

Jak bude zítra?

Skoro jako dneska.

Dan Lessner

Osnova

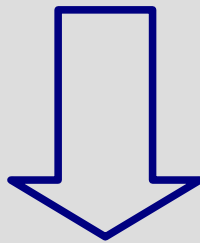
O čem budu mluvit

- Motivace, popis problému
- Vhodná data
- Použitá architektura sítě
- Zajímavá zjištění
- Kde je zakopaný pes?
- Tady
- Co teď s tím
- Další zajímavá zjištění
- Jak dál
- Závěr

Motivace, popis problému

Zápočet

1. Chci vědět, co zítra na sebe
2. Nemám ve sklepě superpočítač
3. Vím jak bylo dnes
4. Moji přátelé v Evropě to vědí taky
5. Nechci s tím mít příliš práce



Pokus o predikci vývoje časové řady na základě několika posledních členů. Nesimuluju, používám empirické metody.

Vhodná data

Aby to fungovalo

Hodí se o problému
(počasí) něco vědět:

- dnes je skoro jako včera
- „je“ zejména
 - teplo
 - vlhko
 - těžko
 - prach, sluníčko, emise, bio...
- počasí k nám chodí ve frontách
- počasí se různí podle ročního období

Proto vezmu data:

- nedávná
- o průměrné denní teplotě
 - to je dost zjednodušující (pro mě, ne pro sít')
- z různých míst
 - odborník: 800 – 900 km od Klementina
- o roční době

Vhodná data

Konkrétněji

V Tróji sídlí Katedra meteorologie a ochrany prostředí a na ní dvoučlenná Skupina deterministického času.



Tímto děkuji Mgr. Jiřímu Mikšovskému za pomoc se sháněním vhodných dat.

- Na <http://eca.knmi.nl> je k dispozici volně přístupná databáze ECA (European Climate Assessment) s daty ze stovek stanic v Evropě.
 - různá období
 - uvádění teplot v desetínách °C ušetří tečku

Vhodná data

Příprava

- Výběr vhodných stanic k měření (odhadem), stažení časových řad
- Nalezení společného průniku (1966 – 1997)
 - Poslední tři roky jako testovací
- Normalizace teplot: lineárně jako z $[-40;40]$ na $[-1;1]$ (přenosová funkce bude \tanh)
- Zavedení informace o ročním období
 - separace sezón
 - čtyři nové vstupy
 - sinus a cosinus dne v roce
- Příklad vstupu: 0.831; 0.556; 0,185; -0,008; 0,105; 0,015; 0,128; 0,0003; 0,025; 0,188; 0,0005; 0,14

Použitá architektura sítě

Jak na to

- Požadovaná přesnost předpovědi: 1°C
 - Není to moc, ale menší rozdíl stejně nepoznám
 - Navíc se teplota během dne stejně dost mění
 - Průměrný rozptyl denní teploty v Klementinu (1905 – 2005) činí víc než 7°C
 - Taky je tu otázka přesnosti samotných dat
- Zkoušel jsem různé varianty
 - různé učící algoritmy a jejich parametry
 - jedna nebo dvě skryté vrstvy s až čtyřnásobkem počtu vektorů ze vstupní vrstvy
 - 10|17|9|1, 10|13|4|1, 10|10|10|1, 10|20|1, 10|10|1, 10|5|1, 10|3|1, **10|1 !!!**

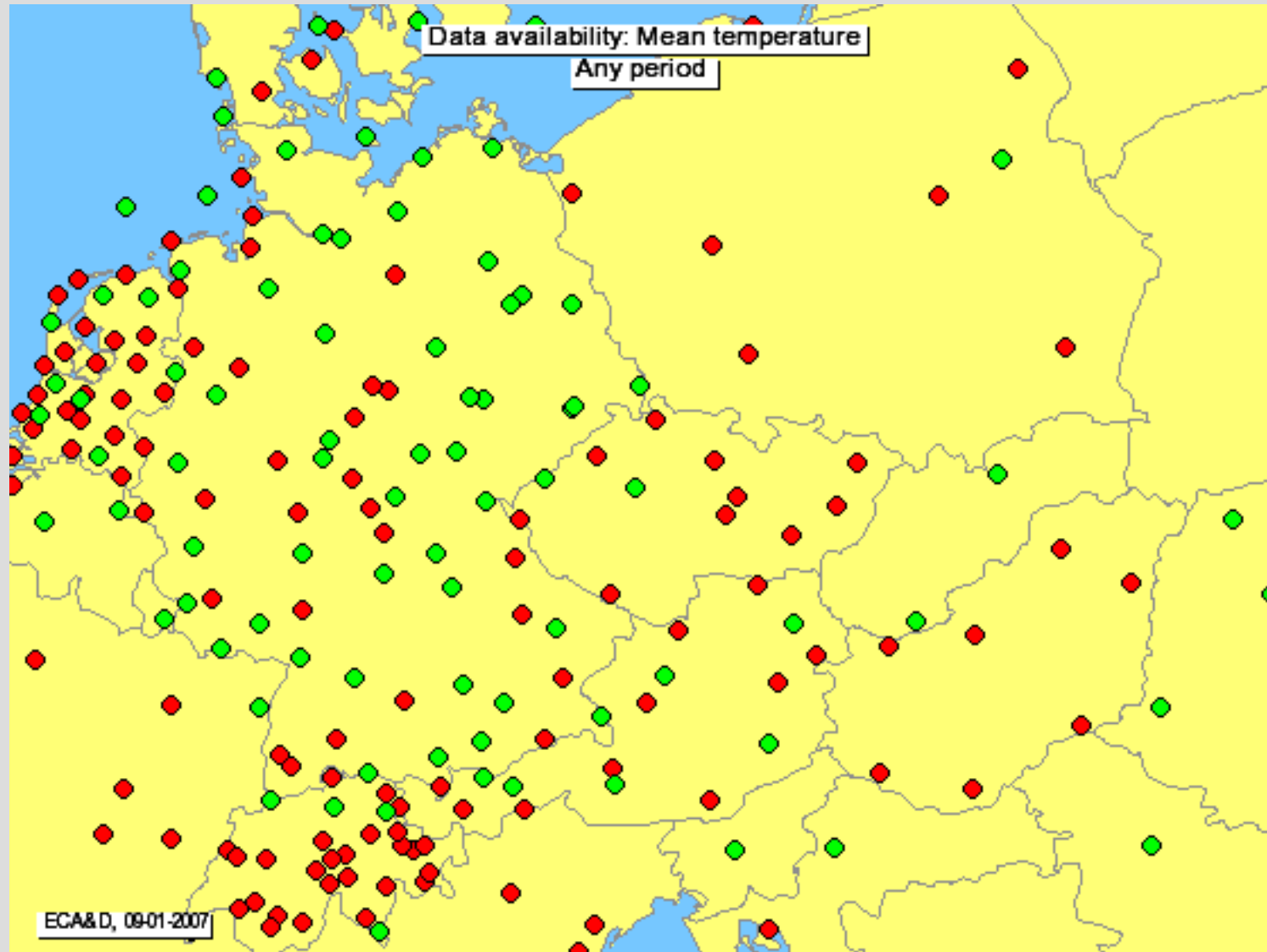
Zajímavá zjištění

Něco tu smrdí

- Záleží na inicializaci
- Nejlépe se síť učila s LM
 - Za 17 epoch bylo hotovo
 - Celkem dobře vycházelo i trainrp
- Přesnost by šla hnát výš, ovšem s řádovým nárůstem počtu epoch
 - Třeba na 400
 - To není nic špatného, jiné problémy se tak učí běžně.
- Na konkrétní architektuře příliš nezáleželo
 - Opravdu až příliš... co je tedy obtížného na předpovědi počasí?

Kde je zakopaný pes?

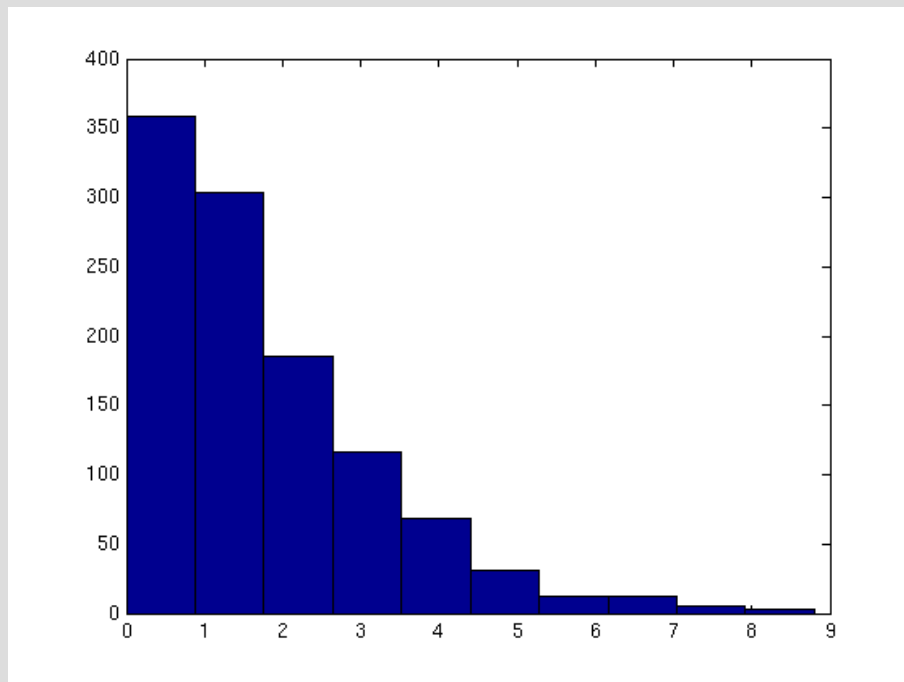
10 532 000 km²



Tady

Zítřa bude SKORO jako dneska.

Kdyby v Klementinu dnešek znamenal zítřek:



Obléct se na denní průměr o deset stupňů jinde může někoho stát zdraví. A to bohatě stačí třeba dvakrát za rok.

Ale sít' tyhle případy mezi stovkami ostatních ignoruje.

A podobně vypadá i rozložení chyb sítě.

(je jen o něco strmější)

Co teď s tím?



- Zvýším šance sítě přidáním informací ze včerejška (ale zato uberu polovinu míst)
 - nová data (ne)zapomenu normalizovat
- Snížím práh citlivosti na dlouhé učení
- Nebudu měřit střední kvadratickou odchylku, ale absolutní hodnotu chyby (funkce mae)

Další zajímavá zjištění

Odkud vítr vane

- Délka učení poskočila z desítek na stovky epoch
- Takže na velikosti sítě už dost záleží
- Co se týče zvrátů, moc jsem si nepomohl
 - Dokážu naučit síť „velmi“ dobře na malých množinách, ale pro neznámá data je nepoužitelná (chyba vzroste třeba na stonásobek)
 - To „velmi“ je mnohem lepší než v předchozích pokusech (maximální chyba 2°C, výskyt v desetinách procenta)
 - Ale nedokážu říct, jestli by šlo stejný výsledek získat pro větší vstupy
 - Začíná být zřejmé, co je složitého na předpovědi počasí
- Opravdu se může stát, že síť leze po plošině a najednou narazí na údolí chybové funkce

Jak dál

Musím taky spát

- Kudy by šlo pokračovat
 - Podívat se na váhy – která místa mají na zítřejší teplotu nejmenší vliv (pak je ubrat a urychlit učení)
 - Přidat kvalitativně nová data na vstup (tlak, vlhkost)
 - Začít znovu, lépe
 - od začátku se zaměřit na vyhledávání zvrátů v počasí (že zítra bude jako dnes, odhadne často každý trouba)
 - Průběžně sít' přizpůsobovat podle každodenního pozorování – počasí a jeho změny se vyvíjí i dlouhodobě
 - funkce adapt

Závěr

Problém vyřešen

Jak bude zítra?
Skoro jako dneska, ale
ráno se radši podívám
z okna.

