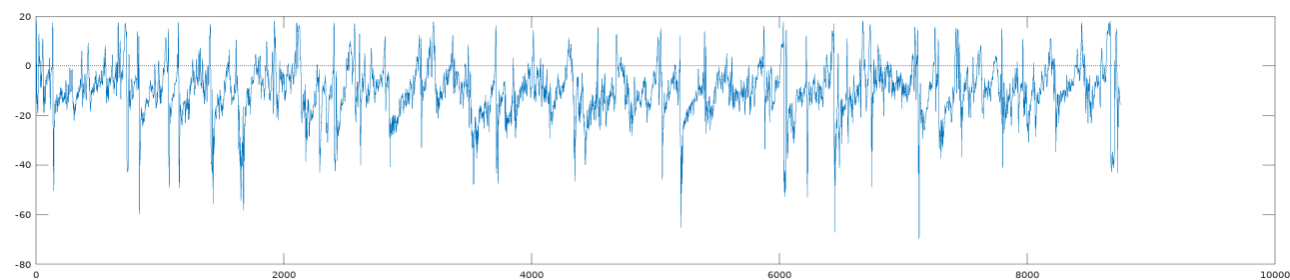
- Model C našiel 9 búrok na pozíciách:

147, 745, 1078, 3722, 5208, 6044, 6232, 6450, 7129



- Model D našiel 7 búrok na pozíciách:

145, 741, 1078, 2417, 6042, 6449, 7128



6. Porovnanie:

- Pozície predikovaných poklesov Dst-index (v prvom riadku sú nájdené poklesy v pôvodných dátach):

Pôvodne			836	1076					3718	5204	6039	6225	6447	7126	
Model A	142	740		1077	1423	1677	2422	3218	3720	5206	6040	6227	6449	7126	7979
Model B	144	739		1076	1431	1659			3733	5206	6041		6449	7128	
Model C	147	745		1078					3722	5208	6044	6232	6450	7129	
Model D	145	741		1078			2417				6042		6449	7128	

Ak berieme za správne identifikovanie búrky iba hodnoty ± 3 hodiny od skutočného poklesu tak jednotlivé modely môžeme vyhodnotiť nasledujúcou tabuľkou:

	Správne nájdené búrky	Nenájdené búrky	Nesprávne nájdené búrky
Model A	7	1	7
Model B	6	2	4
Model C	7	1	2
Model D	4	4	3

- Stredné kvadratické odchýlky u jednotlivých modelov boli:

Model A	9,3593
Model B	8,9572
Model C	10,570
Model D	10,584