

Proteinový vědec Jan Velecký:

Strojové učení může zazářit tam, kde už lidská představivost nestačí

Jan Velecký je výzkumník z centra RECETOX. Ve své disertační práci se zabývá metodami strojového učení pro proteinové inženýrství. V centru působí v rámci Loschmidtových laboratoří, konkrétně v týmu Stanislava Mazurenka. V rozhovoru vysvětluje, čím se proteinové inženýrství zabývá, proč je důležité věnovat při zkoumání proteinů pozornost právě jejich rozpustnosti, nebo v čem může být strojové učení pro lidské poznání klíčové.

11. 12. 2024 Sabina Vojtěchová Rozhovor Výzkum Osobnost



Jak jsi se dostal k proteinovému inženýrství?

Od mala jsem měl zvědavý vztah k počítačům, který mě přivedl na technické lyceum ve Zlíně, kde jsem studoval informační technologie. Poté jsem pokračoval na Fakultu informačních technologií Vysokého učení technického v Brně. Po bakaláři mi došlo, že by se mi líbilo informatiku spojit s něčím víc hmatatelným, co nežije jen v počítači. Jedna z možností byla bioinformatika, které jsem se začal věnovat v rámci magisterského studia. Už tehdy jsem se také začal zaměřovat na zkoumání proteinů, konkrétně jsem si pro diplomovou práci zvolil téma rozpustnosti proteinů. A tomu se věnuji doteď.

Mohl bys trochu přiblížit, na co se obor proteinového inženýrství vlastně zaměřuje?

Proteinové inženýrství je zjednodušeně návrh nebo úprava proteinů. Někdy vytváříme úplně nové proteiny, případně, a to je běžnější, ty stávající pozměňujeme. Velmi často postupujeme na základě mutací, tedy cílených změn, které vědec v proteinu provádí. Abychom mohli proteiny cíleně měnit a upravovat, potřebujeme je nejdříve zkoumat a zjišťovat, jak fungují. Ve svých počátcích obor cílil na vytváření takových mutací, které by "poškodily" jeho funkci. Když se nám totiž povede protein skrze mutaci vypnout, odhalíme současně, která jeho část je pro jeho funkci nezbytně důležitá. I proto se proteinovému inženýrství v jeho počátcích přezdívalo proteinový terorismus.

Jak si můžu takové záměrné mutování proteinu prakticky představit?

Protein je lineární molekula, řetězec aminokyselin zakódovaný genem. Podle genu jej organismus vyrobí a protein se následně sbalí (složí), a tvar pak určí jeho funkci. My tvar nemusíme znát, ale můžeme si hrát například se složením aminokyselin. Typicky tedy mutací

genu vyměníme jeden typ aminokyseliny za jiný. Máme 20 typů aminokyselin, můžu tedy třeba místo cysteinu vložit do řetězce například alanin, který je hodně malý a neutrální, takže je pro naše účely ideální. Když po jeho vložení protein přestane fungovat, tak víme, že tento cystein je důležitý pro jeho funkci nebo nějakou pro nás klíčovou vlastnost.

Čím to, že se ze všech vlastností zaměřuješ zrovna na rozpustnost?

Rozpustnost patří mezi základní vlastnosti a pro produkci nových proteinů je stěžejní. Právě ona rozhoduje o tom, zda bude možné protein vytvořit v dostatečném množství nebo kvalitě a za přijatelnou cenu. Má to přímé implikace v lékařství i v průmyslu, na bázi proteinů se hledají léky například na mozkovou mrtvici nebo jsou obsaženy v pracích přípravcích. Lepší rozpustnost zkrátka úzce souvisí s lepší výtěžností, a tak i efektivnější výrobou. Já se konkrétně zabývám metodami, na jejichž základě lze změnu rozpustnosti proteinu odhadnout. To znamená, že tam, kde bychom obvykle potřebovali pro změření člověka a materiál, použijeme počítač a rozpustnost změříme *in silico*, což ušetří finanční prostředky i čas. Například u zmíněných terapeutik na mozkovou příhodu bychom tak mohli zjistit rozpustnost zavčas a nemuseli se pouštět do experimentů, které by nedávaly smysl. Když to tedy shrnu, snažím se měřit a předvídat změny vlastností proteinů, především rozpustnosti, po jejich mutaci.

Máš k oboru i nějaký osobní vztah?

Bioinformatika mě spíš zaujala tím, že jsem v ní viděl potenciál do budoucna, přišla mi jednoduše zajímavá. Díky tomu, že jsem si už na VUT za vedoucího diplomky vyhlédl svého tehdejšího pedagoga Jiřího Hona, který na fakultě dělal doktorát, jsem se skrze něj dostal také do Loschmidtových laboratoří, kde jsem následně pod Dr. Stanislavem Mazurenkem začal pracovat na doktorátu.

Můžeš trochu vysvětlit, co to jsou Loschmidtovy laboratoře?

Loschmidtovy laboratoře se výzkumně věnují proteinovému inženýrství. Pojmenovány jsou po Johannovi Josefu Loschmidtovi, česko-rakouském vědci 19. století, zakladateli moderní chemie. Laboratoře jsou tvořeny čtyřmi týmy, dvěma teoretickými, zaměřenými na výpočetní metody, a dvěma experimentálními, které pracují v laboratoři. Na naší práci je specifické, že máme sice konkrétní proteiny, na které se zaměřujeme, ale ty nám slouží spíše jako prototypy pro metody, které vytváříme. Vyvíjíme několikero softwarových nástrojů, nejen pro proteinové inženýry, ale i pro lékaře či biochemiky, pomocí kterých můžou proteiny analyzovat a využívat je ve svých postupech. Naříklad velmi úspěšný a používaný nástroj je EnzymeMiner, který dokáže najít proteiny s hledanou funkcí na základě již prostudovaného proteinu. Díky němu jsme schopni zrecyklovat a využít něco, co vzniklo v přírodě v průběhu evoluce. Naše práce je tedy především o vývoji metod a nástrojů.

Kam tedy v rámci zaměření Loschmidtových laboratoří spadá tvoje práce?

Zkoumání rozpustnosti je součástí optimalizace proteinů. Když vytvoříme nový lék, ale výtěžnost je tak nízká, že to nestačí ani na experimenty v laboratoři, můžeme ji díky proteinovému inženýrství zvýšit. Zmiňoval jsem lék na cévní mozkovou příhodu, další dobrý příklad je inzulin. Ten měl zpočátku velký problém s rozpustností. Původně se získával z vepřového pankreatu, ale dnes jsme ho schopni vyrábět v buňkách uměle a pak extrahovat,

díky čemuž je finančně dostupný. Predikce rozpustnosti je ale důležitá i v diagnostice. Spousta nemocí je způsobena vznikem mutace v genetickém kódu konkrétního jedince, která zásadně sníží rozpustnost některého proteinu. Nerozpustné proteiny mohou být i toxické, nástroj má tedy potenciál i v personalizované medicíně.

Jestli to chápu dobře, obecně tedy zapojení IT do vědy šetří čas i finance. Co ale tvému oboru přináší konkrétně využívání strojového učení?

Řekněme, že se všemi problémy, které člověk řeší, si lze dobře poradit, pokud jsou představitelné ve dvou až čtyřech dimenzích. Jakmile má však dimenzí třeba sto, najít vzájemné závislosti a vysvětlit jevy a chování, které k nim vedou, rozhodně není triviální. Zde pak může zazářit strojové učení, které nemá omezení v představitivosti a dokáže najít i to, co je nám skryté.

Jak si můžete být jistí, že jsou informace, které takto získáte, spolehlivé?

Máme samozřejmě nástroje na ověřování spolehlivosti. Základem je, že učíme model strojového učení z dat, u kterých víme, k jakému výsledku vedou. Modelu však neukazujeme všechna data, která máme. Ta, která si ponecháme stranou, nám pak slouží k otestování modelu. Strojové učení je i o experimentování. Není to tak, že se navrhne model a ten ihned funguje.

A díky dnešní dostupnosti výpočetního hardwaru uvádíme do praxe, co bylo dříve nepředstavitelné. Vždyť cena na jednotku výpočtu poklesla biliarkrát od 50. let, kdy se pokládaly základy oboru strojového učení.

Kam obor dál směřuje a jaké má ambice?

Je zde celá plejáda nevyřešených problémů. Poslední velká vyřešená výzva byla predikce struktury proteinů. Teď už víme, jaký bude tvar proteinové molekuly na základě jejího genetického kódu. To je důležitý pokrok, protože jak jsem řekl, funkce proteinu závisí na tvaru. A tak si myslím, že teď uvidíme vyřešených v rychlém sledu mnoho problémů molekulární biologie. Kdybych to měl dát do časové osy, tak dřív jsme proteiny mutovali, abychom zjistili, jak fungují. Dnes je přetváříme. V budoucnu pak budeme vytvářet běžně zcela nové proteiny, což bude podpořeno nejen rozvojem strojového učení, ale i robotiky a jiných vysokokapacitních metod, které nám umožní sběr dat pro strojové učení.

Myslíš si, že může rozvoj AI ohrozit některá pracovní místa ve vědecké sféře?

Podle mě vědcům zlepší rozvoj AI život, protože jim usnadní práci, která je repetitivní a nudná. Nemyslím si, že AI má spadeno na nějaké konkrétní pozice. Korektoři a překladatelé asi budou potřeba méně. Na jiných pozicích to naopak přidá na efektivitě, a tak i vlastně na práci. Kdo v laboratoři nyní charakterizuje pět variant proteinu týdně, bude jednou možná dělat pět tisíc.

Jaké je tvá osobní vize a vědecké ambice?

Byl bych rád, kdyby moje práce obor proteinového inženýrství posunula někam dál, tak aby na tom obor mohl dále stavět. Nyní se chci soustředit hlavně na dokončení doktorátu. Do budoucna by se mi líbilo ve výzkumu zůstat, zároveň však chci trochu experimentovat. Mohla

by to být zkušenost z výzkumu v komerční sféře, či aplikace nabytých znalostí ve strojovém učení do jiné disciplíny. Hlavně nepřestávat rozšiřovat si obzory.