
Zlatý grál, proteiny a softwarové inženýrství. David Hoksza převzal Cenu Bernarda Bolzana v oboru informatika

Zlatý grál, proteiny a softwarové inženýrství. David Hoksza převzal Cenu Bernarda Bolzana v oboru informatika

Zatímco starověcí učenci filozfovali při chůzi olivovým hájem, jak praví staré knihy, Davidu Hoksza chůze nestačí, nejradyji přemýšlí ve zrychleném režimu organismu, tedy při běhu. A přináší mu to výsledky, nejprve v podobě magisterského diplomu z MFF UK v době studií, a naposledy před pár týdny, kdy mu Nadační fond Bernarda Bolzana udělil výroční cenu za rok 2010 v oboru informatika. RNDr. David Hoksza, Ph.D., z Katedry softwarového inženýrství MFF UK ji převzal na zasedání vědecké rady MFF UK 9. února 2011.



**RNDr. David Hoksza, Ph.D.,
Katedra softwarového
inženýrství, MFF UK**

Co je konkrétně předmětem práce „Density-based classification of protein structures using iterative TM-score,“ za kterou jste právě obdržel výroční cenu Bernarda Bolzana?

S rozvojem bioinformatiky a vědy o proteinech se zvyšuje i tlak na efektivní podobnostní vyhledávání v databázích proteinových struktur a sekvencí nebo obecně proteinů. Uvedu modelový případ – jsem chemik nebo biolog a zjistím určitou strukturu proteinu nějaké bílkoviny a chci vědět, k čemu ta bílkovina slouží. Protože vím, že funkčně podobné proteiny mají i podobnou strukturu, vezmu tuto proteinovou strukturu s neznámou funkcí a prohledám databázi proteinových struktur se známou funkcí. A když najdu strukturně co nejpodobnější proteiny, dostanu nápovědu, jaká by mohla být funkce mého nově nalezeného proteinu.

A tím se zabývala moje disertace – vyhledáváním v takových databázích. Navrhl jsem určitý model, jak reprezentovat proteinové struktury, a model vhodný k tomu, aby se na něm dobře prováděly podobnostní operace. Prakticky to vypadá tak, že si navrhnu nějaký model a na něm nadefinuji operaci podobnosti, abych dokázal říct o dvou proteinových strukturách, potažmo modelech, které je reprezentují, jak jsou si podobné. To pak můžu využít při efektivním prohledávání databází proteinových struktur.

K čemu, kromě usnadnění vyhledávání a zefektivnění operací v databázích, bude výsledek vaší práce sloužit v širší praxi?

Obecně se podobnost proteinových struktur může použít jak pro zjištění funkce nově nalezeného proteinu, tak i v širším kontextu, jako je například „zlatý grál“ proteomiky, což je predikce proteinové struktury ze sekvence. Řekněme, že máme nějakou bílkovinu, známe její sekvenci a chceme vědět, co by se stalo, kdybychom nějakou aminokyselinu zaměnili za jinou, jak by se změnila její proteinová struktura. A tady se velmi výrazně používá modelování podobnosti proteinových struktur. Takže podobnost proteinových struktur se dá použít jako částečný modul v algoritmech pro predikci proteinové struktury.

Proč je predikce proteinové struktury „zlatým grálem“ proteomiky?

Zlatým grálem je proto, protože má neobyčejně významné uplatnění v biologii, lékařství a především ve farmacii ve vývoji léčiv. Modelový příklad, který jsem před chvílí zmínil, se uvádí nejčastěji, protože určení funkce nové struktury se rovná faktickému použití metody 1 :1, zatímco když budu dělat predikci proteinové struktury, bude to jen modul v celém aparátu predikce proteinových struktur.

Má vaše práce, tak jak vy sám ji vidíte, nějaké slabiny či nedořešená místa nebo je to naopak vyřešená a dokončená metoda?

Slabiny má v....

A můžeme je tu říkat?

Můžeme, já jsem o tom psal i ve své disertaci. Chtěl bych, aby byla efektivnější, co se týče rychlosti. Není pomalá, ale její slabina je v tom, že s růstem databází proteinových struktur se tato metoda bude zpomalovat, protože je

sekvenční, rychlost není logaritmická vzhledem k velikosti databáze. Ale pracujeme na tom s mými kolegy ve výzkumné skupině SIRET, (SImlarity RETrieval) research group (SRG), a zabýváme se právě efektivními metodami indexování/ vyhledávání ať už indexování ve smyslu rychlého vyhledávání v databázích buď na obecné úrovni, kdy vyvíjíme struktury pro rychlé vyhledávání, nebo tím, že děláme částečně aplikovaný výzkum, kde se snažíme tyto metody uplatnit v určité specifické doméně, třeba v biologii, což dělám já.

Až nějaký lékařský tým dostane Nobelovu cenu za výzkum proteinů, vaše jméno tam zřejmě nebude uvedeno a nikdo z nich si ani neuvedomí, jak měli díky vaší metodě usnadněnou práci a zkrácenou cestu omylů a hledání, o čase nemluvě.

Ano, to je pravda, ale stejně jako když někdo bude ke své práci používat excel a získá světovou slávu, tak vývojářům excelu se uznání také nedostane.

Vás právě ocenila Nadace Bernarda Bolzana, což je skvělé zahájení vědecké kariéry. Jak došlo k tomu, že jste zvolil právě tohle téma? Byla to objednávka z praxe, třeba zadání od biologů, nebo máte k biologii osobní vztah?

K biologii nemám vztah ani záporný, ale neměl jsem ani výrazně kladný - že bych si jako malý hrál s pipetou, to zase ne, ale samozřejmě mě zajímá, jak ty věci fungují. Jednoduše, vybral jsem si ho jako téma své disertační práce. Vystudoval jsem na Katedře softwarového inženýrství MFF UK datové inženýrství, což jsou zjednodušeně řečeno databáze, a chtěl jsem dál pokračovat v doktorském studiu. Snažil jsem se ale najít spíš oblast aplikovaného výzkumu, abych se nezabýval datovými strukturami jen ve vzduchoprázdnu. A doc. RNDr. Tomáš Skopal, Ph.D., který vedl mou disertační práci, zrovna navrhl téma Podobnostní vyhledávání v biologických databázích. A protože on sám právě pracuje na indexovacích strukturách obecně, moje práce byla to o tom zkusit je aplikovat, aby byly někde využité, protože to jsou dobré, efektivní struktury. Takže k tématu jsem se dostal spojením těchto dvou domén a stálo mě to hodně úsilí a hodně studia se do problematiky proteinů dostat.

Tak že teď máte jako vedlejší specializaci proteinové struktury?

... (smích)... Skoro, i když biologové nebo chemici by s tím určitě nesouhlasili, protože já nikdy nemůžu dosáhnout takové hloubky znalostí, jaké mají oni a ani to nepotřebuji. Snažím se abstrahovat od specifičnosti domény a pak pracovat s těmi modely, které nad tím vytvořím, takže jdu jen tak hluboko abych dokázal efektivní modely vytvořit a nad nimi pracovat.

Mluvil jste o abstrakci, hledání vyšších obecných struktur a podstaty funkcí mnohého. Kde je v takové badatelské práci čistých idejí prostor pro vášeň? - vždyť ani matematik jí nezůstane ušetřen.

Myslím, že bez vášně se vědecká práce dělat nedá. Nevím, jestli se tomu tak dá říkat, ale podle mě je to o tom, že člověk rád rozebírá věci a dokáže se do problému intenzivně ponořit, nemyslí na ostatní svět a v jistém smyslu je od světa odtržen. Já například rád přemýšlím, když si jdu zaběhat. V tu chvíli nic nevnímám a často si přemýšlím nad matematickými problémy. Začalo to v době magisterského studia, když jsem před zkouškami chodil běhat a učil jsem se, ne ve smyslu biflování, ale pochopení věcí. Člověk se úplně odstříhne od reality a promílá si to v hlavě - a to mi pomáhá přímo úžasně.

Děkuji za rozhovor.

(Marie Kohoutová)