

Přínos učících se metod statistického rozpoznávání obrazů při hledání faktorů konkurenceschopnosti českých podniků

Jiří ŠPALEK* – Ondřej ČÁSTEK**

Contribution of Statistical Pattern Recognition Methods in Search of Competitiveness Factors of Czech Companies

Abstract

The submitted paper concentrates on the methodical aspects of measuring the relationship between potential competitiveness factors and the corporate competitiveness. We employ methods of statistical pattern recognition, particularly the sequential forward flow search algorithm (SFFS). The algorithm is applied on data from 432 companies. For these companies there was known their financial performance and there was up to 683 (each company) potential factors of this performance in our database. The text therefore summarizes the known approaches, describes the SFFS algorithm and proves its contribution to this field of research. An undeniable advantage of this method is its low demands on data: it does not require the normality or an a priori model. Also, it is able to evaluate relationships among many variables at once in acceptable time frame. The article presents the drawbacks of this method as well.

Keywords: corporate competitiveness, statistical pattern recognition, factors of competitiveness, sequential forward flow search

JEL Classification: C49, L25, M10

Úvod

Problematika hledání faktorů hospodářské výkonnosti podniků představuje pro teorii podnikového hospodářství i pro teorii managementu aktuální a atraktivní téma. Existuje bezpočet prací, které se danou problematikou zabývají. Jen

* Jiří ŠPALEK, Masarykova univerzita, Ekonomicko-správní fakulta, Katedra veřejné ekonomie, Lipová 41a, 602 00 Brno, Česká republika; e-mail: spalek@econ.muni.cz

** Ondřej ČÁSTEK, Masarykova univerzita, Ekonomicko-správní fakulta, Katedra podnikového hospodářství, Lipová 41a, 602 00 Brno, Česká republika; e-mail: castek@econ.muni.cz

Molina, Barrios del Pino a Rodriguez (2004) jmenují přes tři desítky prací v této oblasti a jejich výčet není zdaleka úplný. Také v České i Slovenské republice se jedná o často řešenou úlohu, můžeme jmenovat Kloudovou, Dobeše a Stříže (2005), Hučku (2005), Nemcovou (2006), Nečadovou, Broňovou a Průchu (2007), Kitu, Konštiaka a Kitu (2008) a mnohé další, a to nezmiňujeme problematiku úzce související – měření konkurenceschopnosti podniků (např. Kislingerová a kol., 2006; Rajčániová, 2006). Nicméně způsoby zmocnění se dané tematiky, aplikace přístupů a metod, důvody řešení, věrohodnost výsledků i způsob jejich uplatnění, se v jednotlivých pracích výrazně liší.

Tudíž ačkoli je tato oblast předmětem zájmu výzkumníků již desítky let,³ panuje zde značná rozmanitost jak způsobů měření podnikové výkonnosti, tak toho, na jaké potenciální faktory se zaměřují, a dokonce i způsobů měření souvislosti mezi příčinami a důsledky. V předloženém článku bychom se chtěli zaměřit právě na způsoby měření souvislosti mezi potenciálními faktory hospodářské výkonnosti a hospodářskou výkonností a představit použitou metodu, jež je do jisté míry unikátní nejen v rámci České republiky, ale pravděpodobně i celosvětově.⁴ Jelikož se naše datová analýza opírá o výsledky empirického šetření konkurenceschopnosti českých podniků,⁵ máme k dispozici rozsáhlou datovou bázi. Tato báze obsahuje kvantitativně vyjádřené vyhodnocení hospodářské výkonnosti vybraných podniků na straně jedné a potenciální faktory ovlivňující výkonnost těchto podniků na straně druhé.

V našem případě hospodářskou výkonnost považujeme za míru konkurenceschopnosti podniku. Tradiční pojetí konkurenceschopnosti je, že se jedná o schopnost dosáhnout tržního úspěchu (Michalet, 1981; Mathis, Mazier a Rivaud-Danset, 1988). Tato schopnost by se měla projevit udržením nebo zvýšením objemu tržeb v porovnání s vývojem trhu (Molina, Barrios del Pino a Rodriguez, 2004). Námi použitá data vychází z podobné logiky, a to, že konkurenceschopný podnik „dokáže obsluhovat trh“ (Cellini a Soci, 2002). To se pak odráží ve finančních výsledcích podniku; jak vysvětlili Slaný a kol. (2006), nejlepšími měřítky konkurenceschopnosti je buď zisk na jednotku produkce, nebo rentabilita aktiv. Právě rentabilita aktiv byla vybrána jako vhodné vyjádření hospodářské výkonnosti podniku. Její velkou výhodou je jednoduchost a nenáročnost na potřebná data, na rozdíl od propracovanějších metod (např. EVA, CVA, RAVE, PM apod.). Nevýhodou ovšem je značná možnost zkreslení. Tato nevýhoda byla zde minimalizována a) použitím druhého ukazatele: růstu aktiv

³ Například R. P. Rumelt publikoval empirickou studii na toto téma již v roce 1981.

⁴ I po podrobném studiu dostupné literatury, kterou měl výzkumný tým k dispozici, nám není znám žádný zdroj, který by se podobnou aplikací statistických metod (v tak velkém rozsahu) zabýval.

⁵ Informace o tomto výzkumu jsou uvedeny zejména v publikacích Blažek (2007; 2008).

a b) použitím údajů za pět let u obou ukazatelů pro každý podnik. Více o další práci s těmito ukazateli lze nalézt v části 1.2.

V příspěvku nejprve shrnujeme povahu a zdroj dat, jež jsme měli k dispozici, poté se věnujeme použité metodě, přičemž hlavní důraz klademe na osvětlení principu fungování učících se metod statistického rozpoznávání obrazů. Následně se na vybraných výsledcích snažíme dokázat přínos této metody pro hledání faktorů hospodářské výkonnosti. Ačkoli některé z těchto výsledků přináší zajímavý pohled na faktory ovlivňující konkurenceschopnost českých podniků, jejich obsah podrobně nediskutujeme. Důvodem nám je, mimo omezený rozsah článku, i zaměření článku na metodické aspekty vyhodnocování konkurenceschopnosti podniků obecně.

1. Data

Základem našich analýz jsou data pocházející z hlediska rozsahu unikátního empirického šetření, které provedlo Centrum výzkumu konkurenční schopnosti české ekonomiky při ESF MU v Brně (CVKS). Šetření se opírá dotazník, jehož struktura vychází ze stakeholderského modelu⁶ podniku. Tato data byla dále doplněna o údaje charakterizující hospodářskou výkonnost podniků, které jsme čerpali z veřejných databází, zejména z databáze Albertina.

1.1. Základní a výběrový soubor podniků

Základní soubor šetřených podniků byl omezen na: společnosti s ručením omezeným a akciové společnosti se sídlem v České republice, z odvětví zpracovatelského průmyslu a stavebnictví s 50 a více pracovníky.

Po vyloučení podniků v likvidaci, konkurzu, nebo soudní exekuci, základní soubor v době empirického šetření čítal 4 483 subjektů, z nichž 2 817 mělo z hlediska komplexnosti a kvality účetních informací požadovanou úroveň. Tento soubor se stal pro předmětný výzkum *souborem základním*. Empirického šetření se zúčastnilo 432 podniků, což je 15,33 % základního souboru.

1.2. Primární a sekundární analýza dat a jejich příprava

Na základě strukturovaného dotazníku a s využitím databáze Creditinfo bylo za každý podnik k dispozici 683 hodnot sledovaných proměnných.⁷

⁶ Jelikož příspěvek není orientován na diskusi metodiky sběru dat, nezdůvodňujeme námi zvolený přístup v podobě vnímání podniku jako místa interakce významných stakeholderů. Tento přístup vychází z Freemana (1984) a dalších; například Post, Preston a Sachs (2002) uvádějí, že stakeholderský „pohled“ vidí organizaci „jako bázi pro analýzu a řízení čtených a rozdílných vztahů, které vyvstávají v rámci tohoto uspořádání“ atd.

• **Vytvoření skupin podniků dle jejich hospodářské výkonnosti**

Studované podniky bylo nejprve nutno seskupit do skupin dle hospodářské výkonnosti. Za vysvětlující proměnné byly zvoleny dva účetní ukazatele v pěti-leté retrospektivě – rentabilita aktiv a růst aktiv. Margolis a Walsh (2002) našli ve 122 zkoumaných studiích 70 různých způsobů měření podnikové finanční výkonnosti.⁸

Návratnost (investic, tržeb, aktiv) je přitom jedním z nejčastějších. Rentabilitu aktiv jako způsob měření finanční výkonnosti podniku zvolili pro obdobné studie například Abor a Biekpe (2007), Huselid (1995) nebo Hansen a Wernerfelt (1989). Tento ukazatel jsme doplnili růstem aktiv pro minimalizování vlivu daňové optimalizace. Ze stejného důvodu byla použita data za pět po sobě jdoucích let.⁹

Výběrový soubor byl strukturalizován na základě vybraných ukazatelů shlukovou analýzou. Ta byla provedena metodou nejbližších středů (*K-means cluster analysis*), která rozčlenila podniky do relativně homogenních skupin (shluků) podle minimální mezishlukové vzdálenosti jednotlivých členů shluku. Pro vyjádření této vzdálenosti byla využita euklidovská metrika.¹⁰

Výsledkem shlukové analýzy a navazujících operací bylo rozčlenění podniků do tří skupin:¹¹

- *podniky hospodářsky výkonné* (označené jako A, 205 podniků) s alespoň jedním z ukazatelů nad průměrem a druhým nezáporným;
- *podniky středně hospodářsky výkonné* (skupina B, 185 podniků), u nichž hodnoty obou ukazatelů byly pod průměrem výběrového souboru, ale nezáporné;
- *podniky hospodářsky nevýkonné* (skupina C, 42 podniků), u kterých hodnoty obou uvedených ukazatelů byly záporné.¹²

⁷ Proměnné měly nejrůznější povahu, která se odvíjela od typu otázky, na jejímž základě byla zjišťována. Zdrojová datová báze tak obsahuje jak intervalové proměnné, tak ordinální, či dokonce nominální proměnné. Tato skutečnost, ač ji dále nediskutujeme, nepochybně může mít vliv na dosahované výsledky.

⁸ Obecně využití účetních údajů podporují i Allouche a Laroche (2005), kteří nacházejí mnohem silnější vztah účetních ukazatelů a faktorů konkurenceschopnosti, než je tomu u ukazatelů tržních.

⁹ Alternativní výsledky pro jinou kombinaci ukazatelů uvádějí Suchánek a Špalek (2007).

¹⁰ Metrika má podobu $\text{distance}(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$, kde x je konkrétní hodnota i -té souřadnice vektoru ukazatelů za podnik X a y je konkrétní hodnota i -té souřadnice vektoru ukazatelů za podnik Y .

¹¹ Další možné kombinace, jako např. hodnota jednoho ukazatele záporná a druhá nezáporná, se nevyskytly.

¹² Z prostorových důvodů neuvádíme podrobnější popis metodiky shlukování. Tento popis, včetně ekonomické interpretace, je uveden v Šiška (2008).

• Redukce počtu vstupních proměnných

Pro účely statistického rozpoznání obrazů bylo nutno množinu proměnných (potenciálních příznaků) vhodně připravit. Daný obraz (podnik) je možno popsat celou řadou charakteristik, které můžeme označit jako množinu D_o . Ne všechny tyto proměnné (příznaky) jsou však pro samotný výběr nejinformativnějších příznaků vhodné. Důvodem může být například jejich vzájemná korelace, stejně tak celá řada z nich může být pro daný problém redundantní či irelevantní. Pro jejich redukci byly využity tradiční statistické metody, opírající se zejména o primární analýzu dat, stejně jako o bivariační a korelační analýzu. Spolu s těmito statistickými metodami byla pro redukci využita též věcná analýza vycházející z ekonomických aspektů zkoumaných souvislostí.

Výsledkem byla množina D_I obsahující tyto proměnné, které následně vstoupily do dále prezentovaných experimentů s metodami statistického rozpoznávání obrazů.

2. Metoda

Jak jsme již poznamenali, náš výzkum hledání faktorů konkurenceschopnosti se opírá o využití metod statistického rozpoznávání obrazů (*statistical pattern recognition*). Tato metoda dosud nebyla pro tyto účely využita. V tabulce 1 shrnujeme postupy použité ve více než sto studiích pro měření souvislosti mezi různými potenciálními faktory konkurenceschopnosti a různým vyjádřením konkurenceschopnosti.

T a b u l k a 1

Porovnání používaných metod měření závislosti podnikové výkonnosti na potenciálních faktorech

Metoda	Příklad studie	Nezávislé/kontrolní proměnné
Korelace	Andrews, 2010	19/7
Parciální korelace	Arthur, 1994	1/4
T-test	Artiah, 2010	5/0
ANOVA	Smith, 1989	1/0
OLS	Chauvin, 1993	2/4
Multiregresní analýza	Powell, 1997	26/0
Panelová regrese	Abor, 2007	11/3
Strukturální modelování	Yilmaz, 2005	19/3
See5 algoritmus (rozhodovací strom, učící se metoda)	Molina, 2004	20/0

Pramen: Vlastní zpracování.

Obecně je patrné, že tyto studie se opírají především o klasické statistické metody, jako jsou korelace či regrese. Všechny tyto metody vycházejí, mimo další silné předpoklady, jako je například normalita dat, z existence apriorního

modelu vzhledem k němuž jsou data vyhodnocována. Z hlediska našeho výzkumu jsou nejzajímavějšími metodami strukturální modelování, a zejména učící se metoda rozhodovacího stromu. Avšak i tato metoda vychází z požadavku na apriorní stanovení jednoho faktoru jako výchozího.

Naopak výhodou námi zvolených metod statistického rozpoznávání obrazů oproti již uvedeným spočívá mimo jiné i v tom, že nevycházejí ze žádné apriorní představy o povaze vztahů mezi daty, stejně tak nepředpokládá jejich normalitu. Cílem těchto metod je klasifikovat obrazy (reprezentace objektů reálného světa, v našem případě podniky) do konečného, obvykle nevelkého počtu tříd.¹³ V případě dvou tříd, což je nejčastější, se hovoří o tzv. dichotomické klasifikaci. Metoda vychází z předpokladu, že každý obraz do jedné ze tříd jednoznačně patří, což je jeho klasifikační vlastnost.

Základním předpokladem však je, že tyto jiné měřitelné údaje (v terminologii metody statistického rozpoznávání obrazů příznaky) se skutečným přiřazením obrazu do třídy alespoň statisticky souvisí. Úloha tedy vyžaduje místo analytického přístupu tzv. učící se přístup, který je založen na myšlence, že dostatečná informace potřebná k provedení klasifikace či rozpoznání obrazu je obsažena v datech popisující zkušenosti z minulosti.

Pro realizaci učících se metod je tedy nutno využít tzv. trénovací množinu. Množina musí sestávat z obrazů se známou klasifikací. Výsledkem řešení je nalezení výše uvedených příznaků a nalezení pravidla, jak dle těchto příznaků přiřazovat jednotlivé obrazy do jednotlivých tříd.

V našem případě lze za obrazy považovat podniky z výběrového souboru. Za třídy byly považovány skupiny podniků, vytvořené na základě shlukové analýzy dle vybraných finančních ukazatelů – rentability aktiv a růstu aktiv. Za příznaky byly považovány ty proměnné z výchozího souboru charakteristik (proměnných), získaných za sledované podniky prostřednictvím dotazníku, resp. z databáze, které mají, ve vzájemných souvislostech, podstatný vliv na zařazení podniku do některé z uvedených skupin. Jde tedy o faktory ovlivňující hospodářskou výkonnost.

Námi použitá metodika se opírá o tzv. suboptimální vyhledávací strategie. Suboptimálních metod vyhledávání byla postupně publikována celá řada, zřejmě nejznámějšími a dnes nejcitovanějšími metodami jsou metody sekvenčního plovoucího hledání podmnožin (*sequential floating search*) – viz Pudil, Novovičová a Kittler (1994).¹⁴ Plovoucí hledání zahrnuje ve skutečnosti dvě metody, a i když

¹³ Jak uvedeme dále, v dané úloze jsme podniky členili do různého počtu tříd, a to dle hodnot dvou finančních ukazatelů. Počet tříd varioval od relativně velkého počtu tříd – 13 – až po dichotomní klasifikaci.

¹⁴ Konkrétně jsme využili algoritmus vyvinutý pracovištěm Salome, jež je společným pracovištěm Ústavu teorie informace a automatizace Akademie věd ČR a Fakulty managementu VŠE.

obě střídají přidávání a odebrání příznaků do/z pracovní množiny, můžeme je rozlišit jakožto dva různé algoritmy podle převažujícího směru hledání, a to:

- dopředný algoritmus, který je známý jako sekvenční dopředný plovoucí výběr (*sequential forward floating selection* – SFFS),
- zpětný algoritmus, který je známý jako sekvenční zpětný plovoucí výběr (*sequential backward floating selection* – SBFS).

Oba algoritmy dohromady jsou známé jako plovoucí metody (*floating methods*), protože výsledná dimenzionalita po jednotlivých krocích algoritmu se nemění monotónně, ale ve skutečnosti „plave“ nahoru a dolů. Kombinací původních (pouze po krocích optimálních) metod SFS (*sequential forward selection*) a SBS (*sequential backward selection*) se plovoucí metody blíží optimálnímu řešení. Jejich hlavním rysem je, že například při zpětném plovoucím výběru (SBFS) již jednou vyřazené příznaky mohou být později v procesu výběru znovu mezi vybrané příznaky zařazeny, pokud to zlepší hodnotu informativnosti. Informativností rozumíme procentní vyjádření přesnosti, se kterou jsou podniky zařazeny do správného shluku.

V některých případech (např. velkého množství primárních příznaků, jako je tomu v případě předmětné úlohy) je možno řešení (nejinformativnější podmnožinu příznaků dimenzionality d) ještě „doladit“ použitím oscilačního algoritmu (viz Somol a Pudil, 2000, s. 406 – 409). Oscilační vyhledávání lze považovat za proceduru „vyšší úrovně“, která používá jiné metody výběru příznaků jako subprocedury hlavního průběhu vyhledávání. Celý koncept se vyznačuje vysokou flexibilitou a umožňuje modifikace pro různé účely.

Ferri a kol. (1994, s. 403 – 413) a Kudo se Sklanskym (2000) několika testy dokázali, že algoritmus SFFS patří mezi nejvýkonnější způsoby výběru klasifikačních příznaků ve statistickém rozpoznávání obrazů. S ohledem na již uvedené skutečnosti tedy byla pro analýzu faktorů hospodářské výkonnosti vybrána právě metoda dopředného plovoucího vyhledávání.

Pro hodnocení informativnosti¹⁵ testovaných množin proměnných použijeme klasifikační metodu *k-Nearest Neighbors*, která řadí vzorky (podniky) do jednotlivých tříd (kategorií důležitosti) podle k nejbližších sousedů. Právě počet nejbližších sousedů k je volitelným parametrem experimentů. Protože není možné jednoznačně rozhodnout, jaká velikost k by měla být použita, budeme experimenty provádět s variabilním nastavením k , a to 1, 3 a 5. Více o metodě statistického rozpoznávání obrazů lze nalézt například v Pudil (2008).

¹⁵ Informativnost lze zjednodušeně považovat za míru úspěšnosti metody při hledání množiny faktorů konkurenceschopnosti. Její hodnota udává pravděpodobnost, s jakou jsou obrazy (podniky) dle vybraných příznaků zařazovány do správné třídy. Tedy hodnota informativnosti 50 % říká, že pouze polovina podniků je zařazena do třídy, do níž skutečně patří.

3. Výsledky

Následující sekce představuje souhrn některých – dle našeho názoru nejilustrativnějších – výsledků. Výsledky prezentované v tomto příspěvku si kladou za cíl především ukázat praktickou použitelnost a přínos zvolené metodiky pro studovanou úlohu, tedy pro hledání typické konfigurace faktorů vedoucích k vytvoření specifických typů hospodářské výkonnosti podniků. Z těchto důvodů si zde neklademe za cíl zobecňovat výsledky experimentů (v podobě množiny nalezených faktorů konkurenceschopnosti), ani podrobně diskutovat jejich ekonomický obsah.

Pro ověření aplikovatelnosti metodiky statistického rozpoznávání obrazů na řešenou úlohu a důvěryhodnosti dosažených výsledků, tj. nalezených faktorů, které ve vzájemné interakci ovlivňují hospodářskou výkonnost podniků, byla uskutečněna řada experimentů. Zadáni pro jednotlivé experimenty vyplývala z kombinace volby disponibilních variant této metodiky a z volby vstupů.

Z variant metod statistického rozpoznávání obrazů, byla vybrána, jak jsme již výše uvedli, *metoda sekvenčního dopředného plovoucího výběru* (SFFS). Tato metoda dává zpravidla nejlepší výsledky,¹⁶ a proto byla použita ve všech níže prezentovaných experimentech.

Pro hodnocení informativnosti testovaných množin proměnných byla použita klasifikační metoda *k-Nearest Neighbors* (*kNN*). Experimenty byly prováděny s variabilním nastavením hodnot k , na 1, 3 a 5. Po každém experimentu byla vybrána jako nejvhodnější ta velikost k , pomocí které bylo dosaženo nejinformativnější množiny proměnných. V dalším textu budou tři varianty volby k označeny jako metody 1NN, 3NN a 5NN. Volba variant vstupů se týkala volby počtu a charakteru tříd (skupin), do kterých byly podniky dle své hospodářské výkonnosti zařazeny.

Jak jsme již uvedli, za kritérium pro vzájemné srovnávání a hodnocení výsledků jednotlivých variant byla zvolena hodnota *informativnosti* udávající pravděpodobnost, s jakou jsou obrazy (podniky) dle vybraných příznaků zařazovány do správné třídy.

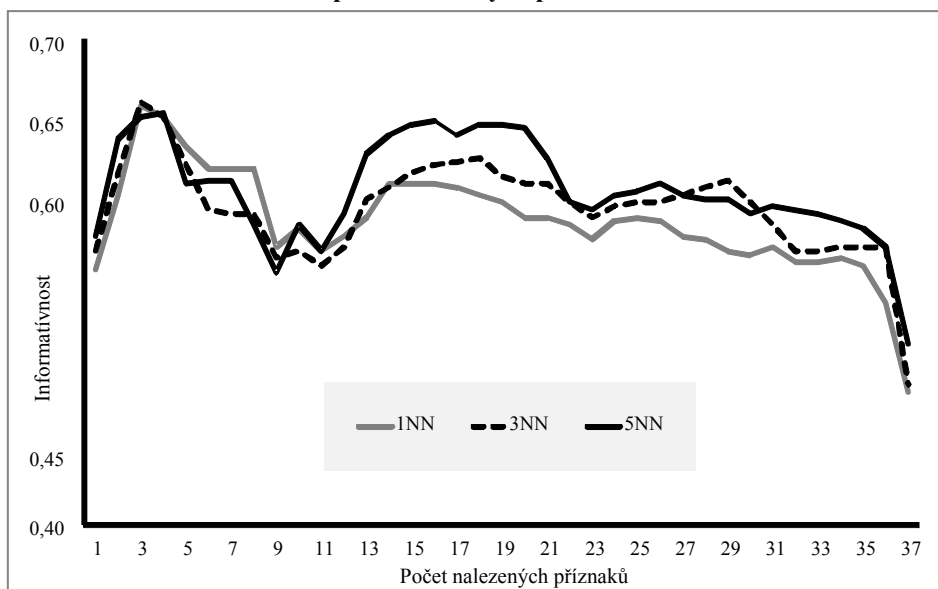
3.1. Základní experiment

Do základního experimentu vstoupila množina D_I obsahující 37 vybraných proměnných (viz část 1.2). Závislost mezi informativností (v našem případě hodnocené pomocí pravděpodobnosti správné klasifikace) a počtem příznaků (proměnných), které daná metoda vybrala v jednotlivých krocích jako faktory hospodářské výkonnosti, je názorně uvedena v grafu 1.

¹⁶ Viz Pudil (2008). Tato skutečnost byla do jisté míry ověřena i při našich dalších experimentech, které však pro omezený rozsah článku již neuvádíme.

Graf 1

Závislost informativnosti na počtu nalezených příznaků



Pramen: Vlastní zpracování.

Jsou zde porovnány výsledky při uplatnění metody 1NN, 3NN a 5NN. Z grafu 1 je zřejmé, že různé počty „nejbližších sousedů“ zřetelně ovlivňují dosažená maxima hodnoty informativnosti. Při 1NN je nejvyšší hodnota informativnosti 0,6590, při 3NN 0,6613 a při 5NN 0,6544.

Graf 1 také ukazuje typický vývoj hodnoty informativnosti v závislosti na počtu nalezených n -tic faktorů, a to bez ohledu na počet *nejbližších sousedů*. Hodnota informativnosti zaznamenává nejprve rychlý růst a dosažení globálního maxima při množině tří až čtyř proměnných, poté následuje její pokles. Pokles hodnoty informativnosti po dosažení optima se však nesnižuje s růstem množiny proměnných ani plynule, ani bez výjimky. Jak je možno pozorovat, úbytek informativnosti neprobíhá monotónně, ale dochází k určitým „zhoupnutím“, spojeným s existencí lokálních maxim. Právě hledání a posouzení těchto lokálních maxim bylo dalším krokem experimentu.

Náš zájem se totiž nesoustředil na analýzu nalezených globálních maxim. Důvod vyplývá z povahy řešené úlohy. Nelze totiž předpokládat, že tak komplexní a složitý jev, jakým je hospodářská výkonnost podniků jistě je (Lipovatz, Mandaraka a Mourelatos, 2000; Haenecke, 2002; a další), závisí pouze na kombinaci hodnot pouhých tří až čtyř proměnných. Rovněž tak by bylo obtížné až nemožné určitý typ hospodářské výkonnosti těmito několika proměnnými důvěryhodně vysvětlit. Toto vědomí podporuje fakt, že lokální maxima jsou do určité

míry hierarchicky podřízena nalezenému globálnímu extrému. Jinými slovy, informace obsažená ve zmíněných 3-4 proměnných charakterizujících globální maximum, je s minimální ztrátou¹⁷ obsažena i v nalezené skupině faktorů vyplývající z lokálního maxima. Metoda tak umožňuje kombinaci čistě statistické analýzy a určité míry heuristického či *rule of thumb* přístupu.

U metody 1NN dosahují hledaného lokálního maxima hned tři množiny proměnných, a to množiny o velikosti 14, 15 a 16 proměnných, ovšem při hodnotě informativnosti pouze 0,6106. Metodou 3NN dosahuje lokálního maxima množina o velikosti 18 proměnných a její informativnost je 0,6267. Konečně metodou 5NN dosahuje lokálního maxima množina o velikosti 16 proměnných s informativností dokonce 0,6498, což není, převedeno na procenta, ani o půl procentního bodu horší výsledek, než nejlepší množina touto metodou dosažená (čtveřice proměnných s informativností 0,6544), a zhruba o jeden procentní bod horší výsledek, než nejlepší informativnost vůbec dosažená s touto 37 člennou množinou proměnných (0,6613).

Výsledky je možno porovnat v tabulce 2.

T a b u l k a 2

Porovnání globálních maxim a vybraných lokálních maxim

Klasifikační metoda	Globální maximum		Vybrané lokální maximum	
	hodnota informativnosti	počet nalezených faktorů	hodnota informativnosti	počet nalezených faktorů
1NN	0.6590	3	0.6106	14
3NN	0.6613	3	0.6267	18
5NN	0.6544	4	0.6498	16

Pramen: Vlastní zpracování.

S ohledem na uvedené skutečnosti byly identifikovány jako faktory hospodářské výkonnosti podniků proměnné obsažené v 16-ti členné množině vytvořené dle metody SFFS 5NN, s hodnotou informativností 0,6498. Můžeme tedy říci, že na základě známých hodnot těchto 16 proměnných jsme schopni 65 % podniků zařadit do správné skupiny hospodářské výkonnosti.

Informativnost můžeme považovat za obdobu míry determinace, již známé z jiných metod měření souvislosti mezi proměnnými. Zatímco korelační analýzou nebo testováním rozdílů v průměrech či rozptylech můžeme zjišťovat míru souvislosti mezi dvěma nebo několika málo proměnnými, výhodou statistického rozpoznávání obrazů je měření míry souvislosti mezi mnoha proměnnými najednou.

¹⁷ Tato ztráta je dána rozdílem hodnoty kritéria informativnosti pro globální a příslušný lokální extrém.

T a b u l k a 3

Nejinformativnější proměnné při nastavení SFFS 5NN, lokální maximum

Faktor	Název	Vliv na hospodářskou výkonnost dle bivariační analýzy	Koeficient
1.	Rozpětí řízení	neprokázán	eta ²
2.	Součást koncernu	neprokázán	
3.	Typ vlastnické struktury	0.02*	
4.	Původ vlastníka	neprokázán	
5.	Softwarové aplikace – Moduly MRP	neprokázán	R
6.	Velikost	neprokázán	
7.	Podíl technicko-hospodářských pracovníků	0.121**	
8.	Fluktuace	neprokázán	
9.	Podíl pohyblivé složky mzdy	0.147**	R
10.	Výše prostředků na zaměstnanecké výhody	neprokázán	eta ²
11.	Podnikatelská strategie	neprokázán	
12.	Podíl importu na dodávkách	neprokázán	
13.	Certifikát ISO 14000	0.06***	
14.	Právní forma podnikání	0.01***	eta ²
15.	Odvětví	0.06***	eta ²
16.	Přidaná hodnota na zaměstnance	0.235**	R

Poznámky: R – Pearsonovo R, jedna hvězdička označuje statistickou významnost na hladině 10 %, dvě hvězdičky 5 % a tři hvězdičky 1 %.

Pramen: Vlastní zpracování.

Při řešení dílčího úkolu stejnou metodou a hlubším zkoumáním vybraných proměnných pro následnou interpretaci výsledků například vyplynulo, že výběrový soubor, jakkoli byl základní soubor zúžen, je z pohledu finanční výkonnosti stále značně heterogenní. Ačkoli výzkumný tým testoval potenciální faktory finanční výkonnosti korelacemi, parciálními korelacemi apod., nebylo pro 683 proměnných efektivní testovat všechny možné kombinace. Metoda SFFS však pomohla omezit tuto množinu a právě při její detailnější analýze vyplynula na povrch mnohá nová a zajímavá fakta, například silná pozitivní závislost důležitosti zaměstnanců na míře korupce ve stavebnictví ($R = 0,33$) a slabá negativní ve zpracovatelském průmyslu ($R = -0,15$), přičemž v nerozděleném výběrovém souboru byla tato souvislost téměř nulová ($R = -0,08$). Dále byla testována souvislost důležitosti zaměstnanců a finanční výkonnosti atp.

Další předností byly nacházené míry souvislosti. Při zkoumání více proměnných současně byly tyto podstatně silnější, než při zkoumání dílčích souvislostí. Jak ukazuje tabulka 3, míry dílčích souvislostí zřídka překročily 0,2, což můžeme označit za souvislost triviální až nízkou (DeVaus, 2002). Hodnotu informativnosti 0,6498 tak můžeme považovat za významné zlepšení při hledání faktorů konkurenceschopnosti.

Vysvětlení je jednoduché a již jsme je jednou použili: komplexní a složitý jev, jakým hospodářská výkonnost podniků jistě je, nemůže záviset pouze na hodnotě jedné nebo několika málo proměnných.

3.2. Další experimenty

Za účelem posouzení vlivu změn vstupních veličin na výsledky výběru množiny faktorů hospodářské výkonnosti podniků byly provedeny další experimenty. Šlo o sledování toho, jak je informativnost ovlivňována změnami ve struktuře tříd. Bylo experimentováno se strukturou

- třinácti skupin podniků tak, jak byly seskupeny dle shlukové analýzy;
- tří skupin podniků, vytvořených agregací zmíněných třinácti skupin;
- tří variant dvou dichotomních skupin podniků.

V případě ad c) byly vytvořeny varianty dichotomních skupin tak, že z daných skupin A, B a C šlo vždy střídavě o analýzu jedné z nich vůči sjednocení zbývajících dvou.

T a b u l k a 4

Vztah mezi informativností a změnou struktury podniků

Struktura podniků dle hospodářské výkonnosti	Globální maximum			Vybrané lokální maximum		
	klasifikační metoda	hodnota informativnosti	počet faktorů	klasifikační metoda	hodnota informativnosti	počet faktorů
13 skupin	5NN	0.3086	2	5NN	0.2815	15
3 sk. (A, B, C)	3NN	0.6613	3	5NN	0.6498	16
2 sk. (A vs B + C)	5NN	0.7413	19	5NN	0.7413	19
2 sk. (B vs A + C)	1NN	0.7229	4	1NN	0.6905	19
2 sk. (C vs A + B)	3NN	0.9284	4	3NN	0.9215	20

Pramen: Vlastní zpracování

V tabulce 4, kde jsou zachyceny vždy nejvyšší hodnoty globálních a vybraných lokálních maxim informativnosti, je vidět, že u 13 skupin dosahuje globální maximum tohoto kritéria hodnoty jen 0,3086, u tří skupin (A, B, C) hodnoty 0,6613 a nejlepší varianta ze tří dichotomních skupin (C versus A + B) dokonce hodnoty 0,9284.

Za pozornost stojí skutečnost, že při porovnání tří variant dichotomních skupin je zřejmé, jak ohraničenost, resp. specifičnost daných tříd (skupin) ovlivňuje informativnost. Nejvyšší hodnotu informativnosti vykazují vybrané příznaky (faktory) v rámci dichotomie „podniky se zápornou rentabilitou a záporným růstem oproti ostatním podnikům“ (C versus A + B), v pořadí druhou hodnotu informativnosti vykazují vybrané příznaky při opačném pohledu, tedy při dichotomii „nadprůměrné podniky oproti ostatním podnikům“ (A versus B + C), a nejhorší hodnotu informativnosti pak vykazují vybrané příznaky při variantě, kdy podniky se střední hospodářskou výkonností jsou stavěny vůči sjednocení proti sobě stojících skupin podniků – nadprůměrných a ztrátových (B versus A + C).

3.3. Diskuse výsledků

Naše výsledky¹⁸ ukazují poměrně komplexně širší problémů, které vznikly při aplikaci uvedené metodiky. Jedním z nich je určitá míra fluktuace ve výsledcích, která se zvyrazňuje u lokálních maxim. Ukázkou těchto fluktuací pro námi studovanou množinu potenciálních faktorů D_i uvádíme v příloze. Dále vzniklo podezření, že metoda upřednostňuje dichotomní a trichotomní proměnné, což bylo pozorováno v prvním provedeném experimentu (výsledky zde neprezentovány). To může být dáno faktem, že u dichotomní proměnné je rozdíl mezi dvěma rozdílnými hodnotami maximální (celá délka intervalu $0 - 1$),¹⁹ u trichotomní proměnné (ne nominální) je rozdíl mezi dvěma hodnotami polovinu intervalu, což je podstatně více, než u dalších zahrnutých proměnných – ordinálních s pětibodou škálou nebo u intervalových. Stejně tvrzení jako pro dichotomní proměnné platí i pro nominální proměnné, ale takových nebylo v původní matici mnoho.

Modifikace základní úlohy prokazují, že největšího nárůstu hodnoty kritéria bylo dosaženo v případech, kdy podniky byly rozděleny do dvou skupin a tyto skupiny se od sebe výrazně odlišovaly – čím výrazněji, tím vyšší nárůst hodnoty kritéria. Dalším poznatkem je, že pokles počtu podniků výsledky destabilizuje a dochází k rozsáhlým fluktuacím v neinformativnějších proměnných (viz příloha). Ukazuje se tak, že je třeba dodržet doporučený minimální poměr počtu vzorků (podniků) k počtu příznaků (proměnných) (viz Jain a Chandrasekar, 1982).

Závěr

Hledání faktorů, které ovlivňují výkonnost, a tedy i úspěšnost podniku, patří k nejzajímavějším, a současně i nejobtížnějším úlohám v podnikové ekonomice. Příspěvek seznamuje s možnostmi, které při hledání faktorů konkurenceschopnosti českých podniků nabízí využití algoritmu sekvenčního dopředného plovoucího výběru.

Naším cílem však nebylo prezentovat a podrobně diskutovat nalezené faktory konkurenceschopnosti českých podniků. Pro tyto účely v textu odkazujeme na jiné texty, které náš výzkumný tým publikoval (především Blažek a kol., 2008). Vzhledem k metodickému zaměření našeho textu se spíše snažíme o představení zmíněné metody a jejímu srovnání s dosud využívanými a publikovanými metodami. I výsledky našich experimentů potvrzují, že metody rozpoznávání obrazů mohou značně zvýšit validitu nalezených faktorů.

¹⁸ Následující závěry nejsou založeny pouze na zde prezentovaných výsledcích, ale na podstatně rozsáhlejší souboru experimentů (v to zahrnující změny vstupních proměnných nebo jejich počtu, změny vstupních podniků nebo jejich počtu a pořadí, různá vyjádření finanční výkonnosti podniku atp.). Z důvodu omezeného prostoru uvádíme pouze nejvýznamnější výsledky.

¹⁹ Hodnoty všech proměnných byly normovány do intervalu $(0 - 1)$.

Zatímco bivariačními analýzami byly nalezené míry souvislosti zřídka silnější než 0,2; pomocí statistického rozpoznávání obrazů jsme našli kombinaci faktorů determinující zařazení téměř 65 % do správné skupiny dle hospodářské výkonnosti.

Použití metody s sebou pochopitelně přináší některá úskalí. Jedná se například o volbu nejvhodnějšího vyjádření finanční výkonnosti podniků či o převod proměnných na vyhovující škály, případně agregování více proměnných do jedné apod. Klíčovým se též ukázalo tzv. prokletí dimenzionality, tedy příliš nízký počet vzorků pro daný počet příznaků (poměr počtu podniků k počtu proměnných), nebo naopak, k danému počtu vzorků příliš vysoký počet příznaků. Nedodržení minimálního doporučeného poměru 10:1 vedlo ke značným fluktuacím neinformativnějších proměnných. Nicméně i tak máme zato – a náš článek by k tomuto názoru měl přispět, že metody rozpoznávání obrazů jsou v této problematice velkým přínosem. Jejich aplikace navíc není regionálně omezená, neboť zde prezentovaná metodika je bez větších obtíží aplikovatelná i v jiných zemích než v České republice.

Literatura

- ABOR, J. – BIEKPE, N. (2007): Corporate Governance, Ownership Structure and Performance of SMEs in Ghana: Implications for Financing Opportunities. *Corporate Governance*, 3, č. 7, s. 288 – 300.
- ALLOUCHE, J. – LAROCHE, P. (2005): A Meta-analytical Investigation of The Relationship Between Corporate Social And Financial Performance. *Revue de Gestion des Ressources Humaines*, 57, s. 18 – 41.
- ANDREWS, R. – BOYNE, G. A. (2010): Capacity, Leadership, and Organizational Performance: Testing the Black Box Model of Public Management. *Public Administration Review*, May/June, s. 443 – 454.
- ARTHUR, J. B. (1994): Effects of Human Resource Systems on Manufacturing Performance and Turnover. *Academy of Management Journal*, 37, č. 3, s. 670 – 687.
- ARTIACH, T. – LEE, D. – NELSON, D. – WALKER, J. (2010): The Determinants of Corporate Sustainability Performance. *Accounting and Finance*, 50, s. 31 – 51.
- BLAŽEK, L. a kol. (2007): Konkurenční schopnost podniků: (primární analýza výsledků empirického šetření). 1. vyd. Brno: Masarykova univerzita, 302 s. ISBN 9788021044562.
- BLAŽEK, L. a kol. (2008): Konkurenční schopnost podniků. Analýza faktorů hospodářské úspěšnosti. Brno: Masarykova univerzita, 210 s. ISBN 978-80-210-4734-1.
- CELLINI, R. – SOCI, A. (2002): Pop Competitiveness. *BNL Quarterly Review*, 220, s. 71 – 101.
- DE VAUS, D. (2002): Analyzing Social Science Data. 50 Key Problems in Data Analysis. 1. vyd. London: Sage, 401 s. ISBN 0-7619-5937-8.
- FERRI, F. J. – PUDIL, P. – HATEF, M. – KITTLER, J. (1994): Comparative Study of Techniques for Large-scale Feature Selection. In: GELSEMA, E. S. – KANAL, L. N. (eds): *Pattern Recognition in Practice IV*. Amsterdam: Elsevier Science B. V.
- FREEMAN, R. E. (1984): *Strategic Management: A Stakeholder Approach*. Boston: Harper Collins.
- HAENECKE, H. (2002): Methodenorientierte Systematisierung der Kritik an der Erfolgsfaktorenforschung. *Zeitschrift für Betriebswirtschaft*, č. 2.
- HANSEN, G. S. – WERNERFELT, B. (1989): Determinants of Firm Performance: The Relative Importance of Economic and Organizational Factors. *Strategic Management Journal*, 10, s. 399 – 411.
- HUČKA, M. (2005): Správa společnosti a konkurenceschopnost: Existuje vazba? *Acta Oeconomica Pragensia*, 13, č. 4, s. 154 – 166.

- HUSELID, M. A. (1995): The Impact of Human Resource Management Practices on Turnover, Produktivity, and Corporate Financial Performance. *Academy of Management Journal*, 38, č. 3, s. 635 – 672.
- CHAUVIN, K. W. – HIRSCHEY, M. (1993): Advertising, R & D Expenditures and the Market Value of the Firm. *Financial Management*, 22, č. 4.
- JAIN, A. K. – CHANDRASEKAR, B. (1982): Dimensionality and Sample Size Considerations in Pattern Recognition Practice. *Handbook of Statistics*, 2, s. 835 – 855.
- KISLINGEROVÁ, E. – SYNEK, M. – NEUMAIEROVÁ, I. – KLEČKA, J. (2006): Nové metody měření a řízení výkonnosti firem v globalizující se ekonomice. [Sborník výstupů grantového projektu GA ČR.] Praha: Oeconomica, 215 s. ISBN 80-245-1011-1.
- KITA, J. – KONŠTIAK, P. – KITA, P. (2008): Miesto nákupu v trhovej orientácii podnikov v Slovenskej republike. *Ekonomický časopis/Journal of Economics*, 56, č. 4, s. 370 – 383.
- KLOUDOVÁ, J. – DOBEŠ, K. – STRÍŽ, P. (2005): Faktory ovlivňující situační postavení firmy na trhu. *Ekonomie a Management*, 8, č. 1, s. 134 – 141.
- KUDO, M. – SKLANSKY, J. (2000): Comparison of Algorithms that Select Features for Pattern Classifiers. *Pattern Recognition*, 33, s. 25 – 41.
- LIPOVATZ, D. – MANDARAKA, M. – MOURELATOS, A. (2000): Multivariate Analysis for the Assessment of Factors Affecting Industrial Competitiveness: The Case of Greek Food and Beverage Industries. *Applied Stochastic Models in Business and Industry*, 16, s. 85 – 98.
- MARGOLIS, J. D. – WALSH, J. P. (2002): Misery Loves Companies: Whither Social Initiatives by Business? [Working Paper.] Boston: Harvard University.
- MATHIS, J. – MAZIER, J. – RIVAUD-DANSET, D. (1988): *La Competitivité Industrielle*. Paris: Dunod.
- MICHALET, C. (1981): *Competitiveness and Internationalisation*. Paris: Mimeo OCDE.
- MOLINA, M. A. – BARRIOS DEL PINO, I. – RODRIGUEZ, A. C. (2004): Industry, Management Capabilities and Firms' Competitiveness: An Empirical Contribution. *Managerial and Decision Economics*, 25, s. 265 – 281.
- NEČADOVÁ, M. – BREŇOVÁ, L. – PRŮCHA, Š. (2005): Konkurenceschopnost firem zpracovatelského průmyslu – komparace výsledků primárního výzkumu a cílů Národního programu reformy ČR. *Acta Oeconomica Pragensia*, 15, č. 2, s. 55 – 70.
- NEMCOVÁ, E. (2006): Priemyselná politika Európskej únie a konkurenceschopnosť. *Ekonomický časopis/Journal of Economics*, 54, č. 8, s. 803 – 815.
- POST, J. E. – PRESTON, L. E. – SACHS, S. (2002): Managing the Extended Enterprise: The New Stakeholder view. *California Management Review*, 45, č. 1, s. 6 – 28.
- POWELL, T. C. – DENT-MICALLEF, A. (1997): Information Technology As Competitive Advantage: The Role of Human, Business, and Technology Resources. *Strategic Management Journal*, 18, č. 5, s. 375 – 405.
- PUDIL, P. (2008): Teoretické základy využití statistických metod rozpoznávání pro vyhledání faktorů hospodářské úspěšnosti podniků. In: *Konkurenční schopnost podniků, analýza faktorů hospodářské úspěšnosti*. 1. vyd. Brno: Masarykova univerzita, 210 s. ISBN 9788021047341.
- PUDIL, P. – NOVOVIČOVÁ, J. – KITTLER, J. (1994): Floating Search Methods in Feature Selection. *Pattern Recognition Letters*, 15, s. 1119 – 1125.
- RAJČÁNIOVÁ, M. (2006): Konkurenceschopnosť a možnosti jej merania. In: *International Scientific Days: Competitiveness in the EU – Challenge for the V4 Countries*. Nitra: SPU, s. 310 – 313.
- RUMELT, P. R. – WENSLEY, R. (1981): In Search of the Market Share Effect. *Academy of Management Proceedings*, s. 1 – 5.
- SLANÝ, A. a kol. (2006). *Konkurenceschopnost České ekonomiky: (vývojové trendy)*. 1. vyd. Brno: Masarykova univerzita, 377 s. ISBN 80-210-4157-9.
- SMITH, K. G. – GUTHRIE, J. P. – CHEN, M. (1989): Strategy, Size and Performance. *Organization Studies*, 10, č. 1, s. 63 – 81.
- SOMOL, P. – PUDIL, P. (2000). Oscillating Search Algorithms for Feature Selection. *Proceedings of the 15th International Conference on Pattern Recognition*. Los Alamitos: IEEE Computer Society.
- SUCHÁNEK, P. – ŠPALEK, J. (2007): Financial Measuring of the Czech Republic Companies Competitiveness. *The Business Review*, 9, č. 1, s. 78 – 85.

- ŠÍŠKA, L. (2008): Analýza finanční výkonnosti respondentů empirického šetření CVKS. [Working paper 10/2008.] Brno: Centrum výzkumu konkurenční schopnosti české ekonomiky, 34 s. Dostupné na: <<http://is.muni.cz/do/1456/soubory/oddeleni/centrum/papers/wp2008-10.pdf>>.
- YILMAZ, C. – ALPKAN, L. – ERGUN, E. (2005): Cultural Determinants of Customer- and Learning-oriented Value Systems and their Joint Effects on Firm Performance. *Journal of Business Research*, 58, s. 1340 – 1352.

Příloha

Fluktuace vybíraných proměnných u lokálních maxim

Číslo proměnné	Název	1NN Lokální max.	3NN Lokální max.	5 NN Lokální max.
0	Důležitost věřitelů			
1	Důležitost státu	x		
2	Důležitost komunity			
3	Rozpětí řízení		x	x
4	Součást koncernu		x	x
5	Typ vlastnické struktury			x
6	Původ vlastníka	x	x	x
7	Vlastníci jako součást top managementu	x	x	
8	Stav hmotného majetku			
9	Softwarové aplikace – Moduly MRP	x	x	x
10	Softwarové aplikace – Moduly CRM			
11	Softwarové aplikace – Moduly SCM	x		
12	Velikost podniku	x	x	x
13	Podíl pracovníků s VŠ vzděláním			
14	Podíl technicko-hospodářských pracovníků			x
15	Fluktuace		x	x
16	Podíl pohyblivé složky mzdy		x	x
17	Výše prostředků na zaměstnanecké výhody	x	x	x
18	Výše prostředků na vzdělávání	x	x	
19	Podnikatelská strategie	x	x	x
20	Stabilita odběratelů			
21	Podíl exportu na tržbách			
22	Specifičnost produktů	x		
23	Stabilita dodavatelů			
24	Podíl importu na dodávkách			x
25	Specifičnost dodávek		x	
26	Důležitost kritérií výběru dodavatele		x	
27	Důležitost certifikátu jakosti při výběru dodavatele			
28	Důležitost souladu politiky CSR dodavatele při výběru dodavatele			
29	Míra angažovanosti v CSR	x		
30	Vlastnictví kodexů		x	
31	Certifikát ISO 14000	x	x	x
32	Certifikát OHSAS 18001	x		
33	Kraj			
34	Právní forma podnikání	x	x	x
35	Obor podnikání		x	x
36	Přidaná hodnota na zaměstnance		x	x

Pramen: Vlastní zpracování.