

Predikcia úpadku spoločností s ručením obmedzeným využitím metód pre rozpoznanie odľahlých bodov

Peter Gnip¹, Martin Zoričák² a Peter Drotár¹

¹Katedra počítačov a informatiky, Technická univerzita v Košiciach

²Katedra financií, Technická univerzita v Košiciach

Abstrakt. Spoločnosti pôsobiace v rámci obchodného a priemyselného odvetvia sa môžu vplyvom nepriaznivej finančnej situácie, alebo nevhodného obchodovania, dostať do finančných ťažkostí, ktoré neskôr vyústia do celkového úpadku spoločnosti. Analyzovali sme dáta obsahujúce tisíce záznamov spoločností s ručením obmedzeným (s.r.o) pôsobiacich na Slovensku v rôznych odvetviach hospodárstva v období rokov 2013-2016. K nastolenému problému sme pristupovali ako k problému rozpoznania odľahlých hodnôt (outliers), pričom bola použitá metóda podporných vektorov pre detekciu odľahlých bodov (OneClassSVM). Dáta pozostávali z 20 štandardných ekonomických ukazovateľov. V prvotnej analýze sme sa zamerali na predikciu úpadku s.r.o. na základe účtovných údajov z jedného roku a kombináciou dvoch po sebe idúcich rokov. Dosiagnutá presnosť predikcie bola od 60,56% do 77,91 % v závislosti od roku v ktorom sme uvažovali výsledný stav spoločnosti a roku z ktorého boli čerpané ekonomické ukazovatele.

1 Úvod

Predikovanie úpadku firiem je téma, ktorou sa autori zaoberajú posledných vyše 80 rokov. Schopnosť čo najpresnejšie predikovať úpadok firmy na základe údajov z predchádzajúcich období je cenná pre manažment spoločnosti, investorov, veriteľov ako aj banky. Opodstatnenosť témy zdôrazňuje počet publikovaných vedeckých článkov v danej oblasti s využitím veľkého počtu metód. Približne od začiatku 30. rokov do 60. rokov autori uvádzali metódy na základe pomerových ukazovateľov, kde aplikovali len jeden takýto ukazovateľ [1]. Postupne začali vznikať modely s viacerými ukazovateľmi, a postupne bol pridávaný väčší počet pomerových ukazovateľov do modelov, z ktorých sú najznámejšie diela Altmana 1968 s použitím piatich faktorov [2], resp. Boritza a Kennedyho 1995 so štrnástimi faktormi [3]. Frekventovaná metóda bola diskriminačná analýza, ktorú následne začali nahrádzať iné štatistické metódy ako napríklad regresná analýza (logit, probit). V rámci štatistických metód boli aplikované metódy na redukciiu dimenzionality, konkrétne faktorová analýza a analýza hlavných komponentov. S rozvojom techniky a prístupom k stále rozsiahlejším databázam, autori aplikovali tzv. inteligentné metódy - metódy strojového učenia a dolovania dát (data mining) v oblasti predikovania úpadkov. Medzi tieto metódy patrí využitie neurónových sietí, rozhodovacích stromov, metódy podporných vektov.

*J. Steinberger, M. Zima, D. Fiala, M. Dostal, M. Nykl (eds.)
Data a znalosti 2017, Plzeň, 5. - 6. října 2017, pp. 187-191.*

Predikcia úpadku spoločností s ručením obmedzeným využitím metód pre rozpoznanie odľahlých bodov

rov (SVM), genetických algoritmov, lineárne programovanie, kvadratické programovanie či data envelopment analysis[4][5]. Problém identifikácie spoločnosti pred bankrotom z pohľadu inteligentných metód spočíva v správnej binárnej klasifikácii firmy do rizikovej (bankrotujúcej) skupiny alebo nerizikovej (nebankrotujúcej) skupiny. Na základe takto postaveného problému je potom možné testovať modely z viacerých hľadísk, ako presne vedia zatriediť firmy do správnej kategórie.

Vzhľadom na charakter problematiky, často dochádza k problému nevyvázenej dátovej sady, pričom počet zdravých firiem mnohonásobne presahuje počet bankrotujúcich. Niektoré z predchádzajúcich štúdií [4],[6] boli založené na dátach obsahujúcich niekoľko stoviek firiem, pričom tieto boli rovnomerne rozdelené medzi spoločnosťami v úpadku a zdravé spoločnosti. Ako naznačujú dáta, získané zo štandardných výkazov finančnej správy, takéto rozdelenie dát je v praxi nereálne, nakoľko počet spoločností v úpadku predstavuje len zlomok z celkového počtu spoločností. Preto je potrebné k predikcii úpadku pristupovať ako k problému predikcie nevyvážených dát (*imbalanced data*) [7].

V ďalšom bližšie opíšeme použité dáta, metodológiu, ako aj prvotné výsledky, t.j. presnosť s akou dokážeme identifikovať upadajúce spoločnosti.

2 Dáta

Analyzovali sme dáta obsahujúce tisíce záznamov spoločností s ručením obmedzeným (s.r.o) pôsobiacich na Slovensku v rôznych odvetviach hospodárstva v období rokov 2010-2016. Odvetvia, zahrnuté v dátovej sade sú: maloobchod, poľnohospodárstvo, priemysel, stavebníctvo a hotely. V tejto štúdii sa zameriame výhradne na dáta z oblasti priemysel. Ku každej s.r.o sú k dispozícii dáta z účtovných uzávierok za posledné tri roky, pred rokom evaluácie R (rok v ktorom sa hodnotil stav spoločnosti zdravá/úpadok). Spoločnosť je hodnotená ako zdravá ak neukončila v danom roku činnosť. Pod upadajúcou/bankrotujúcou spoločnosťou rozumieme spoločnosť, na ktorú bolo v predmetnom roku evaluácie vyhlásené konkurzné konanie. Konkrétne počty s.r.o v úpadku a zdravých spoločností hodnotených v jednotlivých rokoch sú uvedené v Tab. 1. Dáta sú získané z databázy Bisnode Magnusweb, kde je možné exportovať dáta vo formáte *csv*, tieto boli ďalej spracovávané použitím programovacieho jazyka *Python* a jeho modulov.

Tabuľka 1. Charakteristika použitých dát

rok evaluácie R	2013	2014	2015	2016
počet zdravých/bankrotujúcich	4094/30	4482/30	5046/28	5959/16

Databáza Bisnode Magnusweb umožňuje priamy export 20 štandardných ekonomických ukazovateľov, ktoré sme využili pre predikciu úpadku. Tieto štandardné ukazovatele pozostávajú menovite z: rentabilita aktív (ROA), rentabilita vlastného imania (ROE), rentabilita tržieb (ROS), okamžitá likvidita (L1), pohotovú likvidita (L2), bežná likvidita (L3), doba obratu aktív (DOZ), doba obratu kr. pohľadávok (DOKP),

doba obratu kr. záväzkov (DOKZ), doba obratu zásob (DOZ), celková zadlženosť (CZ), zadlženosť vlastného imania (ZVI), finančná páka (FP), návratnosť investícií (NI), krytie cudzích zdrojov (CZ), finančné krytie cudzích zdrojov (FKCZ), krytie stálych aktív (KSA), zadlženosť voči bankám (ZB), mzdy / tržby (MT), mzdy / pridaná hodnota (MPH). Predmetné ukazovatele obsahujú informáciu o finančných tokoch a majetkových pomeroch v rámci firiem pri zohľadnení veľkostných rozdielov medzi firmami.

3 Metodológia

K nastolenému problému sme pristupovali ako k problému rozpoznania odľahlých hodnôt (*outliers*), pričom bola použitá metóda podporných vektorov pre detekciu odľahlých bodov (*OneClassSVM*) [8], konkrétne jej implementácia v scikit-learn [9].

Chýbajúce dáta boli imputované doplnením stredných hodnôt vybraného ukazovateľa. Na natrénovanie klasifikátora bolo použitých 80% vzoriek zdravých spoločností. Samotné ukazovatele boli škálované na nulovú strednú hodnotu a jednotkový rozptyl. Výkonnosť klasifikátora bola testovaná na vzorkách upadnutých spoločností a zvyšných 20% vzorkách zdravých spoločností. Tieto experimenty boli opakované päť krát s náhodným výberom tréningových dát a výsledky boli priemerné.

Môže dôjsť k dvom typom chýb, chyba prvého typu je identifikácia bankrotujúcej firmy ako zdravej a chyba druhého typu je, ak sa zdravá firma identifikuje ako bankrotujúca, preto sme okrem celkovej presnosti predikcie hodnotili aj správne pozitívne hodnoty (TPR, z angl. *true positive rate*) a správne negatívne hodnoty (TNR, z angl. *true negative rate*)

4 Numerické výsledky

V prvotnej analýze sme odhadovali hroziaci bankrot s.r.o, pričom boli použité ekonomické ukazovatele z dát z predchádzajúcich troch rokov jednotlivo, t.j. pre s.r.o, bankrotujúcu v roku 2016, boli na predikciu použité dáta z roku 2015 (*R-1*), potom samostatne z roku 2014 (*R-2*) a samostatne z roku 2013 (*R-3*). Obdobne aj pre spoločnosti bankrotujúce v rokoch 2013-2015. Následne bola predikcia realizovaná využitím dát z dvoch po sebe idúcich rokov, t.j. pre s.r.o bankrotujúce v 2016, boli pre predikciu použité najprv dáta z rokov 2014 a 2015 (*R-2 & R-1*) a následne dáta z rokov 2013 a 2014 (*R-3 & R-2*). Výsledky predikcie sú uvedené v Tab. 2.

Z uvedených hodnôt presnosti predikcie môžeme pozorovať, že dáta z rokov bližšie k roku evaluácie sú výpovednejšie. Na základe údajov z obdobia *R-1* bola dosiahnutá presnosť predikcie na úrovni 77,91 % pre rok *R=2015*. To je v súlade s očakávaním, nakoľko môžeme predpokladať, že s blížiacim sa bankrotom sa v účtovných uzávierkach výraznejšie prejavujú finančné problémy.

Na druhej strane, spojením dát z dvoch rokoch sme očakávali aj zvýšenie presnosti predikcie. V tomto smere sa však presnosť výrazne nezmenila. Maximálna presnosť predikcie v tomto prípade dosahovala 72,41 % pre rok evaluácie *R=2015* a dáta z rokov

Predikcia úpadku spoločností s ručením obmedzeným využitím metód pre rozpoznanie odľahlých bodov

Tabuľka 2. Presnosť predikcie úpadku v jednotlivých rokoch

Rok evaluácie (R)	Rok z ktorého pochádzajú dáta použité na predikciu				
	R-1	R-2	R-3	R-1 & R-2	R-2 & R-3
	73,58±2,84	67,21±2,19	68,08±1,57	71,88±3,37	73,51±2,57
	TNR: 64,52±9,48	TNR: 77,42±8,56	TNR: 56,67±5,67	TNR: 70,97± 6,99	TNR: 53,33± 7,12
	TPR: 69,28± 5,92	TPR: 70,33± 8,59	TPR: 67,84± 4,86	TPR: 68,56± 9,33	TPR: 78,51± 8,29
	74,25±2,07	66,55±2,14	62,50±1,30	67,58±2,80	61,91±0,80
	TNR: 80,65± 6,42	TNR: 41,94± 5,28	TNR: 53,33± 11,18	TNR: 74,19± 6,74	TNR: 43,33± 4,33
	TPR: 77,69± 4,56	TPR: 79,88± 6,59	TPR: 75,00± 10,17	TPR: 79,70± 4,06	TPR: 88,29± 5,07
	77,91±2,64	72,28±1,05	66,41±2,33	72,41±2,67	69,12±1,41
	TNR: 73,08± 10,58	TNR: 67,86± 5,71	TNR: 57,14± 6,93	TNR: 61,54± 6,21	TNR: 57,14± 11,52
	TPR: 71,09± 8,22	TPR: 83,18± 4,96	TPR: 78,41± 5,30	TPR: 81,04± 3,99	TPR: 70,69± 10,39
	67,07±4,08	60,56±4,52	62,80±2,09	67,58±3,56	68,18±3,06
	TNR: 14,29±12,14	TNR: 50,00± 14,64	TNR: 68,75± 12,33	TNR: 67,76±12,14	TNR: 62,50± 10,90
	TPR: 95,59±6,98	TPR: 65,67± 9,96	TPR: 65,70± 10,77	TPR: 70,86±8,15	TPR: 69,30± 7,54

2014 a 2013. Dôvodom môžu byť samotné dáta, kde pridaním údajov z ďalšieho roku nie je pridaná nová informácia, iba tá, ktorá je už obsiahnutá v dátach. Ďalšou hypotézou je, že po spojení údajov z dvoch rokov nedochádza k zlepšeniu, nakoľko klasifikátor nie je schopný dáta dostatočne zmapovať. Ďalšie vyšetrovanie bude predmetom ďalšieho výskumu.

Ako môžeme na základe údajov v Tab. 2 pozorovať, klasifikátor dosahuje pomerne odlišné výsledky v jednotlivých rokoch. V presnosti predikcie je rozdiel až 10% v predikcii bankrotu s.r.o., ktoré upadli v roku 2015 a v roku 2016. To naznačuje, že v prípade databáz, ktoré sú využívané na predikciu úpadku je vhodnejšie analyzovať dáta z viacerých rokov.

5 Záver a ďalšie smerovanie výskumu

V tomto príspevku sme prezentovali prvotné výsledky predikcie úpadku s.r.o. s maximálnou úspešnosťou 77,91%. Výsledok je značne ovplyvnený rokom z ktorého sa dáta čerpajú. Avšak tieto výsledky naznačujú, že úpadok je predikovateľný.

V ďalšej práci sa zameriame na aplikáciu iných metód pre predikciu odľahlých hodnôt, a aj metód pre klasifikáciu s učiteľom (*supervised*), využitie metód pre podzorkovanie majoritnej triedy a podobne. Ďalšou sľubnou oblasťou je návrh nových

ukazovateľov, ktoré dokážu zachytiť finančnú situáciu výstižnejšie ako tradičné finančné ukazovatele, čo môže značne uľahčiť úlohu klasifikačnému algoritmu.

PodĎakovanie: Táto práca bola podporená Agentúrou na podporu výskumu a vývoja projektom číslo APVV-15-0358.

References

1. Bellovary, Jodi L., Don E. Giacomino, and Michael D. Akers. "A review of bankruptcy prediction studies: 1930 to present." *Journal of Financial education* (2007): 1-42.
2. Altman, Edward I. "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy." *The journal of finance* 23.4 (1968): 589-609.
3. Boritz, J. Efrim, and Duane B. Kennedy. "Effectiveness of neural network types for prediction of business failure." *Expert Systems with Applications* 9.4 (1995): 503-512.
4. Zhao, D., Huang, C., Wei, Y., Yu, F., Wang, M., Chen, H.: An Effective Computational Model for Bankruptcy Prediction Using Kernel Extreme Learning Machine Approach. *Computational Economics* 49(2), 325-341 (2017).
5. Wang, L., Wu, C.: Business failure prediction based on two-stage selective ensemble with manifold learning algorithm and kernel-based fuzzy self-organizing map. *Knowledge-Based Systems* 121, 99-110 (2017).
6. Pietruszkiewicz, W.: Dynamical systems and nonlinear Kalman filtering applied in classification. *Cybernetic Intelligent Systems*, 2008. CIS 2008. 7th IEEE International Conference on London, UK.
7. Zhou, Ligang. "Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods." *Knowledge-Based Systems* 41 (2013): 16-25.
8. Manevitz, L., Yousef, M.: One-Class SVMs for Document Classification. *Journal of Machine Learning Research* 2, 139-154 (2001).
9. Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, *JMLR* 12, pp. 2825-2830, 2011.