

Úlohou je pre daný signál a sekvencie k-merov, rozdeliť signál na súvislé úseky, ktoré prislúchajú jednotlivým k-merom. Snaha o minimalizovanie času behu algoritmu, a zároveň minimalizovanie priemernej chyby - štvorec rozdielu očakávaného a reálneho signálu v bode.

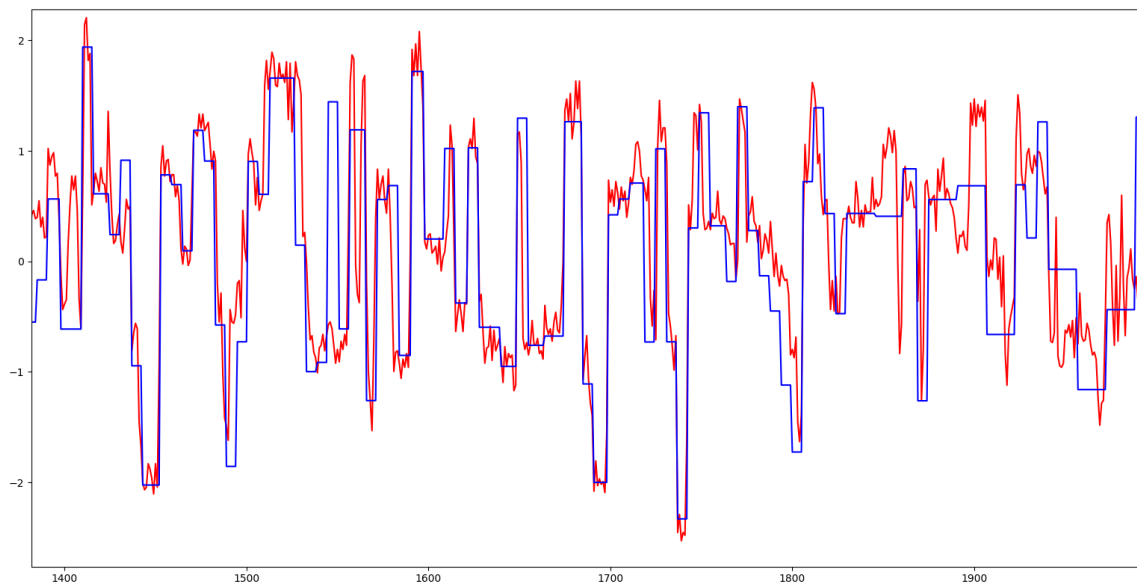
Návrh algoritmu: pre daný interval  $(l;r)$  - možné dĺžky signálu ktoré prislúchajú k-meru. Potom  $l, r$ , nájde optimálne rozdelenie na úseky pomocou Dynamického programovania. Časová zložitosť  $O(n^2)$  -> pomalé

Nápad na vylepšenie, zlepšenie rýchlosti za cenu presnosti: V každom kroku DP si necháme iba najlepších *width* stavov, ostatné zahodíme.

Následne som tento algoritmus implementoval, najprv v pythone, neskôr v c++ pre zefektívnenie rychlosti. Na tréningových dátach som skúsil rôzne parametre pre algoritmus (interval  $(l;r)$ , a *width*).

Taktiež som implementoval paralelizáciu, pre zrýchlenie výpočtu pre súbory ktoré obsahujú viacero nezávislých read-ov, čo je štandard.

ukážka časti predikovaného signálu (modrá) a reálneho (červená):



Návrhy na zlepšenie do budúcnosti:

- nevyhodnocovať stavy DP len na základe aktuálnej hodnoty, ale nadhliadnuť aj kúsok ďalej
- zahrnúť vo funkcii ktorá porovnáva riešenie aj štruktúru intervalov