

„Byl jsem zkrátka ve správnou dobu na správném místě,“ říká doktorand FEI Marek Pecha

28.6.2023 - | Vysoká škola báňská - Technická univerzita Ostrava

Jeho disertační práci na téma vývoj řešičů a jejich implementace pro úlohy strojového učení vedou dva školitelé, doc. David Horák z FEI a Dr. Richard Mills z Argonne National Laboratory v USA. V rámci svého doktorského studia se Marek kromě jiného dostal k aplikaci strojového učení pro lokalizaci požárů na Aljašce. Ptáte se jak? Odpověď na tuto otázku a ne jenom na ni najdete v článku!

Po nástupu na doktorské studium na Katedře aplikované matematiky FEI VŠB-TUO dostal Marek za úkol rozšířit PERMON toolbox o modul pro distribuované strojové učení založené na technice zvané Support Vector Machines (SVM). *„Jednoho dne přišel David a řekl Vaškovi Haplovi a mně, že by rád vyzkoušel, jak obstojí algoritmy profesora Zdeňka Dostála, pro řešení klasifikačních úloh. Tak jsme se do toho s Vaškem pustili. V průběhu mého druhého ročníku dostal Vašek místo postdoca na ETH Zurich, ale základní funkcionalitu jsme udělali ještě spolu. Já jsem pak odcestoval na tříměsíční stáž do Edinburghu, kde jsem v práci na tomto modulu pokračoval. Ten jsme nakonec pojmenovali PermonSVM,“* vysvětluje Marek Pecha na začátek. Před dvěma lety se Davidovi Horákovi a Markovi Pechovi ozval Richard Mills z Argonne National Laboratory v USA, jeden hlavních vývojářů knihovny PETSc, na které je PERMON postavený. *„Napsal nám, že se Zachem Langfordem z Oak Ridge National Laboratory pracuje na lokalizaci lesních požárů ze satelitních snímků na Aljašce a že by rád využil PermonSVM pro natrénování modelů. Po několika měsících intenzivní spolupráce jsme měli první výsledky, které jsme prezentovali na konferenci AGU'21. Naše spolupráce se prohloubila natolik, že se minulý rok David s Richardem dohodli, že moji disertační práci povedou oba: David jako hlavní školitel, Richard jako školitel specialista. Byl jsem zkrátka ve správnou dobu na správném místě,“* usmívá se doktorand. Ten modely trénuje na velkém množství dat, díky čemuž dokáže predikovat výskyt požáru. Jak to tedy celé funguje? *„Nejprve musíme natrénovat model. V našem případě se jedná o sémantický segmentační model pro multispektrální obrázky (satelitní snímky). Jsou to typy klasifikačních modelů, které přiřadí každému pixelu v obrázku kategorii. V našem případě, zda došlo na části území reprezentovaném pixelem k požáru či nikoliv. Jenom pro představu, jeden pixel představuje území o rozloze 500x500m²,“* vysvětluje Marek. Kromě prostorové informace pracuje mezinárodní tým také s informací časovou, takže se na pixely dívá v podstatě jako na časové řady v rámci jednoho roku. *„Abychom takový model mohli natrénovat, potřebujeme si vytvořit tzv. trénovací sadu. Pro výběr a stažení satelitních snímků máme napsaný vlastní software využívající Earth Engine a Cloud Platform od Googlu.“*

Jakmile mají satelitní snímky stažené, provede se vyčištění dat. V praxi to znamená, že odstraní ty části obrazů, kde byly mraky, nebo ty části satelitních snímků, které byly poškozeny kvůli chybám na senzorech. Projekt používá standardní techniky strojového učení, tým tedy musí provést transformaci dat. *„Poté, co máme satelitní snímky transformované, si je rozdělíme do trénovacího a testovacího data setu. Vzhledem k objemu dat používáme k natrénování modelů superpočítač Summit, který je 5. nejvýkonnější na světě, a trénovací proces akcelerujeme pomocí grafických karet NVidia,“* říká Marek Pecha. Kvalitu modelu si s kolegy ověřují pomocí testovacího data setu. Provedou také vizualizaci výsledků, která slouží k ověření toho, zda predikce dává smysl, např. zda model nepredikoval požár uprostřed jezera. *„V tuto chvíli dosahujeme přesnosti kolem 80 % (F1 metrika). Což můžeme považovat za kvalitní model pro takovýto typ aplikace. Predikční vlastnosti našich*

modelů nakonec srovnáváme s jinými přístupy např. náhodnými lesy. Máme také rozpracovanou implementaci pomocí neuronových sítí,” přibližuje. Svůj přístup diskutují i s klimatology v návaznosti na monitoring klimatických změn na Aljašce.

S rozvojem sofistikovanějšího monitoringu a vzdáleného průzkumu Země, zlepšováním technologií pro archivaci velkých dat a rostoucí popularitou strojového učení se stále více využívají pokročilejší nástroje pro sledování přírodních rizik v reálném čase, systémy včasného varování nebo vývoj nástrojů pro odhalování rizik.

Marek Pecha se podílí i na projektu v rámci Strategie AV21 České republiky, kde má na starosti malý tým. *„Od začátku roku s mými dvěma kolegy Janou Rušajovou a Bohdanem Rieznikovem pracujeme na Ústavu geoniky AV ČR v Ostravě na projektu pro automatickou detekci a klasifikaci typu zemětřesení v reálném čase opět pomocí strojového učení,”* pokračuje v povídání Marek. Jinými slovy se snaží ze seismografického záznamu rozpoznat, k jakému typu zemětřesení došlo, jestli došlo k důlnímu otřesu, explozi nebo tektonickému jevu. *„Testujeme různé přístupy a naše nejlepší modely dosahují úspěšnosti kolem 98% (F1 metrika),“* dodává.

A to ho baví na doktorském studiu, že se podílí na zajímavých projektech a může po několikaleté odmlce (zapříčiněné covidem) cestovat na konference a stáže. A jaké jsou jeho plány do budoucna? *„Mně se líbí model propojení akademického prostředí a komerční sféry, tak jak jej mají v Americe či v západní Evropě. Do budoucna bych docela rád, kdyby se mi povedlo zůstat jednou nohou na akademické půdě a druhou nohou v komerční sféře. K tomu však musím nasbírat ještě docela dost zkušeností,”* uzavírá Marek Pecha.

<https://www.vsb.cz/magazin/cs/detail-novinky?reportId=45579>