

Velká finanční data na Fakultě aplikovaných věd

5.12.2022 - Jan Pospíšil, Martina Batková | Západočeská univerzita v Plzni

O jaká data se jedná a jak je plánujete využít ve svém výzkumu?

Eurex je mezinárodní elektronická burza skupiny Deutsche Börse AG, která nabízí především obchodování s evropskými deriváty. Jedná se o největší evropský trh s opcemi a futures kontrakty. Rozhraní Eurex Enhanced Order Book Interface (EOBI) poskytuje kompletní údaje o obchodech a informace o knize objednávek, které zobrazují každou viditelnou objednávku a kotaci pro nejlíkvidnější futures kontrakty na Eurexu. EOBI data budou primárně využita ve výzkumné činnosti zaměřené na studium tzv. modelů rough frakcionální stochastické volatility (RFSV).

Mohl byste široké veřejnosti vysvětlit, o jakých modelech mluvíte?

Ve financích je volatilita míra kolísání řady obchodních cen v čase. Stochastické modely volatility jsou takové, ve kterých má rozptyl ceny náhodný charakter, říkáme, že volatilita je sama o sobě stochastický proces. Uvažujeme novou generaci modelů stochastické volatility, tzv. modely rough volatility (dalo by se asi přeložit jako hrubé volatility, ale nepoužívá se to), kde je okamžitá volatilita řízena tzv. rough frakcionálním Brownovým pohybem. Tento rough stochastický proces s krátkou pamětí je zobecněním standardního Brownova pohybu. Ačkoli se RFSV modely ukázaly jako vynikající při popisu nízkofrekvenčních tržních dat, o rough volatilitě ve vysokofrekvenčních intradenních datech a o jejím vývoji ve velkých datech z objednávkových knih v delších časových horizontech se toho ví velmi málo. Techniky používané při kalibraci modelů rough volatility jsou navíc stále neefektivní z hlediska rychlosti, přesnosti, robustnosti a citlivosti, aby mohly být široce využívány v praxi.

Co ve vašem případě znamenají "velká data"? Jak velký je soubor dat?

Mluvíme o miliardách zpráv z knihy objednávek denně. Hrubá data mají několik stovek gigabajtů denně a roční soubor dat je v řádu několika desítek až stovek terabajtů. Zhruba stejná kapacita bude zapotřebí také k dekodování a zpracování dat.

To je opravdu velké množství dat. Kde a jak plánujete data zpracovávat a co z nich chcete získat? Plánujeme využít výpočetní a úložné kapacity české národní e-infrastruktury pro výzkum a vývoj e-INFRA CZ, zejména vysoce náročné výpočty budou probíhat v gridové infrastruktuře MetaCentrum, které jsme součástí. Netriviální množství času bude samozřejmě věnováno porozumění datům a rekonstrukci zkoumaných objednávkových knih, tj. těžko nyní předvídat výsledky, máme určitou představu a mohu asi říci, že naším cílem je dodat akademickému a finančnímu sektoru správné přesné nástroje a techniky spolu se snadno uchopitelnými algoritmy. Mimo jiné se budeme snažit aplikovat např. moderní attention mechanismy ze strojového učení a vyvinout inovativní, vysoce přesné modely založené na hlubokých neuronových sítích. Zde bych rád ocenil pomoc svého spoluautora Kamila Ekšteina z katedry informatiky a výpočetní techniky.

Plánujete do svého výzkumu zapojit i studenty?

Ano, samozřejmě. Vhodné podmnožiny rozsáhlého souboru EOBI dat budou použity pro účely výuky mého předmětu Stochastické finance a také pro jednotlivé studentské projekty včetně bakalářských, magisterských a doktorských prací.

V neposlední řadě bych rád využil této příležitosti a poděkoval společnosti Deutsche Börse za poskytnutí přístupu k jejím datům. Pevně věřím, že reálná tržní data výrazně zvýší kvalitu naší akademické a výzkumné činnosti.

<https://info.zcu.cz/clanek.jsp?id=4967>