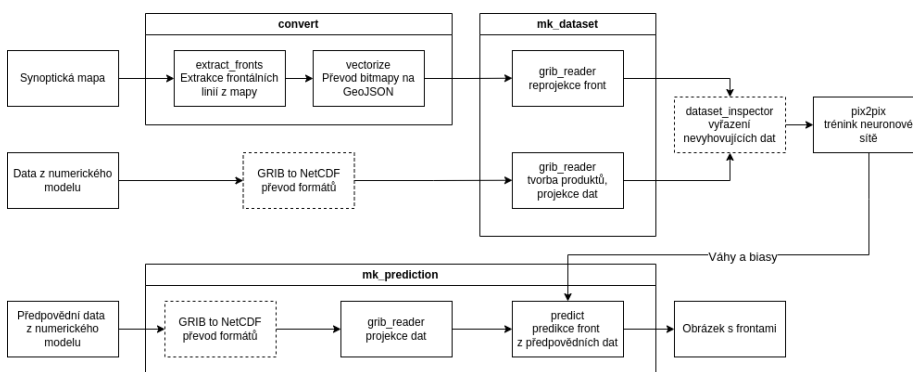


1 Synopse

1.1 Úvod

Tato práce představuje inovativní přístup k predikci polohy a typu atmosférických front prostřednictvím kombinace numerických modelů pro předpověď počasí (NWP), vypočítaných meziproductů a pokročilých metod strojového učení – konkrétně generativní neuronové sítě. Cílem je vytvořit otevřený systém, který bude využitelný jak odborníky, tak dalšími navazujícími systémy, které vyžadují strojově popsanou synoptickou situaci z téměř libovolných modelových výstupů. Architektonicky se skládá se většího množství skriptů provázaných univerzálních utilit.



Obrázek 1: Schéma architektury softwaru

1.2 Numerická vstupní část

Pro trénink modelu, která bude mít za úkol preedikovat místa s potenciálním výskytem atmosférických front je potřeba sehnat velké množství relevantních dat. Jelikož budeme chtít, aby byla predikce tvořena z dat některého volně dostupných modelů, musíme využít data co nejvíce podobná a zároveň je potřeba obsáhnout pestré množství situací - zde byla vybrána doba 10 let. Bohužel klasické běhy NWP modelů se nearchivují dostatečně dlouho. Z toho důvodu byla zvolena reanalýza počasí ERA5 od ECMWF, kterou lze získat z portálu *Copernicus climate change service*. Ten umožňuje stažení vybraných parametrů i za dlouhé období. Stažený balíček dat je pak ve standardizovaném formátu pro modelová data *GRIIB*. Tento formát má však své nevýhody, protože není strukturovaný tak, aby z něj bylo možné rychle vybírat parametry v náhodně organizovaných časech a v omezeném území. Proto byl pro práci zvolen formát *HDF5*, do nějž byla všechna data konvertována.

1.3 Reálné situace

Dalším zdrojem vstupních dat jsou ručně kreslené mapy analýzy synoptické situace, konkrétně od ČHMÚ. Ty jsou určeny pro profesionály a proškolenou veřejnost, tudíž jsou distribuovány formou obrázků s mapou a symboly. Pro potřeby systému byl vytvořen extraktor, který z těchto bitmap po georeferenci

extrahuje pomocí metod strojového vidění předměty zájmu – tedy fronty včetně typů a středy tlakových níží. Jelikož se jednotlivé utility uvnitř frontanimu řídí unixovou filozofií, jedná se o samostatně spustitelný skript s jasným vstupem a výstupem ve formátu GeoJSON. Výstupy lze tedy využít i například v geografických informačních systémech.

1.4 Ukazatelé výskytu front

Aby bylo možné zaručit spolehlivější detekci front modelem, byl před tréninku i predikci vložen mezistupeň, který má za úkol vytvořit sadu produktů indikujících frontální rozhraní a případně jeho charakter. Tyto produkty byly vybrány na základě dvou parametrů – jestli se používají meteorologem při manuální subjektivní analýze synoptické situace a jestli jsou dostatečně spolehlivé pro indikaci.

V současnosti se používají tyto produkty:

1. Horizontální teplotní gradient
2. Termální frontální parametr
3. Proudění jet streamu ve výšce 300 hPa
4. Frontogenetická funkce
5. Q-vektory

Pro zjednodušení jsou všechny produkty složeny do jedné matice kompozitu a při přípravě datasetů přiřazeny k nejbližší profesionálem zakreslené situaci a to jak dle časového razítka, tak i podle výrazných útvarů, zejména hlubších tlakových níží.

1.5 Neuronová síť

Pro predikci samotných front byl použit koncept neuronové sítě zvaný pix2pix, což je generativní adversariální neuronová síť určená pro transformaci jednoho obrázku na druhý, která umožňuje poměrně rozmanitou škálovatelnost. Skládá se ze dvou dílčích neuronových sítí – z generátoru a diskriminátoru, přičemž samotný trénink je prováděn jako vzájemné soupeření těchto dvou částí. Generátor má za úkol vytvářet nové obrázky z meteorologických dat, zatímco diskriminátor má za úkol odhalovat, jak moc je pravděpodobné, že je výstup produktem generátoru. Tento proces je prováděn iterativně, dokud není dosaženo co nejvyšší kvality výstupů. Prvně jsou tedy použita meteorologická data, na jejichž základě je generátorem vytvořena predikce frontálních linií. Tento výstup je zároveň s reálnou situací předán generátoru, který vytvoří úsudek. Na tomto úsudku je provedena křížová entropie oproti poli jedniček a je tak vyčíslena ztráta. Tato ztráta je pak spolu se ztrátou mezi generovanou a reálnou situací, která byla vyčíslena díky Hausdorffově vzdálenosti, využita k aktualizaci vah generátoru pomocí optimizéru. Samozřejmě probíhá i průběžná aktualizace diskriminátoru na základě ztráty mezi jeho úsudky se vstupem reálných a falešných dat.

1.6 Hodnocení výsledků

Pro hodnocení výsledků byla zpracována metodika, která používá způsoby, jež se používají i při hodnocení úspěšnosti předpovědí počasí. Proces spočívá v použití testovacího datasetu, ze kterého jsou vygenerovány predikce. Tyto predikce jsou poté zhodnoceny pomocí průměrné Hausdorffovy vzdálenosti reálných a predikovaných front. Poté jsou situace rozděleny do tří verifikačních kategorií, které vycházejí z teorie detekce signálu na Zásach, Chybu a Falešný poplach. Z počtu situací v jednotlivých kategoriích jsou vypočítána verifikační kritéria, která už udávají objektivní informaci o stavu modelu. Jsou udávány hodnoty Přesnost, Zkreslení, Pravděpodobnost detekce a Poměr falešných poplachů.

	Fronta detekována	Fronta nedetekována
Fronta existuje	Zásah (HIT)	Chyba (MISS)
Fronta neexistuje	Falešný poplach (FA)	Správné vyloučení

Tabulka 1: Přehled verifikačních kategorií

Tabulka níže ukazuje, že na vzorku 282 situací byla nejúspěšnější predikce okluzních front, která je zhruba o 10 procent vyšší než u front studených a teplých. Z hodnoty zkreslení je vidět, že má model obecně problém s podhodnocením výskytu front, tedy nedetekuje je v dostatečné míře. V současné době probíhá další vývoj, který díky augmentaci datasetu a úpravám produktů vykazuje postupné zlepšování.

Prvek	A	BIAS	POD	FAR
Studená fronta	0,46	0,88	0,60	0,33
Teplá fronta	0,48	0,92	0,62	0,33
Okluzní fronta	0,57	0,89	0,69	0,22
Všechny typy	0,52	0,90	0,65	0,28

Tabulka 2: Vypočítaná verifikační kritéria

1.7 Využití

Výstupy systému lze prezentovat buď přímo vykreslením do lidmi čitelné mapy, přičemž by v tomto případě sloužily jako alternativa map vydávaných meteorologickými službami. Přidanou hodnotou by byl volitelný časový krok, případně možnost tvorby interaktivních map.

Vhodnějším využitím jsou ale navazující systémy, které budou pracovat se synoptickou situací ve formě objektů v prostoru. To by umožňovalo klasifikaci situací a otevřelo by to možnost automaticky volit pro určitá území vhodné modely s vhodnou korekcí. Výsledkem by mohlo být částečné zlepšení přesnosti střednědobé předpovědi počasí.