

## **Kladivo na investovanie.**

Na spracovanie dát pre vlastný tradingový model môžu inštitúcie využívať existujúcu sieť bežných počítačov

1.10.2003, TREND príloha FINANCIÉ 2003 / Martin Babiak

Zarobiť viac peňazí vyžaduje mať prevahu, rýchlejšie a kvalitnejšie spracovať čo najväčšie množstvo informácií. Rastúci výkon počítačov spojený s pokrokom vo vývoji analýz bude znamenať čoraz významnejšiu konkurenčnú výhodu. Informačné technológie si podmanili finančné trhy.

Zatiaľ však prenikajú skôr do organizácie obchodovania a komunikácie. Samo investovanie si ľudia nechávajú pre seba.

Investičné rozhodnutia sú len výnimočne podchytené rozsiahlejším výskumom s využitím pokročilých metód. Ak sa aj v praxi opierajú o nejaké analýzy, tak väčšinou len o tie najjednoduchšie, z ktorých väčšina bola známa už v časoch Čiernych piatkov.

### **Štandardné služby a špecializované programy**

Počítač sa môže premeniť na agresívny prostriedok investovania. Pomerne pestré analýzy (grafy a základné ukazovatele) pre potreby investovania ponúkajú informačné služby od renomovaných agentúr, ako sú Reuters či Bloomberg, alebo od početných internetových investičných portálov. Dávajú ich k dispozícii ako štandardnú službu, pričom tie kvalitnejšie ponúkajú možnosť automatického vykonania príkazu pri výskyte určitého signálu.

Na trhu existuje aj množstvo špeciálneho investično-analytického softvéru. Za všetky možno spomenúť napríklad Metastock, ktorý je v tejto brandži nepísaným štandardom. Jeho ponuka analýz je nepomerne rozsiahlejšia, pozostáva z desiatok ukazovateľov technickej analýzy, každý s množstvom nastavitelných parametrov a podmienok generovania signálov.

Takýto programový balík býva vybavený veľkou databázou historických dát s dlhými časovými radmi. Metastock obsahuje aj investičný systém, ktorý je schopný bez vonkajšieho zásahu celkom slušne zarábať. Je vybavený zároveň nástrojmi na overenie investovania simulovaním na historických dátach.

### **Predbehnúť Metastock**

Napriek týmto výhodám už väčšina čitateľov zrejme krúti hlavou. V simulácii to síce fungovať môže, ale ak sa v praxi vrhnú tisíce Metastockov navedených tými istými dátami a analýzami na tú istú akciu v tom istom čase, ceny sa celkom určite nebudú vyvíjať v súlade s pôvodným očakávaním. Zviest' sa na takejto vlne synchronizácie nálad väčšinou znamená pre investora likvidáciu vyhládok na zisk.

Špekulant v snahe predísť davovému správaniu, zvýšiť zisk a predbehnúť jednocelové „metastockové“ aplikácie zvyčajne dospeje k vlastnej sústave investičných pravidiel (tzv. tradingový model). Tie sa zvyknú opierať o väčšie množstvo ukazovateľov, všímať si zvláštne situácie, dať priestor aj menej používaným a vlastným indikátorom.

Investor pri vývoji vlastného tradingového modelu môže okrem svojich skúseností využiť širokú paletu štatisticko-matematických metód, modelov a ukazovateľov, ktorými výskum

podoprie. V zásade platí, že s rastom množiny zvládnutých metód a rozsahu výskumu stúpa aj šanca na vyvinutie kvalitného modelu.

### **Vývoj tradingových modelov**

Za naformulovanou, samostatne zarábajúcou sústavou pravidiel vyvinutých s využitím pokročilých analýz a technológií bude stáť rozsiahly výskum, ktorý sa dá rozdeliť do niekoľkých krokov. Prvý – získanie dostatočného množstva historických trhových dát netreba komentovať.

Dôležitosť tejto fázy chápe každý, podobne ako aj to, že dáta musia vyhovovať charakteru investovania. Iné údaje budú potrebné na denné investovanie do akcií, iné pre deriváty a iné na „minútové“ špekulovanie na pohyb devízových kurzov.

Potom prichádza na rad množstvo analytických metód, ktoré rátajú čiastkové ukazovatele. Aby sa analýzy dali pri programovaní výskumu zavolať s rôznymi parametrami, musí byť ich zvládnutie aspoň na používateľskej úrovni. Investor si pritom môže vyberať spomedzi bežne používaných metód v brandži, ale nič nebráni vyskúšaniam aj tých menej používaných.

Analýzy si možno požičať aj z ostatných vedných odborov: umelá inteligencia, automatizované tvorby matematických modelov alebo rozličné optimalizácie. Sem patrí aj automatizované spätné vyhodnocovanie signálov, ktoré produkujú primárne analýzy. Táto schopnosť je pre hromadné spracovanie údajov rovnako dôležitá ako schopnosť analyzovania.

### **Vlastný výskum**

Ak má investor jasno v druhoch a rozsahu analýz, je možné začať programovať sám výskum. Ten má najmenej dve algoritmicke oddelené fázy: výpočty ukazovateľov s automatickým vyhodnotením signálov a spätné vyhľadávanie nastavení analýz produkujúcich maximum signálov s minimom falošných poplachov.

Prvú fázu opisuje príklad: investora zaujíma 30 technických ukazovateľov. Dôveryhodnosť signálov každého z nich otestuje na tisíckach rôznych nastavení. Okrem toho si vybral niekoľko ďalších, pre finančnú brandžu „avantgardných“ metód, pričom výsledky bude zaznamenávať do databázy (tzv. dátového skladu).

Nájdenie kombinácie ukazovateľov generujúcich najlepšie predpovede alebo najvyšší zisk je úloha druhej fázy výskumu. Problém spočíva v množstve údajov, ktoré treba spracovať: ak investor vzorovým balíkom analýz preženie napríklad tisíc akcií, dostane obrovské množstvo surových výsledkov. Z nich bude drvivá väčšina nepoužiteľná vata, v ktorej bude musieť hľadať dôležité vedomosti ako ihlu v kope sena.

Využije pritom kombináciu prístupov – overovanie pravidiel, ktoré má zo svojich skúseností, výpočty pravdepodobností, spustenie aplikácie hĺbkovej analýzy dát -- DataMiningu alebo rôzne optimalizácie. Získané informácie sa priamo odrazia v tradingovom modeli.

Prvá vlna overovania modelu zvyčajne prebieha na historických časových radoch. Tu si investor urobí predstavu o ziskovosti a spoľahlivosti daného modelu, prípadne zistí, či model nevyhľadáva príliš rizikové tituly a podobne. Ak prejde, prichádza na rad testovanie v reálnom prostredí pri evidencii ziskov. V podstate treba model napojiť priamo na dáta z informačných zdrojov.

## **Zjednodušenie programovania**

Vývoj tradingového modelu stojí a padá na schopnosti naprogramovať výskum. Ak by mal používateľ programovať samotné analýzy, vyžadovalo by to veľa času a množstvo štúdia. Takýto projekt by dlho hltal peniaze bez jedinej vykonanej analýzy.

Urýchlenie vývoja a presun koncentrácie od detailného programovania smerom k výskumu znamená použitie niektorého z univerzálnych analytických nástrojov. Na trhu ich je viac a nami vybrané (STATISTICA, Mathematica, MatLab a SAS) predstavujú len časť ponuky. Prakticky všetky takéto systémy sú skladačky: obsahujú jadro (základnú aplikáciu) a množstvo rozširujúcich modulov, knižníc a toolboxov s funkciami zo špecializovaných oblastí.

Pridanie toolboxu k jadru znamená, že k programovacím nástrojom (príkazom a analýzám) jadra pribudnú nové príkazy, funkcie a analýzy obsiahnuté v toolboxe. Potom stačí už len pridať knižnice s finančnými analýzami a sú k dispozícii bez zdĺhavého programovania. Aj keď väčšina analytických systémov začínala na inom ihrisku ako financie, ku každému existuje množstvo finančných a investičných knižníc vrátane rozširujúcich aplikácií, napríklad pre risk manažment.

## **Obrovský výpočtový výkon**

Pridanie finančných knižníc k systému zrýchli len formuláciu výskumného programu, nie však sám výpočet. Príklad v odseku Vlastný výskum naznačuje, že investor narazí na potrebu vykonať veľké množstvo analýz, a teda na potrebu mimoriadneho výpočtového výkonu. Ak výskum nemá trvať roky, ale dni, v zásade sa núkajú dve možnosti. Buď si zaobstarať veľmi výkonný počítač (superpočítač), alebo výpočty zveriť množstvu bežných počítačov (grid computing).

Výhodu superpočítačov – vysoký výpočtový výkon – treba zaplatiť, pričom prístup k jeho výkonu je iný ako v bežnom počítači. Laikovi stačí vedieť, že analytický systém musí byť schopný „rozhadzovať“ výpočty medzi viacerými procesormi (tzv. parallel computing). Inak by všetko bežalo na jednom a ostatné „kremíkové doštičky“ za pár miliónov korún by zaháľali.

Grid computing je hitom ostatných rokov. Ak má napríklad banka desiatky počítačov, pričom miera využitia výkonu procesora je pri bežnej práci len pár percent, zvyšok sa dá zamestnať na takéto rozsiahle analýzy. Okrem toho, cez noc a víkendy sa nepracuje, čo znamená len odpisy za vypnutú techniku. Nevýhodou gridu je potreba licenčného zabezpečenia.

Ak má analýza bežať na stovke počítačov, jednopočítačová licencia analytického systému nestačí. Tento problém, prípadne s problémom, že analytický systém nepodporuje grid, sa dá s trochou práce vyriešiť. Analýzy sa môžu dokonca zrýchliť.

Každý analytický systém totiž obsahuje sieťové funkcie a toolbox na prepis výskumnej aplikácie do jazyka C++. C je veľmi rýchly jazyk (rýchlosť analytických systémov sa mu v lepšom prípade len približuje), ale hlavne sa naň už v drvivej väčšine licenčné obmedzenia nevzťahujú.

V zásade teda stačí celý výskum preložiť do „céčka“ a z neho do priamo spustiteľnej formy a sieťovými funkciami (nesmú sa používať gridové, tam problém s licenciami znovu vyskočí) zabezpečiť posielanie dát, zadaní a výsledkov medzi počítačmi. Do výskumu sa v takomto prípade môžu bezpečne zapojiť tisíce počítačov.

## Pokročilé analýzy

Ekonomom, ale aj iným, ktorí počas štúdia štatistiky, matematiky a programovania veľmi neublížili, prídu vhod novšie analytické technológie a technológie umelej inteligencie. Ide o rôzne samovytváracie matematické modely, napríklad neurónové siete alebo aplikácie na identifikáciu systému. Za zmienku stojí aj DataMining, ktorý sa uplatní hlavne pri hľadaní závislostí v surových výsledkoch prvej fázy výskumu. Ak ich používateľ hľadá „vlastnoručne“, musí všetko naprogramovať a dotaz do systému sa dá prirovnať k sústavnému sledu otázok typu: „Čo sa stane, keď...?“

Dataminingové prostredia už majú kladené týchto otázok prefabrikované a ich otázky sú vďaka rôznym heuristikám lepšie ciele. Otázka používateľa do systému sa dá prirovnať k formulácii: „Nájdí niečo a povedz to!“, čo je ich nepochybná pridaná hodnota.

Na dvere klopú aj analýzy textov, ktoré dolujú vedomosti z neštruktúrovaných textových informácií (napríklad burzových správ). TextMining sa zatiaľ berie ako nadstavba DataMiningu.

Ak si uvedomíme, že tok trhových dát možno prirovnať k ľubovoľnému inému signálu, nemusí byť na škodu skúsiť aplikáciu analýzy iných vedných odborov. Signal Processing, Wavelet, Pattern Recognition alebo Fuzzy regulácia sú len zlomkom „avantgardných“ analýz, ktoré sú jednak dostupné bez nutnosti špeciálneho rozsiahleho štúdia (teda by mal postačiť lepší kurz) a navyše sú lepšie prepracované ako väčšina ekonomických ukazovateľov.

## Prehľad vybraných analytických systémov pre tvorbu vlastných tradingových modelov

Analytický systém	STATISTICA	Mathematica	MatLab	SAS
Stručná charakteristika	Výrobca zdôrazňuje kvalitu neuronových sietí zo svojej produkcie a dokladá ju víťazstvami v početných nezávislých testoch. Kúpa STATISTICA DataMineru znamená dostať k dispozícii všetky analytické nástroje systému. Podľa testu na stránke StatSoftu je ich datamining dvakrát rýchlejší ako riešenie SASu. Základné finančné analýzy sú priamo v STATISTICA Base module, ostatok sa dá buď naprogramovať, alebo, podobne ako všetky systémy zoznamu, umožňuje volanie funkcií z toolboxov iných výrobcov	Výrobca zdôrazňuje robustnosť analýz, teda nenáchylnosť na ich metodologické pasce, ktorá výrazne spresňuje výsledky. Od ostatnej verzie sa výpočtové jadro výrazne zrýchliilo, takže je podľa výrobcu porovnateľné s C++. Výhodou pre neprogramátorov môže byť schopnosť vkladania vzorcov v symbolickom zápise a možnosti programovania podobajúce sa na inteligentné jazyky ako napr. Prolog alebo LISP	MatLab sa mnohých segmentoch berie ako priemyselný štandard s obrovskou penetráciou na trhu. Od verzie 6 sa jeho výpočtové jadro podstatne zrýchliilo hlavne v klasických programátorských technikách a výrobca ho prirovnáva k jazyku C++. Produktové portfólio je marketingovo výborne riešené, takže odborník, ktorý vie, čo chce robiť, ale nevie, čo na to použiť, to zistí už pri letmom preletení cenníkov. Zlé jazyky tvrdia, že to v MathWorks kvôli maximalizácii príjmov rozmenili na drobné. V kontexte s cenou má MatLab veľmi vysoké zľavy pre školstvo	SAS – štatistický analytický systém. Rozsiahla aplikácia, ktorej kvalita spôsobila, že sa ocitol na zozname výboru COCOM – výrobkov so zakázaným vývozom za železnú oponu. Väčšinu pre financie zaujímavých analýz robí pár modulov, ktorých označenia naznačujú len veľmi málo. Napriek tomu, že je ako jediný zastúpený priamo na Slovensku, väčšina informácií sa nachádza mimo SR a domovskom portále. SAS Institute a ako jediný výrobca z tabuľky koncentruje takmer výhradne na ponuku ucelených riešení, pričom predaj jednotlivých modulov realizuje len sporadicky
Vybrané referencie vo finančníctve		NBS a približne 15 licencií v českých bankách, súhrnne zhruba 150 inštitucionálnych partnerov, z toho približne 30 na Slovensku	ČNB, NBS, Komerční banka, Česká pojišťovna, spolu zhruba 400 inštitucionálnych partnerov, z toho približne 100 na Slovensku	viacero slovenských bánk, ďalšie v jednaní
Základný modul	STATISTICA Base	Mathematica Professional	MatLab	Base SAS

Knižnice finančných a iných potrebných analýz		Finance Essentials, Time Series, Derivates Expert, UnRisk, Operatonal Research, Industrial Optimization	Statistics Toolbox (tbx), Optimization tbx, Financial tbx 1 a 2, GARCH Financial Time Series tbx, Financial Derivates tbx, Fixed Income tbx, Curve Fitting tbx	SAS/Stat, SAS/ETS, SAS/OR, SAS/Risk Management
DataMining	STATISTICA DataMiner		Witness Miner	Enterprise Miner, Text Miner
Neurónové siete a samoidentifikácia správanía sa sytémov	STATISTICA Neural Networks	Neural Network	Neural Network tbx, System Identification tbx	áno
Avandgardné analýzy		Wavelet Explorer, Fuzzy Logic	Wavelet tbx, Fuzzy Logic tbx	
Vstup a výstup dát do databáz	áno	áno	Database tbx	SAS/Access
Špeciálna komunikácia napríklad s konzolami Reuters, Bloomberg	prostredníctvom produktu STATISTICA Text Miner	áno	Datafeed tbx	SAS/Connect
Vlastný programovací jazyk	STATISTICA Visual Basic	áno	áno	áno
Sieťové funkcie	áno	áno	áno	áno
Podpora Grid Computing	áno (DataMiner)	áno (gridMathematica)	nie	áno
Podpora Parallel Computing	áno (DataMiner)	áno (Parallel Computing Toolkit)	nie	áno
Prekadač do C++	áno	áno	áno	áno
Výrobca	StatSoft	Wolfram Research	Math Works	SAS Institute
Distribúcia	StatSoft CR s.r.o., Praha	Elkan, s.r.o., Praha	Humusoft, s.r.o., Praha	SAS Institute GmbH, ozzs, Bratislava
URL	www.statsoft.cz	www.elkan.cz	www.humusoft.cz	www.sas.com