Modeliranje dinamike fizikalnih sustava naprednim metodama dubokog učenja

Filip Mirković
Mentor: izv. prof. dr. sc. Davor Horvatić

Prirodoslovno matematički fakultet, Fizički odjsek
Bijenička cesta 28, 10000 Zagreb

Dinamički sustavi i opis problema

Dinamički sustav

Definicija
Dinamički sustav je uređena trojka prostora stanja $\mathcal{S}$, *vremenskim* skupom $\mathcal{T}$ i funkcije

$$\mathcal{F} : \mathcal{S} \times \mathcal{T} \longrightarrow \mathcal{S} \tag{1}$$

- Dinamički sustav je kontinuran ako vrijedi: $\mathcal{T} \subseteq \mathbb{R}$
- U tom slučaju $\mathcal{F}$ je obična ili parcijalna diferencijalna jednadžba

Evolucija stanja $\boldsymbol{\phi} \in \mathcal{S}$ dana je

$$\partial_t \boldsymbol{\phi} = \mathcal{F}_{\boldsymbol{\eta}}\big(\mathsf{x}, \boldsymbol{\phi}, \nabla_x \boldsymbol{\phi}, \nabla_x^2 \boldsymbol{\phi}, \boldsymbol{\phi} \nabla_x \boldsymbol{\phi}, \ldots\big) \tag{2}$$

- $\boldsymbol{\eta}$ je vektor parametara jednadžbe. Npr. ω u harmoničkom oscilatoru.

Opis problema

Želimo:
Za svaki $\phi_0 \in \mathcal{S}$ poznavati buduće ϕ_t za proizvoljno velike t.

Priječi nas:

1. $\mathcal{F}$ je nelinearna
2. Dinamika je kaotična
3. $\mathcal{S}$ je visokodimenzionalan
4. **x** je visokodimenzionalan
5. $\mathcal{F}$ je djelomično ili potpuno nepoznat
6. Sve od navedenog

Sretni smo ako:
Za veliki broj $\boldsymbol{\phi_0}$ uspijevamo precizno procijeniti ϕ_t do nekog konačnog $T \in \mathcal{T}$.
Iznos T određen je pojedinim problemom.

Tipična rješenja

Konstruckija minmalnog modela

- Rješenje nije skupo
- Temelji se na teoriji, iskustvu i podatcima
- Model opisuje samo općenite značajke dinamike
- Numerika je i dalje potrebna

Numeričko rješavanje diferencijalnih jednadžbi

- Dinamika treba biti potpuno poznata
- Proces je skup
- Visoka dimenzionalnost prostora postaje velika prepeka

Modeliranje iz podataka

Nije potrebno poznavati dinamiku

Često je potrebno puno podataka

Transformerski model i mehanizam pozornosti

Transformerski model

Uvod

- Originalno nastali 2017. u području obrade prirodnog jezika (NLP) [14]
- Pokazali se izuzetno uspješnim, npr. GPT3 chatbot je transformer
- Brzo se šire na ostala prodručja strojnog učenja, npr. računalni vid

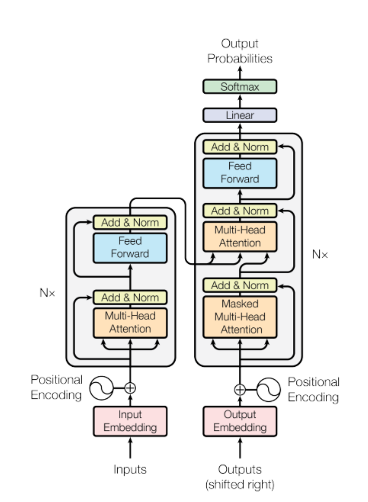
Od prevođenja teksta do modeliranja dinamike

- U obradi prirodnog jezika riječi se često reprezentiraju vektorima
- Uz vektorske reprezentacije prevođenje teksta postaje nastavljanje niza vektora
- U računalima dinamika fizikalnih sustava je diskrtizirana i u formi niza vekora

Transformerski model - arhitektura

- Koder-Dekoder struktura
- Koder lijevo, dekoder desno
- U modeliranju dinamike koristimo samo dekoder

Pozornost
Ključna komponenta transformera je višestruka pozornost (*eng.* Multi-Head Attention)



Slika 1: Arhitektura originalnog transformera. Preuzeto is [14].

Motivacija
Neka je $\{\xi_1, ...\xi_n\}n$ opaženi niz vektora stanja koji predstavlja evoluciju sustava kroz vrijeme.

- Ako je dinamika poznata (i možemo baratati jednadžbama), jedina informacija koju trebamo za nastavak vremenskog niza je ξ_n
- Ako ne poznamo dinamiku korisno je modelu dati informaciju o c prethodnih stanja
- Time dobivamo kontekstni interval $\mathcal{C} = \{\xi_{n-c}, ..., \xi_n\}$
- Iz konekstnog intervala konstruiramo kontekstni vektor **c** kojeg pružamo modelu kao dodatnu informaciju o protekloj evoluciji

Ideja: Kontekstni vektor **c** sadrži informaciju o dalekosežnim vremeskim ovisnostima.

Pojednostavljena pozornost - recept:
1.) Računamo sličnost između ξ_n i svih vektora $\xi_l \in \mathcal{C} \subset \mathbb{R}^d$. Mjera sličnosti je *normirani* skalarni produkt

$$s_l = \frac{\xi_n \cdot \xi_l}{\sqrt{d}} \tag{3}$$

2.) Iz sličnosti dobivamo koeficijente pozornosti

$$\alpha_l = \mathrm{softmax}(\boldsymbol{s})_l = \frac{e^{s_l}}{\sum_{i=1}^{C} e^{s_i}} \tag{4}$$

3.) Kontekstni vektor je

$$\boldsymbol{c} = \sum_{l=1}^{C} \alpha_l \xi_l \tag{5}$$

Ostatku modela se pružaju informacije o ξ_n i $\boldsymbol{c}$ za stvaranje predikcije o ξ_{n+1}.

Mehanizam pozornosti [3] - Potpuna pozornost

Unapređenja pojednostavljene pozornosti:
Uvodimo tri neuronske mreže $\mathcal{F}_v$, $\mathcal{F}_q$, $\mathcal{F}_k$ takve da

$$\mathcal{F}_v(\xi_i) = \boldsymbol{v}_i \in \mathbb{R}^d \tag{6}$$

$$\mathcal{F}_q(\xi_i) = \boldsymbol{q}_i \in \mathbb{R}^{d_k} \tag{7}$$

$$\mathcal{F}_k(\xi_i) = \boldsymbol{k}_i \in \mathbb{R}^{d_k} \tag{8}$$

Veličine $\boldsymbol{v}$, $\boldsymbol{k}$, $\boldsymbol{q}$ zovemo vrijednost, *ključ* i *upit*. Računamo ih za svaki $\xi \in \mathcal{C}$.

Ostatku modela pružaju se infromacije o $\boldsymbol{v}_n$ i $\boldsymbol{c}$ za stvaranje predikcije o ξ_{n+1}.

Potpuna pozornost - recept:
1.) Sličnosti računamo između svih ključeva i zadnjeg upita $\boldsymbol{q}_n$.

$$s_l = \frac{\boldsymbol{q_n} \cdot \boldsymbol{k}_l}{\sqrt{d_k}} \tag{9}$$

2.) Koeficijente pozornosti računamo na isti način

$$\alpha_l = \mathrm{softmax}(\boldsymbol{s})_l = \frac{e^{s_l}}{\sum_{i=1}^{c} e^{s_i}} \tag{10}$$

3.) Kontekstni vektor računamo u prostoru *vrijednosti*

$$\boldsymbol{c} = \sum_{i=1}^{c} \alpha_i \boldsymbol{v}_i \tag{11}$$

Mehanizam pozornosti [4] - Primjer

Zašto uvodimo $\mathcal{F}_v$, $\mathcal{F}_k$ i $\mathcal{F}_q$?

- Dodavanjem mreža povećavamo ekspresivnost modela.
- Želja nam je da mreže nauče bitne značajke stanja ξ.
- Značajke su reprezentirane vektorima $\boldsymbol{q}$, $\boldsymbol{k}$ i $\boldsymbol{v}$.
- Bez korištenja $\mathcal{F}_v$, $\mathcal{F}_k$ i $\mathcal{F}_q$, sličnost pridodaje svim komponentama ξ jednaku važnost.

Mehanizam pozornosti [5] - Primjer

Primjer:
Promotrimo gibanje točkaste čestice u potencijalu 2-D harmoničkog oscilatora

$$V(x, y) = ax^2 + \epsilon y^2$$

Neka je $\epsilon << a$ te vektor stanja $\xi = (x, y, \dot{x}, \dot{y})$.

Kada bismo računali mjere sličnosti između dvaju vektora, običan skalarni produkt pridodaje jednaku važnost koordinatama x i y.

Zbog manje krutosti potencijala u y smijeru y kooridana je većeg raspona, međutim dinamika je osjetljivija na perturbacije u x i $\dot{x}$.

Primjerice, uvođenjem linearih operatora $\mathcal{F}_v = \mathbb{1}$, $\mathcal{F}_q = \mathcal{F}_k = [w_{ij}]$, gdje su $w_{xx} > w_{yy}$ te $w_{i,j} = 0$ za $i \neq j$ zaobilazimo ovaj problem.

Mehanizam pozornosti [6] - Paralelizacija

Najskuplja komponenta izrade dubokog modela je treniranje

Paralelizacija
Tijekom treniranja računamo $\boldsymbol{v}$, $\boldsymbol{q}$ i $\boldsymbol{k}$ za svaki $\xi \in \mathcal{C}$ te iz njih radimo matrice V, Q i K koje imaju c redova, jedan za svaki vektor iz $\mathcal{C}$.

Sličnosti tada možemo izraziti matričnim množenjem koje se efikasno izvršava na grafičkim karticama.

$$S = \frac{QK^T}{\sqrt{d_k}} \tag{12}$$

Analogno dobivamo matricu pozornosti:

$$A = \mathrm{softmax}(S) \tag{13}$$

Normalizacija u softmax se provodi po stupcima matrice S.

Mehanizam pozornosti [7] - Maskiranje

Problem:

- Za neki vektor $\xi_i \in \mathcal{C}$ izračunat ćemo koeficijente pozornosti za sve preostale ξ_j uključujući i buduće ξ_j, $j > i$.
- Ako računamo stanje ξ_n zadnji vektor na kojeg model smije obratiti pozornost je ξ_{n-1}.

Rješenje: Maskiranje

- Sve elemente iznad glavne dijagonale matrice S postavimo na $-\infty$ (u praksi koristimo jako neki jako negativan broj)
- Time je koefijcijent pozornosti 0 za te elemente.

Mehanizam pozornosti [8] - Uvođenje vremenskog uređaja

Problem

- Tijekom treniranje model obrađuje sve vektore iz $\mathcal{C}$ istovremeno
- Postojeći vremenski uređaj ovime je narušen

Kodiranje vremenskog uređaja

- Neka je $\xi_{pos\,i}$ i-ta komponenta vekotra na poziciji *pos* u kontekstnom intervalu.
- Toj komponenti dodajemo vrijednost $PE_{pos\,i}$

$$PE_{pos\,i} = \begin{cases} \sin(pos/10000^{\frac{i}{d}}) & i \text{ je paran} \\ \cos(pos/10000^{\frac{i}{d}}) & i \text{ je neparan} \end{cases} \tag{14}$$

Funkcija gubitka transformera

Neka je $\{(\mathcal{C}_i, \widetilde{\mathcal{C}_i}) \quad i = 1 \dots D\} = \mathcal{D}$ skup podataka. S time da je $\mathcal{C}_i = \{\xi_j \mid j = n \dots n + c\}$ ulazni kontekstni inteval, a $\widetilde{\mathcal{C}_i} = \{\xi_j \mid j = n + 1 \dots n + c + 1\}$ konekstni interval kojeg model treba predvidjeti iz $\mathcal{C}_i$.

Funkcija gubita je

$$\mathcal{L} = \sum_{i=1}^{D} \sum_{j=1}^{c} ||\xi_j^i - \widehat{\xi_j^i}||_2^2 \quad \xi_j^i \in \widetilde{\mathcal{C}_i} \tag{15}$$

Simbol $\widehat{\xi}$ označava sanja koja predviđa model.

Vektorske reprezentacije

Učenje vektorskih reprezentacija sustava [1]

Koji sustavi nisu opisani vektorom?

- Sve što je dosada objašnjeno funkcionira ako je stanje sustava $\boldsymbol{\phi}$ reprezentirano vektorom ξ.
- Bitna klasa dinamički sustava koji nisu vektori su polja, npr. brzina strujanja fluida, elektromagnetsko polje, valne funkcije...

Naivan pristup

Recimo da je $\phi^t(x, y)$ dvodimenzionalno skalarno polje, npr. iznos brzine toka fluida.

U računalu polje je spremljeno u obliku matrice ϕ^t_{ij}, gjde $i = 1 \ldots N$, $j = 1 \ldots M$.

Naivno možemo reprezentirati ϕ^t_{ij} vektorom tako da matricu *razvučemo* po jednoj dimenziji, recmo po stupcima.

Dobivamo vektor $\vec{\phi}^t_n$, $n = 1 \ldots N \cdot M$, međutim komponente originalne matrice ϕ^t_{ij} i ϕ^t_{ij+1} su u vektoru $\vec{\phi}^t$ udaljene za N redova.

Učenje vektorskih reprezentacija [2]

Narušena lokalnost

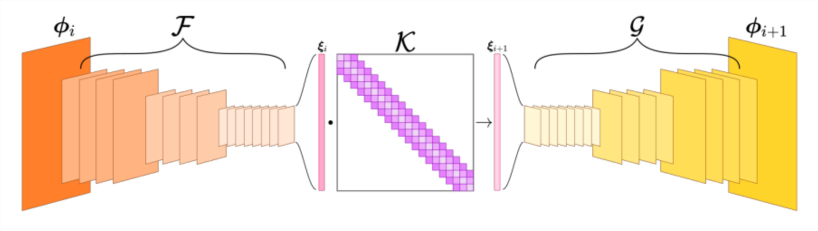
Budući da je evolucija sustava opisana parcijalnom diferencijalnom jednadžbom, na evoluciju ϕ^t_{ij} utječu stanja poput $\phi^t ij + 1$ i $\phi^t i + 1, j$.

U vektoru $\vec{\phi}^t$ ta su stanja jako *daleko* jedna od drugog.

Dinamika na mnogostrukosti

- Osim narušene lokalnosti problem s naivnom vektorskom reprezentacijom jest što je vektorski prostor koji sadrži $\vec{\phi}$ prevelik.
- Skup svih matrica tipa $N \times M$ sadrži puno više od rješenja Navier-Stokesovih jednadžbi koje opisuju tok fluida, npr. Matrica kojoj su komponente slučajni brojevi isto je sadržana u $\mathbb{R}^{N \times M}$.
- Dinamika danog sustava odvija se na nekoj mnogostrukosti, a ne čitavom $\mathbb{R}^{N \times M}$.

$\implies$ Uvodimo neuronske mreže kojima učimo prikladnu vektorsku reprezentaciju sustava iz podataka.



Slika 2: Koder-Koopman-Dekoder struktura. Preuzeto iz [5].

Koder-Kooman-Dekoder struktura (DKD)

1. Koder $\mathcal{F}$ preslikava stanje sustava $\phi \in \mathcal{S}$ u prikladnu vektorsku reprezentaciju $\xi \in \Xi$
2. Dekoder $\mathcal{G}$ rekosntruira stanje sustava ϕ iz reprezentacije ξ. On je inverz kodera.
3. Koopmanov $\mathcal{K}$ operator uređuje prostor reprezentacija.

Učenje vektorskih reprezentacije [4] - Koder i dekoder

Koder
Za generiranje vektorskih reprezentacija koristimo neuronsku mrežu $\mathcal{F} : \mathcal{S} \longrightarrow \Xi$.

Čest izbor slojeva kodera su konvolucijski slojevi. Konvolucijski slojevi uče lokalne značajke stanja ϕ [2] [7].

Dekoder
Nakon što obavimo predikcije nad vektorima iz Ξ želimo od njih ponovno rekonstruirati stanja sustava.

Za to koristimo dekoder $\mathcal{G} : \Xi \longrightarrow \mathcal{S}$.

Kada je to moguće, arhitektura dekodera je zrcalna arhitekturi kodera.

Učenje vektorskih reprezentacija [5] - Koopmanov operator

Problem

- Ništa ne prijeći sustav kodera i dekodera da nasumično ne razbacaju $\mathcal{F}(\phi)$ po Ξ
- Želimo da originalna dinaimka bude na neki način prisutna u Ξ
- Problem zaobilazimo tako da stanja $\boldsymbol{\phi}^t$ i $\phi^{t+\Delta t}$ preslikamo u bliska stanja ξ^t i $\xi^{t+\Delta t}$, s obzirnom na euklidsku normu.

Koopamnov operator

Stanja ϕ možemo preslikati u beskonačno dimenzionalni vektorski prostor u kojem sukcesivna stanja generiramo grupnim translacijama [10].

$$\xi(t) = e^{Ht}\xi(t_0) \tag{16}$$

Operator $\mathcal{K}$ zocemo Koopmanovim operatorom. Detalji teorije Koopmanovog operatora su u [3], [8], [1].

Učenje vektorskih reprezentacija [6] - Koopmanov operator

Operator translacija u vremenu razvijamo do linearnog člana

$$\xi(t_0 + \Delta t) = (1 + H\Delta t)\xi(t_0) = \mathcal{K}\xi(t_0) \tag{17}$$

Operator $\mathcal{K}$ zovemo Koopmanov operator. Pretpotavljamo ga u obliku jedinične matrice koja ima korekcije reda $\mathcal{O}(\Delta t)$.

Služi za strukturiranje latentnog prostora Ξ.

Učenje vektorskih reprezentacija [7] - Trening DKD sustava

Funkcija gubitka

Funkciju gubitka $\mathcal{L}_{DKD}$ dijelimo na tri dijela: $\mathcal{L}_{DKD} = \mathcal{L}_R + \mathcal{L}_D + \lambda \mathcal{L}_{reg}$

1. Rekonstrukcijski gubitak

$$\mathcal{L}_R = \sum_{i=1}^{D} \sum_{j=0}^{T} \mathrm{MSE}\big(\phi_j^i, \mathcal{G} \circ \mathcal{F}(\phi_j^i)\big) \tag{18}$$

Osigurava da su $\mathcal{F}$ i $\mathcal{G}$ aproksimativno inverzni.

2 . Dinamički gubitak

$$\mathcal{L}_D = \sum_{i=1}^{D} \sum_{j=1}^{T} \mathrm{MSE}\big(\phi_j^i, \mathcal{G} \circ \mathcal{K}^j \mathcal{F}(\phi_0^i)\big) \tag{19}$$

Osigurava da Koopmanov operator generira sukcesivna stanja, odnosno uređuje prostor Ξ.

Regularizacijski gubitak

$$\mathcal{L}_{reg} = ||\mathcal{K}||_2^2 \tag{20}$$

Regularizira oblik Koopanovog operatora [9].

Gissingerov sustav

Gissingerov sustav [1] - Jednadžbe gibanja

Gissingerov sustav

- Uveden kao minimalna aproksimacija magneto-hidrodinamičkih jednadžbi [6].
- Stanje sustava je opisano trima modovima: dipolnim D, kvadrupolnim Q i brzinskim V.
- Namjenjen da opiše obrtaje polova Zemljinog magnetskog polja - prejednostavan.

Jednadžbe gibanja

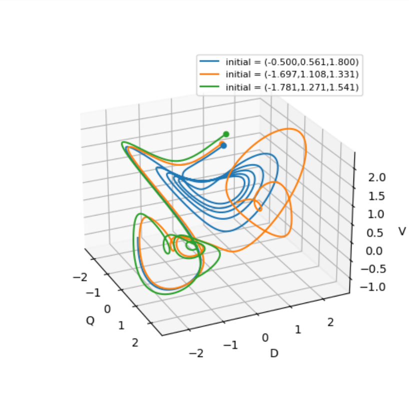
$$\dot{Q} = \mu Q - VD \tag{21}$$

$$\dot{D} = -\nu D + VQ \tag{22}$$

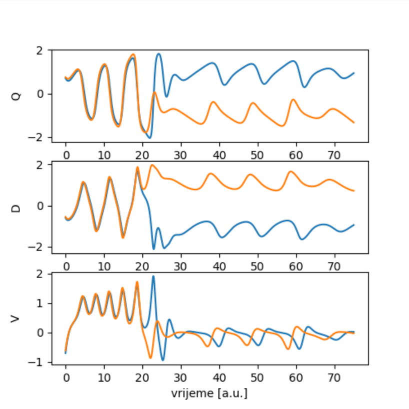
$$\dot{V} = \Gamma - V + DQ \tag{23}$$

Posjeduje bogatu dinamiku koja postaje kaotična za određene parametre μ, ν, Γ. Sustav je analiziran u kaotičnom režimu za $\mu = 0.119$, $\nu = 0.1$, $\Gamma = 0.9$

Gissingerov sustav [2] - Primjer kaosa



Slika 3: Bogata dinamika Gissingerovog sustava. Narančasta krivulja doživljava jedan kaotičan obrtaj dipolnog moda *D*.



Slika 4: Dvije krivulje počinju iz jako bliskih početnih uvijeta te naglo divergiraju jedna od druge.

Gissingerov sustav [3] - Lyapunovljev eksponent

Lyapunovljev eksponent

- Daje mjeru brzine divergencije bliskih trajektorija
- Za kaotične atraktore brzina divergencije je eksponencijalna
- Koristimo ga za ocjenu kvalitete modela

Kratak izvod

Neka je stanje sustava definirano varijablama x_μ, a njihova evolucija

$$\dot{x_\mu} = f_\mu(\mathsf{x})$$

Tajektoriji x dodajemo infinitezimalu smetnju $\delta\mathsf{x}$.

$$\dot{x_\mu} + \delta\dot{x_\mu} = f_\mu(\mathsf{x} + \delta\mathsf{x})$$

Razvojem u red dobivamo jednadžbu za evoluciju smetnje $\delta\mathsf{x}$

$$\delta x_\mu = \sum_\nu \partial_\nu f_\mu(\mathsf{x})\delta x_\nu \tag{24}$$

Jakobijan $[\partial_\nu f_\mu]$ izračunat je duž nesmetane trajektorije.

Gissingerov sustav [4] - Lyapunovljevo vrijeme

Integriramo li jednadžbu (24) divergenciju trajektorija ocjenjujemo veličinom

$$d(t) = \sqrt{\sum_{\nu} \delta x_{\nu}(t)^2}$$

Lypaunovljev eksponent definiramo

$$\lambda = \lim_{t \to \infty} \lim_{d(0) \to 0} \frac{1}{t} \log\left(\frac{d(t)}{d(0)}\right)$$

Lyapunovljevo vrijeme definirano je kao inverz Lyapunovljeva eksponenta i daje vremensku skalu na kojoj sustav postaje kaotičan [13] [12]. Za Gissingerov sustav smo Lyapunovljevo vrijeme procijenili numerički postupkom definiranim u [11].

$$\tau_L = 13.4$$

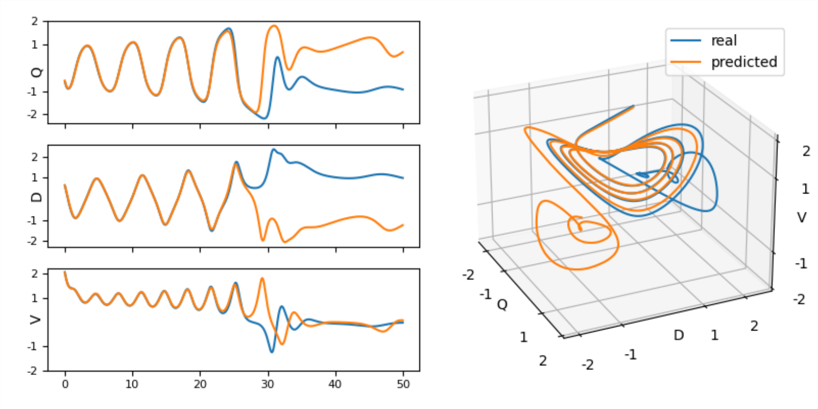
Rezultati

Rezultati [1] - Transformer protiv kaosa

Rezultati

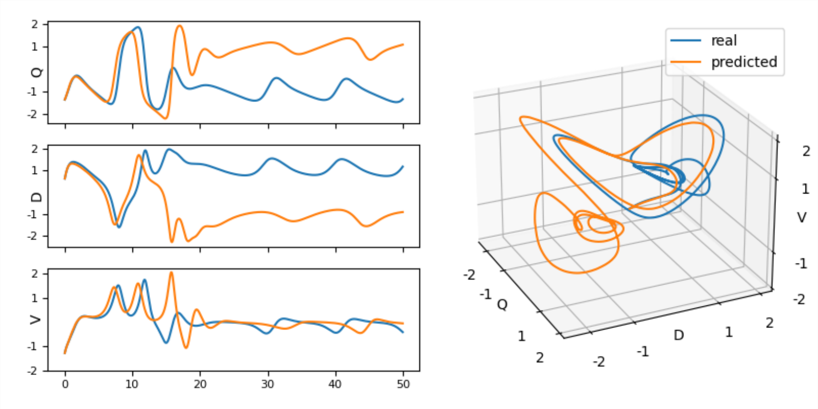
- Transformer uspjeva naučiti općenite krakteristike faznog prostora čime sam postaje kaotičan [4].
- Model smatramo uspješnim, ako daje precizne predikcije barem dva Lyapunovljeva vremena.
- Predikcije se za večinu početnih uvjeta poklapaju sa stvarnim rješenjima na vremenskoj skali dva Lyapunovljeva vremena.
- Postoje regije u na atraktoru gdje je model izrazito uspješan ili izrazito neuspješan.
- Na kraju kaos uvijek prevlada.

Rezulati [2] - Tipična uspješnost



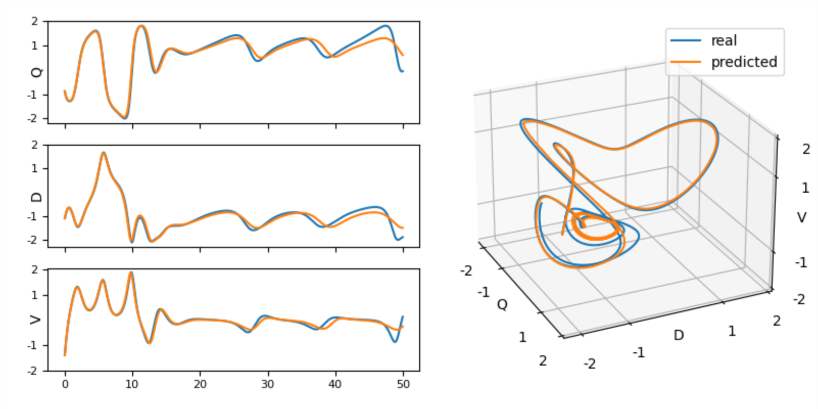
Slika 5: Predikcija modela i stvarna dinamika sustava prate se na skali dva Lyapunovljeva vremena. Model zadržava ovakvu preciznost za većinu početnih uvjeta na atraktoru.

Rezulati [3] - Niska uspješnost



Slika 6: Postoje i nazigled nasumične točke u kojima model gubi preciznost iznimno brzo $\sim 0.5\tau_L$.

Rezulati [4] - Velika uspješnost



Slika 7: Iznimno dugotrajna preciznost modela $\sim 3\tau_L$.

Zaključak i diskusija

Zaključak i diskusija [1] - Prednosti i mane

Prednosti modela

- Arhitekutra trasformera pokazuje se uspješnom u modeliranju dinamike fizikalnih sustava.
- Može se koristiti u situacijama kada sve dinamičke varijable nisu poznate ili osmotrive.

Mane modela

- Potrebne su velike količine podataka o fenomenu.
- Kaos uvijek prevlada i predikcije postanu neprecizne.
- Loša interpretabilnost

Zaključak i diskusija [2] - Moguća unaprjeđenja

Unaprjeđenja modela

- Trenirnanje DKD sustava uz transformer (eng. finetuning)
- Bolje kodiranje vremenskog uređaja
- Ugrađivanje fizikalnih spoznaja o problemu

Hvala na pažnji!

Pitanja?

Literatura i

Petar Bevanda, Stefan Sosnowski, and Sandra Hirche.
Koopman operator dynamical models: Learning, analysis and control.
Annual Reviews in Control, 52:197–212, 2021.

Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković.
Geometric deep learning: Grids, groups, graphs, geodesics, and gauges, 2021.

Akshunna S. Dogra and William T Redman.
Optimizing neural networks via koopman operator theory.
2020.

Jaime Lopez Garcia and Angel Rivero Jimenez.
Phase space learning with neural networks, 2020.

Nicholas Geneva and Nicholas Zabaras.
Transformers for modeling physical systems.
Neural Networks, 146:272–289, feb 2022.

C. Gissinger.
A new deterministic model for chaotic reversals.
The European Physical Journal B, 85(4), apr 2012.

Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville.
***Deep Learning*.**
MIT Press, Cambridge, MA, USA, 2016.
http://www.deeplearningbook.org.

Di Luo, Jiayu Shen, Rumen Dangovski, and Marin Soljačić.
Koopman operator learning for accelerating quantum optimization and machine learning, 2022.

- Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre G.R. Day, Clint Richardson, Charles K. Fisher, and David J. Schwab.
  A high-bias, low-variance introduction to machine learning for physicists.
  Physics Reports, 810:1–124, may 2019.
- Igor Mezic.
  Koopman operator, geometry, and learning, 2020.
- Thomas S. Parker and Leon O. Chua.
  Practical Numerical Algorithms for Chaotic Systems.
  Springer-Verlag, Berlin, Heidelberg, 1989.
- Steven H. Strogatz.
  Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering.
  Westview Press, 2000.

- Michael Tabor.
  Chaos and Integrability in Nonlinear Dynamics-An Introduction.
  Wiley, 1989.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin.
  Attention is all you need.
  In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.