

Dubinska analiza vremenskih nizova

Dubinska analiza podataka
9. predavanje

Pripremio: izv. prof. dr. sc. Alan Jović
Ak. god. 2023./2024.

Sadržaj

- Uvod i terminologija
- Komponente analize vremenskih nizova
- Algoritmi za predikciju i klasifikaciju
 - Pomični prosjek i eksponencijalno zaglađivanje
 - ARIMA
 - PROPHET
 - CIF
 - HIVE-COTE v2
- Modeliranje vremenskog niza ekstrakcijom značajki

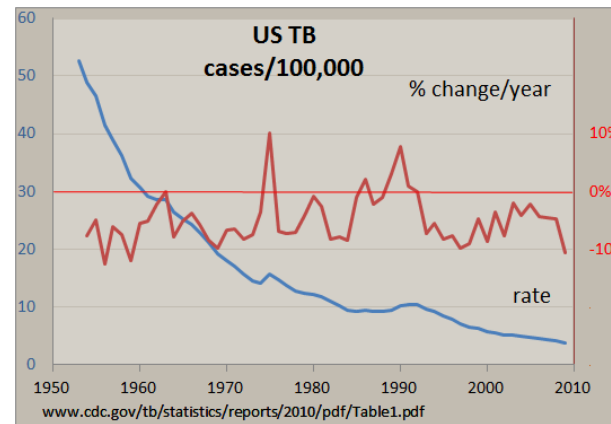
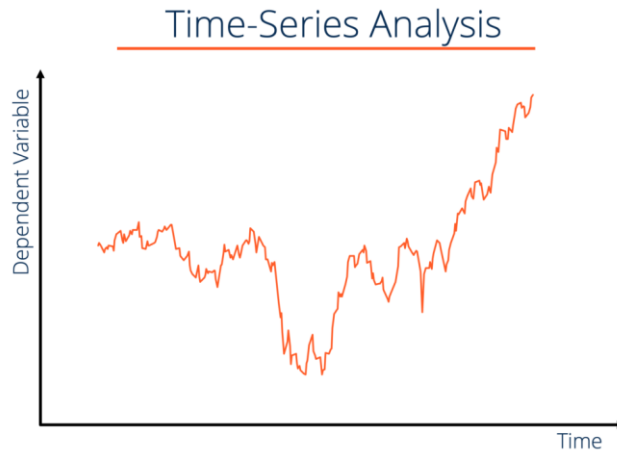
Uvod i terminologija

Dubinska analiza vremenskih nizova

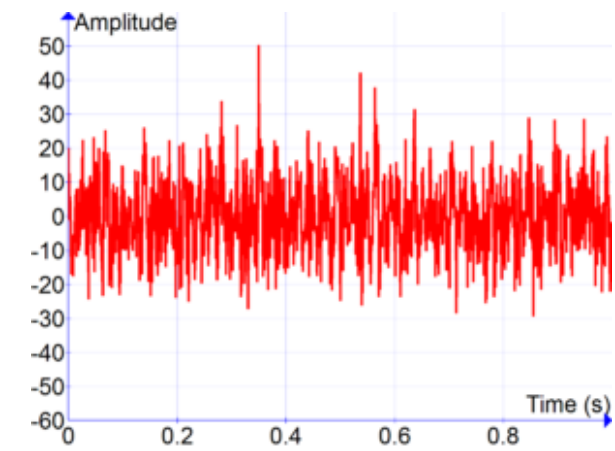
- Engl. *time series data mining, time series analysis*
- Područje dubinske analize podataka koje analizira **vremenske nizove (slijedove, sekvence) podataka** s ciljem pronalaska zanimljivih obrazaca
- Vremenski niz (češći naziv u matematici i fizici) se u elektrotehnici i srodnim područjima poistovjećuje s **jednodimenzionalnim signalom**
 - Otud područje analize vremenskih nizova odgovara području **obrade signala** (engl. *signal processing*)
- Vrlo istraženo i široko primijenjeno područje dubinske analize podataka
 - Upit „Time series analysis” u Google Scholaru vraća oko **2 milijuna** znanstvenih i stručnih članaka

Vremenski niz podataka

- Vremenski niz je kolekcija vrijednosti koja je dobivena slijednim mjerenjima tijekom vremena
- U analizi vremenskih nizova fokus je na **diskretnim** mjerenjima dobivenima u **jednolikim** vremenskim intervalima
- Dubinska analiza vremenskih nizova potekla je iz ljudske sposobnosti uočavanja obrazaca u grafovima prikazanima na različitim vremenskim skalama



Dubinska analiza podataka



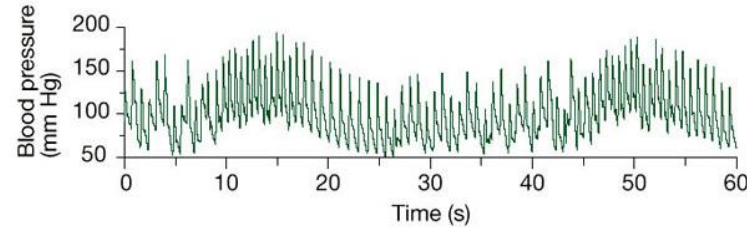
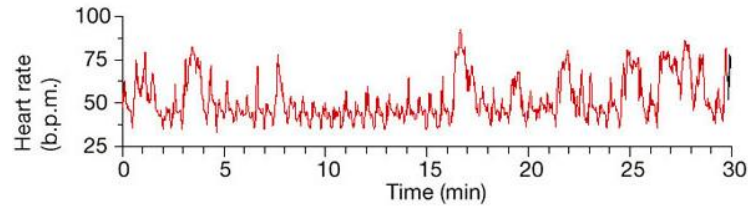
Glavni zadaci analize vremenskih nizova

- **Upit prema sadržaju** (engl. *query by content*) – pronalazak vremenskih nizova **najsličnijih** (prema nekog kriteriju) zadanom vremenskom nizu
- **Detekcija anomalija** (engl. *anomaly detection*) – otkrivanje **neuobičajenih podnizova** u vremenskom nizu
- **Otkrivanje motiva** (engl. *motif discovery*) – otkrivanje svih **tipičnih, nepreklapajućih podnizova** (tzv. motiva) u duljem vremenskom nizu
- **Predviđanje vrijednosti** (engl. *prediction, forecasting*) – predviđanje **sljedećih k vrijednosti niza** koje su najizglednije da će se pojaviti
- **Grupiranje** (engl. *clustering*) – pronalazak određenog broja **grupa** sličnih vremenskih nizova (ili sličnih podnizova jednog niza)
- **Klasifikacija** (engl. *classification*) – na temelju značajki vremenskog niza, pridruži se nekom neoznačenom vremenskom nizu jedna od predefiniranih **oznaka klase**
- **Segmentacija ili reprezentacija** (engl. *segmentation, summarization, representation*) – za izvorni vremenski niz izgradi se njegov **nižedimenzijski model** takav da ga on blisko **aproksimira**

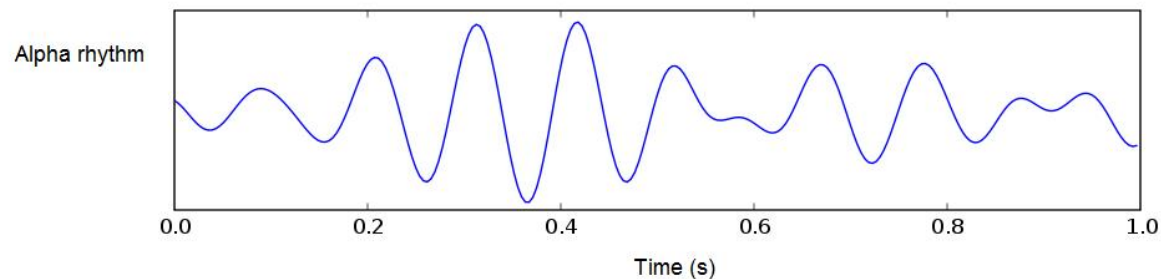
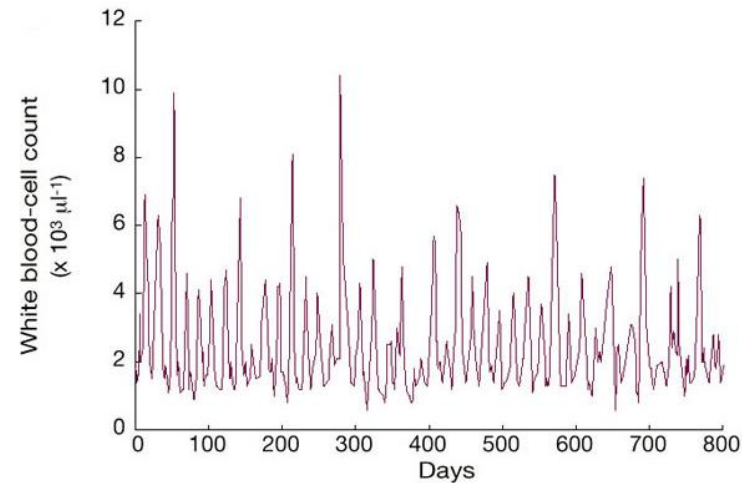
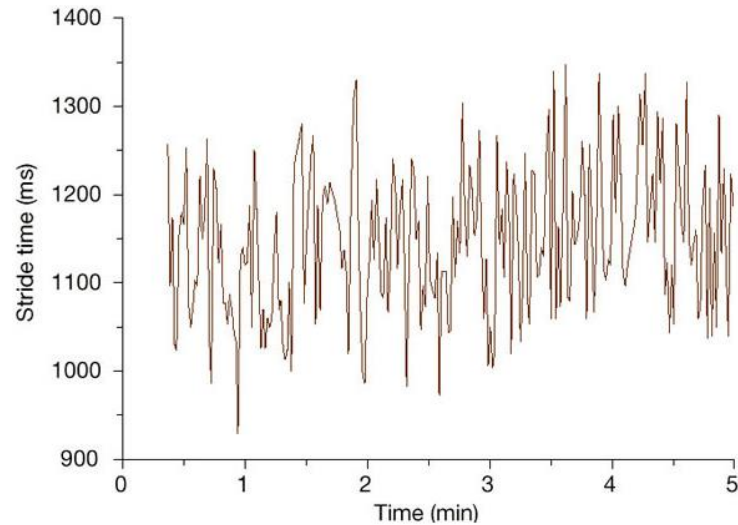
Primjena dubinske analize vremenskih nizova

- Velik broj primjenskih područja i specifičnih primjena:
 - Predviđanje vrijednosti u **ekonometriji** (prodaja, dionice, kriptovalute, prihodi) i **meteorologiji** (vremenska prognoza, klimatološki modeli)
 - Grupiranje i klasifikacija poremećaja rada ljudskog organizma u **biomedicini** (analiza biomedicinskih signala), često u dijagnostičke svrhe
 - Upit prema sadržaju, otkrivanje motiva i detekcija anomalija u **astronomiji** (pronalažak nebeskih tijela, pronalažak zanimljivih obrazaca u signalima iz svemira)
 - Analiza, karakterizacija i predviđanje kvarova (npr. za prediktivno održavanje), detekcija anomalija u **industriji** (klasičnoj i IoT uređajima)
 - Detekcija anomalija i predviđanje prometa u **računalnim mrežama**

Primjena dubinske analize vremenskih nizova



Biomedicinski vremenski nizovi



Izvor: A. Jović, Dubinska analiza biomedicinskih vremenskih nizova zasnovana na računalnom radnom okviru za izlučivanje značajki, doktorski rad, FER, 2012.

Terminologija

- **Vremenski niz** T je uređeni slijed od n realnih brojeva:

$$T = (t_1, \dots, t_n), t_i \in \mathbf{R}$$

- Vremenski niz rezultat je opažanja nekog procesa pri čemu su mjerenja vrijednosti t_i napravljena u **vremenskim trenucima** s **pravilnim** (jednolikim, uniformnim) ili **nepravilnim** (nejednolikim, neuniformnim) međusobnim razmakom
- Za pravilna mjerenja za svaki niz definirana je **stopa uzorkovanja ili otipkavanja** (engl. *sampling rate*) vremenskog niza, izražena najčešće u hercima
 - Npr. niz uzorkovan frekvencijom od 100 Hz znači da je svakih 10 ms napravljeno jedno mjerenje
- Vremenski niz može biti **univarijatan** (engl. *univariate*) prema gornjoj definiciji ili **multivarijatan** (engl. *multivariate*), što znači da unutar istog vremenskog raspona sadržava više (N) univarijatnih vremenskih nizova

Terminologija

- Vremenski niz je konačan ako je mjerenje već provedeno i cijeli niz nam je dostupan ili polubeskonačan ako nam **u toku podataka** (engl. *streaming*) pristiže na obradu
- Za vremenski niz T duljine n , podniz S je niz duljine $m \leq n$ koji se sastoji od neprekinutih mjerenja iz T :

$$S = (t_k, t_{k+1}, \dots, t_{k+m-1}), \text{ za } 1 \leq k \leq n - m + 1$$

- Skup svih podnizova duljine m od T označava se sa S_T^m
- Za lakšu pohranu, veliki skupovi podataka vremenskih nizova se organiziraju u bazu podataka vremenskih nizova (engl. *time series database*, kraće: DB)

Terminologija

- **Mjera sličnosti** $D(T, U)$ između vremenskih nizova T i U je funkcija koja prima dva niza na ulazu i vraća udaljenost (nenegativan broj) između tih nizova
- **Mjera sličnosti podniza** T u nekom nizu S , $D_{podniz}(T, S)$ definira se kao:

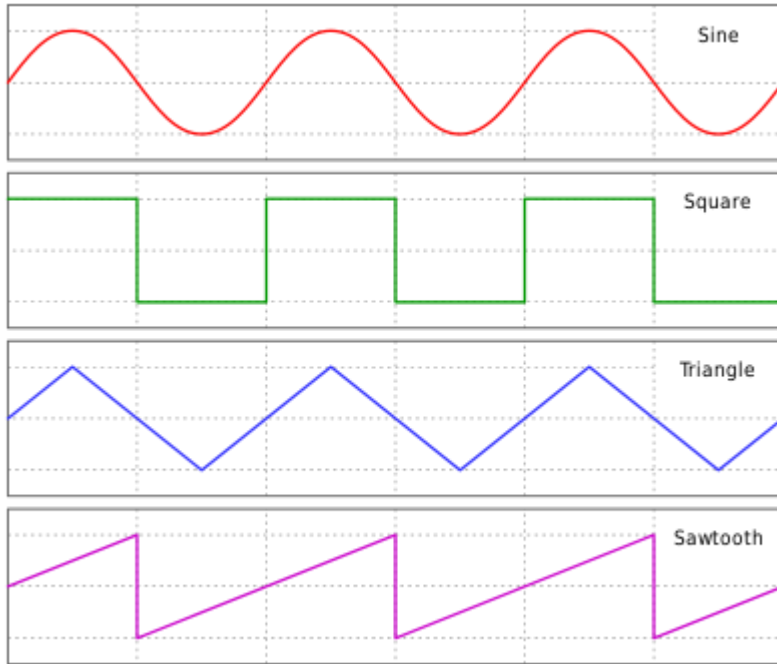
$$D_{podniz}(T, S) = \min(D(T, S')), \text{ za } S' \in \mathcal{S}_S^{|T|}$$

- Predstavlja udaljenost između podniza T i njemu najbližeg podniza S' u nizu S

Terminologija

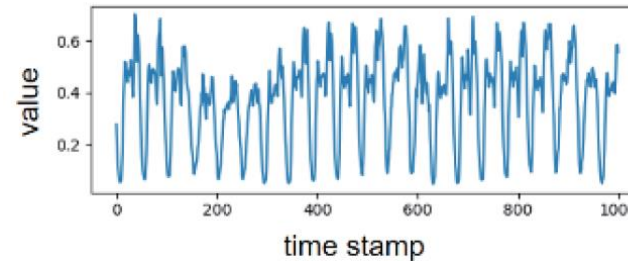
- Vremenski niz može biti **stacionaran** (engl. *stationary*) ili **nestacionaran** (engl. *non-stationary*)
 - Ako se statistička svojstva **srednje vrijednosti**, **varijance** i **autokovarijance** ne mijenjaju između podnizova U i V za bilo koji pomak m između tih podnizova, onda je niz stacionaran, a inače je nestacionaran
 - Najčešći razlog nedostatka stacionarnosti je **postojanje trenda srednje vrijednosti** u podacima
- Nestacionaran vremenski niz može biti **periodički** (sezonalni, ciklostacionarni), ili **neperiodički**
 - Periodičnost je **ponavljanje podniza u nizu s periodom p** koji se obično **vizualno uočava** u vremenskom nizu, a očitava se i kao **jedna frekvencija s najvećom spektralnom gustoćom snage** u frekvencijskoj domeni – npr. sinusoida je periodički nestacionarni niz
 - Neperiodički niz nema period p u kojem bi se podniz ponavljao – npr. govor je neperiodički niz

Terminologija



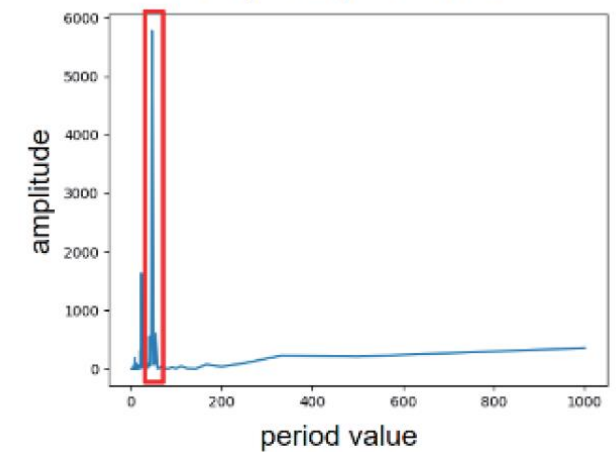
Periodički vremenski nizovi

Time Domain



FFT

Frequency Domain



https://en.wikibooks.org/wiki/Signals_and_Systems/Periodic_Signals#Sinusoidal_wave

J. B. Kao, and J. R. Jiang, “Anomaly detection for univariate time series with statistics and deep learning,” in *Proc. of 2019 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE 2019)*, pp. 404-407, October 2019

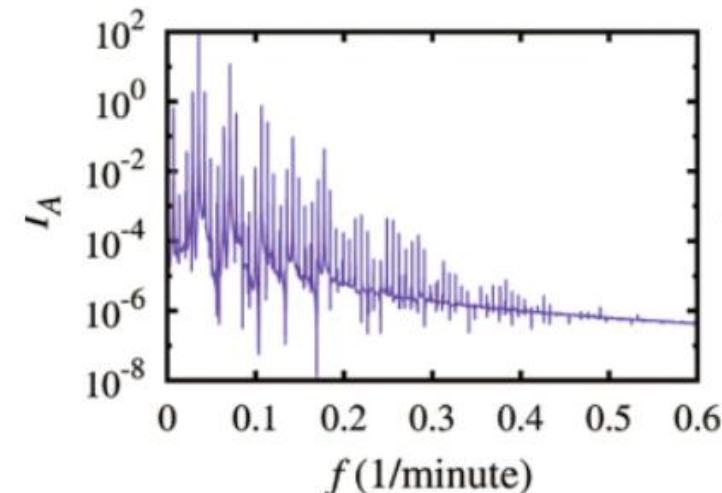
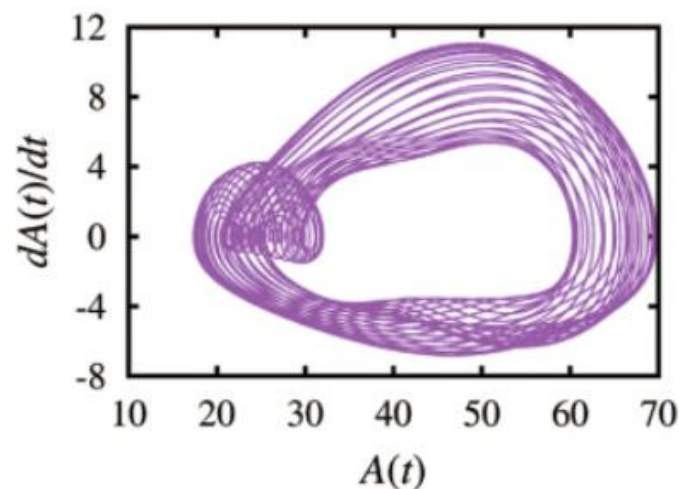
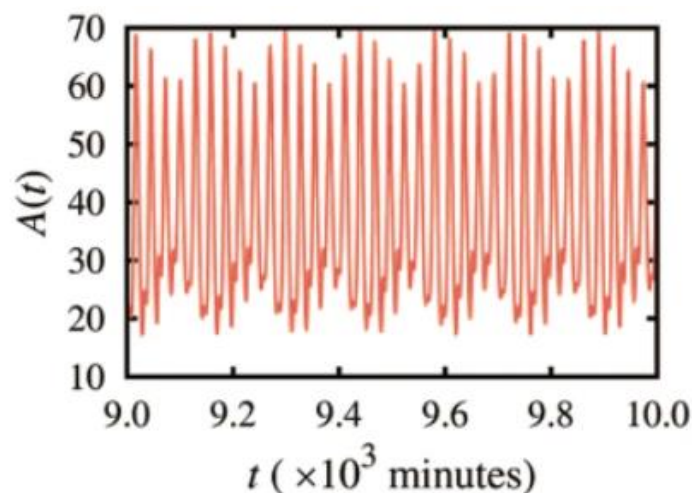
Terminologija

- Postoji značajan raspon mogućnosti između periodičkih i neperiodičkih nizova
- Uobičajena daljnja podjela je na:
 - **Skoro periodički niz:** periodičnost je uočljiva, ali niz nije dovoljno pravilan da bi u potpunosti bio periodički (razmaci između sličnih podnizova ne iznose točno p)
 - npr. normalni srčani ritam je medicinski primjer skoro periodičkog niza (međutim s fraktalnim svojstvima)
 - **Kvaziperiodički niz:** podniz u nizu se ponavlja s **nepredvidljivim** periodom, obično ima više istaknutih frekvencija
 - npr. El Niño–Southern Oscillation (ENSO) je meteorološki primjer kvaziperiodičkog niza gdje visoki tlak zraka i povišene temperature mora u novijoj povijesti imaju period negdje između 4 i 12 godina s maksimumom na 5 godina
 - **Kaotičan niz:** niz koji je **neperiodički**, **deterministički** – moguće mu je odrediti buduću vrijednost (ali samo u teoriji), **ograničenih vrijednosti** i **osjetljiv na početne uvjete** – mala promjena na početku značajno mijenja trajektoriju niza
 - npr. biomedicinski vremenski nizovi u nekim situacijama iskazuju **neke značajke kaosa** (EKG, EEG, EMG)
 - **Slučajan (stohastički) niz:** niz koji je neperiodički i nedeterministički (u praksi teško se razaznaje od kaotičnog)
 - npr. bijeli šum

Terminologija

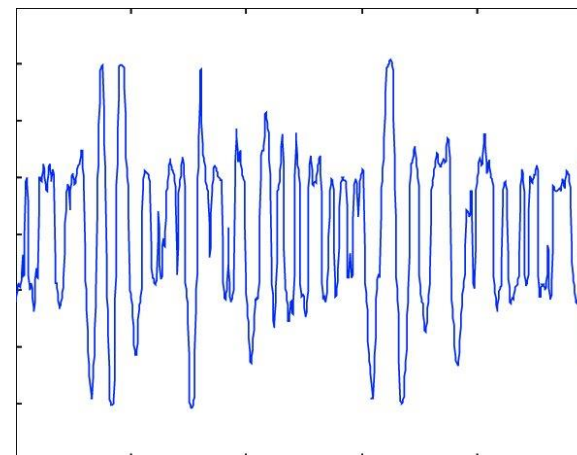
Kvaziperiodički vremenski niz

Suzuki, Y. *et al.* Periodic, Quasi-periodic and Chaotic Dynamics in Simple Gene Elements with Time Delays. *Sci. Rep.* **6**, 21037; doi: 10.1038/srep21037 (2016)



Kaotičan vremenski niz

Lorenzov atraktor u 3D ugrađenom faznom prostoru kao rezultat rješenja Lorenzovih diferencijalnih jednačbi



Slučajan niz

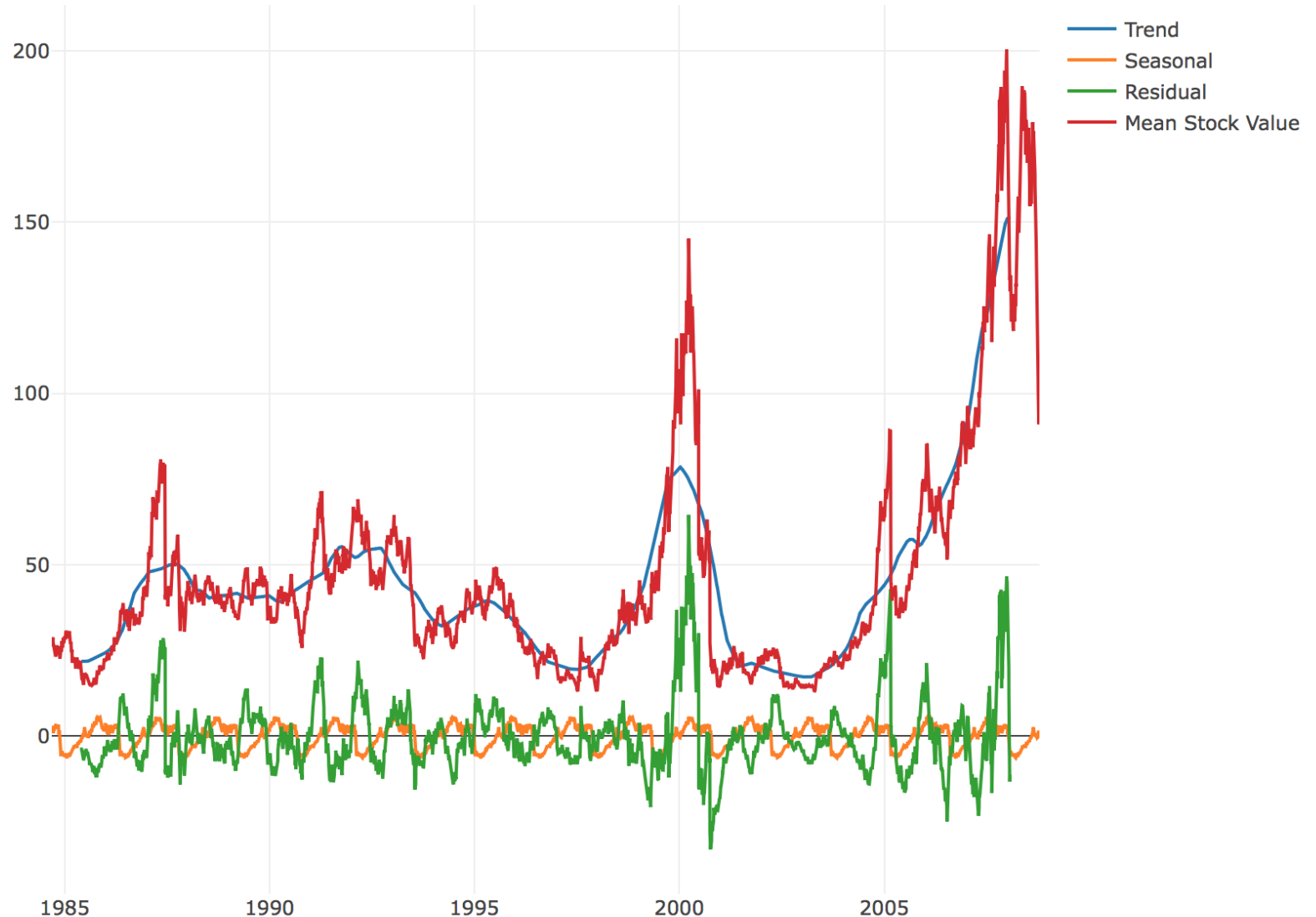
https://www.researchgate.net/publication/258381468_Fast_Detection_of_Weak_Singularities_in_a_Chaotic_Signal_Using_Lorenz_System_and_the_Bisection_Algorithm/figures?o=1

https://en.wikipedia.org/wiki/Lorenz_system#/media/File:A_Trajectory_Through_Phase_Space_in_a_Lorenz_Attractor.gif

Terminologija

Primjer: vrijednost dionice
opisana pomoću tri komponente:
trenda, sezonalnosti i rezidualne
komponente (šum)

<https://towardsdatascience.com/trend-seasonality-moving-average-auto-regressive-model-my-journey-to-time-series-data-with-edc4c0c8284b>



Komponente analize vremenskih nizova

Komponente analize vremenskih nizova

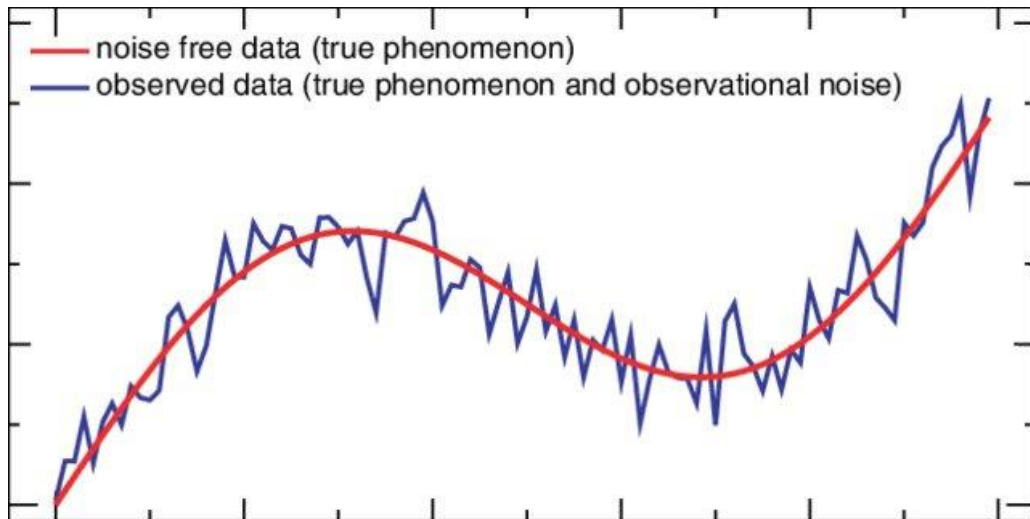
Komponente zajedničke većini zadataka analize vremenskih nizova:

- **Predobrada vremenskog niza** – tehnike koje pripremaju niz za daljnju analizu
- **Reprezentacijske metode** – naglašavanje bitnih karakteristika vremenskog niza na jezgrovit način
- **Mjere sličnosti** – razmatraju se mjere koje određuju kada su nizovi (ili podnizovi) slični
- **Indeksiranje** – učinkovita organizacija podataka o vremenskim nizovima u bazi podataka kako bi se brže pristupalo pojedinim dijelovima niza (ne obrađujemo)

Predobrada vremenskog niza – uklanjanje šuma i skaliranje

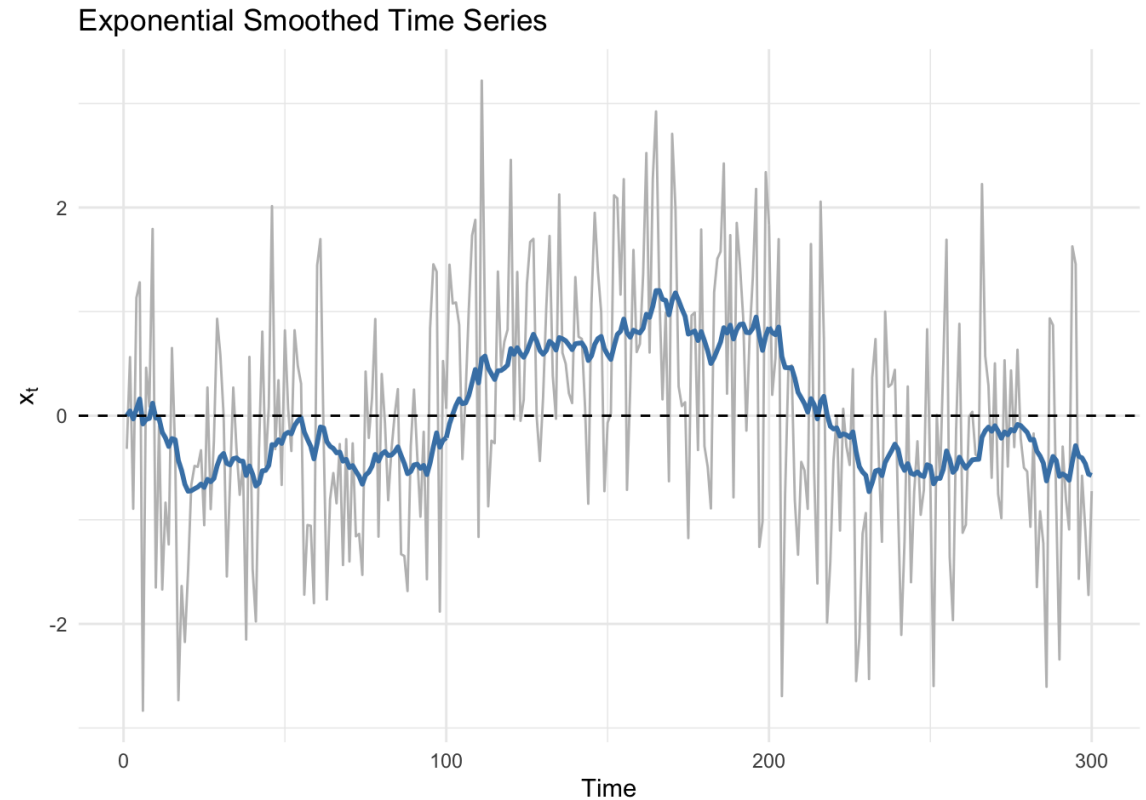
- **Uklanjanje šuma** – zadatak je ukloniti različite izvore šuma (buke, slučajnosti) koji djeluju na niz
 - **Digitalni filtri** – npr. pojasno propusni filter (engl. *bandpass filter*), filter pomičnog prosjeka (engl. *moving average*), filter eksponencijalnog zaglađivanja (engl. *exponential smoothing*)
 - **Korištenje nekih transformacija podataka**, npr. analiza nezavisnih komponenti (engl. *independent component analysis*, ICA) kako bi se zadržale samo glavne karakteristike signala, a uklonio šum
- **Skaliranje** – *min-max* normalizacija ili standardizacija z-skaliranjem
 - Korisno za kasnije uspoređivanje dvaju nizova, grupiranje ili klasifikaciju
 - Potreban je oprez pri primjeni ako je značajan dio signala šum kako se šum ne bi preglasao

Predobrada vremenskog niza – primjeri



Preuzeto s:

https://www.researchgate.net/publication/220273347_A_Comparative_Study_of_Information_Criteria_for_Model_Selection/figures?lo=1



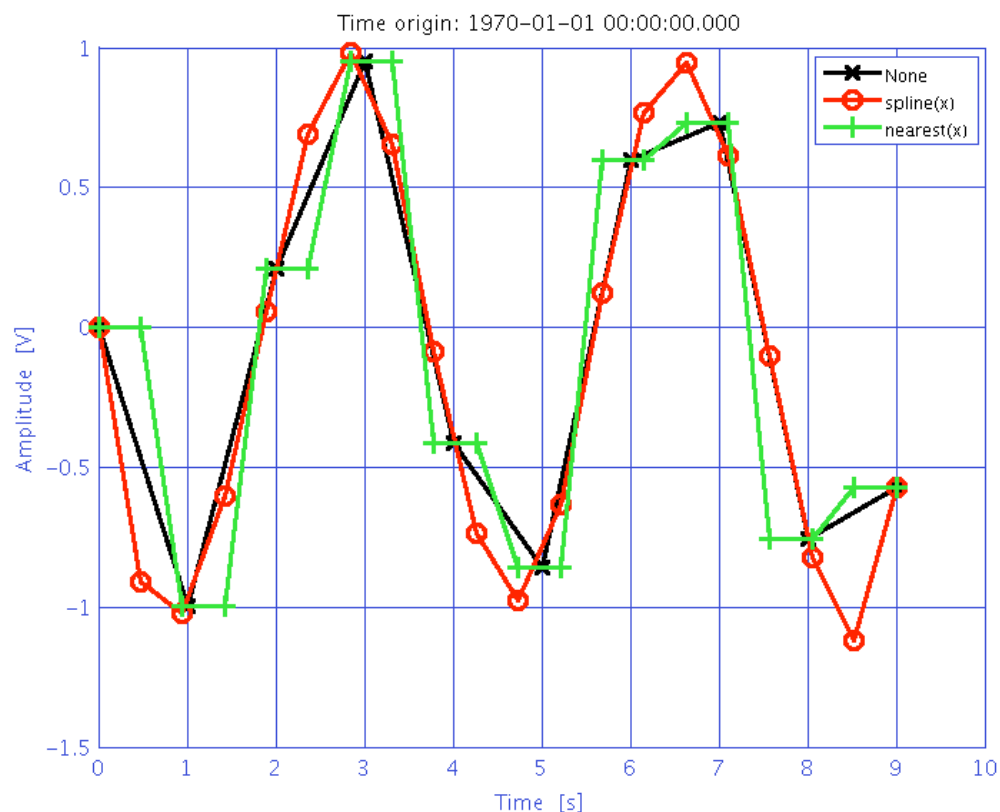
Preuzeto s:

<https://bookdown.org/rdpeng/timeseriesbook/filtering-time-series.html>

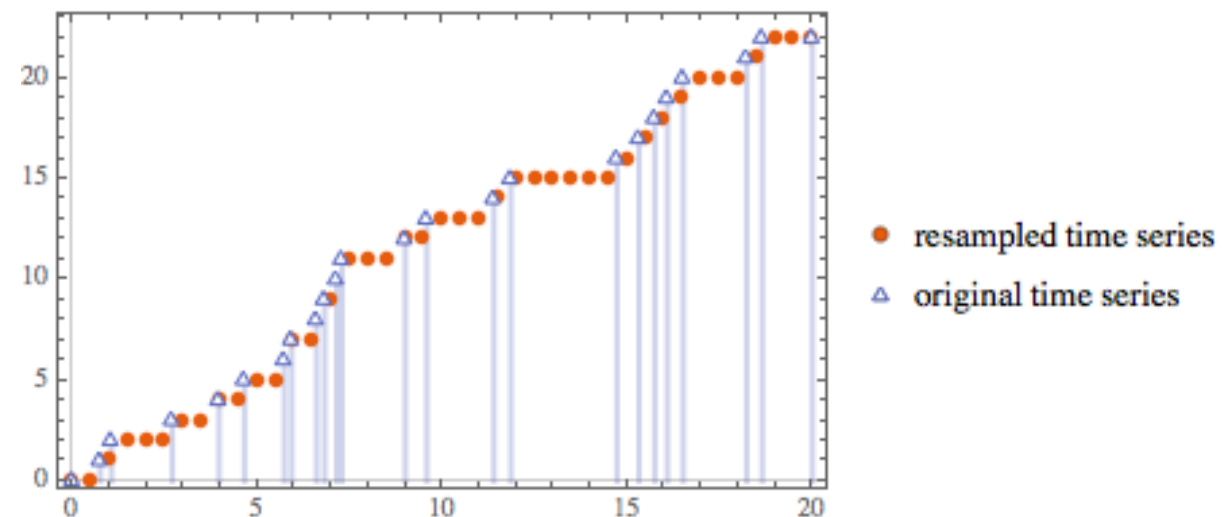
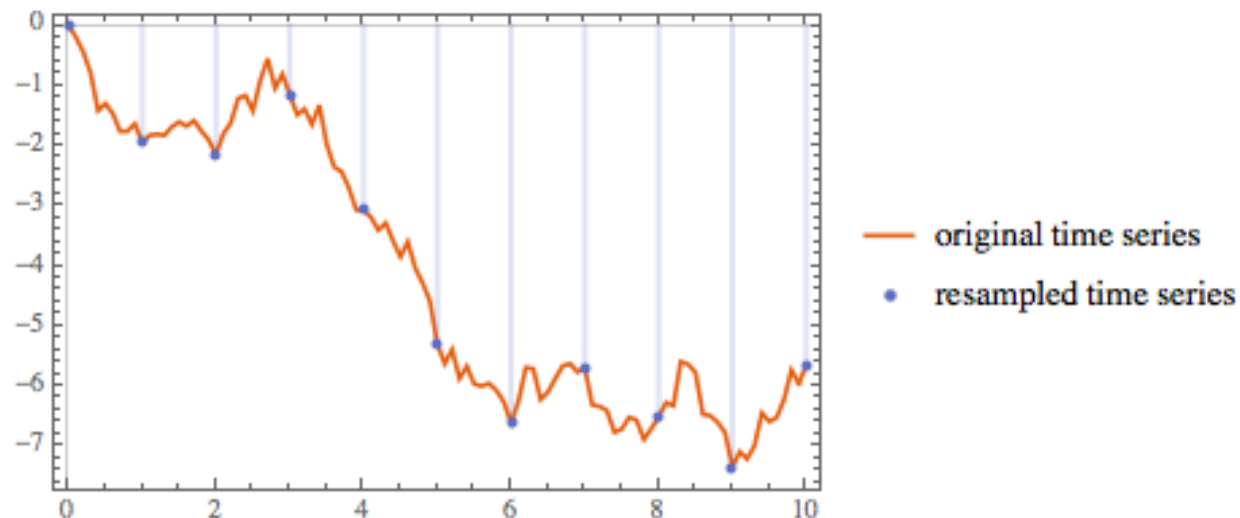
Predobrada vremenskog niza – ponovno uzorkovanje i interpolacija

- **Ponovno uzorkovanje** (ili jednoliko vremensko iskrivljavanje) (engl. *resampling, uniform time warping*)
 - Npr. umjesto 100 Hz uzorkovanja niz se preuzorkuje (poduzorkuje) na 50 Hz
 - Korisno za dobivanje više vremenskih nizova iste duljine – poboljšana usporedivost i mogućnost zajedničkog korištenja
 - U praksi često je **smanjenje frekvencije uzorkovanja** (engl. *downsampling*) kako bi se smanjila dimenzionalnost problema
- **Interpolacija** – zasniva se na procesu pronalaska odgovarajuće kontinuirane funkcije na temelju diskretnih točaka vremenskog niza s ciljem nadomještanja nepoznatih vrijednosti (tijekom ili neovisno o ponovnom uzorkovanju)
 - Najčešće interpolacije: **konstantna, najbliži susjed, linearna, kvadratna, kubni spline**
 - Mogu biti i složenije: Lagrangeova polinomna, Hermiteova polinomna, Newtonova polinomna...

Predobrada vremenskog niza – primjeri



Preuzeto s: <https://www.elisascience.org/ltpda/usermanual/ug/interp.html>



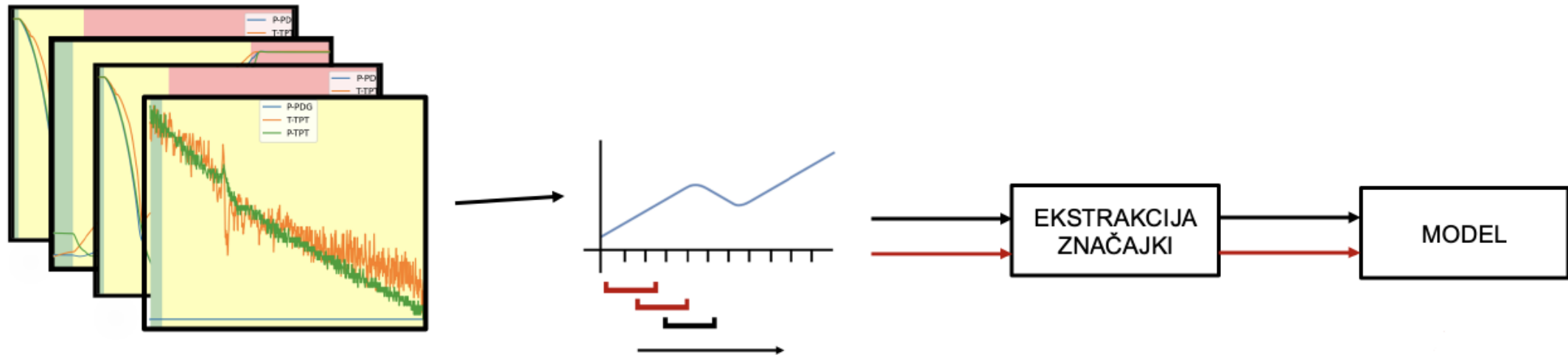
Preuzeto s: <https://www.wolfram.com/mathematica/new-in-10/time-series/resample-time-series.html>

Predobrada vremenskog niza – podjela u prozore

- **Podjela u prozore (engl. *windowing*)**

- Podjela niza od n mjerenja u **prozore** (podnizove) širine $k \ll n$ mjerenja
- Vrlo važan pretkorak za algoritme grupiranja, klasifikacije, predviđanja i za ostale zadatke
- Smanjenje broja točaka smanjuje računske zahtjeve i povećava preciznost izvedbe zadataka
- Podjela u prozore može biti **ovisna o periodičnosti** (npr. prozor odgovara širini perioda p) i **drugim značajkama niza** (npr. prozor od $k/2$ točaka s obje strane nekog detektiranog oblika) ili neovisna o značajkama niza
- Podjela u prozore može biti **s preklapanjem prozora** (engl. *window overlapping*) ili bez preklapanja
- Podjela u prozore nešto otežava vrednovanje algoritma za neki zadatak
 - Za klasifikacijske i predikcijske probleme treba paziti koje prozore uzeti za učenje, a koje za testiranje modela

Predobrada vremenskog niza – primjer



- Primjer: detekcija žutog (prijelaznog) stanja čim prije – indikacija budućeg kvara opreme (crveno stanje)
- Podjela ulaznog niza na prozore širine k neovisno o značajkama niza s preklapanjem slijednih prozora širine $k/2$, prozori su **pomični** (engl. *sliding window*) – **niz se analizira redom po prozorima**
- Ekstrakcija značajki u prozoru i izrada klasif. modela (npr. slučajna šuma) (klasa 0 – zeleno, klasa 1 – žuto)
- Detekcija se smatra pouzdanom ako su tri uzastopna prozora predviđena kao žuto

Reprezentacijske metode

- Izvlačenje najvažnijih informacija iz vremenskog niza
- Zahtjevi na reprezentacijsku metodu vremenskog niza:
 - Značajna redukcija dimenzionalnosti (broja točaka) vremenskog niza
 - Naglasak na temeljnim karakteristikama niza i na lokalnoj razini i na globalnoj razini (cijelom nizu)
 - Niski vremenski zahtjevi na računanje reprezentacije
 - Neosjetljivost na šum ili implicitno uklanjanje šuma u metodi
- Klasifikacija reprezentacijskih metoda:
 - **Neadaptivne s obzirom na podatke** (engl. *nondata adaptive*)
 - **Adaptivne s obzirom na podatke** (engl. *data adaptive*)
 - **Zasnovane na modelu podataka** (engl. *model based*)

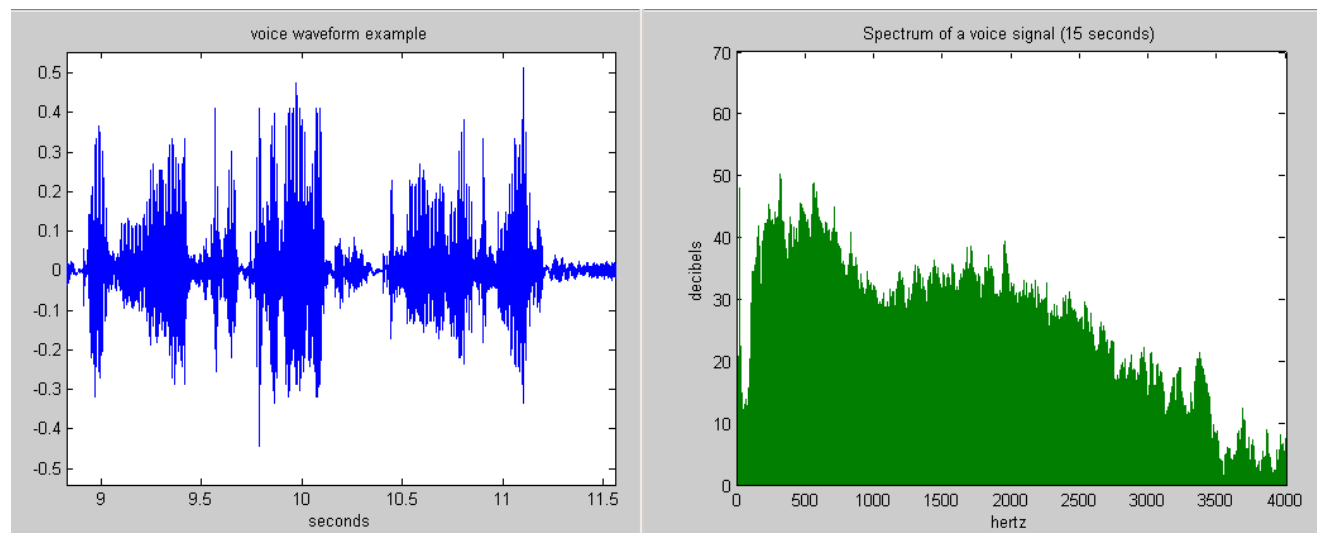
} Još se nazivaju „transformacijama”
vremenskog niza

P. Esling, C. Agon. Time-Series Data Mining. ACM Computing Surveys, Vol. 45, No. 1, p. 12, 2012

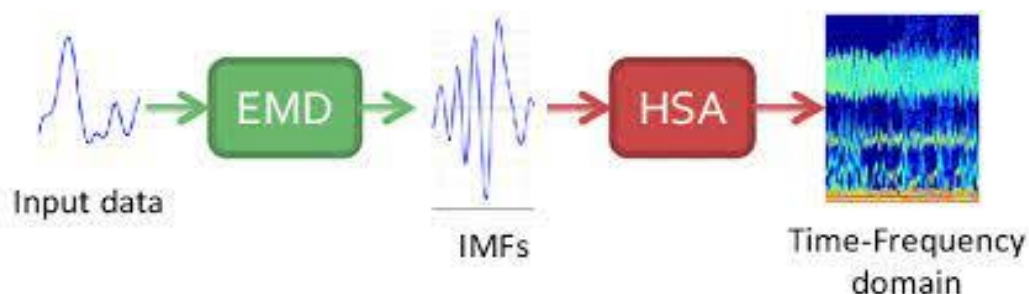
Neadaptivne reprezentacijske metode

- **Parametri transformacije podataka ostaju isti** za svaki niz, neovisno o njegovoj prirodi
- Primjeri:
 - **DFT** – diskretna Fourierova transformacija (npr. algoritmom brze Fourierove transformacije, engl. *Fast Fourier Transform*, FFT) – rezultat su koeficijenti sinusoida koji tvore niz u frekvencijskoj domeni
 - **DWT** – diskretna transformacija valićima (engl. *wavelets*) – skalirana i pomaknuta funkcija valića majke (engl. *mother wavelet*) daje višerezolucijsku dekompoziciju niza (niže frekvencije se gledaju na duljem vremenskom intervalu i obratno)
 - **PAA** – aproksimacija agregacijom po dijelovima (engl. *piecewise aggregate approximation*) – niz se standardizira, podijeli na prozore i reprezentira samo sa srednjom vrijednosti prozora
 - **HHT** – Hilbert Huangova transformacija – razlaže niz (najčešće nestacionarni i nelinearni) u velik broj IMF-funkcija (engl. *intrinsic mode function*), trenutne frekvencije IMF-funkcija uzimaju se za reprezentaciju

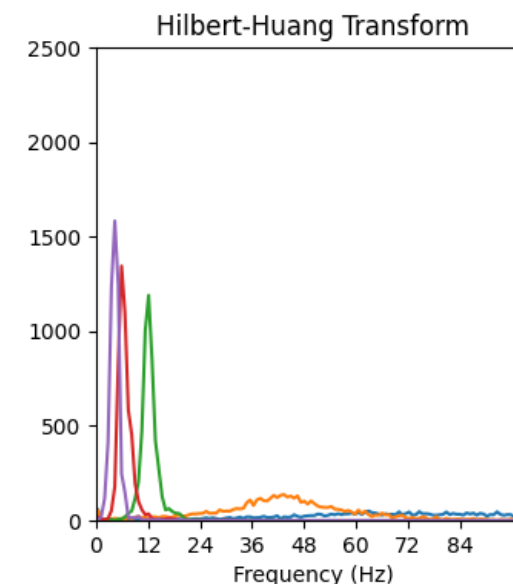
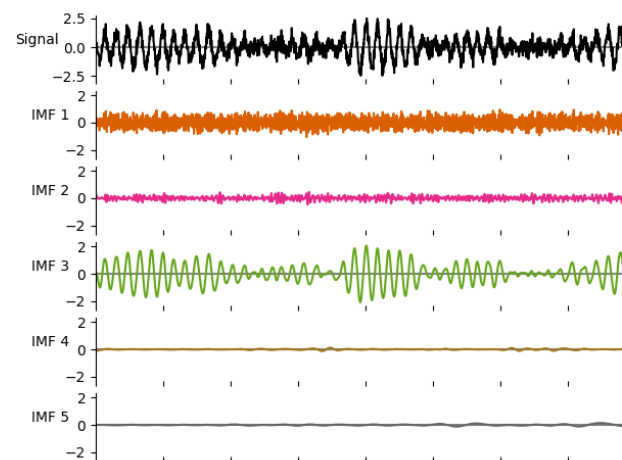
Neadaptivne reprezentacijske metode



- Spektralna gustoća snage (po frekvencijama) dobivena iz koeficijenata sinusoida nakon **DFT**-a
- https://en.wikipedia.org/wiki/Spectral_density#/media/File:Voice_waveform_and_spectrum.png



https://www.researchgate.net/publication/221367144_Parallelizing_Hilbert-Huang_Transform_on_a_GPU/figures?lo=1



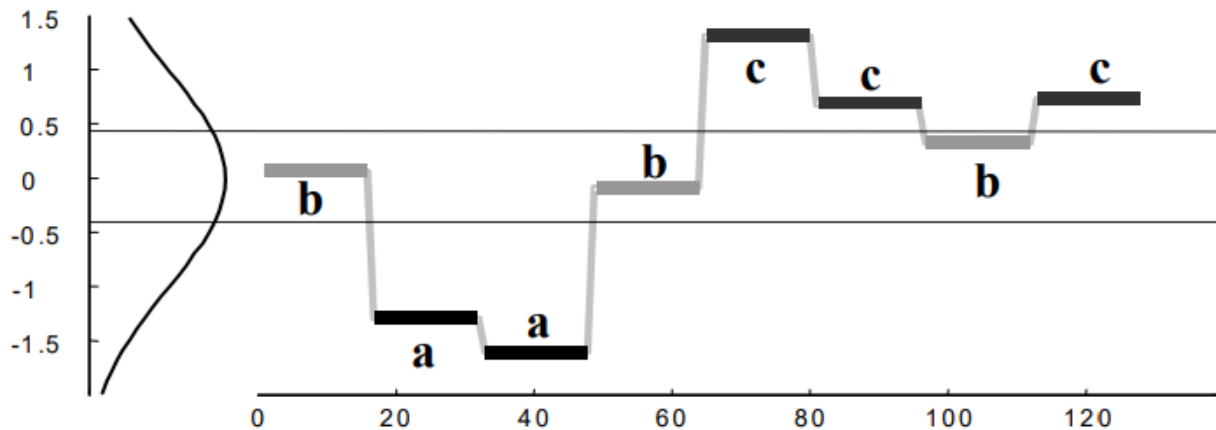
https://emd.readthedocs.io/en/v0.3.1/auto_tutorials/02_spectrum_analysis/plot_tutorial1.html

Adaptivne reprezentacijske metode

- **Definiraju parametre transformacije** koji će biti zadržani u reprezentaciji **ovisno o konkretnim podacima**
- Većina neadaptivnih metoda mogu se prilagoditi da postanu adaptivne
 - Npr. razmatranje samo određenih frekvencijskih pojaseva za reprezentaciju niza kod DFT-a
 - Npr. razmatranje istaknutih vrhova u Hilbertovom spektru (pomak trenutačne frekvencije IMF-funkcija u vremenu) za reprezentaciju niza kod HHT-a
- **Metode simboličke reprezentacije** su većinom adaptivne
 - Vremenski niz **diskretiziraju u niz simbola** (određene abecede)
 - Diskretizacija (broj diskretnih razina i preslikavanje na razine) ovisi o značajkama niza
 - Omogućuju implicitno uklanjanje šuma zbog smanjenja kompleksnosti niza, olakšavaju klasifikaciju i opis kompleksnosti (npr. računanjem entropije)

Adaptivne reprezentacijske metode

- **Primjeri metoda simboličke reprezentacije**
 - **SAX** (engl. *Symbolic Aggregate approXimation*) – kao osnovu ima PAA, diskretizira srednje vrijednosti u prozorima u abecedu s jednakom frekvencijom simbola
 - omogućuje usporedbe univarijatnih nizova, grupiranje i klasifikaciju
 - značajno smanjuje dimenzionalnost niza i memorijske zahtjeve DB

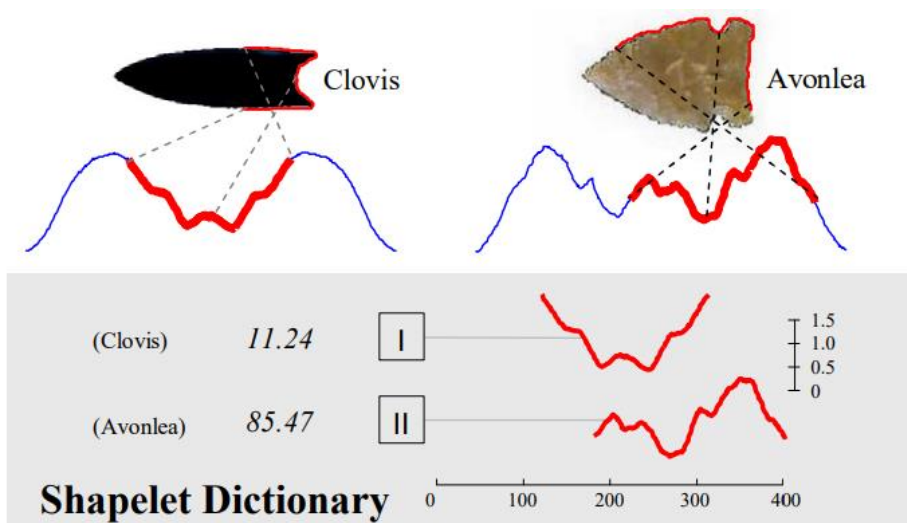


SAX

- J. Lin, E. Keogh, S. Lonardi, B. Chiu. A symbolic representation of time series, with implications for streaming algorithms. In Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. ACM New York, 2–11, 2003.

Adaptivne reprezentacijske metode

- Primjeri metoda simboličke reprezentacije
 - ***Shapeleti*** (engl. *shapelets*) – podnizovi koji su **maksimalno reprezentativni za neku klasu** i stoga mogu u potpunosti diskriminirati klase
 - metoda pretpostavlja da su objekti različitih klasa ugrađeni u vremenski niz i da je ugradnji najbliži njihov karakteristični *shapelet* (manji dio cijelog oblika)
 - uči se rječnik *shapeleta* koji se onda pretražuje prema sličnosti u nizu



Shapeleti

- L. Ye, E. Keogh. Time series shapelets: A new primitive for data mining. In Proceedings of the 15th, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 947–956, 2009.

Reprezentacijske metode zasnovane na modelu

- Pretpostavljaju da je opaženi niz **rezultat nekog modela koji se nalazi u pozadini** i koji ga generira, zadatak je pronaći parametre tog modela kao reprezentaciju
- Metode zasnovane na modelu:
 - **Modeliranje ekstrakcijom značajki** (engl. *feature extraction*) – pretpostavka je da neki broj unaprijed definiranih i potom izlučenih značajki iz vremenskog niza **u dovoljnoj mjeri opisuje model** u pozadini
 - **Modeliranje učenjem značajki** (engl. *feature learning*) – pretpostavka je da učenje značajki niza u dubokim neuronskim mrežama **u dovoljnoj mjeri opisuje model** u pozadini
 - **ARMA modeli** (engl. *autoregressive moving average*) (i slični) – pretpostavlja se da se iduća vrijednost vremenskog niza može modelirati na temelju prozora prethodnih vrijednosti
 - i dr.

Mjere sličnosti

- Mjera sličnosti $D(T, U)$ između vremenskih nizova T i U omogućava provođenje različitih zadataka nad vremenskim nizovima
- Važno je da je **mjera robustna** s obzirom na skaliranje, vremensko iskrivljavanje, šum i stršeće podatke
- Primjeri mjera sličnosti:
 - **DTW** – *Dynamic Time Warping*, Berndt and Clifford, 1994.
 - mjera koja može pronaći podudaranje različitih podnizova vremenskog niza tako što omogućuje iskrivljavanje vremenske osi (*warping*)
 - optimalno poravnavanje je definirano kao najkraći put iskrivljavanja W u matrici udaljenosti
 - **CDM** – *Compression-Based Distance Measure*, Keogh et al. 2004.
 - jedna od mjera udaljenosti zasnovanih na kompresiji – **one analiziraju koliko je moguće dva niza zajedno komprimirati**, gdje se veća sličnost pokazuje u većem stupnju kompresije

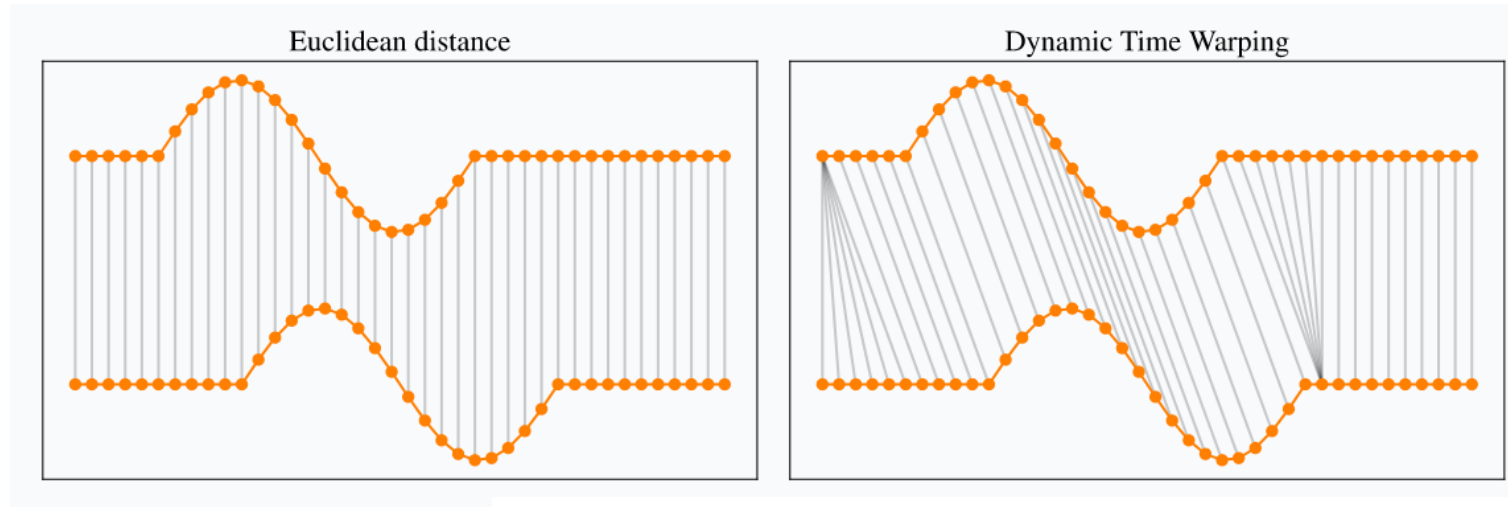
D. Berndt, J. Clifford. Using dynamic time warping to find patterns in time series. In Proceedings of the AAAI-94 Workshop on Knowledge Discovery in Databases. 229–248, 1994.

E. Keogh, S. Lonardi, C. Ratanamahatana. Towards parameter-free data mining. In *Proceedings of 10th ACM International Conference on Knowledge Discovery and Data Mining*. 206–215, 2004.

Mjere sličnosti – DTW

Elastičnost preslikavanja niza S u T kod DTW-a

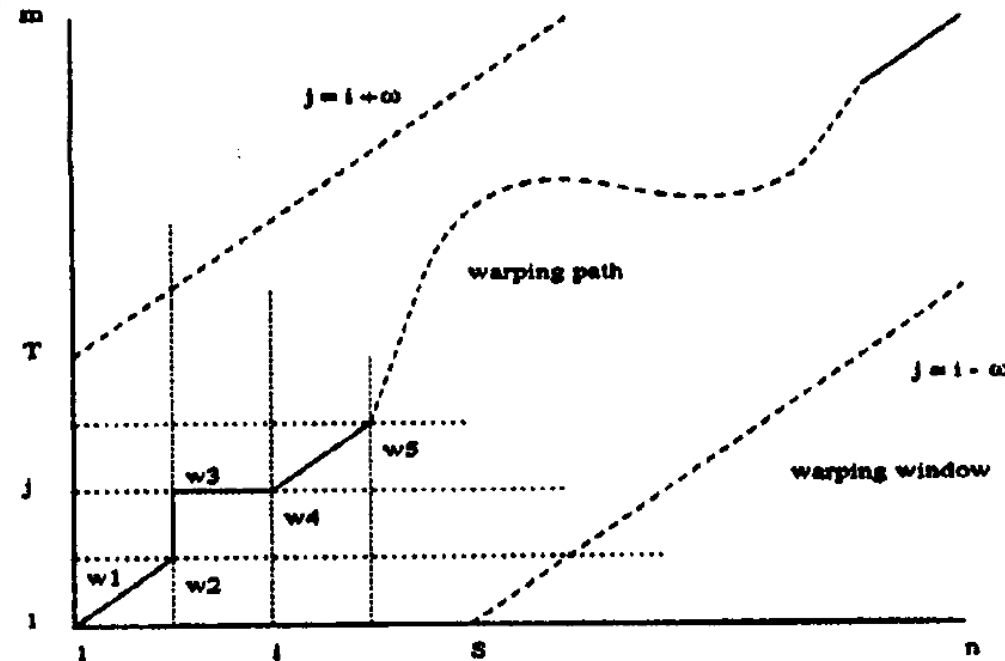
Izvor: <https://rtavenar.github.io/blog/dtw.html>



$$S = (s_1, \dots, s_n), s_i \in \mathbb{R}$$

$$T = (t_1, \dots, t_m), t_i \in \mathbb{R}$$

$$W = (w_1, \dots, w_n), w_i \in \mathbb{R}$$



W je slijed točaka u mreži koji definira odnos preslikavanja niza S u T , npr. w_3 ovdje bi označavao da se s_2 preslikao u t_3

Algoritmi za klasifikaciju i predikciju

Jednostavni predikcijski algoritmi: Pomični prosjek i eksponencijalno zaglađivanje

- Engl. *Simple Moving Average*
- Vrijednost **pomičnog prosjeka** računa se kao prosjek N prošlih vrijednosti, s jednakom težinom za svaku prošlu vrijednost

$$X_t = \frac{\sum_{i=1}^N X_{t-i}}{N}$$

- Engl. *Exponential Smoothing, Exponential Moving Average*
- Vrijednost **eksponencijalnog zaglađivanja** računa se sljedećom rekurzijom:

$$S_t = \begin{cases} X_0, & \text{za } t = 0 \\ (1 - \alpha)S_{t-1} + \alpha X_t, & \alpha \in [0, 1] \end{cases}$$

- Osim za predikciju sljedeće vrijednosti, EMA se koristi i za predobradu (zaglađivanje) vremenskog niza radi redukcije šuma

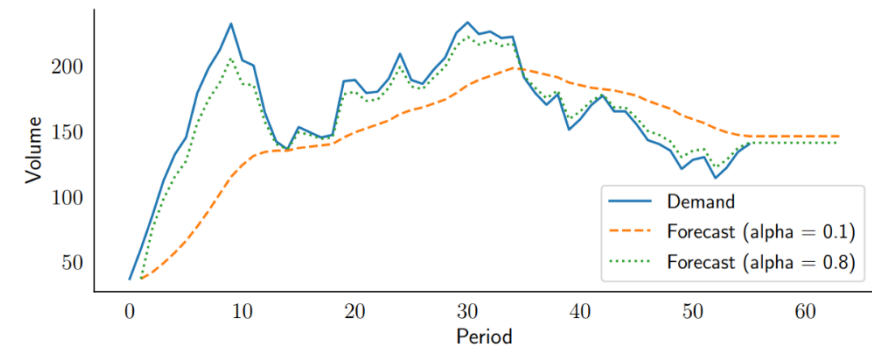


Figure 3.2: Simple smoothing

<https://www.linkedin.com/pulse/simple-exponential-smoothing-nicolas-vandeput/>

Predikcijski algoritam: ARIMA

- Engl. *Auto-Regressive Integrated Moving Average*
- ARIMA je statistički parametarski model namijenjen za analizu i predviđanje vremenskih nizova
- Predikcija ovisi o parametrima (p, d, q) modela **ARIMA** (p, d, q) :
 - **AR** (*Auto-Regressive*) – parametar p - **autoregresija** je zavisnost buduće vrijednosti promatranog niza o određenom broju prethodnih vrijednosti niza. Za $p = 2$, vrijednost u t_{i+1} ovisi o vrijednostima t_i i t_{i-1}
 - **I** (*Integrated*) – parametar d – **integracija** (zapravo **diferenciranje**) označava da se razmatra d -ta razlika između vrijednosti niza, a ne same vrijednosti. Time se uklanja trend (i postiže stacioniranje u slučaju nestacionarnosti srednje vrijednosti). Npr. za $d = 1$, nova vrijednost u t_{i+1} ovisi o razlici $t_i - t_{i-1}$
 - **MA** (*Moving Average*) – parametar q – **pomični prosjek** razmatra ovisnost između vrijednosti niza i rezidualne pogreške na vrijednostima niza ε_i modelom **utežanog pomičnog prosjeka**. Npr za $q = 3$, vrijednost u trenutku t_{i+1} ovisi o srednjoj vrijednosti niza (ako nije normiran na srednju vrijednost), rezidualnoj pogrešci u t_{i+1} i rezidualnim pogreškama u prethodna tri trenutka: ε_t , ε_{t-1} i ε_{t-2}

ARIMA

- Podešavanjem gore navedenih parametara konstruira se linearni regresijski model
- Vrijednost 0 može se koristiti za parametar, što ukazuje da se određeni element modela ne koristi
- Općeniti implicitni oblik modela $ARIMA(p, d, q)$ za vremenski niz prethodno normiran na srednju vrijednost μ :

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1 - L)^d X_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t$$

- Ovdje je X_t vrijednost niza u trenutku t , L operator kašnjenja (engl. *lag*), $L^i = X_{t-i}$, ϕ_i su parametri autoregresijskog dijela modela, θ_i su parametri dijela pomičnog prosjeka, a ε_t su rezidualne pogreške koje se računaju kao razlika X_t i modela prethodnih pogreški

ARIMA

- Neki specijalni slučajevi jednadžbe:
 - ARMA = ARIMA(p,0,q): $X_t = c + \sum_{i=1}^p (\phi_i X_{t-i}) + \sum_{i=1}^q (\theta_i \varepsilon_{t-i}) + \varepsilon_t$, c je konstanta (intercept) AR modela
 - AR = ARIMA(p,0,0): $X_t = \varepsilon_t + \sum_{i=1}^p (\phi_i X_{t-i})$
 - I = ARIMA(0,1,0): $X_t = X_{t-1} + \varepsilon_t$ - slučajni hod (engl. *random walk*)
 - MA = ARIMA(0,0,1): $X_t = \theta_1 \varepsilon_{t-1} + \varepsilon_t$
 - ARIMA (0,0,0): $X_t = \varepsilon_t$ - bijeli šum (engl. *white noise*)
- Predikcija se radi tako da pretpostavimo da tražimo nepoznati podatak X_{t+1} na temelju preostalih, već poznatih podataka

ARIMA

- Kako izabrati p , d i q na temelju podataka?
- Najbolje koristeći Akaikeov informacijski kriterij (AIC) za procjenu pogreške predviđanja
 - Izgradi se skup modela i istraži se koji model ima najnižu AIC vrijednost
 - $AIC = -2 \cdot \log(L) + 2(p + d + q)$, L je izglednost podataka u nizu
 - Napomena: AIC ne uzima u obzir stupanj diferencijacije (d) ARIMA modela

$$L(y_t | \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_t - \hat{y}_t)^2}{2\sigma^2}}$$

- gdje je θ dana kombinacija parametara (npr. $p = 1$, $q = 1$, $d = 0$) za predviđenu opservaciju y_t , ako se radi predviđanje više opservacija, onda je $L(\text{podaci} | \theta) = L(y_1 | \theta) * L(y_2 | \theta) * \dots * L(y_n | \theta)$

ARIMA

- Proširenja ARIMA modela:
 - **SARIMA** – Season ARIMA – uzima u obzir i komponentu perioda ponavljanja (sezonalnosti) u modelu, proširuje osnovna tri parametra s još **četiri dodatna**: **P: sezonalni stupanj AR**, **D: sezonalni stupanj razlike**, **Q: sezonalni stupanj pomičnog prosjeka** i **m: broj točaka u jednom periodu**
 - **SARIMA(p,d,q)(P,D,Q)m**, npr. za SARIMA(p,d,q)(2,0,0)12 model bi ukazivao na sezonalnost od 12 točaka u periodu (česti slučaj za godišnju sezonalnost za mjesečne podatke), gdje model uzima u obzir prethodne dvije opservacije s pomakom perioda: $X_{t-(m*1)}$ i $X_{t-(m*2)}$
 - **SARIMAX** – SARIMA + eXogeneous factors (polje dodatnih varijabli koje utječu na niz)
 - **VARIMA** – *Vector* ARIMA – modeli pogodni za predikciju na multivarijatnim vremenskim nizovima, predviđa se vektor vrijednosti na temelju vektora vrijednosti iz prethodnih koraka, koeficijenti AR i MA dijela jednadžbi su u općem slučaju matrice, npr. za VAR(1) model s dvije varijable:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix}$$

Korstanje, J. (2021). The SARIMAX Model. In: Advanced Forecasting with Python. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-7150-6_8

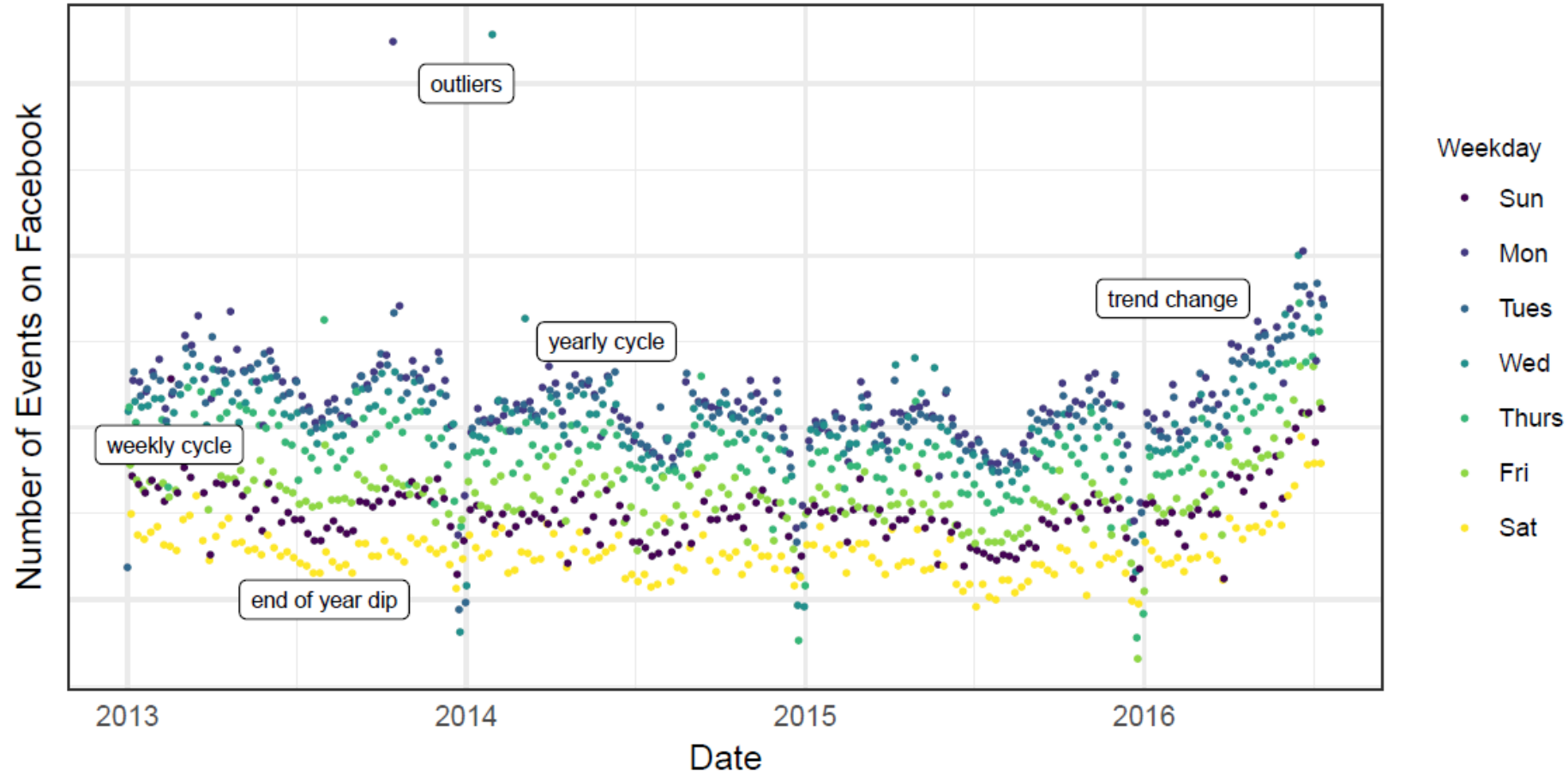
Predikcijski algoritam: PROPHET

- Facebook, Inc., 2017.
- PROPHET je generalizirani aditivni model za predikciju vrijednosti vremenskih nizova s tri glavne komponente modela: trendom, sezonalnošću i praznicima, prema izrazu:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t$$

- $g(t)$ (engl. *growth*) je po dijelovima linearna ili logistička funkcija za modeliranje **nesezonalnih promjena u nizu**
- $s(t)$ (engl. *seasonal*) je funkcija **periodičkih promjena** (sezonalnosti) u modelu
- $h(t)$ (engl. *holidays*) je funkcija **utjecaja praznika** tijekom jednog ili nekoliko dana, s nepravilnim razmacima, koju zadaje korisnik
- ε_t je **pogreška** (Gaussova razdioba) koja modelira one promjene koje nisu obuhvaćene ostalim komponentama
- PROPHET pokušava „fitati” i linearne i nelinearne funkcije za komponente gornje jednadžbe tako da najbolje odgovaraju opaženom vremenskom nizu

PROPHET



- Tipično ponašanje za poslovne podatke. Uočljiva sezonalnost na godišnjoj razini, na razini dana u tjednu, praznici (kraj godine) i stršeće vrijednosti, također uočiti promjenu trenda u 2016.

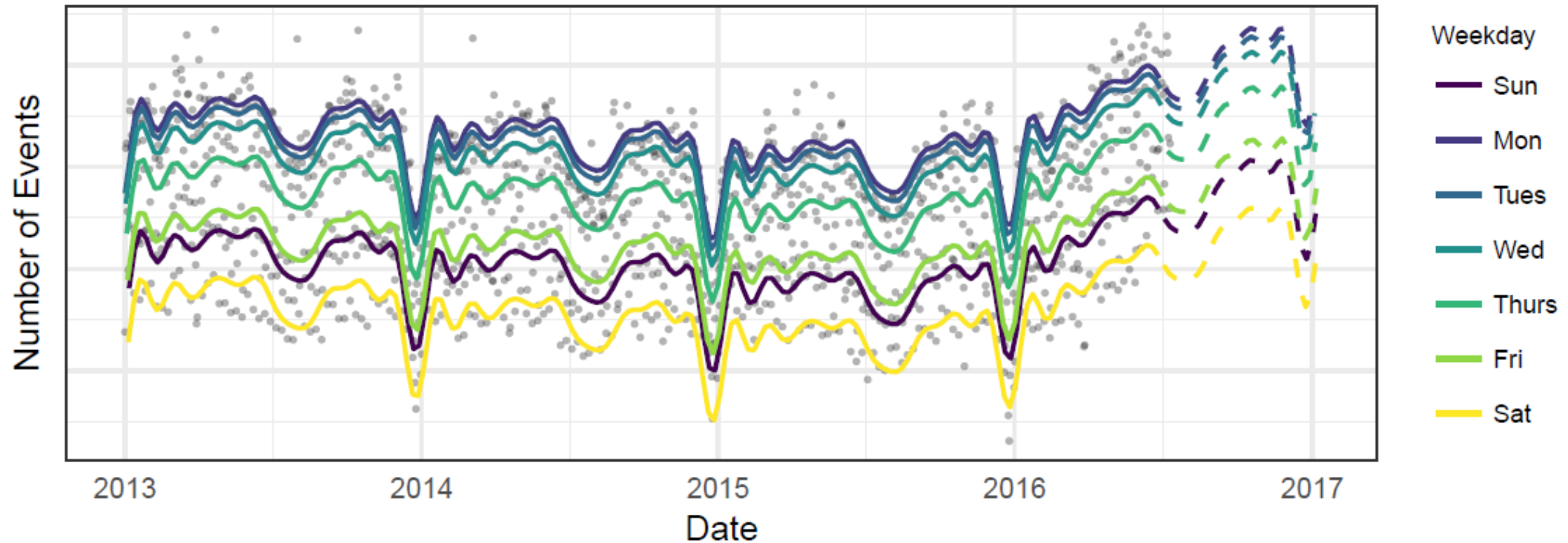
PROPHET

- Primjer uzimanja u obzir sezonalnosti $s(t)$ korištenjem Fourierovog niza:

$$s(t) = \sum_{n=1}^N \left(a_n \cos \left(\frac{2\pi nt}{P} \right) + b_n \sin \left(\frac{2\pi nt}{P} \right) \right)$$

- Potrebno je optimirati $2N$ parametara: $[a_1, b_1, \dots, a_N, b_N]$ za svaku vrstu sezonalnosti (npr. $P = 365,25$ za godišnju, $P = 7$ za tjednu), dobar se pokazao $N = 10$, a $2N$ parametara generira se po normalnoj razdiobi

PROPHET



- Ukupno predviđanje. Crtkane linije su one predikcije za koje u trenutku pisanja članka još nije bilo podataka.

Klasifikacijski algoritam: CIF

- **Kanonska intervalna šuma** (engl. *Canonical Interval Forest*), *Middlehurst et al. 2020*
- Napredni klasifikacijski ansambl stabala odluke učenih na **određenim značajkama** iz slučajno odabranih intervala vremenskih nizova
 - **Kanonske karakteristike vremenskih nizova (catch22)** (engl. *Canonical Time Series Characteristics*, *Lubba et al. 2019*)
 - **Tri sumarne statistike** intervala (srednja vrijednost, standardna devijacija i nagib)
- Podržava klasifikaciju multivarijatnih vremenskih nizova
- Usporediv po točnosti s drugim klasifikatorima vremenskih nizova (TS-CHIEF, STC, InceptionTime, ROCKET), vremenski i prostorno vrlo učinkovit

Middlehurst M, Large J, Bagnall A (2020) The canonical interval forest (CIF) classifier for time series classification. In: Proceedings of the IEEE international conference on big data

Lubba C, Sethi S, Knaute P, Schultz S, Fulcher B, Jones N (2019) catch22: canonical time-series characteristics. Data Min Knowl Disc 33(6):1821–1852

CIF – značajke catch22

- Značajke dobivene sustavnim istraživanjem i redukcijom skupa od **7658** značajki za općenite vremenske nizove dostupnih za izlučivanje u radnom okviru **hctsa** (*Highly Comparative Time Series Analysis toolbox*, Fulcher and Jones, 2017)
 - Istraživanje zasnovano na velikom broju skupova iz repozitorija UCR (preko 1000 skupova podataka vremenskih nizova)
 - Uklonjene značajke osjetljive na srednju vrijednost i varijancu, one koje se nisu mogle izračunati na više od 80% skupova podataka (npr. imaju negativne vrijednosti, ponavljajuće vrijednosti...)
 - Dalje se koristila redukcija u tri koraka:
 - Stratificirana unakrsna validacija sa stablom odluke za svaku značajku, zadržane ako su bile bolje od definiranog praga za balansiranu točnost
 - Hijerarhijsko grupiranje nad korelacijskom matricom značajki da se uklone redundantne značajke
 - Grupe značajki sortirane po balansiranoj točnosti, izabrana po jedna značajke iz svake od **22** grupe uzimajući u obzir složenost izračunavanja, interpretabilnost i balansiranu točnost
- Rezultantnih 22 značajke pokrivaju široki spektar osnovnih statistika, korelacija i entropija

B. D. Fulcher and N. S. Jones, "hctsa: A computational framework for automated time-series phenotyping using massive feature extraction," Cell systems, vol. 5, no. 5, pp. 527–531, 2017.

CIF – opis algoritma

- Koristi se 25 značajki (22 od catch22, 3 sumarne statistike)
- Svako stablo **poduzorkuje** $a = 8$ od ovih 25 značajki, svi čvorovi razmatraju istih a značajki
- Svako stablo razmatra $k = \sqrt{d} \cdot \sqrt{m}$ slučajno odabranih intervala, gdje je d dimenzionalnost problema (broj neovisnih vremenskih nizova), a m je duljina vremenskog niza (fiksna i jednaka za sve nizove)
 - Mogućih intervala u nizu duljine m je $m \cdot (m+1) / 2$
 - Položaj i duljina intervala u nizu kao i dimenzija niza se izabire **slučajno**
- Gradi se šuma od r stabala (default: 500)

CIF – pseudokod

Algorithm 1 buildCIF(A list of n cases of length m with d dimensions, $\mathbf{T} = (\mathbf{X}, \mathbf{y})$)

Parameters: the number of trees, r , the number of intervals per tree, k , and the number of attributes subsampled per tree, a (default $r = 500$, $k = \sqrt{d} \cdot \sqrt{m}$, and $a = 8$)

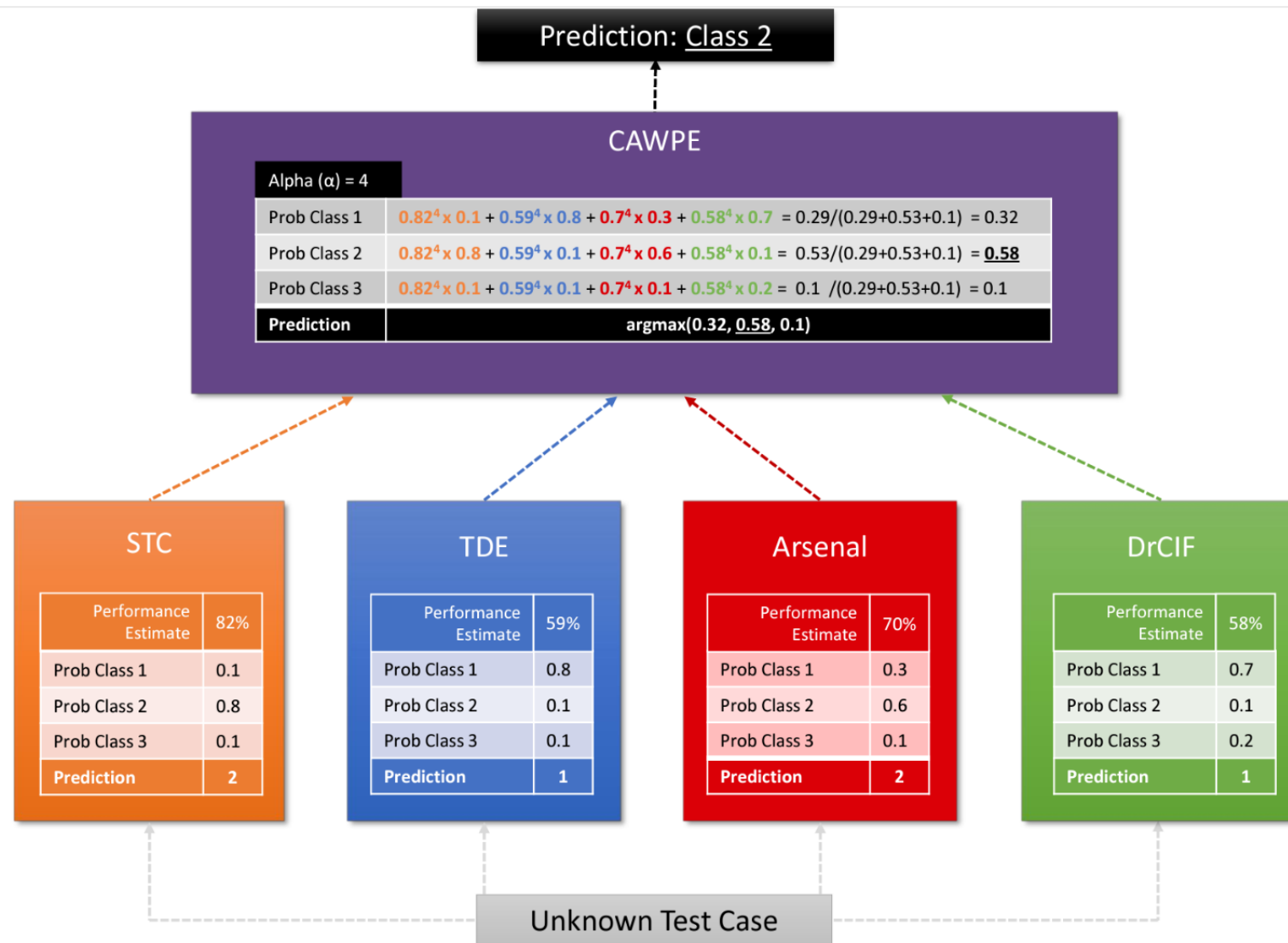
- 1: Let $\mathbf{F} = (\mathbf{F}_1 \dots \mathbf{F}_r)$ be the trees in the forest
 - 2: **for** $i \leftarrow 1$ to r **do**
 - 3: Let \mathbf{S} be a list of n cases ($s_1 \dots s_n$) with $a \cdot k$ attributes
 - 4: Let \mathbf{U} be a list of a randomly selected attribute indices ($u_1 \dots u_a$)
 - 5: **for** $j \leftarrow 1$ to k **do**
 - 6: $b = \text{rand}(1, m - 3)$
 - 7: $l = \text{rand}(b + 3, m)$
 - 8: $o = \text{rand}(1, d)$
 - 9: **for** $t \leftarrow 1$ to n **do**
 - 10: **for** $c \leftarrow 1$ to a **do**
 - 11: **if** $u_c \leq 22$ **then**
 - 12: $s_{t,a(j-1)+c} = \text{c22Feature}(\mathbf{u}_c, \mathbf{X}_{t,o}, \mathbf{b}, \mathbf{l})$
 - 13: **else**
 - 14: $s_{t,a(j-1)+c} = \text{tsfFeature}(\mathbf{u}_c, \mathbf{X}_{t,o}, \mathbf{b}, \mathbf{l})$
 - 15: $\mathbf{F}_i.\text{buildTimeSeriesTree}([S, \mathbf{y}])$
-

Klasifikacijski algoritam: HIVE-COTE v2

- *The Hierarchical Vote Collective Of Transformation-based Ensembles*, Middlehurst et al. 2021.
- *State-of-the-art* klasifikator za općenite vremenske nizove (engl. *time series classification*, TSC)
- Glavne značajke: meta-ansambl koji uključuje **4** komponente – pojedinačnih ansambala klasifikatora vremenskih nizova od kojih se svaki individualno (i istovremeno) uči na podacima, koristi eksponencirano utežano većinsko glasanje za donošenje odluke o klasi
- **Prednosti: vrlo točna klasifikacija vremenskih nizova** iz širokog područja primjene (ispitan na 112 univarijatnih i 26 multivarijatnih skupova iz UEA arhive za TSC), točnost usporediva ili bolja od svih ostalih predloženih pristupa
- **Nedostatci:** nešto sporiji od nekih konkurentskih metoda

Middlehurst M, Large J, Flynn M, Lines J, Bostrom A, Bagnall A (2021) HIVE-COTE 2.0: a new meta ensemble for time series classification. Machine Learning 110:3211{3243.

HIVE-COTE v2 – donošenje odluke



Izvor: Middlehurst M, Large J, Flynn M, Lines J, Bostrom A, Bagnall A (2021) HIVE-COTE 2.0: a new meta ensemble for time series classification. Machine Learning 110:3211{3243.

HIVE-COTE v2 – komponente – ideja

- ***Shapelet Transform Classifier (STC)***
 - Pretražuje niz za moguće visokokvalitetne shapelete, primijenjuje transformaciju značajki takvu da značajke predstavljaju udaljenosti do zadržanih shapeleta, gradi rotacijsku šumu na transformiranim značajkama
- ***Arsenal***
 - Anbambl koji se sastoji od k ROCKET klasifikatora s varirajućim parametrima kernela
- ***Temporal Dictionary Ensemble (TDE)***
 - Ansambl 1-NN klasifikatora, transformira svaki niz u histogram broja pojavljivanja riječi (simboličkog podniza). Za to se koristi prozor duljine w na čitavom nizu, svaki prozor se diskretizira u riječ duljine l na temelju abecede veličine α , gradi k klasifikatora varirajući ulazni skup i hiperparametre
- ***Diverse Representation Canonical Interval Forest (DrCIF)***
 - Osnovni klasifikator je stablo kao kod CIF-a, značajke su uzete iz većeg broja slučajno pozicioniranih intervala iz osnovnog niza, prvih razlika i periodograma, po 7 statističkih značajki i 22 catch-22 značajke za svaku od tri reprezentacije, gradi k klasifikatora koji se razlikuju po poziciji i duljini intervala

Middlehurst M, Large J, Flynn M, Lines J, Bostrom A, Bagnall A (2021) HIVE-COTE 2.0: a new meta ensemble for time series classification. Machine Learning 110:3211{3243.

Klasifikacijski algoritmi – performance

- Usporedba najboljih algoritama za TSC

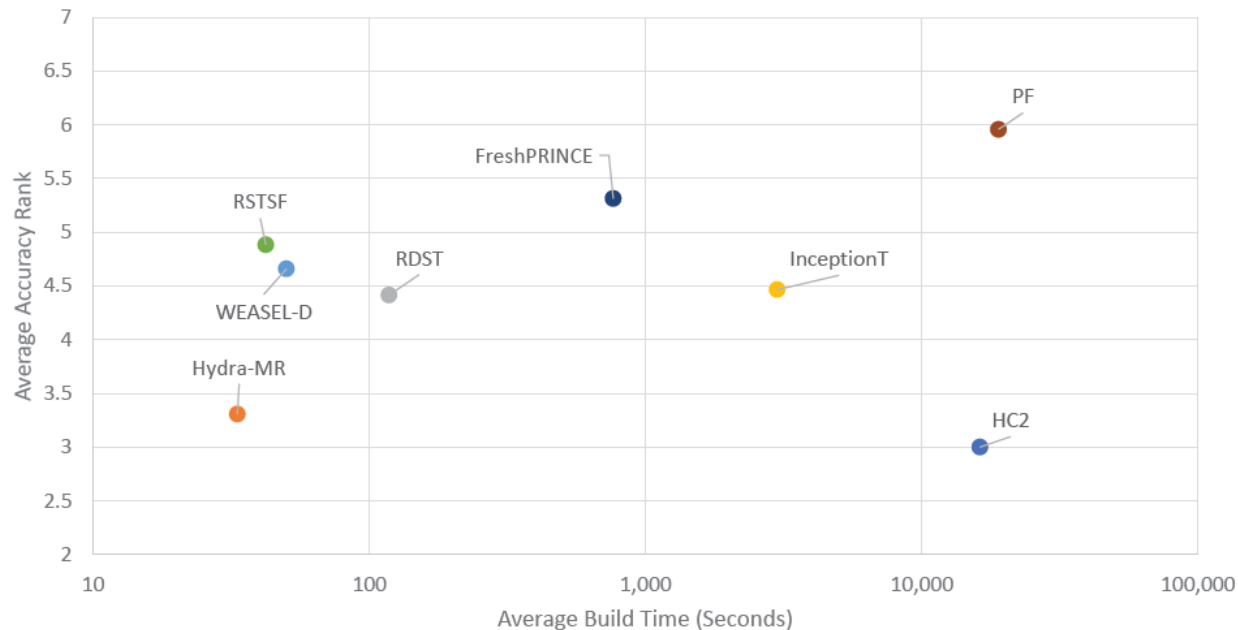


Table 18 Average rank of classifiers on 30 resamples of 142 TSC problems split by problem type.

	DEVICE (11)	ECG (7)	IMAGE (34)	MOTION (27)
HC2	2.000 (1)	2.143 (1)	3.441 (2)	2.759 (1)
Hydra-MR	2.455 (2)	2.571 (2)	2.912 (1)	3.056 (2)
InceptionT	4.455 (4)	4.286 (3)	4.676 (5)	3.833 (4)
RDST	3.909 (3)	4.571 (5)	3.971 (4)	3.722 (3)
WEASEL-D	5.364 (7)	4.286 (4)	3.706 (3)	4.741 (5)
RSTSF	5.091 (6)	5.429 (6)	5.324 (6)	5.926 (6)
FreshPRINCE	4.909 (5)	5.571 (7)	5.676 (7)	6.019 (8)
PF	7.818 (8)	7.143 (8)	6.294 (8)	5.944 (7)
	SENSOR (35)	SIMULATED (12)	SPECTRO (12)	
HC2	3.414 (1)	3.333 (2)	2.167 (1)	
Hydra-MR	3.957 (2)	3.083 (1)	3.792 (3)	
InceptionT	4.200 (4)	3.917 (4)	5.958 (8)	
RDST	4.871 (5)	5.667 (7)	5.083 (5)	
WEASEL-D	4.900 (6)	6.833 (8)	4.083 (4)	
RSTSF	4.071 (3)	5.167 (6)	3.417 (2)	
FreshPRINCE	5.171 (7)	3.750 (3)	5.667 (6)	
PF	5.414 (8)	4.250 (5)	5.833 (7)	

Izvor: Middlehurst, M., Schäfer, P., & Bagnall, A. (2024). Bake off redux: a review and experimental evaluation of recent time series classification algorithms. Data Mining and Knowledge Discovery, 1-74.

Modeliranje niza ekstrakcijom značajki

Modeliranje vremenskog niza ekstrakcijom značajki

- Moguće je izračunati različite **značajke vremenskog niza** koje će ga reprezentirati te ih potom iskoristiti za učenje klasifikatora (za klasifikacijske probleme) ili regresora (za regresijske probleme)
- U ovom pristupu važno je da broj izlučenih značajki m bude **značajno manji** od broja točaka u nizu n kako bi se smanjila dimenzionalnost problema, ali **ne premali** kako bismo pokrili sve potrebne informacije
- U literaturi je poznat velik broj značajki vremenskih nizova (mjere se u tisućama)
 - Linearne značajke (najčešće: vremenska i frekvencijska domena)
 - Nelinearne značajke (najčešće: kaos, kompleksnost)
 - Entropijske značajke
 - Značajke dobivene na temelju grafova i mreža
 - Prostorne značajke (za neke vrste multivarijatnih nizova)

Modeliranje niza ekstrakcijom značajki

- Primjer radnog okvira za ekstrakciju značajki iz općih vremenskih nizova: **hctsa** (Matlab)
 - B.D. Fulcher and N.S. Jones, 2017.
 - 7000+ pojedinačnih značajki, osim ekstrakcije podržava i neke dodatne operacije s vremenskim nizovima (npr. normalizaciju, grupiranje, klasifikaciju, generiranje niskodimenzionalne reprezentacije)
 - <https://github.com/benfulcher/hctsa>
- Primjer radnog okvira za ekstrakciju značajki iz općih vremenskih nizova: **tsfresh** (Python)
 - M. Christ *et al.*/ Blue Yonder GmbH, 2016.
 - 77 značajki, 10 minimalnih statističkih značajki; osim uobičajenih statističkih značajki, uključuje spektralne značajke te neke značajke entropijske značajke
 - <https://tsfresh.readthedocs.io/en/latest/>
- Primjer radnog okvira za ekstrakciju značajki iz specifičnih vremenskih nizova: **HRVFrame** (Java)
 - A. Jović and N. Bogunović, 2011.
 - Opće i domenski specifične značajke za niz otkucaja srca, preko 40 metoda, neke računaju veći broj pojedinačnih značajki
 - <http://www.zemris.fer.hr/~ajovic/hrvframe/hrvframe.html>

Modeliranje niza ekstrakcijom značajki

- Pregled općih i domenskih značajki za analizu EEG-a (detaljan ali i dalje neiscrpan prikaz)

Group	Feature name	Abbr.	Group	Feature name	Abbr.	Group	Feature name	Abbr.	Group	Feature name	Abbr.
Time-domain	Mean		Frequency-domain	θ/β		Entropies	Shannon entropy		Undirected spt.	Imaginary component of Coh	
	Median			$(\theta+\alpha)/(\alpha+\beta)$			Rényi's entropy			Phase-lag index	PLI
	Variance			γ/δ			Tsallis entropy			Weighted phase lag index	wPLI
	Standard deviation			$(\gamma+\beta)/(\delta+\alpha)$			Kraskov entropy	KE		Debiased weighted PLI	dwPLI
	Skewness			Reflection coefficients			Spectral entropy	SEN		Pairwise phase consistency	PPC
	Kurtosis			Partial correlation coefficient			Quadratic Rényi's SEN	QRSEN		Generalized synchronization	
	Zero-crossing rate	ZCR		Wavelet coefficients			Response entropy	RE		Synchronization likelihood	SL
	Number of waves			Phase coupling			State entropy	SE	Directed spt.	Mutual information	MI
	Wave duration		Nonlinear	Hurst exponent	H		Wavelet entropy	WE		Mutual information in freq.	MIF
	Peak amplitude			Rényi scaling exponent			Tsallis wavelet entropy	TWE		Cross-RQA	
	Instantaneous frequency	IF		Rényi gener. dim. multifractals			Rényi's wavelet entropy	RWE		Correlation length	ξKLD
	Hjorth parameters			Capacity dimension D0	D0		Hilbert-Huang SEN	HHSE		Granger causality	
	Mobility			Information dimension D1	D1		Log energy entropy	LogEn		Spectral Granger causality	
	Activity			Correlation dimension D2	D2	Complex networks	Multiresolution entropy			Phase slope index	PSI
	Complexity			Katz fractal dimension	KFD		Kolmogorov's entropy			Number of vertices	
	K-complex			Petrosian fractal dimension	PFD		Nonlinear forecasting entropy			Number of edges	
	Energy	E		Higuchi fractal dimension	HFD		Maximum-likelihood entropy			Degree	D
Frequency-domain	Mean			Fractal spectrum			Coarse-grained entropy			Mean degree	
	Median			Lyapunov exponents	LE		Correntropy	CoE		Degree distribution	
	Variance			Lempel-Ziv complexity	LZC		Approximate entropy	ApEn		Degree correlation	r
	Standard deviation			Central tendency measure	CTM		Sample entropy	SampEn		Kappa	k
	Skewness			Auto-mutual information	AMI		Quadratic sample entropy	QSE		Clustering coefficient	
	Kurtosis			Temporal irreversibility			Multiscale entropy	MSE		Transitivity	
	Delta	δ		Recurrence rate	RR		Modified multiscale entropy	MMSE		Motif	
	Theta	θ		Determinism	Det		Composite multiscale entropy	CMSE		Characteristic path length	
	Alpha	α		Laminarity	Lam		Permutation entropy	PE		Small worldness	
	Beta	β		Average diagonal line length	L		Rényi's permutation entropy	RPE		Assortativity	
	Gamma	γ		Maximum length of diagonal	Lmax		Permutation Rényi entropy	PER		Efficiency	
	Sigma	σ		Max. length of vertical lines	Vmax		Multivariate PE	MvPE		Local efficiency	
	θ/α			Trapping time	TT		Tsallis permutation entropy	TPE		Global efficiency	
	β/α			Divergence	Div		Dispersion entropy	DisE		Modularity	
	(θ+α)/β			Entropy of recurrence plot	ENTR		Amplitude-aware PE	AAPE		Centrality degree	
						Undirected spt.	Bubble entropy	BE		Closeness centrality	
							Differential entropy	DifE		Eigenvalue centrality	
							Fuzzy entropy	FuzzyEn		Betweenness centrality	
							Transfer entropy	TrEn		Diameter	d
							Coherence			Eccentricity	Ecc
							Partial coherence			Hubs	
Frequency-domain	Mean		Nonlinear	Lyapunov exponents	LE		Phase coherence			Rich club	
	Median			Lempel-Ziv complexity	LZC	Undirected spt.	Phase-locking value	PLV		Leaf fraction	
	Variance			Central tendency measure	CTM		Coherency	Coh		Hierarchy	Th
	Standard deviation			Auto-mutual information	AMI						
	Skewness			Temporal irreversibility							
	Kurtosis			Recurrence rate	RR						

Stancin I, Cifrek M, Jovic A. A Review of EEG Signal Features and their Application in Driver Drowsiness Detection Systems. Sensors (Basel). 2021 May 30;21(11):3786. doi: 10.3390/s21113786

Dubinska analiza podataka



Zaključak

- Dubinska analiza vremenskih nizova ima više ciljeva kojima je zajedničko modeliranje informacija pohranjenih u vremenskom nizu
- Algoritmi implicitno ili eksplicitno razmatraju sličnosti nizova i podnizova
- ARIMA i njezine varijante statistički su alati za modeliranje ponašanja (trenda, pomaka, šuma i sezonalnosti) u vremenskim nizovima te omogućuju predviđanje vrijednosti
- PROPHET je primjer pametnog, inženjerski složenog algoritma za potrebe poslovnog predviđanja
- CIF i drugi klasifikacijski algoritmi nastoje pronaći najučinkovitiji način za klasifikaciju na velikom broju različitih skupova podataka
- Vremenski niz uspješno se modelira korištenjem pristupa s velikim brojem općih i domensko specifičnih značajki