

Vysoká škola technická a ekonomická

v Českých Budějovicích

Pokročilé metody analýz a modelování

Studijní opora pro kombinovanou formu studia

Garant: doc. RNDr. Jaroslav Stuchlý, CSc.

Ústav podnikové strategie

Katedra ekonomiky

Autoři: doc. RNDr. Jaroslav Stuchlý, CSc.,

Ing. Marek Vokoun, Ph.D.

Obsah

1	Anotace.....	5
2	Příprava na přednášky	10
2.1	Představení pokročilých metod analýz a modelování v ekonomickém aplikovaném výzkumu a v podnikové praxi. Přednosti a nedostatky popisné analýzy dat, využívání vizualizovaných dat.....	10
2.2	Představení pravidel logiky a matematického zápisu množin, možnosti zapisování modelů, pravděpodobnost a statistické rozdělení dat.....	13
2.3	Představení parametrického a neparametrického přístupu v modelování. Pravidla regresní analýzy pro průřezová data	17
2.4	Pravidla vícenásobné regresní analýzy pro časové řady	20
2.5	Testování předpokladů konvenční asymptotické teorie pro odhady časových řad metodou nejmenších čtverců, transformace časových řad, časové řady typu $I(0)$ a $I(1)$.	23
2.6	Modelování časových řad pomocí regresní analýzy	26
2.7	Časové řady v ekonomickém výzkumu a podnikové praxi, možnosti pokročilých metod	29
2.8	Pokročilé modelování časových řad pomocí regresní analýzy	32
2.9	Základní principy teorie fuzzy množin a fuzzy logiky	36
2.10	Fuzzy systémy, adaptace ve fuzzy systémech.....	39
2.11	Uvedení příkladů aplikací fuzzy logiky v rozhodování, jako je např. manažerské a investiční rozhodování, predikce atd.....	41
2.12	Modelování časových řad pomocí fuzzy logiky	44
2.13	Umělé neuronové sítě (UNS): Seznámení se se základními pojmy v oblasti umělých neuronových sítí, uvedení pojmu perceptron	47
2.14	Vícevrstvá neuronová síť a její parametry	50
2.15	Uvedení příkladů aplikace UNS pro investiční rozhodování, odhady cen výrobků a množstevní odhady.....	54

2.16	Uvedení příkladů aplikace UNS pro odhad cen nemovitostí, oceňování bonity klienta atd.	57
2.17	Modelování časových řad pomocí UNS	60
2.18	Pokročilé modelování časových řad pomocí UNS.....	63
2.19	Genetické algoritmy (GA): Seznámení se základy genetiky, analogií mezi přírodou a matematickým popisem, umožňující řešení problémů rozhodování	65
2.20	Uvedení použití GA v oblasti optimalizace široké palety problémů – optimalizace investiční strategie, řízení výroby, řezných plánů, aproximace křivek, řešení problému obchodního cestujícího, využití shlukové analýzy apod.	68
2.21	Optimalizační úlohy v GA, schémata, cykly, smyčky	71
2.22	Modelování časových řad pomocí GA.....	74
2.23	Teorie chaosu: Teorie pojednává o možnosti lepšího popisu ekonomických jevů, než je tomu u klasických metod. Je objasněn pojem chaos a řád, fraktál	77
2.24	Uvedení využití teorie chaosu u měřeného sledovaného systému, uvedení příkladů z ekonomického výzkumu a podnikové praxe	81
2.25	Uvedení pojmu datamining, textmining a definování cílů, výběr techniky modelování, zdroje a příprava dat, tvorba modelů, jejich ověření, vyhodnocení, implementace a údržba.....	84
2.26	Uvedení příkladů použití dataminingu v Business Intelligence a pro volbu strategie spolupráce se zákazníkem, direct mailing apod.	87
3	Příprava na semináře	89
3.1	Popisná statistika a analýza dat, vizualizace dat, příklady pravidel logiky a matematického zápisu množin, možnosti zapisování modelů	89
3.2	Počet pravděpodobnosti a statistické rozdělení dat. Pravidla regresní analýzy pro průřezová data	93
3.3	Testování předpokladů konvenční asymptotické teorie pro odhady metodou nejmenších čtverců pro průřezová data a pro časové řady.....	96
3.4	Modelování časových řad pomocí regresní analýzy	100

3.5	Pokročilé modelování časových řad pomocí regresní analýzy	103
3.6	Průběžný test	107
3.7	Procvičování základních principů teorie fuzzy množin a fuzzy logiky	112
3.8	Modelování časových řad pomocí fuzzy logiky	114
3.9	Procvičování základních principů umělých neuronových sítí	116
3.10	Modelování časových řad pomocí UNS	118
3.11	Pokročilé modelování časových řad pomocí UNS.....	120
3.12	Procvičování základních principů genetických algoritmů	122
3.13	Modelování časových řad pomocí GA.....	124

1 Anotace

Období	1. semestr/ 1. ročník
Název předmětu	Pokročilé metody analýz a modelování
Vyučovací jazyk	český
Garant předmětu	doc. RNDr. Jaroslav Stuchlý, CSc.
Garanční ústav	Ústav podnikové strategie
Katedra	Katedra ekonomiky
Vyučující (přednášející)	doc. RNDr. Jaroslav Stuchlý, CSc.
Vyučující (cvičící)	Ing. Marek Vokoun, Ph.D.
Ukončení předmětu	zkouška
Poznámka k ukončení	docházka na semináře 70 % včetně dalších poznámek garanta předmětu
Rozsah	4/2
Počet kreditů	8
Cíle předmětu výstupy z učení	Cílem předmětu je seznámení se s některými pokročilými metodami analýz a technikami modelování za účelem podpory rozhodování v podniku. Předmět si rovněž klade za cíl vysvětlit princip těchto teorií, naučit studenta pracovat s těmito teoriemi a jejich aplikací.
Výstupy z učení	Po úspěšném absolvování předmětu student: 16.1 zná základní pojmy a pravidla fuzzy logiky, 16.2 rozumí fungování umělých neuronových struktur, 16.3 rozumí genetickým algoritmům, 16.4 zná teorii chaosu, 16.5 rozumí dataminingu a textminingu, 16.6 modeluje časové řady pomocí fuzzy logiky, 16.7 modeluje časové řady pomocí umělých neuronových struktur, 16.8 modeluje časové řady pomocí genetických algoritmů, 16.9 zná aplikace modelů časových řad v podnikové praxi.
Osnova předmětu	<u>Přednášky</u> 1. Představení pokročilých metod analýz a modelování v ekonomickém aplikovaném výzkumu a v podnikové praxi. Přednosti a nedostatky popisné analýzy dat, využívání vizualizovaných dat. (16.9) 2. Představení pravidel logiky a matematického zápisu množin, možnosti

	<p>zapisování modelů, pravděpodobnost a statistické rozdělení dat. (16.1, 16.2, 16.3)</p> <p>3. Představení parametrického a neparametrického přístupu v modelování. Pravidla regresní analýzy pro průřezová data. (16.6, 16.7, 16.8)</p> <p>4. Pravidla vícenásobné regresní analýzy pro časové řady. (16.6, 16.7, 16.8)</p> <p>5. Testování předpokladů konvenční asymptotické teorie pro odhady časových řad metodou nejmenších čtverců, transformace časových řad, časové řady typu $I(0)$ a $I(1)$. (16.6, 16.7, 16.8)</p> <p>6. Modelování časových řad pomocí regresní analýzy. (16.6, 16.7, 16.8)</p> <p>7. Časové řady v ekonomickém výzkumu a podnikové praxi, možnosti pokročilých metod. (16.6, 16.7, 16.8)</p> <p>8. Pokročilé modelování časových řad pomocí regresní analýzy. (16.6, 16.7, 16.8)</p> <p>9. Základní principy teorie fuzzy množin a fuzzy logiky. (16.1)</p> <p>10. Fuzzy systémy, adaptace ve fuzzy systémech. (16.1)</p> <p>11. Uvedení příkladů aplikací fuzzy logiky v rozhodování, jako je např. manažerské a investiční rozhodování, predikce atd. (16.1)</p> <p>12. Modelování časových řad pomocí fuzzy logiky. (16.1, 16.6)</p> <p>13. Umělé neuronové sítě (UNS): Seznámení se se základními pojmy v oblasti umělých neuronových sítí, uvedení pojmu perceptron. (16.2)</p> <p>14. Vícevrstvá neuronová síť a její parametry. (16.2)</p> <p>15. Uvedení příkladů aplikace UNS pro investiční rozhodování, odhady cen výrobků a množstevní odhady. (16.7)</p> <p>16. Uvedení příkladů aplikace UNS pro odhad cen nemovitostí, oceňování bonity klienta atd. (16.7)</p> <p>17. Modelování časových řad pomocí UNS. (16.2, 16.7)</p> <p>18. Pokročilé modelování časových řad pomocí UNS. (16.2, 16.7)</p> <p>19. Genetické algoritmy (GA): Seznámení se se základy genetiky, analogií mezi přírodou a matematickým popisem, umožňující řešení problémů rozhodování. (16.3)</p> <p>20. Uvedení použití GA v oblasti optimalizace široké palety problémů – optimalizace investiční strategie, řízení výroby, řezných plánů, aproximace křivek, řešení problému obchodního cestujícího, využití shlukové analýzy apod. (16.3)</p> <p>21. Optimalizační úlohy v GA, schémata, cykly, smyčky. (16.3)</p> <p>22. Modelování časových řad pomocí GA. (16.3, 16.8)</p> <p>23. Teorie chaosu: Teorie pojednává o možnosti lepšího popisu ekonomických jevů, než je tomu u klasických metod. Je objasněn pojem chaos a řád, fraktál. (16.4)</p> <p>24. Uvedení využití teorie chaosu u měřeného sledovaného systému, uvedení příkladů z ekonomického výzkumu a podnikové praxe. (16.4)</p> <p>25. Uvedení pojmu datamining, textmining a definování cílů, výběr techniky modelování, zdroje a příprava dat, tvorba modelů, jejich ověření, vyhodnocení, implementace a údržba. (16.5)</p> <p>26. Uvedení příkladů použití dataminingu v Business Intelligence a pro volbu strategie spolupráce se zákazníkem, direct mailing apod. (16.4)</p>
--	---

	<p><u>Semináře</u></p> <ol style="list-style-type: none"> 1. Popisná statistika a analýza dat, vizualizace dat, příklady pravidel logiky a matematického zápisu množin, možnosti zapisování modelů. (16.6, 16.7, 16.8) 2. Počet pravděpodobnosti a statistické rozdělení dat. Pravidla regresní analýzy pro průřezová data. (16.6, 16.7, 16.8) 3. Testování předpokladů konvenční asymptotické teorie pro odhady metodou nejmenších čtverců pro průřezová data a pro časové řady. (16.6, 16.7, 16.8) 4. Modelování časových řad pomocí regresní analýzy. (16.6, 16.7, 16.8) 5. Pokročilé modelování časových řad pomocí regresní analýzy. (16.6, 16.7, 16.8) 6. Průběžný test. (16.6, 16.7, 16.8) 7. Procvičování základních principů teorie fuzzy množin a fuzzy logiky. (16.1, 16.6) 8. Modelování časových řad pomocí fuzzy logiky. (16.1, 16.6) 9. Procvičování základních principů umělých neuronových sítí. (16.2, 16.7) 10. Modelování časových řad pomocí UNS. (16.2, 16.7) 11. Pokročilé modelování časových řad pomocí UNS. (16.2, 16.7) 12. Procvičování základních principů genetických algoritmů. (16.3, 16.8) 13. Modelování časových řad pomocí GA. (16.3, 16.8) 																									
Organizační formy výuky	přednáška, seminář																									
Komplexní výukové metody	frontální výuka projektová výuka skupinová výuka – kooperace brainstorming kritické myšlení samostatná práce – individuální nebo individualizovaná činnost výuka podporovaná multimediálními technologiemi																									
Studijní zátěž	<table border="1"> <thead> <tr> <th rowspan="2">Aktivita</th> <th colspan="2">Počet hodin za semestr</th> </tr> <tr> <th>Prezenční forma</th> <th>Kombinovaná forma</th> </tr> </thead> <tbody> <tr> <td>Příprava na průběžný test</td> <td>18</td> <td>32</td> </tr> <tr> <td>Příprava na přednášky</td> <td>26</td> <td>0</td> </tr> <tr> <td>Příprava na seminář, cvičení, tutoriál</td> <td>11</td> <td>90</td> </tr> <tr> <td>Příprava seminární práce</td> <td>25</td> <td>0</td> </tr> <tr> <td>Průběžný test</td> <td>2</td> <td>2</td> </tr> <tr> <td>Závěrečný test</td> <td>2</td> <td>2</td> </tr> </tbody> </table>			Aktivita	Počet hodin za semestr		Prezenční forma	Kombinovaná forma	Příprava na průběžný test	18	32	Příprava na přednášky	26	0	Příprava na seminář, cvičení, tutoriál	11	90	Příprava seminární práce	25	0	Průběžný test	2	2	Závěrečný test	2	2
Aktivita	Počet hodin za semestr																									
	Prezenční forma	Kombinovaná forma																								
Příprava na průběžný test	18	32																								
Příprava na přednášky	26	0																								
Příprava na seminář, cvičení, tutoriál	11	90																								
Příprava seminární práce	25	0																								
Průběžný test	2	2																								
Závěrečný test	2	2																								

	Účast na přednáškách	52	0
	Účast na semináři/cvičeních/tutoriálu/exkurzi	26	24
	Příprava na závěrečný test	46	58
	Celkem:	208	208
Metody hodnocení a jejich poměr	závěrečný test 70 % seminární práce 20 % průběžný test 10 %		
Podmínky pro úspěšné absolvování předmětu včetně jejich hodnocení	Celková klasifikace předmětu, tj. body za závěrečný test (70 - 0) + body z průběžného hodnocení (30 - 0): A 100 – 90, B 89,99 – 84, C 83,99 – 77, D 76,99 – 73, E 72,99 – 70, FX 69,99 – 30, F 29,99 - 0.		
Informace učitele	Účast na výuce ve všech formách řeší samostatná vnitřní norma VŠTE (Evidence docházky studentů na VŠTE). Pro studenty prezenční formy studia je na seminářích a cvičeních povinná 70% účast.		
Literatura povinná	<p>COUFAL, J., 2014. <i>Logika a logické myšlení</i>. [Praha]: Vysoká škola ekonomie a managementu. ISBN 978-80-87839-59-1.</p> <p>DOSTÁL, P., 2015. <i>Soft computing v podnikatelství a veřejné správě</i>. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4.</p> <p>HAMPEL, D., V. BLAŠKOVÁ a L. STŘELEČEK, 2016. <i>Ekonometrie 2. 3. přeprac. vyd.</i> Brno: Mendelova univerzita. ISBN 978-80-7509-427-8.</p> <p>VOCHOZKA, M. et al., 2017. <i>Využití neuronových sítí při komplexním hodnocení podniků</i>. Praha: C. H. Beck. ISBN 978-80-7400-642-5.</p>		
Literatura doporučená	<p>DOSTÁL, P., 2012. <i>Pokročilé metody rozhodování v podnikatelství a veřejné správě</i>. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-798-7.</p> <p>FONG, L. Y. a K. Y. SZETO, 2001. Rules extraction in short memory time series using genetic algorithms. <i>The European Physical Journal B-Condensed Matter and Complex Systems</i>. 20(4), 569-572. ISSN 1434-6028.</p>		
Webové stránky	<p>https://is.muni.cz/el/1431/podzim2007/Bi3101/um/skripta.pdf (Učební text k matematickému modelování)</p> <p>http://www1.osu.cz/~zacek/mopop/mopop.pdf (Učební text o modelování podnikových procesů)</p>		
Publikační činnost	<p><u>Garant předmětu a přednášející (doc. RNDr. Jaroslav Stuchlý, CSc.)</u></p> <p>CAHA, Z., J. STUHLÝ, M. RUSCHAK a M. ŠUTA, 2017. MANAŽERSKÉ VZDĚLÁVÁNÍ V MALÝCH A STŘEDNĚ VELKÝCH FIRMÁCH V ČESKÉ REPUBLICI. <i>Mladá věda</i>. 5(9), 1-</p>		

	<p>12. ISSN 1339-3189.</p> <p>STUCHLÝ, J., 2015. <i>Statistické analýzy dat: vysokoškolská učebnice</i>. České Budějovice: Vysoká škola technická a ekonomická v Českých Budějovicích. ISBN 978-80-7468-087-8.</p> <p>ZEMAN, R. a J. STUCHLÝ, 2014. Predicting stock prices in Bank VIG. <i>Logi</i>. 5(2), 116-129. ISSN 1804-3216.</p> <p>ŠOLC, L. a J. STUCHLÝ, 2013. Preference kvality nebo ceny při nákupech spotřebního zboží. <i>Littera Scripta</i>. 6(1), 158-169. ISSN 1802-503X.</p> <p>ZEMAN, R. a J. STUCHLÝ, 2013. Prognozyje modeli dlja akcij «Erste Bank». <i>Vestnik Astrachanskogo gosudarstvennogo techničeskogo universiteta</i>. 2013(2), 104-111. ISSN 2073-5537.</p> <p><u>Cvičící (Ing. Marek Vokoun, Ph.D.)</u></p> <p>VOKOUN, M., 2016. Innovation behaviour of firms in a small open economy: the case of the Czech manufacturing industry. <i>Empirica</i>. 43(1), 111-139. ISSN 0340-8744.</p> <p>NOVÁK, V., M. VOKOUN, F. STELLNER a M. VOCHOZKA, 2016. Institutional analysis of contemporary regional labour market in the Czech Republic. <i>E+M Ekonomie a Management</i>. 19(3), 4-19. ISSN 1212-3609.</p> <p>VOKOUN, M. a F. STELLNER, 2015. Czech economic historians and interdisciplinary approach. <i>Économies et Sociétés. Série "Histoire Économique Quantitative"</i>. 50(6), 857-875. ISSN 0013-0567.</p> <p>STELLNER, F. a M. VOKOUN, 2014. Internet, social sciences and humanities. <i>Human Affairs: Postdisciplinary Humanities & Social Sciences Quarterly</i>. 24(4), 492-510. ISSN 1210-3055.</p> <p>VOKOUN, M., 2013. Cliometrie. In: <i>Nové přístupy k metodologii hospodářských a politických dějin</i>. Praha: Setoutbooks.cz, 127-152. ISBN 978-80-86277-77-6.</p>
Témata diplomových prací	<p>Modelování stochastických procesů v podnikové ekonomice a financích</p> <p>Modelování procesů na finančních trzích</p> <p>Složitější modely teorie zásob</p>

2 Příprava na přednášky

2.1 Představení pokročilých metod analýz a modelování v ekonomickém aplikovaném výzkumu a v podnikové praxi. Přednosti a nedostatky popisné analýzy dat, využívání vizualizovaných dat

Klíčová slova

Statistika, ekonomie, přehledové statistiky, měření

Cíle kapitoly

Cílem je pochopení významu pokročilých metod analýz a modelování v ekonomickém aplikovaném výzkumu a v podnikové praxi. Cílem je pochopit přednosti a nedostatky popisné analýzy dat a využívání vizualizovaných dat.

Výstupy z učení

- 16.9 zná aplikace modelů časových řad v podnikové praxi

Abstrakt

Ekonometrie je součástí společenských věd, představuje měření v ekonomii, resp. je to kvantitativní ekonomická analýza s přímou vazbou na matematicko-statistické metody a obvykle neoklasickou ekonomii. Problémem ekonometrie je absence analýzy věrohodnosti použitých materiálů vzhledem ke kontextu a snaha parametrizovat a linearizovat ekonomické problémy, které jsou z povahy věcí komplexní strukturou (chaos, dynamické systémy, adaptivní systémy apod.). Součástí každého vědeckého článku je proto kapitola materiály a metody, případně data a proměnné. V těchto kapitolách pozorujeme přehledové statistiky (počet pozorování, průměr, směrodatná odchylka, minimum a maximum) a odkud data pocházejí. Avšak neměla by chybět i analýza toho, co daná data doopravdy popisují, jak věrohodná jsou a jak reprezentují populaci. Tyto povinná kritika dat, pramenů a teorií je typická pro historiografii, méně už pro ekonomii.

Proměnné vstupující do ekonometrického modelu mohou být transformované, očištěné (prázdná pozorování), podmíněné (nenulové hodnoty, jen muži apod.), využívající data k tvorbě nových proměnných (interakce, dummy proměnné, kvadratické proměnné, nepozorované proměnné). Transformace dat je obvykle založena na zlepšení interpretace zkoumané problematiky. Např. logaritmická transformace – přirozený logaritmus $\ln(x)$ – se využívá pro procentní interpretaci a pro data, kde je vhodné sledovat pokles či růst. Dále se využívá transformace na jednotku (na kg, na zaměstnance...), procentní podíl na celku, kvadratická transformace (výnosy z rozsahu), rozdíly (růsty mezi obdobími, vůči průměru, od konstanty) a inverzní hyperbolický sinus (podobně jako logaritmická transformace, avšak s možností použít i nulové hodnoty, např. u příjmových studií).

Mezi obvyklé popisné statistiky a analýzy patří: (1) jedna proměnná – střední hodnoty, odchylky, podobnost vůči standardnímu rozdělení, (2) dvě proměnné – vzájemný vztah, podobnost, rozdílnost, korelace, přímková regrese (3) tři a více proměnných – regrese průřezových dat, kontingenční tabulky, regrese panelových dat, regrese časové řady.

Průměrná hodnota se zapisuje jako: $\bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$. Např. jde o pozorování výšky (znak y) lidí (i), kteří reprezentují výběr (n) z populace (N). Označení proměnné se stříškou \hat{y}_i je v ekonometrii vyhrazeno pro označení modu, avšak častěji se využívá pro označení odhadnutých hodnot ekonometrického modelu. Např. pro odhad reziduální složky je obvyklé označení \hat{u}_i . Rezidua modelu jsou označována také jako „u_hat“ (u stříška), na této složce regresního modelu testujeme nedostatky, které by model mohl mít, tj. neplní některé Gauss-Markovy podmínky.

V ekonomickém výzkumu je obvyklé využívat tabulky. Grafy jsou obvykle součástí přílohy, avšak představují důležitý prvek analýzy dat. Histogram pomůže při pohledu na normalitu a rozdělení dané proměnné. Korelogram je pro analýzu autokorelace v časové řadě, zde pohledem a obvykle i doprovodnou statistikou (Durbinův-Watsonův alternativní test) zjišťujeme, zda se mezi pozorováními vyskytuje silná závislost (korelační koeficient blízký jedničce). Např. nabývá-li koeficient utilizace (u proměnné y) hodnoty 0,79 a v předchozím období hodnoty 0,8, resp. hodnoty koeficientů mezi obdobími budou vždy blízké hodnotě 0,8, vztah bude podobný rovnici $y = 1 * y_{t-1} + u$, kde číslo 1 je odhadnutý beta koeficient a y_{t-1} je hodnota y v předchozím období. Mezi hodnotami (a 0,79 a 0,80) a následně i dalšími páry (0,8 a 0,81) je v čase silná závislost. Tato autokorelace vlastnost časových řad, která je sice využitelná, ale typicky nebezpečná, protože svádí k naivní

interpretaci (zdánlivá regrese). Tyto řady je možné zpracovávat a interpretovat při zapojení dalších pokročilých metod, mezi které patří Box Jenkinsonova metoda, fuzzy logika, genetické algoritmy a neuronové sítě.

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 1-6)

VOCHOZKA, M. et al., 2017. *Využití neuronových sítí při komplexním hodnocení podniků*. Praha: C. H. Beck. ISBN 978-80-7400-642-5. (s. 1-10)

Kontrolní otázky

1. Co je podmínkou kvalitního ekonometrického výzkumu?
2. Jaké pokročilé metody ekonomických analýz znáte?
3. Jaké existují datové struktury?
4. Jaké existují vizualizační pomůcky v ekonometrickém výzkumu?
5. Jaké ekonometrické metody se používají v podnikové praxi?
6. Jaké jsou přednosti popisné statistiky?
7. Jaké jsou nedostatky popisné statistiky?
8. K čemu slouží transformace dat?
9. Po jaká data je vhodná logaritmická transformace dat?
10. K čemu slouží kvadratická transformace dat?

Odkaz na praktickou část

3.1 Popisná statistika a analýza dat, vizualizace dat, příklady pravidel logiky a matematického zápisu množin, možnosti zapisování modelů

2.2 Představení pravidel logiky a matematického zápisu množin, možnosti zapisování modelů, pravděpodobnost a statistické rozdělení dat

Klíčová slova

Množina, operace, axiom, věta, důkaz

Cíle kapitoly

Cílem kapitoly je pochopení důležitosti pravděpodobnostního rozdělení, výrokové logiky, množinového zápisu a výstavby matematického vyjadřování (axiómy, definice, věty, důkazy).

Výstupy z učení

- 16.1 zná základní pojmy a pravidla fuzzy logiky
- 16.2 rozumí fungování umělých neuronových struktur
- 16.3 rozumí genetickým algoritmům

Abstrakt

Pro pochopení tzv. statistických rozdělení, křivek a funkcí používáme křivku Gaussovu, Studentova t rozdělení, Fisherova F rozdělení a dalších. Např. u Studentova t a normálního rozdělení pozorujeme asymptotický „blob“, který se nedotýká osy x v bodě $y=0$. Tyto křivky jsou pokaždé jinak „laděné“, závisí na střední hodnotě, směrodatné odchylce a počtu pozorování.

Centrální limitní teorém (Ljapunovova věta) se zabývá následující problematikou: Součet křivek rozdělení vzájemně nezávislých veličin x_i při dostatečně velkém počtu pozorování konverguje k normálnímu rozdělení, a to bez ohledu na tvar těchto rozdělení. Důležitá je také notace, kdy nějaký jev nastane pouze za určité podmínky. K označení „za předpokladu je svíslá čárka „|““. $P(C|B) = 0,5$ pak čteme jako: pravděpodobnost jevu C za vzhledem k podmínkám B je 0,5. Nebo $E(u | x_1, x_2, x_3) = c$ pak čteme jako: střední hodnota (označení E) reziduí (označení u) je rovna konstantě (označení c), za podmínky, že máme

nezávislé proměnné x_1 , x_2 a x_3 . V pravděpodobnosti je velmi důležitá Bayesova věta pro schopnost interpretovat odhadnuté skutečnosti. Ta udává, jak podmíněná pravděpodobnost nějakého jevu souvisí s opačnou podmíněnou pravděpodobností: $P(A|B) = [P(B|A)*P(A)]/P(B)$.

V logice výrokem nazýváme každou formalizovanou (vyslovenou, napsanou) myšlenku, o níž má smysl říci, že je buď pravdivá, nebo nepravdivá. Výrokem je pouze a jen věta oznamovací. Výroky označujeme obvykle velkými písmeny (např. výrok A, výroky B, výrok C) a jejich pravdivostní hodnotu (pravdivost, nebo nepravdivost) označíme symbolem (číslicí) 1 (pravda), resp. 0 (nepravda).

Hypotézou nazýváme výrok, jehož pravdivostní hodnotu neznáme a pokoušíme se ji odvodit logickými operacemi z jiných pravdivých výroků. Z výroků lze vytvářet nové výroky negací, užitím logických spojek a závorek.

Negace výroku A je zcela nový výrok, který vyjadřujeme slovy „neplatí A“. Tuto negaci označujeme symbolem non A. Je-li výrok A pravdivý, je výrok non A nepravdivý; je-li výrok A nepravdivý, je výrok non A pravdivý. Výroky A a non A se vzájemně vylučují, jejich pravdivostní hodnoty jsou opačné.

Logické spojky jsou čtyři: (1) Spojka „a“, kterou označujeme symbolem \wedge ; (2) spojka „nebo“ (\vee); (3) spojka „jestliže – pak“ (\Rightarrow); (4) spojka „právě tehdy, když“ (\Leftrightarrow).

Pomocí logických spojek vytváříme z výroků A, B nové výroky. (1) Konjunkce výroků A, B je výrok, který vyjadřujeme spojkou a; označujeme jej $A \wedge B$, což čteme A a B. Konjunkce je pravdivým výrokem, právě když oba výroky A a B jsou pravdivé. (2) Disjunkce výroků A, B je výrok, který vyjadřujeme spojkou nebo; označujeme jej $A \vee B$, což čteme A nebo B. Disjunkce je pravdivým výrokem, když je pravdivý aspoň jeden z výroků A, B. (3) Implikace výroků A, B (na daném pořadí záleží!) je výrok, který vyjadřujeme slovním spojením jestliže - pak. Označujeme jej $A \Rightarrow B$, což čteme: Jestliže platí A, pak platí B. Implikace je pravdivým výrokem při všech možných pravdivostních hodnotách výroků A, B kromě případu, kdy A je pravdivým a B nepravdivým výrokem. (4) Ekvivalence výroků A, B je výrok, který vyjadřujeme slovním spojením právě tehdy, když. Označujeme jej $A \Leftrightarrow B$, což čteme: A platí právě tehdy, když platí B. Ekvivalence je pravdivým výrokem, pokud výroky A, B jsou oba pravdivé nebo výroky A, B jsou oba nepravdivé.

Výroková forma (formule) představuje formalizovaný zápis, kdy používáme písmena a logické spojky. Tautologie je výroková formule, která nabývá ve všech případech pravdivostní hodnoty 1. Výroková forma $S(x)$ je slovní vyjádření, v němž se vyskytuje proměnná x . Toto vyjádření má tu vlastnost, že se stane výrokem, jen tehdy a tehdy, když proměnnou nahradíme prvky jisté množiny, kterou označujeme D , definiční obor výrokové formy $S(x)$. Množinu všech prvků, pro něž je $S(x)$ pravdivým výrokem, nazýváme oborem pravdivosti výrokové formy $S(x)$ a označujeme jej P . Z výrokových forem, podobně jako z výroků, lze tvořit nové výrokové formy pomocí logických spojek („a“, „nebo“, „jestliže-pak“, „právě tehdy, když“).

Kvantifikátory představují další možnost formalizovaného zápisu. Symbol \forall je tzv. obecný kvantifikátor. Výrok „Pro všechna reálná čísla x, y platí $x + y = y + x$ “, lze formalizovat pomocí obecného kvantifikátoru jako „ $\forall x, y \in \mathbf{R}: x + y = y + x$ “. Symbol \exists je tzv. existenční kvantifikátor. Výrok „Existují taková reálná čísla x, y , že platí $x > y$ “, lze formalizovat ve tvaru „ $\exists x, y \in \mathbf{R}: x > y$ “.

Studijní literatura

Povinná literatura

COUFAL, J., 2014. *Logika a logické myšlení*. [Praha]: Vysoká škola ekonomie a managementu. ISBN 978-80-87839-59-1. (s. 6-16)

HAMPEL, D., V. BLAŠKOVÁ a L. STŘELEČEK, 2016. *Ekonometrie 2*. 3. přeprac. vyd. Brno: Mendelova univerzita. ISBN 978-80-7509-427-8. (s. 8-10)

Kontrolní otázky

1. Co vyjadřuje Centrální limitní teorém?
2. Co v počtu pravděpodobnosti znamená Zákon velkých čísel?
3. Definujte výrok.
4. Co je to hypotéza?
5. Jaké existují logické spojky?
6. Co je to výroková forma?
7. Jaké se v logice používají kvantifikátory?
8. Co je to axióm?
9. Co je to definice?

10. Co je to věta?

Odkaz na praktickou část

3.1 Popisná statistika a analýza dat, vizualizace dat, příklady pravidel logiky a matematického zápisu množin, možnosti zapisování modelů

2.3 Představení parametrického a neparametrického přístupu v modelování. Pravidla regresní analýzy pro průřezová data

Klíčová slova

Distribuční funkce, histogram, rozptyl, podobnost, medián, průměr, odchylka

Cíle kapitoly

Cílem je pochopení a použití parametrického a neparametrického přístupu v modelování.

Cílem je pochopení pravidel regresní analýzy pro průřezová data.

Výstupy z učení

- 16.6 modeluje časové řady pomocí fuzzy logiky
- 16.7 modeluje časové řady pomocí umělých neuronových struktur
- 16.8 modeluje časové řady pomocí genetických algoritmů

Abstrakt

Metody a přístupy ve statistickém modelování lze obecně rozdělit na základě jejich předpokladu o popisném charakteru pozorovaných dat. Tyto přístupy se nazývají parametrický a neparametrický. Parametrické metody vyžadují specifikaci konkrétního rozdělení náhodné veličiny (normální rozdělení, F rozdělení, Poissonova rozdělení apod.). Neparametrické a semi-parametrické metody pak žádné zvláštní předpoklady ohledně statistického rozdělení pravděpodobnosti náhodné veličiny nevyžadují.

Podmínkou pro parametrické přístupy je nejčastěji tzv. B.L.U.E. podmínka (Best Linear Unbiased Estimator) odhadu. Zde je využito Gaussova-Markovova teorému. Tyto podmínky nazýváme klasickými předpoklady obecného v parametrech lineárního regresního modelu (GLM – general linear model). Tyto podmínky jsou důležité pro analýzy typu ANOVA, MANOVA, t-test F-test, jednoduchá regrese (SLR) a vícenásobná regrese (MLR).

Necht' je standardní funkci, která je lineární v parametrech (bety $- b_i$): „ $y = b_0 + b_1x_1 + u$ “, kde u jsou nepozorované faktory, disturbance, chybový člen. Potom označení „ $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1x + u$ “ je odhad pomocí metody nejmenších čtverců (ordinary least squares - OLS, nebo též MNČ). Přímka „začíná“ (pro $y=0$) v bodě úrovnové konstanty $\hat{\beta}_0$ a sklon přímky je $\hat{\beta}_1$. Odhadnutá chybová složka má označení \hat{u} , což lze vyjádřit jako rozdíl skutečnosti znaku y a odhadu \hat{y} . Jednotlivé rozdíly odhadu a skutečnosti, je vhodné mít co nejmenší. Zároveň jejich součet musí být nulový. Je nutné podotknout, že \hat{u} není náhodná složka, je to pouze odhad náhodné složky. Pokud má tento odhad podobné vlastnosti jako očekávaná náhodná složka u pro normální rozdělení, pak produkuje tzv. BLUE odhad.

Gaussovy-Markovovy podmínky pro mnohonásobnou v parametrech lineární regresi jsou následující. Podmínka MLR 1 se věnuje linearitě v parametrech. MLR. 2 je o náhodnosti výběru dat. U průřezových dat se snažíme zajistit reprezentativní datový vzorek, který má dostatečný počet pozorování. U náhodnosti výběru je požadována také nulová sériová korelace (autokorelace, série je zde označení pro časové řady) pro nepozorované faktory v náhodné složce u , tedy i pro její odhad \hat{u} . Zároveň je nutné, aby chybová složka u (a zároveň odhad \hat{u}) byla náhodně distribuovaná a nezávislá na každé proměnné v modelu. Problémy s autokorelací (sériovou korelací) nastávají vždy pro časové řady (time series, série dat). MLR.3 znamená absence perfektní kolinearit. Většina programových balíčků tuto podmínku hlídá. Identické proměnné, nebo proměnné s vysokou vzájemnou korelací jsou z modelu odebrány. Programy hlídají i počet proměnných (k), těch musí být méně, než je počet pozorování (n). Přesněji je to $n \geq k+1$, jde o maticovou logiku a problém tzv. stupňů volnosti. Přičemž je vhodné sledovat i pestrost použitých statistických rozdělení. Při absenci ordinálních proměnných (věk, tržby, počet zaměstnanců apod.) dochází ke zkreslení chyby odhadu parametrů v modelu. MLR.4 je o nulovém podmíněném průměru pro chybovou složku u . Přesněji jde o vyjádření $E(u/x_1, x_2, \dots, x_k) = 0$. Ve skutečnosti není nutné, aby to byla nula, protože v regresi máme přítomnu konstantu, která nulovou hodnotu zajistí. Proto je vhodné konstantu v modelu vždy ponechat. Problémy s podmínkou MLR4 způsobují zejména vynechané proměnné, chyba měření, špatná funkční forma, simultaneita a endogenita proměnných.

Další podmínky (pátá a šestá podmínka) jsou obvyklé už jen pro jednoduchý lineární regresní model (SLR), protože je lze v MLR modelech snadno vyřešit nastavením designu

modelu. SLR je přímková regrese s jednou vysvětlovanou a vysvětlující proměnnou. SLR.5 znamená, že je vyžadována homoskedasticita (konstantní rozptyl) chybové složky u (i odhadnuté složky \hat{u}). Konkrétní zápis je $\text{var}(u/x_1, \dots, x_k) = \sigma^2$. SLR.5 je vhodné u vícenásobné regrese ošetřit zvolením robustních standardních chyb, abychom měli vždy konzistentní odhad s opravenou nestranností. SLR.6 vyžaduje u jednoduché regresní přímky normalitu (u) chybového členu. U vícenásobné regrese je vhodné, aby rezidua měla rozdělení alespoň podobné normálnímu rozdělení. Platí zde předpoklad centrální limitní věty, tj. že toto rozdělení se s počtem pozorování „dovrhná“ do normálního. Výrazné odchylky od normálního rozdělení pak znamenají, že máme nějaký problém v MLR 2 a MLR 4.

Studijní literatura

Povinná literatura

HAMPEL, D., V. BLAŠKOVÁ a L. STŘELEČEK, 2016. *Ekonomie 2*. 3. přeprac. vyd. Brno: Mendelova univerzita. ISBN 978-80-7509-427-8. (s. 10-14)

Kontrolní otázky

1. Jaké existují parametrické metody odhadu?
2. Co je to neparametrický odhad?
3. Co je to tzv. BLUE odhad?
4. Definujte 1. Gausovu-Markovovu podmínku pro vícenásobnou lineární regresi.
5. Definujte 2. Gausovu-Markovovu podmínku pro vícenásobnou lineární regresi.
6. Definujte 3. Gausovu-Markovovu podmínku pro vícenásobnou lineární regresi.
7. Definujte 4. Gausovu-Markovovu podmínku pro vícenásobnou lineární regresi.
8. Definujte 5. Gausovu-Markovovu podmínku pro vícenásobnou lineární regresi.
9. Definujte 6. Gausovu-Markovovu podmínku pro jednoduchou lineární regresi.
10. Co je to náhodná složka u , a jak se liší od tzv. reziduí?

Odkaz na praktickou část

3.2 Počet pravděpodobnosti a statistické rozdělení dat. Pravidla regresní analýzy pro průřezová data

2.4 Pravidla vícenásobné regresní analýzy pro časové řady

Klíčová slova

Gaussovy-Markovovy podmínky, proměnná, chyba odhadu, signifikance, F-test

Cíle kapitoly

Cílem je pochopení pravidel pro vícenásobnou regresní analýzu pro průřezová data a pro časové řady.

Výstupy z učení

- 16.6 modeluje časové řady pomocí fuzzy logiky
- 16.7 modeluje časové řady pomocí umělých neuronových struktur
- 16.8 modeluje časové řady pomocí genetických algoritmů

Abstrakt

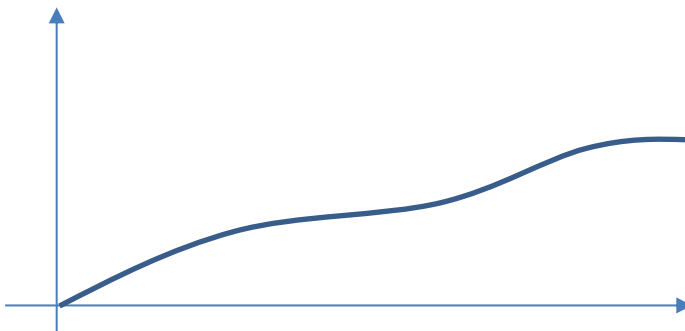
Časové řady představují pozorování jednoho jevu y_i (HDP, objem produkce, příjmy, náklady...aj.) ve více časových (t) okamžicích, označujeme jako y_{it} . Pohledem na jejich průběh v čase jsou rozeznávány časové řady stochastické a deterministické. Stochastické se vyznačují odchylkou od rovnovážné hodnoty, jsou stacionární (ustálené). Tyto řady jsou označovány jako proces I(0), tj. časová řada integrovaná (I) řádu nula (0). V jednorovnicovém vyjádření při konečném zpoždění, jde o autoregresní funkci, kde máme zpožděnou vysvětlovanou proměnnou na pravé straně rovnice (y_{t-1}). Konečné zpoždění jednoho období AR(1) lze vyjádřit jako rovnicí: „ $y_t = \text{konstanta} + \alpha_1 \text{trend} + \alpha_2 y_{t-1} + \varepsilon_t$ “ s trendem, nebo bez trendu: „ $y_t = \text{konstanta} + \alpha_2 y_{t-1} + \varepsilon_t$ “.

Deterministické časové řady se vyznačují předurčeným trendem, mají v sobě zakomponovanou počáteční podmínku a odchylka má povahu tzv. „náhodné procházky“ (random walk). Hodnoty se „nevrací“ k nějaké konstantě či trendu, povaha řady je nestacionární (neustálená). V našem zkoumání budeme nejčastěji tyto řady označovat jako I(1), integrované (I) řádu (1), případně jako I(d) integrované vyššího řádu.

U deterministických (nestacionárních) řad se objevuje problém v MLR.2 a MLR.4. Tento problém nazýváme zdánlivou regresí. Analýza nestacionárních časových řad je

komplikovanější. Obvykle je nutné odstranit problém nestacionarity např. pomocí transformace dat na jejich první diferenci (odečíst od sebe pozorování v čase, např. $\Delta y_t = y_t - y_{t-1}$), avšak tím už sledujeme v odhadu změnovou proměnnou Δy_t místo úrovně proměnné y_t . Vyhledávanou vlastností u těchto řad je možnost sledovat vztah vzhledem k jiné časové řadě. Jde o tzv. kointegraci, kdy je provedena analýza 2 (a více) nestacionárních časových řad (např. HDP a spotřeba), které jsou integrovány stejného řádu a jejich společné namodelované ekvilibrium je časová řada stacionární $I(0)$.

Obrázek 1: Ukázka nestacionární řady



Zdroj: Vlastní zpracování

V těchto případech je možné zvažovat i další analýzu pomocí modelů s korekcí chyby („Error correction models“) a tzv. Grangerovu kauzalitu. Při interpretaci Grangerovy kauzality vždy používáme toto slovní spojení, protože se nejedná o kauzalitu skutečnou. V Grangerově kauzalitě testujeme pouze směr, ve kterém hodnoty v minulosti jedné proměnné ovlivňují hodnoty v následujícím období jiné proměnné. Grangerova kauzalita může být obousměrná a je tu stále velké riziko zdánlivé regrese anebo vynechání klíčové tzv. třetí proměnné z modelu.

Kointegrační analýzu je nutné podložit ekonomickou teorií a historickým kontextem. Je nutné zdůvodnění, proč jsou dané řady vzájemně v ekvilibriu. Pokud nemáme dostatečný teoretický podklad, lze takto odhadnout i nesmyslné vztahy (S&P index a cena oběda v Menze).

Studijní literatura

Povinná literatura

HAMPEL, D., V. BLAŠKOVÁ a L. STŘELEČ, 2016. *Ekonomie 2*. 3. přeprac. vyd. Brno: Mendelova univerzita. ISBN 978-80-7509-427-8. (s. 15-48)

Kontrolní otázky

1. Uveďte příklady časových řad – sérií.
2. Definujte stochastický proces.
3. Definujte stacionární proces.
4. K jakému problému dochází u časových řad při podmínce MLR 2.?
5. K jakému problému dochází u časových řad při podmínce MLR 4.?
6. Co je to jednotkový kořen a jak se testuje?
7. Co je to kointegrace?
8. Jak je definována Grangerova kauzalita?
9. Jaký je vztah mezi kauzalitou a regresí?
10. Co je to zdánlivá regrese?

Odkaz na praktickou část

3.2 Počet pravděpodobnosti a statistické rozdělení dat. Pravidla regresní analýzy pro průřezová data

2.5 Testování předpokladů konvenční asymptotické teorie pro odhady časových řad metodou nejmenších čtverců, transformace časových řad, časové řady typu I(0) a I(1)

Klíčová slova

Stacionarita, jednotkový kořen, bílý šum, náhodná procházka, zdánlivá regrese

Cíle kapitoly

Cílem je pochopení testování předpokladů konvenční asymptotické teorie pro odhady časových řad metodou nejmenších čtverců. Cílem je pochopit transformaci časových řad na časové řady typu I(0).

Výstupy z učení

- 16.6 modeluje časové řady pomocí fuzzy logiky
- 16.7 modeluje časové řady pomocí umělých neuronových struktur
- 16.8 modeluje časové řady pomocí genetických algoritmů

Abstrakt

V této kapitole je představen postup pro rozšířený Dickeyův Fullerův test (Augmented Dickey Fuller – ADF). ADF testování přítomnosti jednotkového kořenu je součástí každé analýzy časových řad. Necht' jsou definovány tyto základní modely vycházející z předpisu pro AR(1): „ $y_t = \text{konstanta} + \alpha_1 \text{trend} + \alpha_2 y_{t-1} + \varepsilon_t$ “. Konstanta je výchozí průměrná úroveň, trend odpovídá zapojení časové proměnné (rok, čtvrtletí apod.) do modelu a y_{t-1} je o jedno období zpožděná vysvětlovaná proměnná y_t . Před dalším zkoumáním a modelováním je nutné otestovat přítomnost jednotkového kořenu v časové řadě.

Rovnice pro rozšířený Dickeyův Fullerův (ADF) test s trendem má podobu: „ $\Delta y_t = \text{konstanta} + \beta_1 \text{trend} + (\beta_2 - 1)y_{t-1} + \sum_{m=1}^k \omega_m \Delta y_{t-m} + u_t$ “. U této rovnice ADF platí předpoklad, že je vhodné sledovat i koeficient trendu ($\beta_1 \text{trend}$). Takový ADF test je vhodné použít pro časové řady, které viditelně rostou, nebo klesají. To samé platí i o počtu přidaných časových zpoždění v předposledním členu rovnice ADF testu $\sum_{m=1}^k \omega_m \Delta y_{t-m}$,

které redukuje problém autokorelace. U ročních časových řad je vhodné přidat minimálně zpoždění dvou let, logicky pak u měsíčních časových řad pak minimálně 24 zpoždění. Pro ADF je model co nejvíce doplnit dalšími exogenními proměnnými.

Rovnice pro ADF test bez trendu má podobu: " $\Delta y_t = \text{konstanta} + (\beta_2 - 1)y_{t-1} + \sum_{m=1}^k \omega_m \Delta y_{t-m} + u_t$ " předpokládá se, že trend v časové řadě nedává smysl, je to například u časových řad, které obvykle nemají jasný klesající či rostoucí trend. Nebo u časových řad, kde např. ekonomická teorie ani nepředpokládá výskyt členu β_1 trend. ADF test je součástí běžných statistických programů a zobrazují se i kritické hodnoty pro vyhodnocení t-testu pro sledovaný koeficient $(\beta_2 - 1)$. Kritické hodnoty jsou striktnější a odlišné od standardního dvoustranného (two-tailed) t-testu pro testování nulové hypotézy H_0 , která říká, že: Zkoumaná časová řada obsahuje jednotkový kořen a není řadou stacionární.

ADF test není silným testem a neposkytuje vždy přesné výsledky a je vhodné vždy používat intuici, data nejprve zobrazit a poté rozhodnout o strategii testování. Pro ujištění se, zda testování jednotkového kořene bylo úspěšné, je vhodné použít i další testy jednotkového kořene např. Phillipsův-Perronův test a verzi ADF testu odhadovanou pomocí generalizovaných nejmenších čtverců (Generalized Least Squares).

Studijní literatura

Povinná literatura

HAMPEL, D., V. BLAŠKOVÁ a L. STŘELEČEK, 2016. *Ekonometrie* 2. 3. přeprac. vyd. Brno: Mendelova univerzita. ISBN 978-80-7509-427-8. (s. 49-54)

Kontrolní otázky

1. Na jakém principu funguje ADF test?
2. Jakým způsobem upravujeme parametry ADF testu, na čem naše taktika testování závisí?
3. Pro jaké časové řady použijeme ADF, kdy testujeme přítomnost jednotkového kořenu bez trendu?
4. Kolik období je vhodné mít pro zpoždění v modelu, pokud jednotkou je týden, abychom předešli problémům s autokorelací?
5. Jaké testy používáme pro ověření přítomnosti autokorelace?

6. Jaký test nám pomůže odhalit sériovou korelaci na konkrétním stupni zpoždění?
7. Popište problémy, které mají časové řady při plnění Gaussových – Markovových podmínek?
8. Co je to striktní exogenita a podmíněná exogenita časových řad?
9. Které metody s kontrolou na autokorelaci lze použít pro odhad stacionárních časových řad.
10. Jaké transformace využíváme pro odhad časových řad?

Odkaz na praktickou část

3.3 Testování předpokladů konvenční asymptotické teorie pro odhady metodou nejmenších čtverců pro průřezová data a pro časové řady

2.6 Modelování časových řad pomocí regresní analýzy

Klíčová slova

Autoregresní model, dlouhodobý efekt, Box-Jenkinsonova metoda

Cíle kapitoly

Cílem kapitoly je osvojení si principů Box-Jenkinsonovy metodologie a pochopení modelování časových řad pomocí regresní analýzy.

Výstupy z učení

- 16.6 modeluje časové řady pomocí fuzzy logiky
- 16.7 modeluje časové řady pomocí umělých neuronových struktur
- 16.8 modeluje časové řady pomocí genetických algoritmů

Abstrakt

Standardní regresní analýza je aplikována na odhad stacionární časové řady. V autoregresních modelech (AR) se zpoždění vysvětlované proměnné objevuje na pravé straně rovnice. Avšak lze zapojit i zpožděné vysvětlující proměnné, které jsou na pravé straně rovnice. Tím se odhaduje efekt minulých hodnot exogenních proměnných (LRP - Long run propensity).

Klasický AR(1) model má podobu: $y_t = \text{konstanta} + \beta'X'_t + \lambda y_{t-1} + u$ pro zapojení jednoho zpoždění (y_{t-1}) vysvětlované proměnné y_t s vektorem exogenních vysvětlujících proměnných X'_t a chybové složky u . Dalším příkladem je zapojení zpoždění vysvětlujících proměnných. $y_t = \text{konstanta} + \beta x_t + \gamma x_{t-1} + u$. V tomto případě lze sledovat souhrnný dlouhodobý efekt (LRP) vysvětlující proměnné. Kolik zpoždění $x_t + x_{t-1} + x_{t-2} + x_{t-3}$ apod. bude v modelu zapojeno je testováno společným F-testem odhadnutých koeficientů. V obou příkladech je možné dále zapojit trend, a to buď prostý lineární $y_t = \text{konstanta} + \beta'X'_t + t + u$, nebo kvadratický t^2 , případně kubický t^3 apod.

Základní transformací pro odhad časových řad je běžná diference $\Delta y_t = y_t - y_{t-1}$, anebo logaritmičká diference $\Delta \log y_t = \log y_t - \log y_{t-1}$, která zajistí odstranění jednotkového kořenu. Problémem je však, že touto transformací ztrácíme informaci o

úrovni sledované proměnné y_t a máme informaci pouze o průběhu změny Δy_t , případně procentní změny u logaritmické difference. Např. nevysvětluje se úroveň HDP v mld. Kč ale absolutní, případně procentní růst HDP. Dále se používá Boxova-Jenkinsonova metoda, která ztotožňuje systematickou část časové řady s částí deterministickou (nestacionarita) a je založena na myšlence, že časová řada může být vždy chápána jako řada stochastického charakteru.

Pro testování autokorelace se používá Durbinův-Watsonův alternativní test, který dokáže identifikovat problém autokorelace. Pokud nás zajímá autokorelace pro vybraná použitá zpoždění, které se použily v modelu, je vhodné použít Breuschův – Godfreyův test využívající Lagrangeovy multiplikátory (LM). Pro stacionární řady bez jednotkového kořene lze využít přímo metodu odhadu, která pracuje se korelací časových řad. První je Praisova – Winstenova metoda, případně Cochranova-Orcuttova metoda AR(1) regrese. Obě metody využívají iterací (pro chybovou složku), a pokud dojde ke konvergenci v parametru ρ , je výsledný model bez autokorelace. Další metodou je např. Feasible generalized least squared (FGLS).

$$\text{Rovnice regrese } y_t = \text{konstanta} + \beta x_t + \gamma x_{t-1} + u_t \quad (1)$$

Rovnice Durbin – Watson d statistiky (2) pro rovnice regrese (1):

$$d = \frac{\sum_{t=1}^{n-1} (\hat{u}_{t+1} - \hat{u}_t)^2}{\sum_{t=1}^n \hat{u}_t^2} \quad (2)$$

Studijní literatura

Povinná literatura

HAMPEL, D., V. BLAŠKOVÁ a L. STŘELEČEK, 2016. *Ekonometrie 2*. 3. přeprac. vyd. Brno: Mendelova univerzita. ISBN 978-80-7509-427-8. (s. 67-76)

Kontrolní otázky

1. Popište rovnici AR(1) modelu.
2. Popište rovnici AR(1) modelu se zpožděnou vysvětlující proměnnou.

3. Sestavte rovnici AR(1) modelu tak, abyste mohli definovat LRP pro 3 období vysvětlující proměnné.
4. Jaké druhy trendu lze použít v AR(1) modelech?
5. Jak lze otestovat stacionaritu AR(1) modelu?
6. Jak lze otestovat přítomnost sériové korelace v AR(1) modelu?
7. K čemu slouží Praisova – Winstenova metoda odhadu?
8. Jak otestuji heteroskedasticitu v AR(1) modelu?
9. Co je hlavním problémem transformace nestacionární časové řady pomocí diferencí?
10. Proč se využívá logaritmická transformace, jaké má praktické výhody?

Odkaz na praktickou část

3.3 Testování předpokladů konvenční asymptotické teorie pro odhady metodou nejmenších čtverců pro průřezová data a pro časové řady

2.7 Časové řady v ekonomickém výzkumu a podnikové praxi, možnosti pokročilých metod

Klíčová slova

Box-Jenkinsonova metoda, stacionární řady, nestacionární řady, ARIMA, GARCH

Cíle kapitoly

Cílem je pochopení modelování časových řady v ekonomickém výzkumu a podnikové praxi při využití pokročilých metod pro nestacionární časové řady.

Výstupy z učení

- 16.6 modeluje časové řady pomocí fuzzy logiky
- 16.7 modeluje časové řady pomocí umělých neuronových struktur
- 16.8 modeluje časové řady pomocí genetických algoritmů

Abstrakt

Stochastický proces je striktně stacionární. Protože striktní stacionaritu je v praxi obtížné ověřovat, je v analýze časových řad zavedena postačující podmínka „slabé stacionarity“. Stochastický proces je slabě stacionární jen tehdy a tehdy, když $\mu_t = \mu$, $\sigma_t^2 = \sigma^2$ pro všechna t a kovarianční a korelační funkce závisí pouze na časové vzdálenosti náhodných veličin.

Invertibilní proces je lineární proces, který umožňuje zápis ve tvaru, kdy jeho současná hodnota je vyjádřena pomocí minulých hodnot ($t-p$) a současné hodnoty bílého šumu (u_t). Invertibilita má stěžejní význam pro konstrukci předpovědí.

Identifikace použití základního nebo pokročilého modelu je počátečním krokem tzv. Boxovy-Jenkinsonovy metodologie. Pro modely stacionárních časových řad jde o autoregresní procesy (AR), procesy klouzavých průměrů (MA), o smíšené procesy AR+MA (ARMA). Modely nestacionárních časových řad identifikují procesy náhodné procházky (Random Walk Process), procesy ARIMA, modely sezonních časových řad, sezonní autoregresní procesy (SAR), sezonní procesy klouzavých průměrů (SMA), smíšené sezonní a nesezonní procesy (SARMA), modely sezonních integrovaných časových řad (SARIMA), modely časových řad s dlouhou pamětí, modely s režimy určenými

nepozorovatelnými veličinami – model MSW (Markov-Switching). Samostatnou kategorií jsou lineární modely volatility v analýze finančních dat, modely ARCH (Autoregressive Conditional Heteroscedasticity), modely GARCH (Generalized ARCH), nelineární modely volatility – modely EGARCH (Exponential GARCH). Druhým krokem tzv. Boxovy-Jenkinsovy metodologie je samotný odhad parametrů modelu a třetím krokem je intenzivní testování – ověřování modelu.

Mezi základní modely stacionárních procesů patří AR(1), což je autoregresní proces prvního řádu. Pokročilou metodou je AR(p), kdy v modelu je více než jedno roční zpoždění: $y_t = \text{konstanta} + \beta * x_t + \gamma * y_{t-1} + \dots + u_t$. Je-li model klouzavých průměrů reziduální složky konečného stupně MA(stupeň), je vždy stacionární. Proces klouzavých průměrů prvního řádu má formu $y_t = a_t - \theta_1 a_{t-1}$, u vyšších stupňů máme více zpoždění proměnné a_t : $y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_p a_{t-p}$. ARMA(1,1) proces je MA(1) proces se zpožděnou vysvětlovanou proměnnou AR(1) dohromady tedy ARMA (AR stupeň p, MA stupeň q): $y_t = y_{t-1} + a_t - \theta_1 a_{t-1}$.

Model ARIMA (p, d, q) se nazývá autoregresním integrovaným procesem klouzavých průměrů řádu p, d, q (z angl. autoregressive integrated moving average). Z modelu ARIMA(p, d, q) se stává po diferenci d -tého řádu model ARMA(p, q), což lze vyjádřit jako $X_t \sim ARIMA(p, d, q) \Leftrightarrow \Delta X \sim ARMA(p, q)$, kde Δ je operátor difference. U konstrukce ARIMA modelu se mj. díky diferenci nepožaduje stacionarita analyzované řady.

Identifikace modelu je první fází výstavby modelu. Úkolem při pohledu na vizualizovanou časovou řadu (graf časové řady) je rozhodnout, jaký typ modelu použít (AR, MA, ARIMA apod.) a explicitně určit i řád modelu. Vlastní identifikace je založena na zkoumání průběhu odhadnuté autokorelační funkce a parciální autokorelační funkce. Chování autokorelační funkce napovídá, jaký typ modelu je vhodné pro danou řadu použít (test jednotkového kořenu). Odhad parametrů modelu je komplikovaný a k jejich provedení je zapotřebí použít naprogramované funkce a výpočetního výkonu počítače. Ověřování modelu pomocí dodatečných testů má za úkol nezamítnout adekvátnost modelu. Existuje více metod ověření modelu, testy „přeparametrizování“ modelu, různé testy autokorelace a odhadnutých reziduí apod.

Studijní literatura

Povinná literatura

HAMPEL, D., V. BLAŠKOVÁ a L. STŘELEČ, 2016. *Ekonometrie 2*. 3. přeprac. vyd. Brno: Mendelova univerzita. ISBN 978-80-7509-427-8. (s. 77-86)

Kontrolní otázky

1. Proč časové řady neplní podmínku tzv. striktní exogenity.
2. Popište podmínky „slabé stacionarity“.
3. Co je to tzv. invertibilní proces?
4. Co je to bílý šum?
5. Jaké pokročilé metody odhadu lze použít pro stacionární časové řady?
6. Jaké pokročilé metody odhadu lze použít pro nestacionární časové řady?
7. Popište vazby mezi AR(1), MA (1) a ARMA(1,1) procesem.
8. Jak se liší ARMA a ARIMA proces?
9. Jaké testy jsou vhodné ve fázi post odhadu časové řady v případě pokročilých metod?
10. Jaké pokročilé metody lze použít pro lineární modely volatility v analýze finančních dat?

Odkaz na praktickou část

3.3 Testování předpokladů konvenční asymptotické teorie pro odhady metodou nejmenších čtverců pro průřezová data a pro časové řady

2.8 Pokročilé modelování časových řad pomocí regresní analýzy

Klíčová slova

Logaritmická transformace, dummy, kvadratická proměnná, proxy, instrument

Cíle kapitoly

Cílem je pochopení pokročilého modelování časových řad pomocí regresní analýzy s využitím různých druhů transformace proměnných a proměnných typu IV a proxy.

Výstupy z učení

- 16.6 modeluje časové řady pomocí fuzzy logiky
- 16.7 modeluje časové řady pomocí umělých neuronových struktur
- 16.8 modeluje časové řady pomocí genetických algoritmů

Abstrakt

Dummy proměnná v MLR modelu umožňuje přidat „přepínač“, který nám upravuje výchozí hodnotu β_0 . Tím dochází k předefinování toho, co obsahuje konstanta v modelu (β_0). Konstanta popisuje výchozí populaci pokud je vysvětlovaná proměnná nulová, je to skupina, vůči které srovnáváme a interpretujeme další odhadnuté koeficienty v regresním modelu. Necht' $y = \beta_0 + \dots + \beta_k x_k + u$... kde $\beta_k x_k$ jsou klasické exogenní kvantitativní proměnné typu příjem, výška, množství a počet let. Pak β_0 je úroňová konstanta vzorku. Např. průměrná výchozí výška zkoumané populace, za jinak nezměněných podmínek: $\hat{y} = 179 + \dots + \beta_k x_k + u$, kde 179 je průměrná výchozí výška zkoumané populace za jinak nezměněných podmínek. Pokud do MLR modelu přidáme proměnnou x_{1_dummy} , která nabývá hodnot 0 a 1, a hodnoty 1, když jde o ženy. Např. $\hat{y} = 175 + 4 x_{1_dummy} + \dots + \beta_k x_k + u$, kde 175 je průměrná výchozí výška mužů ve zkoumané populaci za jinak nezměněných podmínek. Průměrná výchozí výška žen ve zkoumané populaci je o 4 cm vyšší (175+4) za jinak nezměněných podmínek. Obvykle se používá více dummy proměnných v modelu. Úroňovou konstantu lze pak popsat jako průměrnou výchozí hodnotu vysvětlované proměnné, když všechny dummy proměnné se rovnají 0. Problém „dummy trap“ neboli

past dummy proměnných nastane, pokud v modelu máme kolineární dummy proměnnou. Např. x_{1_dummy} (=1 když žena) a zároveň x_{2_dummy} (=1 když muž), pokud muž a žena jsou jediné definované pohlaví, pak dummy proměnná pro ženy plně vysvětluje dummy proměnnou pro muže (přítomnost perfektní kolinearit).

Kvadratická proměnná v rovnici $y = b_0 + b_1z + b_2z^2 + \dots + b_kx_k + u$ se používá s cílem zjistit, zda existují rostoucí ($b_1z + b_2z^2$) nebo klesající ($b_1z - b_2z^2$) výnosy z rozsahu. Cílem je odhadnout „optimální“ hodnotu, resp. hodnotu globálního extrému z^* a parciální efekt:

$$\frac{\partial y}{\partial z} = \beta_1 + 2\beta_2z, \text{ přičemž tento efekt vypočteme jako } z^* = \left| \frac{\beta_1}{2\beta_2} \right|.$$

Logaritmované proměnné v MLR modelu, kombinace $\ln(x_1)$ a $\ln(y)$

- I. Typ **log-level**: $\ln(y) = \beta_0 + \beta_1x_1 + \dots + \beta_kx_k + u$
- II. Typ **level-log**: $y = \beta_0 + \beta_1\ln(x_1) + \dots + \beta_kx_k + u$
- III. Typ **log-log**: $\ln(y) = \beta_0 + \beta_1\ln(x_1) + \dots + \beta_kx_k + u$

Interpretace pro konkrétní hodnotu logaritmicky transformované proměnné je přímočará. Hodnota je dosazena do výsledné rovnice a v případě potřeby převedeme vysvětlovanou proměnnou (typ I. a III.) zpět na tzv. netransformovanou „level hodnotu“ pomocí inverzní funkce k přirozenému logaritmu, tj. jde o exponenciální funkci. Avšak snadnost interpretace je právě v možnosti interpretovat situaci i procentně. Zde je třeba rozlišit tři případy, které mohou nastat.

I. Odhadnutá rovnice (pod koeficienty v závorce je standardní chyba):

$$\hat{y} = \ln(\text{hodinová mzda}) = 1.56 - 0,251*\text{ženy} + 0,0112*\text{věk} + u$$

Pak β_0 je úroňová konstanta a nabývá hodnoty 1,56. Po převedení na exponenciální hodnotu $\exp(1,56304)$ nabývá hodnoty 4,76, což představuje nikoliv aritmetický průměr, ale průměr geometrický, pouze pro muže a to ve věku 0 let. Úroňová konstanta proto zde není pro interpretaci zajímavá. Růst věku o jednotku (zde je jednotka 1 rok) způsobí růst platu v (geometrickém) průměru o $[\exp(0,0111950)-1]$, což je 1,122 %, za jinak nezměněných podmínek. Při velmi malých hodnotách beta koeficientů jsou si hrubé (1,112 %) a skutečné vyjádření (1,122 %) více podobné.

II. Odhadnutá rovnice (pod koeficienty v závorce je standardní chyba):

$$\hat{y} = \text{hodinová mzda} = -4.03 - 1.77*\text{ženy} + 3.47*\ln(\text{věk})$$

Pak β_0 je úroňová konstanta a nabývá průměrné (aritmetické) hodnoty -4,032 dolarů za hodinu pro muže ve věku nula let. Úroňová konstanta v tomto „level-log“ modelu není

nijak zajímavá. Pro případnou interpretaci je nutné zvolit např. úroveň mzdy pro 25leté, nebo jinou konkrétní hodnotu. Odhadnutý efekt věku zde není lineární (i když efekt $3.47 \cdot \ln(\text{věk})$ je lineární v parametru). Díky logaritmování má efekt věku průběh logaritmické funkce. To znamená, že není možné použít absolutní změnu ve věku (o 10 let), ale změnu relativní, např. o 10 %.

Vyjádřeno formálně 10 % změna u věku je hodnota $\ln(1,1)$ a tou vynásobíme odhadnutý koeficient u věku (3.46983). Při 10% změně věku sledujeme v průměru (aritmetickém) růst hodinové mzdy o $\ln(1,1) \cdot 3.46983 = 0,33$ dolaru.

III. Odhadnutý logaritmovaný model hodinové mzdy v dolarech v podobě:

$$\hat{y} = \ln(\text{hodinová mzda}) = -0.921 - 0.268 \cdot \text{ženy} + 0.536 \cdot \ln(\text{věk}) + 0.0783 \cdot \text{vzdělání}$$

Pak β_0 je úrovněová konstanta a nabývá hodnoty -0,92. Po převedení na exponenciální hodnotu $\exp(-0,92)$ nabývá hodnoty 0,3985, což představuje nikoliv aritmetický průměr, ale průměr geometrický pro hodinovou mzdu a to pouze pro muže, ve věku 0 let a s 0 lety vzdělání. Úrovněová konstanta opět není nijak zajímavá a museli bychom arbitrárně vybrat pro interpretaci zajímavější skupinu mužů, např. ve věku 40 let a průměrného počtu 12 let vzdělání, pokud bychom vůči této skupině chtěli srovnávat a dávat příklady.

Odhadnutý efekt věku zde není lineární, má průběh logaritmické funkce. To znamená, že nelze použít absolutní změnu ve věku, ale změnu relativní, např. o 10 %, 5 % anebo 1 %. Při 10% změně věku je v průměru (geometrickém) růst hodinové mzdy $1,1^{0.535984} = 1,0524$ o 5,24 %.

Studijní literatura

Povinná literatura

HAMPEL, D., V. BLAŠKOVÁ a L. STŘELEČEK, 2016. *Ekonometrie 2*. 3. přeprac. vyd. Brno: Mendelova univerzita. ISBN 978-80-7509-427-8. (s. 87-93)

Kontrolní otázky

1. Co je to dummy proměnná?
2. Co je to dummy trap?
3. K čemu slouží v modelu kvadratická proměnná?
4. Popište výpočet globálního extrému kvadratické funkce.
5. Jak interpretujeme tzv. „log – level“ model?

6. Jak interpretujeme tzv. „level – log“ model?
7. Jak interpretujeme tzv. „log – log“ model?
8. Co je to proxy proměnná?
9. Uveďte příklad proxy proměnné.
10. Co je to instrumentální proměnná?
11. Popište podmínky použití instrumentální proměnné.

Odkaz na praktickou část

- 3.4 Modelování časových řad pomocí regresní analýzy

2.9 Základní principy teorie fuzzy množin a fuzzy logiky

Klíčová slova

Množina, logika, fuzzy logiky, věty, inference

Cíle kapitoly

Cílem je pochopení základních principů teorie fuzzy množin a fuzzy logiky. cílem je pochopit konstrukci množin a tzv. funkcí příslušnosti.

Výstupy z učení

- 16.1 zná základní pojmy a pravidla fuzzy logiky

Abstrakt

Fuzzy je slovo pocházející z angličtiny a znamená mlhavý, nejasný, nebo neostrý. Fuzzy logika je také mlhavá, nejasná, neostrá a je na první pohled značně neurčitá. Standardní matematická a výroková logika je věda o zákonech správného myšlení, o zákonech a pravidlech nutných pro vyvozování správných závěrů. Tato logika je postavena na hodnocení výroku (věty, výrazu apod.) a nabývá hodnot ANO a NE. „ANO, toto tvrzení je pravdivé“ nebo „NE, toto tvrzení není pravdivé“.

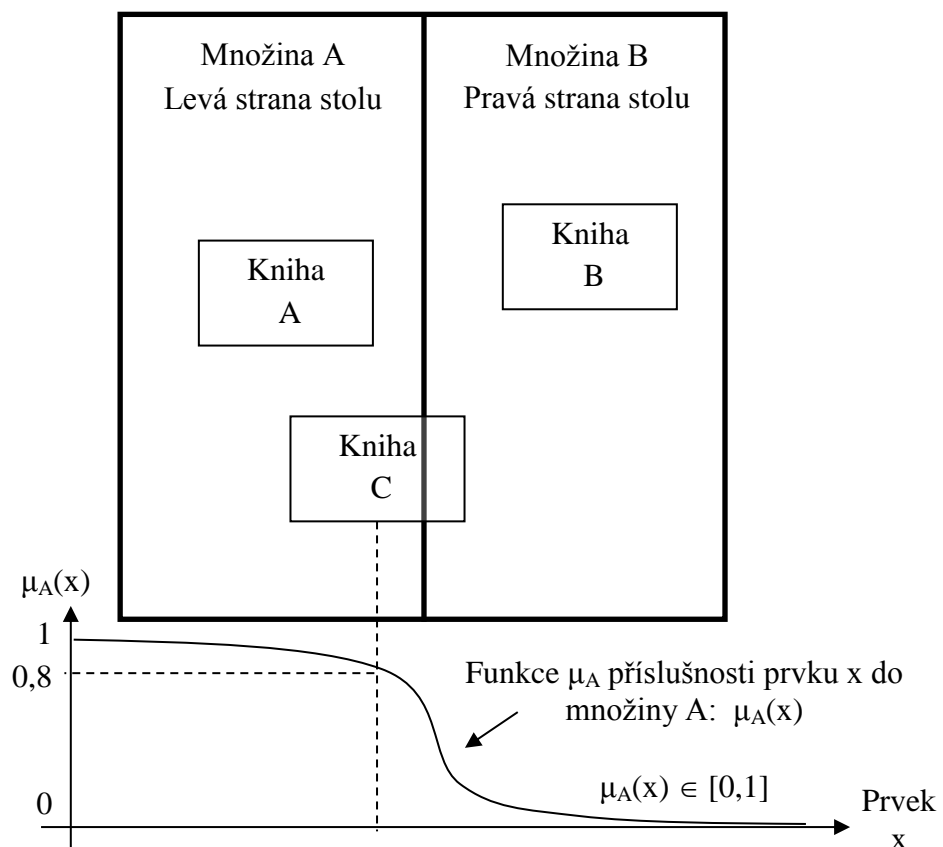
V praxi se však člověk pohybuje v prostředí, kde neurčitost, mlhavost a nejasnost hraje hlavní roli při každodenní komunikaci. Fuzzy logika se proto hodí pro společenské vědy, avšak objevuje se i v exaktních vědách. Fuzzy logika pomáhá při konstrukci například elektronického tlumočnicka, při rozvoji a konstrukci umělé inteligence, která je schopna provést výtah textu, provést diagnózu pacienta, nebo řídit automobil v hustém provozu.

Fuzzy logika staví na tom, že člověk stojící před řešením složitého úkolu neuvažuje a nepočítá ve své mysli pomocí přesně definovaných nula jedničkových hodnot, ale spíše „počítá“ s několika mlhavými pojmy najednou, které nabývají hodnot jako nedaleko, skoro dobře, přijatelně, rychle, málo, vlažně apod. Jde o nenumerické pojmy, které umí vhodně zpracovávat algoritmus neuronové sítě, genetické algoritmy a zejména teorie fuzzy množin.

Mlhavé pojmy je možno popsat matematicky přesně s použitím nástrojů tradiční logiky a následně s nimi v rámci povolených operací pracovat. Výsledkem je schopnost systému využívat rozhodovacích procesů, do kterých vstupují neurčité a mlhavé vstupní údaje.

Standardně vyučovaná matematická teorie množin připouští pro příslušnost prvku do množiny pouze dvě možnosti. Prvek buď do množiny patří (hodnota 1) nebo nepatří (hodnota 0). Příkladem může být poloha knih na čtvercovém stole (viz Obrázek níže). Prvek (kniha) je buď v levé polovině, nebo pravé polovině. Pokud by ležel na rozhraní, pak musíme vytvořit novou množinu, nebo využijeme fuzzy množin a prvku přisoudíme jakousi váhu, pravděpodobnost příslušnosti k systému. Předmět (Kniha C) je na rozhraní. Z fuzzy logiky je z 80 % na levé straně a z 20 % na pravé straně.

Obrázek 2: Fuzzy množina a charakteristická fuzzy funkce náležitosti do množiny A



Zdroj: Vlastní zpracování

Věrnější popis Knihy C tedy je, že do množiny A patří jen částečně. Hodnota funkce $\mu_A(x) = 0,8$, tj. 80% náležitost do množiny A. Charakteristickým rysem fuzzy množiny tedy je, že

příslušnost není dána hodnotou 0 nebo 1, ale celým intervalem od nuly do jedné $\langle 0,1 \rangle$. Charakteristická funkce μ se v teorii fuzzy množin nazývá funkce příslušnosti prvku do množiny.

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 6-15)

Doporučená literatura

DOSTÁL, P., 2012. *Pokročilé metody rozhodování v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-798-7. (s. 1-26)

Kontrolní otázky

1. Co znamená slovo fuzzy?
2. Jak se liší fuzzy logika a tradiční výroková logika?
3. Proč je fuzzy logika vhodná pro společenské vědy?
4. Kde se využívá fuzzy logika v ekonomii?
5. Kde se využívá fuzzy logika v exaktních vědách?
6. Jaký je rozdíl mezi numerickými a nenumerickými pojmy?
7. Co je to fuzzy výrok?
8. Co je to fuzzy funkce náležitosti do množiny?
9. Jakých hodnot nabývá fuzzy funkce náležitosti do množiny?
10. Jakých hodnot by tato funkce náležitosti do množiny nabývala v klasické výrokové logice?

Odkaz na praktickou část

3.4 Modelování časových řad pomocí regresní analýzy

2.10 Fuzzy systémy, adaptace ve fuzzy systémech

Klíčová slova

Inference, integrace, fuzzifikace, operátor

Cíle kapitoly

Cílem je pochopení fuzzy systémů, procesu adaptace ve fuzzy systémech. Cílem je seznámit se s fuzzifikací a základními operacemi s fuzzy množinami.

Výstupy z učení

- 16.1 zná základní pojmy a pravidla fuzzy logiky

Abstrakt

Zatímco v klasické teorii množin charakteristická funkce příslušnosti prvku do množiny nabývá jen dvou hodnot: 0 nebo 1, v teorii fuzzy množin charakteristická funkce příslušnosti do fuzzy množiny nabývá všech hodnot od 0 do 1. Tvar funkce příslušnosti může být různý. Nejčastěji se používají tvary sestavené z lomených přímek. Používají se i hladké průběhy jako je uvedeno v předchozí kapitole pro příslušnost knihy do množiny. V extrémních případech se používá tvar nazývaný singleton, což je případ jednobodové množiny. Rozsah charakteristické funkce příslušnosti do fuzzy množiny se nazývá univerzum.

Pro fuzzy výroky existují operátory A (AND) a NEBO (OR). Využívá se též logická negace a skládání (doplňek množiny $A = 1 - \mu_A$, průnik množiny A a B = $\text{minimum}\{\mu_A(x), \mu_B(x)\}$, sjednocení = $\text{maximum}\{\mu_A(x), \mu_B(x)\}$). Místo implikací se využívá když-potom (IF-THEN) relací. Podmínek a pravidel v relaci může být více. Do relace vstupují fuzzy množiny, podléhají soboru pravidel, dochází k agregaci výsledného souladu s pravidly a formování výstupní fuzzy množiny.

Inference slouží ke konstrukci fuzzy systémů. Ostré hodnoty z báze dat jsou podkladem pro tvorbu fuzzy množin „proces fuzzifikace“. Tyto množiny vstupují do relace (inferenční mechanismus s bází pravidel, který opět využívá bází dat), vzniká nová fuzzy množina, dochází k defuzzifikaci a získání ostré hodnoty.

Tento fuzzy systém si lze představit na příkladu. Představte si, že expert daný problém vyjadřuje slovním popisem typu „jestliže x_1 je malé a současně x_2 je dostatečně pružné, potom veličina y je středně velká“. Tento popis je fuzzy pravidlem a důležitá je zde slovní klasifikace „malý, dostatečně pružný, středně velký“. Tuto expertní znalost je však potřeba „vytěžit“ z experta a naučit i ostatní schopnosti posuzovat realitu pojmy „malý, dostatečně pružný, středně velký“.

Aproximace je další pojem související s fuzzy logikou. Libovolnou spojitou funkci definovanou na kompaktní množině je možno s libovolnou přesností aproximovat fuzzy systémem. to znamená, že fuzzy systém je univerzální aproximátor.

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 16-33)

Kontrolní otázky

1. Jaký je obvyklý tvar funkce příslušnosti?
2. Co je to singleton?
3. Co je to univerzum fuzzy funkce?
4. Definujte doplněk fuzzy množiny.
5. Definujte průnik fuzzy množin A a B.
6. Definujte sjednocení fuzzy množin A a B.
7. Jak je definována implikace ve fuzzy teorii?
8. Co jsou to fuzzy relace?
9. Co je to fuzzifikace?
10. Popište proces defuzzifikace.

Odkaz na praktickou část

3.5 Pokročilé modelování časových řad pomocí regresní analýzy

2.11 Uvedení příkladů aplikací fuzzy logiky v rozhodování, jako je např. manažerské a investiční rozhodování, predikce atd.

Klíčová slova

Finanční trhy, ceny, kolísání, pravděpodobnost

Cíle kapitoly

Cílem je pochopení fuzzy systémů a příkladů aplikací fuzzy logiky v manažerském a investičním rozhodování a při predikci časových řad.

Výstupy z učení

- 16.1 zná základní pojmy a pravidla fuzzy logiky

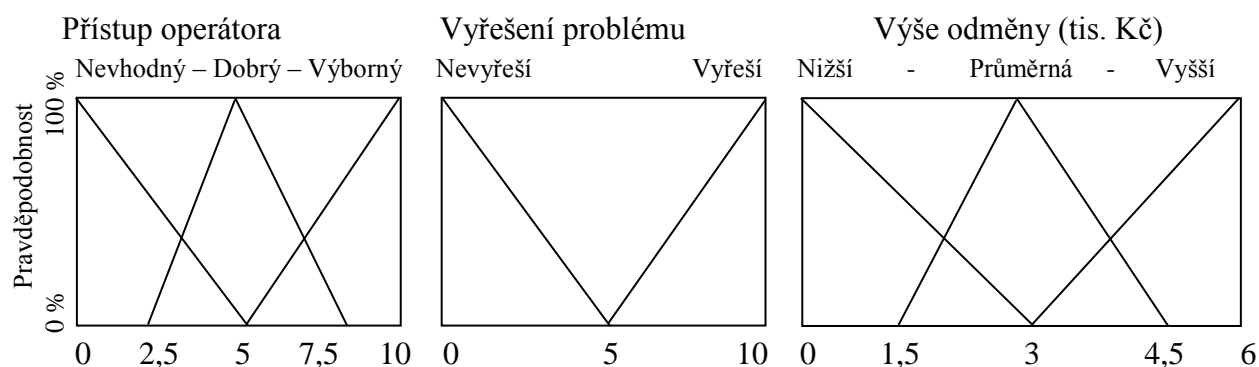
Abstrakt

Fuzzy logika je hojně využívána v ekonomickém výzkumu, a to v oblastech: Predikce ekonomických časových řad, predikce volatility časových řad na kapitálových trzích, zde je možné zpracovávat data o obchodování s akciemi a vytvářet vlastní fuzzy systémy (kurzy měn, ceny derivátů, akcií, dluhopisů, komodit atd.). V podnikovém a data miningu umí fuzzy logika získávat expertní znalosti a informací skryté v chaotickém a na první pohled nesrozumitelném kvantu dat. Dále při řízení vztahů se zákazníky a při řízení zdrojů. Díky tomu je možné, jak snížit náklady, tak zvýšit zisk firmy, a tudíž přispět k alokační efektivnosti firmy.

K modelování příkladů se využívá nástroje MATLAB – počítačový program. Ten obsahuje tzv. toolbox s názvem „Fuzzy Logic“, zde je možné nastavit počet vstupních proměnných, počet výstupních proměnných, označit (identifikovat) proměnné a definovat metodu defuzzifikace. Definování funkce příslušnosti (Membership function) sestává z počtu, tvaru, jména a rozsahu proměnných. Vybrat lze mnoho funkcí a tvarů od klasických „trojúhelníkových tvarů“, Gaussovu křivku a další. Poté je možné definovat inference typu „if-then“ (jestliže „fuzzy výrok“ – pak „fuzzy výrok“).

Fuzzy množiny jsou vhodné tam, kde je více kritérií ohodnocení. Např. hodnocení přístupu operátora zákaznické podpory ke klientovi, které sestává i z jeho schopnost vyřešit konkrétní problém, a následně je hodnoceno na úrovni 0 až 10. Jakou měsíční odměnu k platu dostane, pokud ta je v průměru 3 000,- Kč? Manažer oddělení definuje fuzzy pravidla (jestliže-pak) následovně: Jestliže je přístup operátora nevhodný, potom je odměna nižší. Jestliže je přístup operátora dobrý, potom je odměna průměrná. Jestliže je přístup výborný, potom je odměna vyšší. Jestliže problém nevyřeší, potom je odměna nízká. Jestliže problém vyřeší, potom je odměna vyšší. Funkce příslušnosti jsou následující, pro ilustraci problému je použita klasické trojúhelníková funkce příslušnosti (viz Obrázek níže):

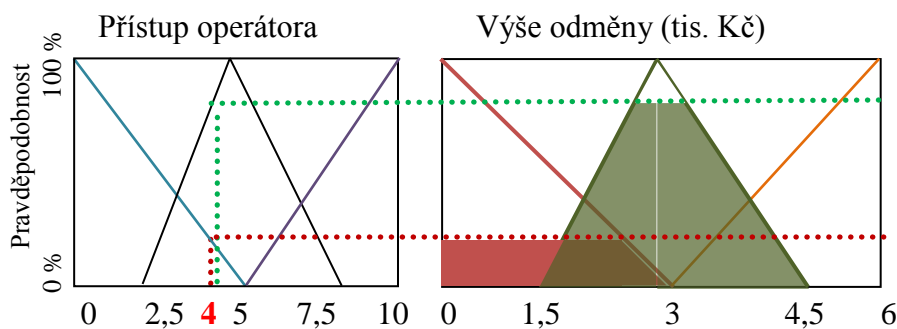
Obrázek 3: Funkce příslušnosti pro hodnocení odměny zákaznické podpory



Zdroj: Vlastní zpracování

Zpracování pravidel pak probíhá postupně, začneme prvním: Jestliže je přístup operátora nevhodný, potom je odměna nižší (viz Obrázek níže).

Obrázek 4: Zpracování prvního pravidla



Zdroj: Vlastní zpracování

Z obrázku je patrné, že operátor může dostat odměnu od nuly do 4,5 tisíce Kč, pokud jeho hodnocení přístupu je na úrovni 4 bodů z 10. V obrázku pro přístup operátora v bodě 4 je protnuta jak funkce pro nevhodný přístup, tak pro dobrý přístup, rozhodně není protnuta fialově znázorněná funkce pro výborný přístup. Výše odměny pak spadá do barevně ohraničeného intervalu, přičemž zde existuje zajímavý průnik mezi hodnotami 1,5 až 3 tisíce Kč. K danému schéma lze přidat i třetí dimenzi, a to je operátorova schopnost vyřešit problém. Takto lze graficky zpracovávat funkce příslušnosti a poté v programech typu MATLAB postupně využít tzv. Mandaniho nebo Sugeno metody defuzzifikace pro výsledné určení výše odměny.

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 34-70)

Kontrolní otázky

1. Jak lze fuzzy logiku použít v dataminingu?
2. Jak byste použili fuzzy logiku pro predikci cen akcií?
3. Jaké počítačové programy lze využít pro modelování fuzzy systémů?
4. Co je to vícekriteriální analýza, kde se používá a proč je vhodné používat v těchto případech fuzzy logiky?
5. Načrtněte fuzzy množiny s trojúhelníkovou funkcí příslušnosti pro hodnocení výkonu zaměstnance.
6. Definujte fuzzy pravidlo pro hodnocení zaměstnance.
7. Graficky zpracujte toto pravidlo.
8. Co je to Mandaniho metoda defuzzifikace?
9. Co je to Sugeno metoda defuzzifikace?
10. Co je to Gaussova křivka?

Odkaz na praktickou část

3.6 Průběžný test

2.12 Modelování časových řad pomocí fuzzy logiky

Klíčová slova

Zpoždění, časový úsek, perioda, vzorce

Cíle kapitoly

Cílem je pochopit proces modelování časových řad pomocí fuzzy logiky a pochopit systém fuzzifikace časových řad.

Výstupy z učení

- 16.1 zná základní pojmy a pravidla fuzzy logiky
- 16.6 modeluje časové řady pomocí fuzzy logiky

Abstrakt

Časové řady lze modelovat pomocí funkcí. První z nich je reprezentace pomocí „hustoty“. Stupeň příslušnosti vyjadřuje, jak moc je pravdivý výrok „prvek X je přibližně r“. Pro základní výpočty a využití fuzzy logiky se využívá toho, že funkce příslušnosti jsou lineární (trojúhelníková fuzzy čísla). Pomocí distribuce vyjadřujeme, jak moc je pravdivý výrok „prvek X je větší než fuzzy číslo A“, případně jak moc je pravdivý výrok „prvek X je menší než fuzzy číslo B“. To umožňuje modelovat ostrá fuzzy čísla, nekonečná fuzzy čísla, číselné intervaly apod.

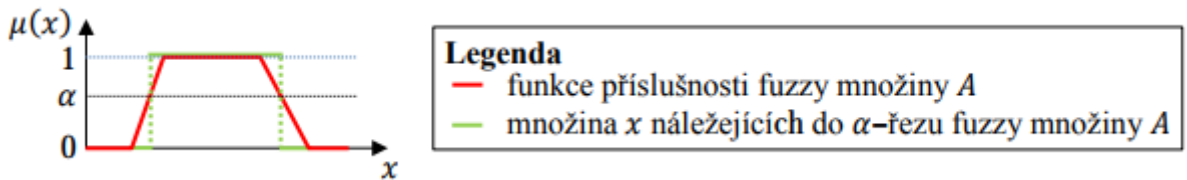
Nejčastěji se fuzzy logiky využívá pro nákup a prodej na finančních trzích. Jsou zde pokyny pro obchodníky, kteří hospodaří s portfolii a zajišťují např. podnik proti cyklickým výkyvům ekonomiky, či výkyvům měnových kurzů. Jde o příkazy typu silné doporučení pro prodej (krátká pozice), spíše prodej, držet, spíše koupě, doporučení koupit (dlouhá pozice). Pravidlo jestliže – pak (if-then) pro fuzzy systém definujeme interně. Např. Jestliže jdou-li ceny akcií dolů, nakupujeme bezpečné dluhopisy firem a výhodné státní dluhopisy.

U časových řad identifikujeme podobně jako u genetických algoritmů bloky a úseky v minulosti, tj. trendy a výkyvy časových řad. Při zapojení dalších pravidel jde o doplnění technické analýzy o další prvky ocenění aktiv na finančních trzích, jako je fundamentální

analýza a psychologická analýza. To však vyžaduje kvalitní přístup k datům z finančních trhů. Dále se využívá kombinace s umělými neuronovými sítěmi, kdy definujeme vstupní a výstupní proměnné, proces učení a hlavní roli hraje expertní nastavení sítě.

Používané tvary funkcí příslušnosti jsou S tvar (Γ -funkce), L tvar, trojúhelník (Λ -funkce), lichoběžník (Π -funkce), goniometrické funkce (sin, cos, sinh apod.), Gaussova funkce (Studentovo t rozdělení, F rozdělení, normální rozdělení), sigmoida a další nelineární funkce. Důležitým termínem pro fuzzy funkce příslušnosti je α -řez (α -cut, strong α -cut). Alfa řez je označení podmnožiny prvků x fuzzy množiny A definované na univerzu X , do které spadají všechny prvky, jejichž stupeň příslušnosti (definován funkcí) je vyšší nebo roven zvolené hodnotě α .

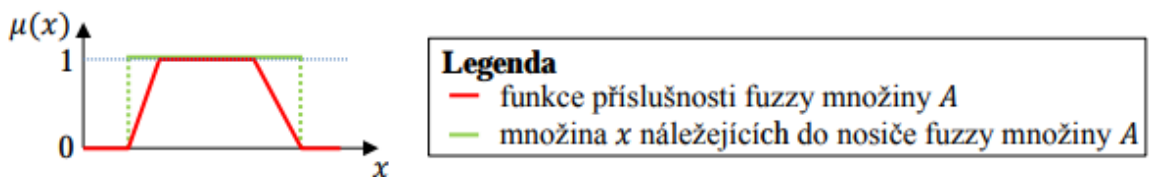
Obrázek 5: Alfa řez fuzzy množiny



Zdroj: Vlastní zpracování

Nosič fuzzy množiny A definované na univerzu X s prvky x je ostrá množina všech prvků s nenulovou příslušností do této fuzzy množiny.

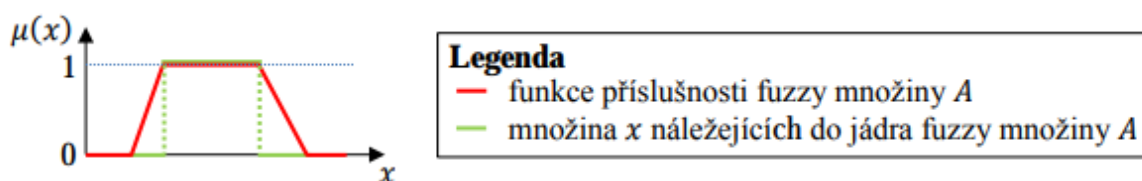
Obrázek 6: Nosič fuzzy množiny



Zdroj: Vlastní zpracování

Jádro fuzzy množiny A definované na univerzu X s prvky x je ostrá množina všech prvků jejichž příslušnost se rovná hodnotě 1, $\alpha = 1$.

Obrázek 7: Jádro fuzzy množiny



Zdroj: Vlastní zpracování

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 70-110)

Doporučená literatura

DOSTÁL, P., 2012. *Pokročilé metody rozhodování v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-798-7. (s. 26-45)

Kontrolní otázky

1. Co je to reprezentace fuzzy funkcí pomocí hustoty?
2. Co je to reprezentace fuzzy funkcí pomocí distribuce?
3. Jaké časové řady obvykle zpracovává fuzzy systém?
4. Jaké další nástroje se používají pro zpracování fuzzy časových řad?
5. Popište Γ funkci příslušnosti a uveďte příklad.
6. Popište Λ funkci příslušnosti a uveďte příklad.
7. Popište Π funkci příslušnosti a uveďte příklad.
8. Popište funkci příslušnosti tzv. sigmoidu a uveďte příklad.
9. Co je to alfa řez?
10. Definujte a nakreslete příklad nosiče fuzzy množiny.
11. Definujte a nakreslete příklad jádra fuzzy množiny.

Odkaz na praktickou část

3.6 Průběžný test

2.13 Umělé neuronové sítě (UNS): Seznámení se se základními pojmy v oblasti umělých neuronových sítí, uvedení pojmu perceptron

Klíčová slova

Neuron, perceptron, zpracování dat, vrstva, síť

Cíle kapitoly

Cílem je pochopit umělé neuronové sítě (UNS) a základní pojmy v oblasti umělých neuronových sítí. Cílem je pochopení pojmu perceptron.

Výstupy z učení

- 16.2 rozumí fungování umělých neuronových struktur

Abstrakt

Umělá neuronová síť je analytický nástroj, který pomocí algoritmů simuluje myšlenkové pochody člověka. Umělá neuronová síť je modelem biologické nervové soustavy, která je postavena na neuronech. Ty představují základní části biologického nervového systému a jsou přenašeči informací. Neurony mají tělo, které obsahuje jádro. Do tohoto těla vedou krátké výběžky, kterým říkáme dendrity. Ty vedou informaci směrem do buňky a z těla pak vystupuje jeden dlouhý výběžek. Tomu říkáme axon, který vede vzruch směrem z buňky. Umělý neuron jako jednotka umělé neuronové sítě je stavěna podobně jako biologický neuron (Vochozka et al. 2017).

Umělý neuron je na rozdíl od biologického definován konečným počtem vstupů. Umělý neuron má pak stejně jako biologický jeden výstup, který je schopen přenášet informaci dál. Umělé neurony podobně jako biologické jsou mezi sebou navzájem propojeny. To, jakým způsobem jsou umělé neurony napojeny, definuje architektura (či topologie) neuronové sítě. UNS je analytickým nástrojem vyhodnocujícím vstupní informace, které vyhodnotí nastavením architektury a poskytuje výstupy. Prvním modelem neuronové sítě

byl tzv. perceptron a jde o nejjednodušší neuronovou síť, která se skládá z jediného výkonného prvku (Vochozka et al. 2017).

UNS se tedy skládá ze vstupních neuronů, ty jsou dále v síti navázány na výstupní neurony. Každému spojení je přiřazena váha. Přenášená informace vstupními neurony je buď binární (ano/ne; 1/0), nebo bipolární (-1, 0, 1). Perceptronová síť je jednovrstvá. Princip adaptace (učení) vstupních neuronů se děje podle tzv. Hebbova pravidla. To je postaveno na minimalizaci Euklidovské vzdálenosti (kvadratické odchylky) požadované odezvy sítě a vnitřního potenciálu neuronu. Kromě Hebbovského učení je zde i princip adaptace s názvem ADALINE dle Widrowse a Hoffa. To je postaveno na kvantifikaci chyby neuronu. Při modelování UNS se využívá tzv. Rosenblattova perceptronu a Minsky-Papertova perceptronu. Tyto modely lze aplikovat na libovolnou aritmetickou nebo logickou funkci (Vochozka et al. 2017).

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 111-139)

VOCHOZKA, M. et al., 2017. *Využití neuronových sítí při komplexním hodnocení podniků*. Praha: C. H. Beck. ISBN 978-80-7400-642-5. (s. 67-80)

Kontrolní otázky

1. Co je to UNS?
2. Jaká je podobnost s biologickou neuronovou sítí?
3. Jaká je rozdílnost oproti biologické neuronové síti?
4. Definujte umělý neuron.
5. Co je to perceptron?
6. Jaká je povaha přenášené informace mezi umělými neurony?
7. Co je to Hebbovské učení?
8. Co je to ADALINE?
9. Jaké perceptronové modely se využívají?
10. Na co lze aplikovat UNS?

Odkaz na praktickou část

3.7 Procvičování základních principů teorie fuzzy množin a fuzzy logiky

2.14 Vícevrstvá neuronová síť a její parametry

Klíčová slova

Skrytá vrstva, trénink, neuron, perceptron, váhy

Cíle kapitoly

Cílem je pochopení principů vícevrstevných neuronových sítí a osvojení si pojmů jako je skrytá vrstva a proces učení.

Výstupy z učení

- 16.2 rozumí fungování umělých neuronových struktur

Abstrakt

Perceptronová síť využívala jednu vrstvu. Vícevrstvá neuronová síť využívá principu „učení s učitelem“ a vychází ze struktury lidského mozku. Ve výzkumu se minimálně využívá třívrstvá neuronová síť s jednou skrytou vrstvou. Tato síť je pak schopna realizovat libovolné zobrazení z prostoru R^n do prostoru R^m . R prostory jsou množiny uspořádaných n -tic reálných čísel. Zobrazení matice R^n do prostoru R^m je reprezentováno maticí A typu (m, n) . Existuje mnoho typů vícevrstevných neuronových sítí. První typ je s tzv. „dopředným šířením“, další typ je tzv. „rekurentní síť“, která je vhodná pro zpracování časových řad (Vochozka et al., 2017).

Struktura třívrstvé L1-L2-L3 neuronové sítě typu perceptron může být například tvořena první vrstvou (L1), tzv. vrstva vstupní. Do této vrstvy je vložen vstupní vzor. Za touto vrstvou následuje jedna či více skrytých vrstev (L2). Počet buněk v těchto vrstvách odpovídá složitosti realizovaného zobrazení $R^n > R^m$. K těmto vrstvám je navíc připojen jeden neuron s pevně danou výstupní hodnotou rovnou 1, reprezentující práh neuronu. Výstupní vrstva (L3) obsahovat jednu či více buněk dle rozměru R^m . Neuronová síť pracuje pouze s číselnými hodnotami. Jak vstupní, tak výstupní hodnoty je vhodné normovat do intervalu $[0; 1]$, případně do rozšířeného intervalu $[-1, 0, 1]$ (Vochozka et al., 2017).

Model s učením backpropagation se vyznačuje tím, že jednotlivé vrstvy sítě jsou plně propojeny vahami, vždy směřujícími od vstupu směrem k výstupu. Na hodnoty vah

jednotlivých spojů mezi umělými neurony nejsou kladena žádná omezení. Obvykle se využívají sítě s jednou skrytou vrstvou. Neurony vstupní vrstvy jsou pasivní, protože pouze zobrazují příslušnou vstupní souřadnici předloženému vzoru. Neurony skryté a výstupní vrstvy jsou neurony aktivní, protože realizují výpočty na základě vstupních vah, které do nich směřují. Každý neuron z těchto dvou vrstev má stejný výpočetní algoritmus pro zpracování vstupu. Na těch je nejprve proveden jejich vážený součet x_j reprezentující vnitřní stav neuronu j z vrstvy J , přičemž váhy w_{ij} jsou zde váhy příslušných spojení mezi neurony i předchozí vrstvy I a sledovaným neuronem j z vrstvy J . Hodnoty tohoto váženého součtu mohou být vysoké a značně se vymykající požadovanému intervalu výstupu $[0; 1]$. Za tím účelem je na vážený součet uplatněna výstupní funkce $s(x)$. Tato funkce je obvykle nelineární a spojitá, například jde o Gaussovu křivku a tzv. sigmoidální funkci $s(x)=1/[1-e^{(-x)}]$. Hodnota výstupní funkce $s(x)$ je potom výstupem neuronu, který je dále šířen prostřednictvím spojů do neuronů vyšší vrstvy. Výstupem celého modelu jsou potom výstupy buněk výstupní vrstvy (Vochozka et al., 2017).

Produkční fáze činnosti sítě (vybavování) vychází z předložení vstupního vzoru na buňky vstupní vrstvy a postupného šíření předložených hodnot prostřednictvím sítě až k vrstvě výstupní. Nejprve po předložení vzoru na vstupní vrstvu jsou tedy vypočítány vážené součty a výstupní funkce vrstvy skryté a na základě jejich výstupů rovněž vážené součty a výstupy vrstvy výstupní. Tím produkční fáze činnosti sítě končí. Po produkční fázi sítě následuje fáze adaptace parametrů sítě (vah jednotlivých spojů) tak, aby se výstup sítě přiblížil výstupu požadovanému. Pro tento krok je k dispozici několik učebních algoritmů, přičemž nejpoužívanějším z nich je algoritmus označovaný jako zpětné šíření chyby (back propagation of error). Algoritmus získal svůj název podle své základní myšlenky: síť nejprve provede vybavování vzoru, po kterém je porovnán požadovaný výstup sítě. Na každé buňce výstupní vrstvy tak vzniká rozdílová chyba (skutečný výstup neuronu minus výstup požadovaný). Tato chyba je následně zpětně šířena od výstupní vrstvy, přes skrytou až směrem k vrstvě vstupní. Proces adaptace vah je charakterizován obecným koeficientem α , který udává míru, s jakou bude příslušná chyba eliminována. Adaptační proces probíhá po každém předložení vstupního vzoru síti ve fázi učení, tedy samostatně pro každý vstupní vzor ze vstupní množiny. Zejména z důvodu obvykle většího počtu vstupních vzorů a rovněž z důvodu existence koeficientu α je i v tomto modelu nutno provádět učení iterativně opakovaným předkládáním vstupní množiny a postupným snižováním chyby na

výstupu modelu. I zde je zásadní, jakým způsobem bude voleno kritérium pro ukončení učebního procesu. Použita byla varianta ukončení učení po dosažení dostatečně malého rozdílu v celkové chybě E mezi vybranými iteracemi. Druhou možností učení vícevrstvé sítě je využití mechanismu RBF (Radial Basis Function). Formálně má síť stejný tvar jako síť s učením backpropagation, vnitřní postupy jsou však zcela jiné. Vstupní vrstva má opět pouze formální význam a umožňuje modelu předložit vstupní vzory (vektory). Výstupy vstupní vrstvy se zpracovávají ve vrstvě skryté, kde je použita nelineární funkce ϕ , kterou je nejčastěji Gaussova funkce. V uvedeném vzorci je důležitý vektor výstupu neuronů vstupní vrstvy, který obsahuje μ_j (střední hodnoty Gaussova rozložení pro j -tý neuron skryté vrstvy) a σ_j (šířka neboli směrodatná odchylka střední hodnoty Gaussova rozložení pro j -tý neuron skryté vrstvy). Právě hodnoty μ_j a σ_j jsou u této RBF sítě nastavitelné hodnoty. Dalším nastavitelným parametrem jsou hodnoty vah mezi skrytou a výstupní vrstvou w_{jk} . V procesu učení je tedy nutné optimalizovat parametry sítě μ_j , σ_j a w_j (Vochozka et al. 2017).

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 140-182)

VOCHOZKA, M. et al., 2017. *Využití neuronových sítí při komplexním hodnocení podniků*. Praha: C. H. Beck. ISBN 978-80-7400-642-5. (s. 80-106)

Kontrolní otázky

1. Jaký je rozdíl mezi jednovrstvou a vícevrstvou UNS?
2. Popište prvky třívrstvé UNS.
3. Co znamená zobrazení z prostoru R^n do prostoru R^m ?
4. Popište model UNS s učením backpropagation.
5. Co je to sigmoidální funkce?
6. Co je to produkční fáze činnosti UNS?
7. Jak probíhá adaptace UNS?
8. Co je to mechanismus RBF (Radial Basis Function)?
9. Jaký je rozdíl mezi RBF a principem UNS s backpropagation?

10. Jaký je rozdíl mezi principem UNS s dopředným šířením a tzv. „rekurentní sítí“?

Odkaz na praktickou část

3.7 Procvičování základních principů teorie fuzzy množin a fuzzy logiky

2.15 Uvedení příkladů aplikace UNS pro investiční rozhodování, odhady cen výrobků a množstevní odhady

Klíčová slova

Samoorganizace, mapa, shluková analýza, časové řady

Cíle kapitoly

Cílem je pochopení příkladů aplikací UNS pro investiční rozhodování, odhady cen výrobků a množstevní odhady.

Výstupy z učení

- 16.7 modeluje časové řady pomocí umělých neuronových struktur

Abstrakt

Stejně jako v ekonometrickém regresním modelu i u umělých neuronových sítí, je vhodné promyslet celý experiment dopředu. Náhodné použití variant standardních modelů neuronových sítí nevede vždy k žádoucím výsledkům. Proto byl navržen vícestupňový mechanismus pro zpracování ekonomických dat. Pro ekonomická data je vhodná umělá neuronová síť s učícím Kohonenovým algoritmem. Použití Kohonenovy mapy spočívá v nastavení pomocí trénovací množiny. 2D vrstva sítě je adaptována pouze na základě vzájemných podobností vstupních vzorů z hlediska jejich souřadnic. V této vrstvě tedy můžeme sledováním počtu vzorů přiřazených k jednotlivým neuronům detekovat shluky neuronů zastupujících vzory odlišné od ostatních. Shluky jsou vcelku jasně identifikovatelné, jejich hranice obvykle tvoří neurony zastupující jen velmi malý počet vstupních vzorů (Vochozka et al. 2017).

V dalším kroku nastavení se celý proces učení opakuje s redukovanou trénovací množinou a novou instancí Kohonenovy mapy. Základní myšlenkou je využití interních mechanismů Kohonenovy mapy (zejména pak systém utlumení často vybíraných neuronů v 2D vrstvě) tak, aby síť v každém kroku zpřesnila své klasifikační schopnosti. Interní mechanismy zde zaručí, že plocha 2D vrstvy bude plně adaptována (všechny neurony budou mít nastavené váhy interního vektoru zastupovaného neuronem) (Vochozka et al. 2017).

Tento proces postupné redukce trénovací množiny je možné chápat i jako postupné zvyšování citlivosti sítě na odchylky v trénovací množině. Tomu odpovídají i experimentální výsledky, kdy se během vícekrokového procesu postupně eliminují zařaditelné vzory a zůstávají vzory přiřazené k neuronům zastupujícím i další vzory jinak kategorizované. Obecně zde můžeme hovořit o určité formě hierarchického učení, jehož výsledkem je posloupnost nastavených Kohonenových map, z nichž každá je učena na menší trénovací množině než předchozí (Vochozka et al. 2017).

Pro aplikaci uvedeného vícekrokového mechanismu jsou zásadní dvě kritéria: kritérium naučení sítě v každém kroku hierarchického učení; a kritérium celkové schopnosti množiny naučených modelů správně klasifikovat trénovací a později i testovací množinu vzorů. V produkční fázi se způsob klasifikace liší podle toho, zda se nacházíme v posledním modelu hierarchické struktury či nikoliv. Pro všechny ostatní modely je v produkční fázi použito kritérium absolutní shody. Testovací vzor je správně zařazen, pokud je přiřazen k neuronu zastupujícímu pouze vzory klasifikované během učení do stejné kategorie. Pro poslední model v hierarchické struktuře je pak použit pravděpodobnostní přístup, kdy je vzor zařazen do kategorie, která v daném neuronu převažovala v procesu učení (Vochozka et al. 2017).

Hierarchický model (více k modelu viz povinná literatura) je charakterizován sadou parametrů. Parametry mají velmi rozdílný charakter (celé číslo, text, číslo s desetinnou částí). Proto bylo rozhodnuto o doplnění modelu o optimalizační vrstvu, schopnou tyto parametry testovat a zvolit jejich nejlepší možnou hodnotu. Optimalizačních nástrojů existuje celá řada, obvykle však jde o uzavřená řešení, často vázaná na jediný typ dat. Vhodné je užití tzv. genetických algoritmů. Tato skupina nástrojů vychází z několika základních a jednoduchých východisek, které (jak již název napovídá) mají svůj základ v přírodě a evolučním vývoji. Prvním pravidlem je, že přežívá nejlepší. Systém generuje různá řešení (v našem případě nastavení parametrů) označovaná jako chromozomy. Tato řešení jsou následně testována. To vede k tomu, že jsou vybrány pouze chromozomy s nejlepším ohodnocením (Vochozka et al. 2017).

Podobně jako v přírodě je nutné nastavit mechanismus generování nových chromozomů, aby se populace mohla vyvíjet směrem k lepším hodnotám. To je zachyceno evaluační funkcí v. Jedná se o proces vybírající z aktuální populace náhodný chromozom, jehož náhodný atribut (hodnota parametru) se změní. Tím vznikne nový chromozom mírně odlišný od svého originálu, který se zařadí do populace. Procento mutace v populaci je

spolu s maximální velikostí populace hlavním parametrem optimalizačního systému. Jeho činnost probíhá v iterativních krocích obsahujících mutaci, vyhodnocení populace a odstranění nejhorších chromozomů. Tyto kroky se označují jako generace. Již po cca 30 generacích jsou další změny jen velmi malé (Vochozka et al. 2017).

Studijní literatura

Povinná literatura

VOCHOZKA, M. et al., 2017. *Využití neuronových sítí při komplexním hodnocení podniků*. Praha: C. H. Beck. ISBN 978-80-7400-642-5. (s. 107-130)

Kontrolní otázky

1. Popište rámec zpracování ekonomických dat pro UNS aplikace.
2. Co jsou to samoorganizující se mapy?
3. Co je to Kohenova mapa?
4. Popište princip trénování UNS.
5. Co je to testovací množina?
6. Co je to koeficient alfa v Kohenově mapě?
7. Co je to evaluační funkce?
8. Jak je možné přidat genetické algoritmy do UNS?
9. Co je to koeficient k f Kohenově mapě?
10. Jak je definována velikost Kohenovy vrstvy?

Odkaz na praktickou část

3.8 Modelování časových řad pomocí fuzzy logiky

2.16 Uvedení příkladů aplikace UNS pro odhad cen nemovitostí, oceňování bonity klienta atd.

Klíčová slova

Finanční analýza, situační analýza, zdraví podniku, hodnocení podniku, rating

Cíle kapitoly

Cílem je pochopení příkladů aplikací UNS pro odhad cen nemovitostí, oceňování bonity klienta a osvojení si pokročilých metod finanční analýzy.

Výstupy z učení

- 16.7 modeluje časové řady pomocí umělých neuronových struktur

Abstrakt

Predikčních modely v ekonomii a podnikové ekonomice využívajících umělé neuronové sítě na rating podniků (bonita, riziko bankrotu). Jde o důležitou informaci pro vlastníky, dodavatele, investory, banky a pojišťovny. Pokud jsou vstupní data věrohodná, pak UNS dokáže rychle a snadno poskytnout odhad toho, jak se bude společnost chovat v recesi nebo při stagnaci hrubého domácího produktu. Modely umělých neuronových sítí předpovídají také obchodní úspěch (tržby z produkce vlastního zboží a služeb, investiční zisky apod.) firem a pomáhají vyhodnocovat start-up projekty. Investoři, banky a pojišťovny využívají převážně statistické techniky (strukturální regresní modely, modely volatility časových řad apod.). Z ostatních technik využívají např. rozhodovací stromy, genetické algoritmy a postupně i umělé neuronové sítě. UNS mají lepší predikční vlastnosti než rozhodovací stromy, diskriminační analýzy a lze je použít i pro predikci vývoje finančního selhání malých a středních podniků (Vochozka et al. 2017).

V závislosti na celosvětové finanční krizi roku 2008 se rozšířilo uplatnění systému umělých neuronových sítí. Model UNS lze použít na předpovědi bankrotu obchodních bank, finanční situaci podniků v jednotlivých hospodářských odvětvích (primární sektor, stavebnictví apod.). Uplatňují se kombinace genetického algoritmu a umělých neuronových sítí. Neuronové sítě byly v minulosti vytvářeny jako jednoúčelový hardware, nebo na míru

dělaný software, který byl schopen řešit jeden problém. V současné době jsou neuronové struktury vytvářeny především v podobě univerzálního software. Tyto programy jsou schopny zvládnout různé úlohy a nabízejí různá nastavení UNS. Tento vývoj je možný díky pokrokem v oblasti výpočetní techniky – výpočetního výkonu (Vochozka et al., 2017).

Bonitní diagnostické modely UNS jsou založeny na ekonomických teoretických poznatcích a jsou často doplněny o empirické poznatky finančních analytiků. Tyto modely klasifikují podnik podle stupně finančního zdraví. Index bonity je založen na diskriminační analýze a skládá se z šesti poměrových ukazatelů (cash flow, cizí zdroje, aktiva/cizí zdroje, zisk/aktiva, zisk/výnosy, zásoby/výnosy, výnosy/aktiva). Záporné hodnoty indexu znamenají bankrotní podnik. Hodnoty indexu v intervalu $(-\infty, -2)$ naznačují, že jde o extrémně špatnou ekonomickou situaci podniku, hodnoty od mínus dvou (včetně) do mínus jedné $(-2, -1)$ naznačují velmi špatnou ekonomickou situaci a hodnoty od mínus jedné do nuly $(-1; 0)$ naznačují špatnou ekonomickou situaci. Naopak bonitní podniky se vyznačují kladnými hodnotami indexu bonity. Hodnoty indexu v intervalu $(0; 1)$ označují podnik v problematické ekonomické situaci, hodnoty v intervalu $(1; 2)$ v dobré ekonomické situaci, hodnoty v intervalu $(2; 3)$ ve velmi dobré ekonomické situaci a hodnoty od tří a výše pak za podnik v extrémně dobré ekonomická situaci.

Grünwaldův index analyzující finanční situaci podniku používá naprosto odlišný systém vah od ostatních bonitních indexů. Dvě ze šesti užitých vah jsou proměnlivé (průměrná úroková míra z úvěrů a zdaněná úroková míra z úvěrů), ostatní jsou konstantní. Užité konstantní váhy přiřazují největší důležitost poměru „cash flow / dlouhodobé úvěry“, dále „čistý pracovní kapitál / zásoby“, „(finanční majetek + krátkodobé pohledávky) / krátkodobý cizí kapitál“ a „zisk / nákladové úroky“ (Vochozka et al. 2017).

Kralickův Quick test je využíván především v německy mluvících zemích. Na základě ukazatelů z jednotlivých oblastí finanční analýzy udává celkové hodnocení všech druhů podniků. Kralickův test je vhodný zejména pro obchodní společnosti, uplatní se ale také ve výrobních společnostech. Původní varianta je koncipována jako jednorozměrný známkovací test, kde vybraným ukazatelům z jednotlivých oblastí finanční analýzy (finanční, likvidační, rentabilní a výnosové) jsou přiřazeny známky, jejichž součet udává celkové hodnocení firmy. Ukazatel Rychlého Kralickova testu K1 se vypočte jako prostý aritmetický průměr známek obdržných za hodnoty jednotlivých ukazatelů. Významný rozdíl mezi původní a modifikovanou variantou Rychlého Kralickova testu je fakt, že

výsledné hodnoty jednotlivých ukazatelů nejsou poměřovány s konkrétními, předem určenými hodnotami. Výsledné hodnoty ukazatelů jsou poměřovány s percentily oborových hodnot. Tamariho model využívá sedmi podílových ukazatelů, které jsou bodově ohodnocovány. Podnik je poté hodnocen dle součtu získaných bodů. Celkový Tamariho index (T_a) je vyjádřen součtem získaných bodových hodnot (Vochozka et al. 2017).

Studijní literatura

Povinná literatura

VOCHOZKA, M. et al., 2017. *Využití neuronových sítí při komplexním hodnocení podniků*. Praha: C. H. Beck. ISBN 978-80-7400-642-5. (s. 131-142)

Kontrolní otázky

1. Které instituce potřebují provádět rating firem a proč?
2. Jaké další metody se k UNS přidávají k zvýšení predikční schopnosti?
3. Na čem jsou založeny bonitní modely UNS?
4. Popište index bonity.
5. Popište Grünwaldův index.
6. Co je to Kralickův Quick test?
7. Definujte Tamariho model.
8. Jaký software lze použít pro modelování UNS?
9. Proč se v minulém století nevyužívali hojně UNS?
10. Jaký je rozdíl mezi rozhodovací stromy a UNS?

Odkaz na praktickou část

3.8 Modelování časových řad pomocí fuzzy logiky

2.17 Modelování časových řad pomocí UNS

Klíčová slova

Bayesovský klasifikátor, algoritmus, pravidla sítě, učení

Cíle kapitoly

Cílem je pochopení procesu modelování časových řad pomocí UNS a osvojení si principů UNS jako je učení s učitelem a učení bez učitele.

Výstupy z učení

- 16.2 rozumí fungování umělých neuronových struktur
- 16.7 modeluje časové řady pomocí umělých neuronových struktur

Abstrakt

Modelování časových řad pomocí UNS využívá klasických vícevrstevných neuronových sítí typu multilayer perceptron. Ty používají hodnoty souřadnic vstupních vektorů pro výpočty vah. Použit je možné i tzv. asociativní paměť (jednovrstvá neuronová síť), či Kohonenovy mapy pro vizualizaci podobnosti vstupních vektorů a jejich shlukování. Při použití technik umělé inteligence obvykle využíváme dostupná data pro nastavení zvoleného typu modelu, tento proces je obecně označován jako učení. Jeho cílem je získat nastavený model, který je schopen určité míry generalizace, tedy adekvátní odezvy i na vzory, na které nebyl učen. I z tohoto pohledu lze nahlížet na vstupní data a jejich možné využití při nastavení modelu. Hovoříme zde o tzv. učení s učitelem a bez učitele (Vochozka et al. 2017).

Učení bez učitele v podstatě znamená, že máme k dispozici pouze nehodnocená vstupní data, která hodláme nějakým způsobem analyzovat. Učení bez učitele se typicky používá v počátečních fázích analýzy dat, kdy zkoumáme strukturu dat, jejich vzájemné vztahy a charakter.

Typickým příkladem je použití shlukové analýzy, např. algoritmu ART, či Kohonenových map. Ty mají na vstupu vzory (vektory), které na základě vnitřních pravidel zařazují do jednotlivých výstupních kategorií. Charakteristika těchto kategorií není předem předeepsána a je výsledkem zpracování. Chod těchto algoritmů je možno ovlivnit jejich parametry, tedy

např. citlivostí na rozdíly ve vstupních datech. Čím bude tato citlivost vyšší, tím více výstupních kategorií či shluků systém vytvoří (Vochozka et al. 2017).

Při učení s učitelem systému předložíme nejenom vstup, ale rovněž požadovaný výstup u jednotlivých vzorů (či příkladů zobrazení vstup – výstup). Tyto dvojice (vstup, požadovaný výstup) algoritmus používá pro své učení, a tedy formování vnitřních pravidel, či nastavení parametrů modelu. Metod na tomto principu je několik různých druhů. Typickým příkladem je neuronová síť s algoritmem učení pomocí zpětného šíření chyby (backpropagation). Příkladem učení s učitelem je také tzv. Bayesovský klasifikátor, který zařazuje vstupní vektory do výstupních kategorií podle výskytu těchto kategorií v okolí příslušného vstupního vektoru. Bylo-li např. v okolí vstupního vzoru klasifikováno 9 příkladů do jedné kategorie a 1 příklad do druhé a obě kategorie jsou ve vstupní množině zastoupeny stejně, potom můžeme říci, že Bayesovský klasifikátor přiřadí zkoumaný vektor do první kategorie s pravděpodobností 90 %. Pro zpracování ordinálních vstupních dat je tento model mírně upraven (Vochozka et al. 2017).

Aby bylo možné ověřit kvalitu nastavení modelu, je nutné ještě před procesem učení rozdělit vstupní množinu příkladů na dvě disjunktní podmnožiny. Trénovací, na které se příslušný systém učí, a testovací, která je systému předložena pro ověření správnosti jeho reakcí i na předem neznámé příklady. Rozdělení do těchto dvou množin (trénovací a testovací) je často provedeno podle rovnoměrného rozložení a obvykle lze použít rozdělení vstupní množiny na tyto dvě části. V případě nedostatku vstupních vzorů je více vzorů umístováno do trénovací množiny, aby bylo zajištěno lepší naučení sítě, zbytek je množinou testovací. Mezi metody zpracování nominálních vstupních dat učením s učitelem lze zařadit rovněž rozhodovací stromy, např. s algoritmem TDIDT. Algoritmy rozhodovacích stromů jsou založeny na vytvoření stromu, v jehož uzlech jsou jednotlivé atributy (souřadnice) vstupního vektoru a hrany z uzlu vycházející označují všechny hodnoty, kterých příslušný atribut nabývá v celé trénovací množině. Koncové (listové) uzly stromu pak označují hodnoty požadovaného výstupu, který je často tvořen jediným atributem. Těchto hodnot může být i více než 2. Pokud bychom vyšli pouze z této definice je zřejmé, že pro jeden úkol je možno vytvořit celou řadu rozhodovacích stromů. Aby byl vytvořený rozhodovací strom sestaven efektivně, vychází se při jeho konstrukci z výpočtu entropie pro jednotlivé vstupní atributy v závislosti na požadovaných hodnotách výstupního atributu. Jinak řečeno jsou nejdříve vybrány ty atributy pro uzly rozhodovacího

stromu, které přinášejí nejvyšší hodnotu diskriminační informace umožňující odlišit od sebe jednotlivé výstupní kategorie vstupních vzorů (Vochozka et al. 2017).

Studijní literatura

Povinná literatura

VOCHOZKA, M. et al., 2017. *Využití neuronových sítí při komplexním hodnocení podniků*. Praha: C. H. Beck. ISBN 978-80-7400-642-5. (s. 143-152)

Kontrolní otázky

1. Popište síť typu multilayer perceptron.
2. Co je to princip učení UNS?
3. Jaký je rozdíl v učení UNS s učitelem a bez učitele?
4. Co je to Bayesovský klasifikátor?
5. Proč rozdělujeme trénovací a testovací vzorek?
6. Jak fungují algoritmy rozhodovacích stromů?
7. Co jsou to vstupní vzory?
8. Jakou transformaci dat využíváme pro UNS?
9. Uveďte příklad aplikace UNS pro manažerské rozhodování.
10. Uveďte příklad aplikace UNS v ekonomickém výzkumu.

Odkaz na praktickou část

3.9 Procvičování základních principů umělých neuronových sítí

2.18 Pokročilé modelování časových řad pomocí UNS

Klíčová slova

Predikce, řízení zdrojů, logistika, prodejní data

Cíle kapitoly

Cílem je pochopení pokročilého modelování časových řad pomocí UNS při použití vícevrstevných sítí.

Výstupy z učení

- 16.2 rozumí fungování umělých neuronových struktur
- 16.7 modeluje časové řady pomocí umělých neuronových struktur

Abstrakt

Ve společenských vědách jsou UNS aplikovány na oblasti ekonomie (makroekonomie, např. predikce GPN a modelu spotřeby; mikroekonomie, např. stanovení burzovních indexů; klasifikace výsledků akciové společnosti; zpracování marketingových studií), sociologie (např. oblast zpracování sociologických informací, predikce a klasifikace sociologických dat), psychologie (např. modulace lidského chování, simulace lidského vnímání, přenosu informací, procesy učení včetně predikce chování jednotlivce i skupin lidí). Dále jsou UNS využity na dílčí problematiky jako marketing a prodej, skladové hospodářství, prodej a výdej zboží, vnitropodniková doprava, klasifikuje i jednotlivé procesy skladování včetně balení výrobků, uvádí využití neuronových sítí i u základních manažerských funkcí podniku, jako je plánování, organizování a kontroly, zabývá se i samotným řízením pracovních operací i celého výrobního procesu (Vochozka et al. 2017).

UNS není možné využívat na veškeré podnikové operace. V některých aplikacích zůstávají přednostní volbou pro jejich řešení algoritmické informační technologie (např. pro provádění numerických výpočtů). Je tedy zřejmé, že umělé neuronové sítě se využívají v případech, kdy není znám ani přibližný, natož přesný algoritmus řešení. V současné době jsou využívány zejména pro konstrukci podpůrných podnikových softwarů. Pro umělé neuronové sítě v podnikovém řízení jsou vhodné takové úkoly, které vyžadují řešení

pomocí adaptivních metod (učící se expertní systémy, predikce časových řad, analýza i syntéza přirozené řeči apod.) (Vochozka et al. 2017).

Nejčastější výtky k umělým neuronovým sítím jsou směřovány k retrospektivnosti použitých dat. Pro predikci se používají starší data, která již nemusí reflektovat aktuální stav. Jako možnou nevýhodou se také jeví skutečnost, že může dojít k tzv. přetrénování sítě, kdy vlivem opakovaného testování ztrácí neuronová síť schopnost klasifikace a generalizace poznatků a výstupů. Tomuto problému se dá předejít použitím různých testovacích setů obsahujících odlišná vstupní data. Tím by měla být zachována samoučící se schopnost neuronové sítě. Nakonec se jako největší problém neuronových sítí může jevit jejich vlastnost označovaná jako tzv. Black box. Neuronová síť je totiž schopna predikovat velkou řadu jevů, ale již není schopna provést klasifikaci svého hodnocení (Vochozka et al. 2017).

Studijní literatura

Povinná literatura

VOCHOZKA, M. et al., 2017. *Využití neuronových sítí při komplexním hodnocení podniků*. Praha: C. H. Beck. ISBN 978-80-7400-642-5. (s. 153-166)

Kontrolní otázky

1. Uveďte příklad aplikace UNS na vybrané makroekonomické časové řadě.
2. Uveďte příklad aplikace UNS pro analýzu mikroekonomického trhu.
3. Uveďte příklad aplikace UNS pro zpracování marketingových studií.
4. Uveďte příklad aplikace UNS pro predikci spotřeby materiálu.
5. Uveďte příklad aplikace UNS pro zhodnocení portfolia.
6. Uveďte příklad aplikace UNS pro simulaci lidského vnímání.
7. Uveďte příklad aplikace UNS pro manažerské rozhodování.
8. Co je to problém retrospektivnosti použitých dat v UNS?
9. Popište, kdy dojde k přetrénování sítě?
10. Dokáže UNS provést klasifikaci svého hodnocení?

Odkaz na praktickou část

3.9 Procvičování základních principů umělých neuronových sítí

2.19 Genetické algoritmy (GA): Seznámení se základy genetiky, analogií mezi přírodou a matematickým popisem, umožňující řešení problémů rozhodování

Klíčová slova

Evoluce, optimalizace, proces, chromozóm, mutace

Cíle kapitoly

Cílem je pochopení genetických algoritmů a seznámení se základy genetiky. Cílem je osvojit si analogií mezi biologickým a matematickým popisem GA, které umožňují řešit náročné problémy v rozhodování.

Výstupy z učení

- 16.3 rozumí genetickým algoritmům

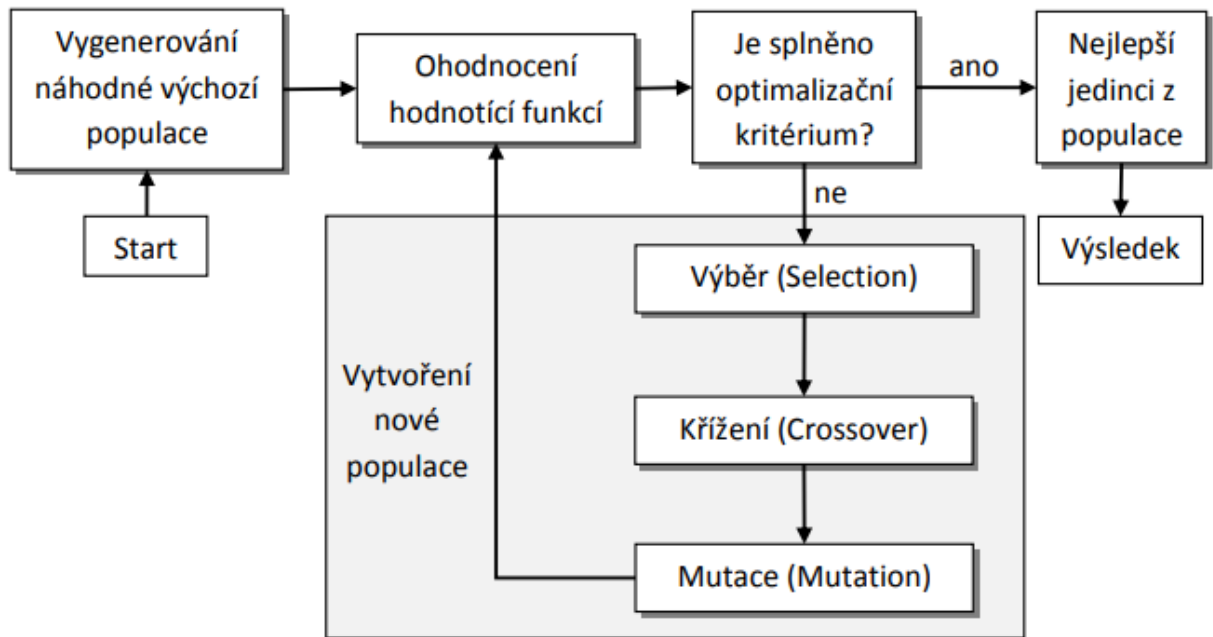
Abstrakt

Genetické algoritmy (GA) jsou založeny na darwinovském principu evoluce. GA napodobují evoluční procesy známé z biologie jako je dědičnost, mutace, přirozený výběr a křížení. Zakladatelem této vědní disciplíny je John Holland. GA slouží pro řešení optimalizačních problémů. Díky tomu, že jsou založeny na principech genetiky a mechanismech přirozeného výběru jsou na rozdíl od matematických optimalizačních metod velmi jednoduché.

Principy a vlastnosti těchto algoritmů jsou definovány pomocí těchto pojmů: chromozóm, gen, populace a fitness hodnota. Chromozóm je řetězec informací, který v sobě nese vlastnosti a chování jedince. Nejčastěji jde o řetězec nul a jedniček, použít můžeme i reálná čísla, matice, vektory i křivky. Gen je nejmenší a dále nedělitelná část chromozómu.

Populace je skupina jedinců popsaných svými chromozómy v rámci jedné generace. Fitness hodnota je číselné vyjádření kvality jedince. Obvykle jde o reálné číslo v rozsahu od 0 do 1 (0–100 %). Činnost genetických algoritmů je velmi jednoduchá, používají se 3 základní operace: selekce (výběr), křížení a mutace.

Obrázek 8: Proces genetického algoritmu



Zdroj: Vlastní zpracování

Inicializace (Start) znamená vytvoření počáteční populace. Ta může být sestavena záměrnou selekcí, nebo může být vygenerována zcela náhodně. Selektce tedy slouží k výběru jedinců z populace. Výběr „rodičů“ by měl respektovat pravidla přirozeného výběru dle Darwinovy teorie. Využívá se vzorce $p_i = f_i / \text{Suma všech } (fitness \text{ hodnot } f_i)$. Díky tomu je pravděpodobnost rodičovství vyšší pro schopnější jedince (dáno hodnotící funkcí – posouzením kvality jedince). Nebo dochází k ořezávání populace, kdy méně schopní jedinci nejsou rodiči vůbec. Uplatňuje se také zcela náhodný výběr, nebo i metoda turnaje náhodně vybrané skupiny jedinců mezi sebou.

Křížení navazuje na selekci. Rodiče si vymění část svého genetického kódu, tedy část chromozómu. Využívá se jednobodového a vícebodového křížení. Mutace je poslední fází. U každého jedince z nové generace procházíme celý chromozóm a s velmi malou pravděpodobností měníme hodnoty některých genů z 0 na 1 a obráceně. Význam mutace je v tom, že se může v dané generaci objevit vlastnost, kterou dosud žádný jedinec neměl a nemohl ji tedy předat potomkům.

Opakováním cyklu se vytvořené potomstvo stává novou generací a starou generaci již nebereme v úvahu. Jedná se o nejjednodušší generační strategii. Celý cyklus se opakuje,

dokud není splněna ukončovací podmínka. Ukončení algoritmu může vyvolat například maximální počet cyklů, nalezení uspokojivé mutace či generace atd.

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 183-197)

Kontrolní otázky

1. Co jsou GA?
2. Kdo se zabýval studiem GA?
3. Čím se zabývá genetika?
4. Co je to chromozóm?
5. Popište proces genetického algoritmu.
6. Co je to výběr?
7. Co je to křížení?
8. Co je to selekce?
9. Co je to mutace?
10. Co je to hodnotící funkce?

Odkaz na praktickou část

3.10 Modelování časových řad pomocí UNS

2.20 Uvedení použití GA v oblasti optimalizace široké palety problémů – optimalizace investiční strategie, řízení výroby, řezných plánů, aproximace křivek, řešení problému obchodního cestujícího, využití shlukové analýzy apod.

Klíčová slova

Délka cesty, křížení, optimum, sub-optimum

Cíle kapitoly

Cílem je pochopení použití GA v oblasti optimalizace široké palety problémů – optimalizace investiční strategie, řízení výroby, řezných plánů, aproximace křivek, řešení problému obchodního cestujícího a shlukové analýzy.

Výstupy z učení

- 16.3 rozumí genetickým algoritmům

Abstrakt

Problém obchodního cestujícího (Travelling Salesman Problem - TSP) je jedním z nejznámějších kombinatorických problémů. Od roku 1759 se úlohou tohoto typu zabýval švýcarský matematik Leonhard Euler, který se zajímal o vyřešení „problému jezdce“. Řešení popisuje způsob, jakým jezdec při své cestě navštíví každé z 64 polí na šachovnici právě jednou.

Problém obchodního cestujícího lze formalizovat a definovat. Je dána množina měst $\{c_1, c_2, \dots, c_N\}$ a pro každou dvojici $\{c_i, c_j\}$ navzájem různých měst je dána vzdálenost $d(c_i, c_j)$. Úkolem obchodního cestujícího je najít uspořádání π všech měst, které minimalizuje hodnotu délky cesty, kterou obchodní cestující ujede, než projede všemi městy v pořadí daném permutací π a vrátí se do výchozího města. Výsledná celková délka cesty je chápána jako cena cesty. Pojem cena tedy můžeme chápat podle potřeby – jako skutečnou vzdálenost, jako čas potřebný k ujetí z města do města nebo jako ekonomické náklady na cestu.

Problém obchodního cestujícího je výpočetně komplikovaný. Je třeba znát třídy výpočtové složitosti. K tomu se využívá pojem Turingův stroj. Třída složitosti P obsahuje všechny úlohy, jejichž řešení lze nalézt deterministickým Turingovým strojem v polynomiálním čase. Třída složitosti NP je množina problémů, které lze řešit v polynomiálně omezeném čase na nedeterministickém Turingově stroji. To je počítač, který umožňuje v každém kroku rozvětvit výpočet na n větví, v nichž se posléze řešení hledá současně. NP-úplné problémy jsou takové nedeterministicky polynomiální problémy, na které jsou polynomiálně redukovatelné všechny ostatní problémy z NP. NP-úplné problémy jsou zpravidla ty nejtěžší úlohy z NP třídy.

Problém obchodního cestujícího patří do třídy NP-úplných problémů. Jednoduchý algoritmus prohledávající všechny cesty musí projít všechny permutace množiny 1 až N , kterých je $N!$. Je zřejmé, že z hlediska času výpočtu je tento postup nepoužitelný již pro relativně malé hodnoty N . Proto se většinou musí udělat nějaký kompromis, kdy řešení je sice horší, ale nalezneme ho v přijatelném čase. K tomu byly navrženy mnohé aproximační algoritmy.

V genetických algoritmech jsou jedinci reprezentováni pomocí řetězců – chromozomů. Každá pozice v řetězci se v analogii s obecnou genetikou nazývá alela a konkrétní symbol v chromozomu se nazývá gen. Historicky nejstarší je binární kódování řetězce (0 a 1) s fixní délkou (např. 0011). Pro toto kódování se využívá metod jako je sousedská reprezentace, ordinální reprezentace a přirozená reprezentace. Při křížení se využívá metody s částečným zobrazením – PMX, cyklický operátor křížení – CX, operátor křížení se zachováním pořadí – OX, operátor křížení s rekombinací hran – ERX (symetrické úlohy), operátor křížení se střídáním pozic – AP. Mutace využívá operátorů založený na posunutí – DM, operátoru mutace založený na výměně – EM, na vložení – ISM, inverzní mutace – SIM, zkracující inverzní mutace – ZIM a další.

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 198-217)

Kontrolní otázky

1. Co je to problém obchodního cestujícího?
2. Jak lze vyjádřit cenu cesty v problému obchodního cestujícího?
3. Co je to Turingův stroj?
4. Popište třídu složitosti P.
5. Popište třídu složitosti NP.
6. Popište třídu složitosti NP – úplná.
7. Co je principem aproximačních algoritmů?
8. Co je to binární kódování řetězce?
9. Popište metodu křížení PMX.
10. Popište metodu mutace DM.

Odkaz na praktickou část

- 3.10 Modelování časových řad pomocí UNS

2.21 *Optimalizační úlohy v GA, schémata, cykly, smyčky*

Klíčová slova

Náhodnost, vyřazení, schopnost přežít, kvalita.

Cíle kapitoly

Cílem je pochopení použití GA v optimalizačních úlohách. Cílem je si osvojit principy schémat, cyklů a smyček v úlohách GA.

Výstupy z učení

- 16.3 rozumí genetickým algoritmům

Abstrakt

Generační přístup je základem pro optimalizační řešení. Zpravidla se používá následující schéma: (1) vytvoříme náhodnou populaci, (2) populaci seřadíme podle účelové funkce (přisuzující kvalitu jedincům), (3) vyřadíme nejhorší jedince, (4) populaci doplníme o jedince, vytvořené křížením a mutací. Jak křížením, tak mutacemi může vzniknout jedinec, který není podle definice problému přípustný. Proto po každém generování proběhne kontrola přípustnosti a nevhodní jedinci se vyřadí.

Evoluční výpočetní techniky (genetické algoritmy, genetické programování, gramatická evoluce) používáme tam, kde ostatní metody umělé inteligence (např. neuronové sítě) nelze použít nebo z nějakých důvodů selhávají (např. doba řešení je nepřijatelná). Klíčové problémy genetických algoritmů jsou (A) popis problému pomocí sady symbolů, (B) stanovení účelové funkce (kvality).

Schéma pro evoluční výpočetní techniky nabízí teorie Markovových řetězců. Tyto řetězce jsou vhodné, neboť populace v evolučních algoritmech závisí pouze na stavu předchozí populace pravděpodobnostním způsobem. Využívá se dimenzionální analýzy, která se pokouší identifikovat důležité vlastnosti a klíčové znaky a závislosti zkoumaného systému. Dále se využívá kvantitativní genetika, která je zaměřena na aditivně separovatelné fitness funkce a nekonečně velké populace. Dalším přístupem je analýza ortogonálních funkcí (Fourierovy, Walshovy a Haarovy funkce), které jsou používány jako nástroj pro vytváření

fitness funkcí. Dále se využívají kvadratické dynamické systémy a statistická fyzika. Pro genetické programování lze využít např. výpočetním systémem MATLAB.

Tabulka 1: Genetický algoritmus – evoluční pojetí kódu

t: = 0	Počátek algoritmu
Initialize G(0)	Vytvoření počáteční generace
Evaluate G(0)	Ohodnocení počáteční generace
do while not Done	Dokud není splněna ukončovací podmínka, proved' cyklus $t := t + 1$
Select G(t) from G(t-1)	Proved' přirozený výběr. Cílem je, aby průměrné ohodnocení nové generace bylo lepší než průměrné ohodnocení generace předchozí.
Crossover G(t)	Aplikuj křížení
Mutate G(t)	Aplikuj mutaci
Evaluate G(t)	Proved' ohodnocení nové generace
loop	Opakuj cyklus

Zdroj: Vlastní zpracování

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 218-230)

Kontrolní otázky

1. Popište proces vytvoření náhodné populace.
2. Popište proces sestavení účelové funkce.
3. Popište proces vyřazení nejhorších jedinců.
4. Popište proces křížení.
5. Popište proces mutací.
6. Pomocí vybraného programového kódu definujte obecně genetický algoritmus.
7. Co jsou to Markovovy řetězce?
8. Jaké funkce jsou ortogonální?
9. Jaký software se využívá pro analýzu pomocí GA?
10. Jaké další techniky a metody lze spojit s analýzou GA?

Odkaz na praktickou část

3.11 Pokročilé modelování časových řad pomocí UNS

2.22 Modelování časových řad pomocí GA

Klíčová slova

Náhodnost, vyřazení, schopnost přežít, kvalita.

Cíle kapitoly

Cílem je pochopení použití GA v úlohách modelování časových řad.

Výstupy z učení

- 16.3 rozumí genetickým algoritmům
- 16.8 modeluje časové řady pomocí genetických algoritmů

Abstrakt

Použití genetických algoritmů lze aplikovat na časové řady. Většinou jde o kombinaci neuronových sítí a genetických algoritmů. Parametrické, lineární a na statistickém rozdělení závislé metody (např. GARCH a ARIMA) mají problém s přesností predikce výnosů finančních instrumentů, neboť jde o data s trendem, sezónností, nelinearitou a především heteroskedasticitou, která se v čase mění. S těmito problémy se velmi dobře umějí vypořádat genetické algoritmy.

Spektrální analýza časových řad využívá Fourierovu transformaci. Časově spojitě periodické signály (data např. z měsíční časové řady) lze pomocí Fourierovy řady rozložit na součet harmonických funkcí. Vzniká spektrální vyjádření časové řady. Nevýhodou Fourierovy transformace je, že nedokáže přesně analyzovat finanční časové řady, které se vyznačují náhlými změnami a celkovou nepředvídatelností.

Díky vlnkové transformaci (pomocí funkcí sin a cos) je možné analyzovat i tyto nepředvídatelné nestacionární signály (časové řady typu „náhodná procházka“ – Random walk). Avšak transformace už neposkytuje rozklad ryze na harmonické funkce.

Genetické algoritmy se využívají i pro nalezení maxima funkce. Jde o tzv. extrémální problém. Graf řešené funkce obsahuje často hned několik lokálních extrémů (lokální minimum, resp. lokální maximum) a jen jeden globální extrém. Lokální extrémy jsou překážky, které musí umět GA překonat. Cílem programování je, aby GA neuvízl

v lokálním extrému a našel globální extrém funkce. Proto je nutné vhodně ošetřit algoritmus v části evoluce populace.

Fong a Szeto (2001) využili GA pro predikci časových řad finančních výnosů aktiva. Predikce je úloha GA, kdy se snažíme nalézt opakující se vzor v časové řadě. Nalezení opakujících se vzorů je předáno další populaci dané GA. Pro každý interval (stejně dlouhý) jsou zaznamenány čtyři hodnoty: cena při zahájení intervalu, maximum ceny daného intervalu, minimum ceny daného intervalu a cena na konci intervalu. poté je z těchto 4 hodnot počítána střední hodnota x_i z logaritmovaného maxima a minima intervalu a poté odchylka v podobě volatility podle rovnice ve tvaru:

$$v(t) = \text{odmocnina} \{ 1/n * \text{suma absolutní hodnoty výrazu: } [x(t_i) - x(t_i - \Delta t)]^2 \}.$$

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 231-243)

Doporučená literatura

FONG, L. Y. a K. Y. SZETO, 2001. Rules extraction in short memory time series using genetic algorithms. *The European Physical Journal B-Condensed Matter and Complex Systems*. **20**(4), 569-572. ISSN 1434-6028. (s. 569-572)

Kontrolní otázky

1. Jaké metody lze použít pro analýzu časových řad?
2. Jaké jsou nevýhody tradiční analýzy časových řad?
3. Jaké jsou výhody GA u časových řad?
4. Co je to Furierova transformace?
5. Jaké nevýhody má Furierova transformace?
6. Co je to vlnková transformace?
7. Co za problém způsobují tzv. lokální extrémy?
8. Jak probíhá predikce časové řady pomocí GA?
9. Uveďte příklad použití GA v mikroekonomickém výzkumu.
10. Uveďte příklad použití GA v podnikové praxi.

Odkaz na praktickou část

3.11 Pokročilé modelování časových řad pomocí UNS

2.23 Teorie chaosu: Teorie pojednává o možnosti lepšího popisu ekonomických jevů, než je tomu u klasických metod. Je objasněn pojem chaos a řád, fraktál

Klíčová slova

Determinace, složitost, jistota, pravděpodobnost, řád, fraktál

Cíle kapitoly

Cílem je pochopení teorie chaosu, která pojednává o možnosti lepšího popisu ekonomických jevů, než je tomu u klasických ekonometrických metod. Cílem je osvojení si pojmů chaos, řád a fraktál.

Výstupy z učení

- 16.4 zná teorii chaosu

Abstrakt

Teorie pojednává o možnosti lepšího popisu nelineárních dynamických systémů, než je tomu u klasických metod. Chování takových systémů (atmosférické jevy, turbulence tekutin, vývoj populací, ekonomické jevy jako hospodářský růst, vývoj podniku apod.) se jeví jako náhodné i přes to, že jsou zcela deterministické. Schopnost předvídat počasí se jevila v minulosti jako nemožná. Celý systém atmosférických jevů je tak složitý, že nejde vytvořit žádnou rozumnou předpověď. Je to chaos. Avšak s rozvojem vědy a techniky je možné sebrat mnohem více dat, vyhodnotit je a pochopit alespoň část tohoto komplexního systému. Jinak řečeno budeme-li vědět více o komplexním systému (chaosu), tak postupně odstraníme zdroj náhodnosti.

S problematikou chaosu je spojen efekt motýlích křídel, který říká, že systém se chová stejně, pokud jeho počáteční konfigurace je stejná. To, jak se chaotický systém chová, závisí na počátečních podmínkách. Změny počátečních podmínek mohou mít za následek velké variace v systému v delším období. Mezi umělé systémy, které tento efekt vykazují, patří například Lorenzův atraktor.

Chaotický pohyb je nejsložitějšího typ chování systému. Aby systém byl chaotický, musí splňovat tyto podmínky: musí být citlivý na počáteční podmínky (efekt motýlích křídel), musí být topologicky tranzitivní, jeho periodické orbity musí být husté. Topologické mixování (tranzitivita) popisuje situaci, kdy systém se v čase rozvine tak, že jakákoliv jeho daná oblast se nakonec překrývá s jakoukoliv jinou součástí systému. Husté periodické orbity v chaotickém systému jsou stav, kdy ke každému bodu v prostoru je možné se (arbitrárně blízko) přiblížit periodickou orbitou. Příkladem je loviště (systém – prostor X), které je plné zvěře (množina D). Tento systém musí být dostatečně hustý, aby byl chaotický. Pokud hodíme na toto loviště libovolně velkou loveckou síť (s libovolně uspořádaným vzorkem sítě), vždy ulovíme zvěř.

Vývoj dynamického systému se znázorňuje ve fázovém prostoru. Každý jednotlivý rozměr představuje jednu stavovou proměnnou dynamického systému. V průběhu vývoje systému lze ve fázovém prostoru pozorovat buď (1) křivku, pokud je čas spojitý, nebo (2) body, pokud je čas diskrétní. Atraktorem se nazývá stav, do kterého daný dynamický systém směřuje. Rozeznáváme několik typů atraktorů: (A) pevné body, (B) periodické nebo kvaziperiodické body, (C) chaotické atraktory, (D) podivné atraktory. Podivný atraktor je z hlediska teorie chaosu nejdůležitějším. Může vzniknout jen tehdy, je-li daný dynamický systém popsán nejméně třemi diferenciálními rovnicemi. Např. Lorenzův atraktor, Rösslerův atraktor, Hénonův atraktor, estetické atraktory (Kings Dream).

Pro pojem fraktál, je nejprve nutné zopakovat, co je to dimenze. Dimenze je počet vektorů báze daného prostoru (počet souřadnic), které potřebujeme k jednoznačnému určení bodu v prostoru. Můžeme také říct, že je to počet rozměrů daného objektu. Např.: dimenze roviny je 2, jinak řečeno potřebujeme dvě souřadnice k určení bodu. Dimenze krychle je 3, má tři rozměry: šířku, výšku a hloubku. Je-li hodnota dimenze celočíselná, pak řekneme, že toto číslo je topologickou dimenzí daného objektu. Topologická dimenze popisuje geometricky hladké objekty.

Fraktální dimenze říká, jak moc je daný útvar členitý. Existuje několik typů fraktálních dimenzí: Hausdorffova dimenze (Hausdorff dimension), mřížková dimenze (box – counting dimension), informační dimenze, korelační dimenze, Ljapunovova dimenze a Kolmogorovova entropie. Poincarého mapa je zobrazení fázového diagramu vyšší dimenze v dimenzi nižší. Jedná se o řez nebo projekci vícerozměrného útvaru, kdy jedna či více proměnných je konstantní. Jev, kdy v systému při malé změně řídicího parametru dojde k

náhlé a výrazné změně vnitřního stavu systému, nazýváme bifurkací. Bifurkační diagram je obrazec, kdy se na osu x vynáší hodnota řídicího parametru a na osu y hodnota vnitřního stavu systému. Existuje několik základních typů bifurkací: (1) Lokální (flip, fold, transkritická, Hopfova, transcritical) a (2) globální (homoclinic, heteroclinic, infinite period, blue sky catastrophe).

Neexistuje žádná definice pojmu fraktál, která by byla matematicky definovatelná. Mandelbrotova definice uvádí, že fraktál je množina, jejíž Hausdorffova dimenze je větší než dimenze topologická, jinak řečeno fraktál je takový útvar, jehož dimenze je neceločíselná. avšak za fraktály považujeme i útvary mající celočíselnou dimenzi. Fraktál lze nejlépe definovat pomocí jeho vlastností. Fraktál je geometrický útvar, který lze rozdělit na části, přičemž tyto části jsou (alespoň přibližně) zmenšené kopie celého útvaru. Tato vlastnost fraktálu se nazývá invariance vůči změně měřítka. Další definicí založenou na vlastnostech je: Fraktál je nekonečně členitý útvar. Fraktály se rozdělují do několika typů. (a) L – systémy, (b) IFS (Iterated Function Systems), (c) TEA (polynomické fraktály), (d) náhodné fraktály.

Náhodné fraktály vznikají obohacením tvorby fraktálu o náhodný prvek. Existuje několik způsobů, jak lze generovat náhodné fraktály: (i) simulace Brownova pohybu, (ii) metoda přesouvání středního bodu, (iii) spektrální syntéza.

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 291-318)

Kontrolní otázky

1. Popište základní princip teorie chaosu.
2. Vysvětlete rozdíl mezi deterministickým chaosem a „čistým“ chaosem.
3. Uveďte příklady chaosu v přírodě.
4. Uveďte příklady chaotických systémů analyzovaných ve společenských vědách.
5. Co je to efekt motýlích křídel?
6. Jaké podmínky má chaotický pohyb?
7. Co je to fázový prostor?

8. Co je to atraktor?
9. Co je to fraktál?
10. Jak lze generovat náhodné fraktály?

Odkaz na praktickou část

3.12 Procvičování základních principů genetických algoritmů

2.24 Uvedení využití teorie chaosu u měřeného sledovaného systému, uvedení příkladů z ekonomického výzkumu a podnikové praxe

Klíčová slova

Samoorganizace, bublina, krize, Elliotovy vlny, časová řada

Cíle kapitoly

Cílem je pochopení možných způsobů využití teorie chaosu u měřeného sledovaného systému. Cílem je osvojit si teorii chaosu na příkladech z ekonomického výzkumu a podnikové praxe.

Výstupy z učení

- 16.4 zná teorii chaosu

Abstrakt

Mezi praktické aplikace teorie chaosu patří např. Samoorganizace. Tu lze definovat jako jev, kdy se samostatné jednotky systému začnou chovat jako celek. Využití je v přírodních i společenských vědách. Ekonomie se zabývá studiem komplexních systémů. Tyto systémy jsou ovlivňovány fluktuacemi vnitřními i vnějšími. Pokud je systém vychýlen (vnitřně/vně), může systém nabývat nových stavů. Protože ekonomika je plná nelineárních vazeb, může dojít k posunu chování až do chaosu. Tento chaos má při dlouhodobějším pohledu (100 let) určitou symetrii. V ekonomii je vhodné studovat výkyvy cen, události známé jako ekonomické bubliny. Ekonomická bublina buď vyprchá a systém zaznamená drobný výkyv, nebo praskne a dojde ke krizi – velkému výkyvu. Ceny v ekonomických bublinách i v následné krizi neodpovídají reálným hodnotám (resp. hodnotám v předcházejícím období stability, na pomyslném potenciálu). Prudký růst i pokles cen pak dokáže znehodnotit bohatství a poškodit celou ekonomiku.

Kdy nastane příští Černý pátek? Teorie chaosu nám bohužel nic takového neřekne. Dokonce nám to neřeknou ani ekonomické teorie, které mnohdy navíc zavádějí

nerealistické podmínky dokonalé racionality investorů a dokonalé předvídatelnosti ekonomiky. Schopnost předpovědět ekonomickou bublinu a její prasknutí neexistuje. Isaac Newton poté, co kleslo jeho bohatství účastí na ekonomické bublině krachu akcií South Sea Company v roce 1720 údajně řekl, že dokáže spočítat pohyby planet na palec přesně, ale nemá ani tušení, jakým směrem požene šílenství davů trh zítra. Ekonomové zatím ani nedokáží předvídat ani racionální chování davů. Těch proměnných, které ovlivňují národní natož globální trh, je zatím velmi mnoho.

Elliotovy vlny vznikly empirickým pozorováním dat New Yorkské burzy. Popis chování této vlny se datuje do období 1935–1947. Elliotovy vlny byly identifikovány jako fraktály a celý jev lze zařadit do studia chaotických systémů. O fraktál se jedná proto, že se tvar této vlny opakuje v ní samé. Dokonce její dobrá znalost nám umožňuje s velkou pravděpodobností určit zlomy v cenovém vývoji a tím minimalizovat riziko obchodování. Celá podstata Elliotových vln vychází z existence dvou fází: impulzní (vývoj ve směru trendu) a korekční (vývoj proti směru trendu). Každá fáze se skládá z “pod-vln”. Díky určité pravidelnosti, lze s vysokou pravděpodobností očekávat změnu ceny opačným směrem, než jakým se ubíral dosavadní trend. Elliotovy vlny bývají velmi často různě deformovány, což se také promítá do budoucího průběhu vývoje ceny.

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 319-342)

Kontrolní otázky

1. Co je to samo-organizace?
2. Co jsou to komplexní systémy v ekonomii?
3. Uveďte příklad nelineárních vazeb v ekonomice.
4. Proč jsou cenové bubliny zajímavé z pohledu teorie chaosu?
5. Co jsou to Elliotovy vlny?
6. Jaká je podstata modelování časové řady pomocí Elliotových vln?
7. Jaké jsou limity teorie chaosu v ekonomickém výzkumu?
8. Uveďte příklad aplikace teorie chaosu ve finanční ekonomii.

9. Uveďte příklad aplikace teorie chaosu v podnikové praxi.
10. Uveďte příklad aplikace teorie chaosu v makroekonomickém výzkumu.

Odkaz na praktickou část

- 3.12 Procvičování základních principů genetických algoritmů

2.25 Uvedení pojmu datamining, textmining a definování cílů, výběr techniky modelování, zdroje a příprava dat, tvorba modelů, jejich ověření, vyhodnocení, implementace a údržba

Klíčová slova

Datové sklady, OLAP, SQL, clustering, data, informace, znalost

Cíle kapitoly

Cílem je pochopení pojmu datamining, textmining. Cílem je získat schopnost definovat výběr techniky modelování, připravit zdroje a data, vytvářet modely a ověřovat jejich smysluplnost.

Výstupy z učení

- 16.5 rozumí dataminingu a textminingu

Abstrakt

Data mining je o získávání užitečných informací ukrytých v datech. Do češtiny se občas překládá jako „dolování“ či „vytěžování“ dat. Datový specialista potřebuje pro provádění dataminingu rozumět jazyku SQL (získat dat z databází), metodám a postupům (ekonometrie, UNS, GA, teorie chaosu), musí umět vizualizovat a prezentovat data a být schopen aplikovat pokročilé metody pro řešení zadaných úloh (testování hypotéz). K tomu se v praxi využívají predikční modely, segmentace, asociační analýzy (clustering), analýza časových řad a text mining. Rozdíl mezi dataminingem a textminingem je ten, že v textminingu vychází vzory z přirozeného jazyka textu spíše než ze strukturovaných databází (proměnné, datové typy).

Rizikem dataminingu je tzv. GIGO (Garbage In, Garbage Out), že i z nesmyslných dat je možné dostat nějaké zdánlivě smysluplné výsledky. Problémem může být zdánlivá regrese, nebo náhoda. To je nutné odhalit dodatečným testem, testem smysluplnosti, uvedením do kontextu, ověřením datové základny, aby nedošlo ke špatné interpretaci výsledků.

Pro uložení dat ve velké organizaci se nejčastěji využívá podnikový informační systém (PIS). Business Intelligence (BI) je jedním ze čtyř základních modulů současných ERP II systémů (anglicky Enterprise Resource Planning, česky plánování podnikových zdrojů). BI je důležitou podporou strategického rozhodování manažerů, neboť poskytuje maticový/multidimenzionální pohled na podnik (kombinace ukazatelů v čase a prostoru).

Typické využití textminingu je při zkoumání konkurence, kdy dochází k analýze webových stránek konkurence. Automatické zpracování obsahů webových stránek slouží jako velmi efektivní způsob získávání významných obchodních informací (cena, slevová akce apod.) a aktivitách konkurentů (nabírání pracovní síly, nové projekty, nové obchodní podmínky, nové způsoby doručení zákaznickovy apod.). Analýza anket, poptávkových průzkumů, odezvy zákazníků (hodnocení produktů a služeb) může objevit soubory slov a termínů (zejména v otevřených otázkách), které respondenti vyžívají při popisu pozitiv a negativ produktu nebo služby.

Užití ve filtrech a při automatické zpracování elektronických zpráv (instant messaging, email apod.). Naprogramovaná textminingová aplikace pomáhá s automatickým tříděním textů a pomáhá přesměrovat zprávy na správná oddělení. Vhodná aplikace může sloužit jako screening příchozích zpráv a vyřadit (přesměrovat) ty, kde se vyskytují neznámá, či vulgární slova.

Obrázek 9: Ukázka OLAP datové kostky

	Březen	Únor	Leden
Židle	212	534	254
Pohovky	21	46	33
Gauče	310	321	200
Kuchyně	120	234	131
Křesla	43	78	55
Svítidla	12	32	21
	Praha	Brno	Ostrava

Zdroj: Vlastní zpracování

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 343-350)

Kontrolní otázky

1. Definujte princip dataminingu.
2. Definujte princip text miningu.
3. Jak se liší data a text mining?
4. Co je to SQL?
5. Co je to OLAP?
6. Proč dochází k problémům v dataminingu?
7. Co je to GIGO?
8. Popište principy business intelligence.
9. Popište součásti ERP II informačních systémů.
10. Uveďte příklady využití textmingu.

Odkaz na praktickou část

3.13 Modelování časových řad pomocí GA

2.26 Uvedení příkladů použití dataminingu v Business Intelligence a pro volbu strategie spolupráce se zákazníkem, direct mailing apod.

Klíčová slova

Scoreboard, výkonnost, cíle, mise, strategie, plány, vyhodnocení

Cíle kapitoly

Cílem kapitoly je pochopení příkladů použití dataminingu. Cílem je pochopit principy Business Intelligence pro vyhodnocení a volbu strategie.

Výstupy z učení

- 16.4 zná teorii chaosu

Abstrakt

BI je množina technologií, postupů a procesů, která organizaci umožňuje získat nové informace na základě vlastních dat. Jsou to statistické a softwarové nástroje či interní procesy, které z obrovského množství firemních dat (datové kostky, data mining) pomohou vytřídit jen ta důležitá. Software dokáže nesourodá data seskupit do ucelených přehledů s důležitými informacemi, vhodně je prezentovat ve formě reportů, dashboardů a scoreboardů. BI se snaží odbourat problém nadbytečného sběru a zpracování dat, čímž se automatizuje práce s obrovským množstvím v podnikovém systému uložených dat.

Moduly s názvem Business intelligence využívají metod (ETL, DWH, OLAP), které dokáží mapovat data v podniku. Mezi přínosy patří zejména detailní a přehlednější výstup o skutečném stavu dosahování cílů, možnost objevení nových souvislostí a slabých míst v procesním řízení, správa dat v podniku prostřednictvím jediného systému a vyvarování se ukládání a desintegraci dat, kvalitní analýzy a možnost měřit výkon formou personalizovaných reportů, dashboardů, forecastů a scorecardů. U menších organizací je výhodou možnost sledovat vývoj online – za chodu. Architektura vhodná pro procesní řízení se hodí i pro moduly business intelligence či ERP

II Podnikové informační systémy. BI software využívá datových skladů a datových úložišť. Nejprve se pomocí datových pump dolují data, která se pak ukládají v tzv. OLAP databázích (Online Analytical Processing), které pojmu i větší množství dat. Data se ukládají specifickým způsobem tak, aby se dala interpretovat několika různými způsoby a častým dotazování přes SQL jazyk.

Direct mailing je neslavná metoda přímého marketingu, která obecně dosahuje velké míry ignorace a snahy zákazníků odhlásit se ze všech odběrů. Problémem je to, že email často necílí na konkrétní potřebu zákazníka a email jej odradí. Snahou moderního Direct mailingu je automatizovat emailovou komunikaci a posílat jen to, co zákazník očekává. Data k odhadu toho, co zákazník očekává, jsou založena na datech o prohlíženém zboží, nedokončených objednávkách, personalizací, anketou, zda se newsletter, nebo reklamní sdělení líbilo.

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 351-360)

Kontrolní otázky

1. Jaké metody využívá Business Intelligence?
2. Popište strukturu ukládání dat ve středně velkém podniku.
3. Co je to direct mailing?
4. Proč je direct mailing z velké části neúspěšný?
5. Co je to kontingenční tabulka?
6. Uveďte příklad použití dataminingu při řízení vztahu se zákazníkem.
7. Uveďte příklad použití dataminingu při řízení vztahu se dodavatelem.
8. Uveďte příklad použití dataminingu při řízení inovací
9. Uveďte příklad použití dataminingu při definování strategie.
10. Uveďte příklad použití dataminingu pro podporu marketingového plánu.

Odkaz na praktickou část

3.13 Modelování časových řad pomocí GA

3 Příprava na semináře

3.1 Popisná statistika a analýza dat, vizualizace dat, příklady pravidel logiky a matematického zápisu množin, možnosti zapisování modelů

Klíčová slova

Test dobré shody, korelace, závislost, ANOVA

Cíle kapitoly

Cílem kapitoly je procvičení popisné statistiky a analýzy dat, vizualizace dat, příkladů pravidel logiky a matematického zápisu množin.

Výstupy z učení

- 16.6 modeluje časové řady pomocí fuzzy logiky
- 16.7 modeluje časové řady pomocí umělých neuronových struktur
- 16.8 modeluje časové řady pomocí genetických algoritmů

Příklad, uvedení vzorového úkolu

1. Statistická analýza kontingenčních tabulek

Dochází tak k dobývání dat a následně i znalostí. Příkladem je např. tabulka o podnicích s webovými stránkami, které umožňují učinit online platbu platební kartou. Tato tabulka vznikla agregací tisíců řádků dat o podnicích z dotazníkového šetření. Výsledkem je pohled na data, kdy jedna osa vyjadřuje odvětví národního hospodářství na druhá osa pak velikost podniku. Nyní je možné testovat několik tezí a z vybraných dat i informací o procentních podílech firem získávat nové znalosti. Např. zda jsou rozdíly mezi malými středními a velkými podniky, zda podíl ovlivňuje jako velikost podniku tak ekonomická činnost podniku, zda závisí na ekonomické činnosti podniku.

Tabulka 2: Podniky s webovými portály s online platbou

	Velikost podniku (počet zaměstnanců)			
	Celkem	10–49	50–249	250+
Celkem	7,5	6,7	9,7	12,0
Zpracovatelský průmysl – C	4,5	4,0	4,3	5,9
Výroba a rozvod elektřiny, plynu, tepla; Zásobování vodou, činnosti související s odpady – D, E	2,7	1,8	2,6	12,7
Stavebnictví – F	1,9	1,7	3,7	2,0
Velkoobchod a maloobchod; opravy a údržba motorových vozidel – G	16,8	14,6	30,5	30,7
Doprava a skladování – H	3,2	1,9	5,0	19,6
Ubytování, stravování a pohostinství – I	10,3	8,4	30,1	46,8
Informační a komunikační činnosti – J	17,3	15,6	20,0	32,8
Profesní, vědecké a technické činnosti – M	4,3	4,1	5,5	7,5
Administrativní a podpůrné činnosti – N	5,3	4,5	6,6	7,3

Zdroj: Český statistický úřad 2018, Veřejná databáze

Tyto otázky dokáže zodpovědět klasická analýza variability dat. Postup (Online Web Statistical Calculators, 2018): Select the number of independent treatments below – zde vybereme počet skupin testované proměnné % podílu podniků s webovými stránkami, které umožňují učinit platbu. Testovat budeme tezi, zda závisí na velikostní kategorii. Máme 3 kategorie (0–49, 50–249, 250+) proto zvolíme $k=3$. Formulář odešleme kliknutím na „Proceed to enter your treatment data“ a nezapomeneme, že nechceme demo data – checkbox pod kategoriemi je prázdný (viz obrázek).

Obrázek 10: První krok ANOVA analýzy

Select the number of independent treatments below:

Select k , the number of independent treatments, sometimes also called samples. Since these are independent and not paired or correlated, the number of observations of each treatment may be different.

This would lead to an input screen with k columns to paste your observation data on various treatments. This calculator is hard-coded for a maximum of 10 treatments, which is more than adequate for most researchers.

$k = 2$ *
 $k = 3$
 $k = 4$
 $k = 5$

Zdroj: vlastní zpracování na základě Online Web Statistical Calculators, 2018

V dalším kroku zadáme hodnoty do připravené tabulky:

Obrázek 11: Druhý krok ANOVA analýzy

$k = 3$ ← this is the number of treatments or samples.

Sample A	Sample B	Sample C
6.7	9.7	12.0
4.0	4.3	5.9
1.8	2.6	12.7
1.7	3.7	2.0
14.6	30.5	30.7
1.9	5.0	19.6

Zdroj: vlastní zpracování na základě Online Web Statistical Calculators, 2018

Tabulka 3: One-way ANOVA pro 3 nezávislé velikostní skupiny podniků

source	sum of squares SS	degrees of freedom vv	mean square MS	F statistic	p-value
treatment	650.1527	2	325.0763	2.7321	0.0831
error	3,212.6220	27	118.9860		
total	3,862.7747	29			

Zdroj: vlastní zpracování

Poté nám program zobrazí výsledky (viz tabulka výše). Z analýzy plyne, že na hladině významnosti 5 % nelze říci, že skupiny se mezi sebou liší. Je to dáno tím, že v datech je příliš mnoho výkyvů. Pokud bychom chtěli data podrobněji analyzovat (dropdown), mohli bychom analyzovat konkrétní odvětví zpracovatelského průmyslu, nebo všechna odvětví apod.

Zadání samostatné práce (úkolů)

Ze stránek Českého statistického úřadu zjistěte v datech o podnicích, zda závisí velikost podniku na tom, zda webové stránky umožňují konfigurovat produkt či službu.

Tabulka 4: Podniky s webovými stránkami umožňujícími konfigurovat produkt či službu

	Velikost podniku (počet zaměstnanců)			
	Celkem	10–49	50–249	250+
Celkem	10,7	9,8	13,6	17,3
Zpracovatelský průmysl – C	10,0	8,7	12,9	11,7
Výroba a rozvod elektřiny, plynu, tepla; Zásobování vodou a činn. související s odpady – D, E	6,7	3,8	7,6	32,7
Stavebnictví – F	4,6	4,3	6,8	6,7
Velkoobchod a maloobchod; opravy a údržba motorových vozidel – G	16,0	14,2	26,6	30,2
Doprava a skladování – H	6,1	5,3	6,6	19,1
Ubytování, stravování a pohostinství – I	16,5	14,4	40,0	49,5
Informační a komunikační činnosti – J	20,7	18,3	24,4	44,9
Administrativní a podpůrné činnosti – N	9,1	9,9	5,8	14,2

Zdroj: Český statistický úřad, 2018 Veřejná databáze

Použité zdroje v příkladu:

Český statistický úřad, 2018. *Veřejná databáze*, [online]. Dostupné z: goo.gl/dctFSq

Online Web Statistical Calculators, 2018. *One-way ANOVA with post-hoc Tukey HSD Test* [online]. Dostupné z: <http://astatsa.com/>

Studijní literatura

Povinná literatura

COUFAL, J., 2014. *Logika a logické myšlení*. [Praha]: Vysoká škola ekonomie a managementu. ISBN 978-80-87839-59-1. (s. 6-16)

HAMPEL, D., V. BLAŠKOVÁ a L. STŘELEČEK, 2016. *Ekonometrie 2*. 3. přeprac. vyd. Brno: Mendelova univerzita. ISBN 978-80-7509-427-8. (s. 8-10)

3.2 Počet pravděpodobnosti a statistické rozdělení dat.

Pravidla regresní analýzy pro průřezová data

Klíčová slova

Normalita, histogram, četnost, data, proměnná

Cíle kapitoly

Cílem je procvičení počtu pravděpodobnosti, statistického rozdělení dat a pravidel regresní analýzy pro průřezová data.

Výstupy z učení

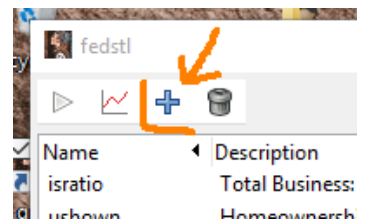
- 16.6 modeluje časové řady pomocí fuzzy logiky
- 16.7 modeluje časové řady pomocí umělých neuronových struktur
- 16.8 modeluje časové řady pomocí genetických algoritmů

Příklad, uvedení vzorového úkolu

1. Příklad testování normality v programu Gretl na příkladu využití výrobní kapacity.

Gretl příklad

1. Spustíte Gretl, File > Databases > Gretl Native > dvojklik na „fedstl“. V tabulce vyberte proměnnou cumfn „Capacity utilization...“ a klikněte na modré plus vedle popelnice.
2. Zavřete aktivní okno, zůstane jen výchozí s přehledem proměnných (const a cumfn)
3. Vyberte pravým myšítkem název proměnné „cumfn“ a zadejte volbu „Frequency distribution“.
 - Předpokládáme, že využití výrobních kapacit US ekonomiky mezi lety 1972 až 2015 bude mít normální rozdělní řekněme kolem 80 % (naivní odhad) +/- 10 %, tj. že by to mohlo mít normální rozdělní.
4. V nabídce už jen vyberte Test against normal distribution.



5. Naběhnou dvě okna – Graf zhodnotíme pohledem - využití výrobních kapacit má až na pár „děr“ a skoro normální rozdělení (v jednoduchém překladu „kolem průměrné využití kapacity 78 % je v průměru US ekonomika, pak to jde dolu, hodnoty jako 65 % je extrém s nízkou pravděpodobností“), jak jsme předpokládali.
6. Graf posuneme doprava nahoru, ať nám nepřekáží a podíváme se na to samé v textové podobě. Hvězdičky také naznačují, že to je cca normální rozdělení. Počet kategorií i šířka jsou určeny arbitrárně, Gretl pro výpočet používá odmocninu z počtu pozorování. Zavřeme textovou podobu histogramu.
7. Pravím myšítkem se podíváme na summary statistics. Co známe a už víme, co znamená? Průměr, medián (hodnota uprostřed intervalu), minimum, maximum, standardní odchylka – jak hodně se pozorování „v průměru liší od průměru“, C.V. je variační koeficient a ten jen jinak uvádí standardní odchylku jako procentní podíl průměru (5,98 %), Skewness – skutečný blob je oproti tomu nakreslenému normálnímu nižší (záporná hodnota). Kurtosis – skutečný blob je nakloněn více napravo (pozitivní hodnota) od normálního rozdělení. Percentily – hodnoty na ose X (50 % percentil je např. medián), Range je rozdíl mezi maximem a minimem (88,3-63,9).
8. Test normality: Gretl uvádí Doornik-Hansen test – podívejte se do grafu nahoru: Test statistics for normality. Vidíme, že statistika je vysoká, a proto i pravděpodobnost takové hodnoty bude velmi nízká. Pravděpodobnost té statistiky (p-value v hranaté závorce) je 4.06 %. Tj. na obvyklé hladině významnosti 5 % můžeme prohlásit proměnou za normální (mající normální rozdělení).
9. Jak otestovat proměnou na normalitu rovnou, bez histogramu? Vyberte proměnou (zmodrá řádek). Variable v menu a normality test.

Zadání samostatné práce (úkol)

Z vybraných databází přítomných v programu Gretl otestujte normalitu vybrané proměnné s využitím histogramu.

Studijní literatura

Povinná literatura

HAMPEL, D., V. BLAŠKOVÁ a L. STŘELEČ, 2016. *Ekonomie* 2. 3. přeprac. vyd.
Brno: Mendelova univerzita. ISBN 978-80-7509-427-8. (s. 10-48)

3.3 Testování předpokladů konvenční asymptotické teorie pro odhady metodou nejmenších čtverců pro průřezová data a pro časové řady

Klíčová slova

Gaussovy-Markovovy podmínky, regrese, přímka, metoda odhadu

Cíle kapitoly

Cílem je procvičení a testování předpokladů konvenční asymptotické teorie pro odhady metodou nejmenších čtverců pro průřezová data a pro časové řady.

Výstupy z učení

- 16.6 modeluje časové řady pomocí fuzzy logiky
- 16.7 modeluje časové řady pomocí umělých neuronových struktur
- 16.8 modeluje časové řady pomocí genetických algoritmů

Příklad, uvedení vzorového úkolu

1. Analyzujte vztah mezi mírou nezaměstnanosti a mírou inflace

Regresní přímku lze použít na vztah mezi mírou nezaměstnanosti (u2ratensa) a mírou inflace CPI pro městské spotřebitele (cpiaucns). Pro účely dalšího testování je přidána proměnnou měna v oběhu (currcir). Vše je k nalení v nativní databázi programu Gretl. Jak bude vypadat tento vzta? Proč ho zkoumáme? Mezi inflací a nezaměstnaností přece ekonomové vidí vztah, tak se podívejme, zda tam nějaký je a zda jej dokáže odhalit OLS a zůstat přitom „B.L.U.E.“.

- Použijeme regresní přímku (SLR), ta je lineární v parametrech MLR1 Ale pokud náhodou projdeme přes MLR 4, musíme ještě zvažovat SLR 5 a SLR 6.
- Náhodnost výběru. Výběr není náhodný, máme časovou řadu, kde se vyskytuje problém autokorelace. Pravděpodobně nesplníme podmínku MLR2.
- Absence perfektní kolinearit MLR3 pravděpodobně splníme, mezi oběma proměnnými neočekáváme silnou korelaci.

- MLR.4: Nulový podmíněný průměr pro u splníme přítomností konstanty, avšak budeme mít asi problém s vynecháním proměnných, které jsou pro vysvětlení inflace klíčové. Pravděpodobně bude problém i se specifikací modelu.
- SLR 5: Odhad asi nebude mít konstantní rozptyl - Homoskedasticita.
- SLR 6: Neočekáváme, že by normální složka byla náhodná, vzhledem k podezření na více problémů v modelu, můžeme otestovat rezidua z regrese a podívat se, zda mají normální rozdělení.

GRETl – Model > OLS... > cpiaucns vložíme jako dependent a do regresorss ke konstantě přidáme u2ratensa a zaškrtneme robust standard errors.

Tabulka 5: OLS model závislosti míry nezaměstnanosti a inflace, 1967:01-2015:07 (T = 583)

	<i>Coefficient</i>	<i>Std. Error</i>	<i>t-ratio</i>	<i>p-value</i>	
const	90.7199	18.4824	4.9085	<0.0001	***
u2ratensa	12.9657	5.4336	2.3862	0.0173	**

Mean dependent var	131.5407	S.D. dependent var	64.15422
Sum squared resid	2245695	S.E. of regression	62.17092
R-squared	0.062487	Adjusted R-squared	0.060874
F(1, 581)	5.693987	P-value(F)	0.017342
Log-likelihood	-3233.964	Akaike criterion	6471.928
Schwarz criterion	6480.664	Hannan-Quinn	6475.333
rho	0.999288	Durbin-Watson	0.003874

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + u \rightarrow \text{CPI} = 90,72 + 12,97 * \text{Míra nezaměstnanosti} + u$$

Zdroj: vlastní zpracování

Poznámka: Závislá: cpiaucns, HAC standard errors, bandwidth 6 (Bartlett kernel)

- Ekonomicky jsme spíše očekávali při růstu míry nezaměstnanosti pokles inflace. Takže už tady něco nehraje, model pravděpodobně bude trpět nedostatky.
- Ekonometricky testujeme druhou podmínku – pohledem na Durbin-Watson statistiku

DW - H_0 : chybová složka netrpí autokorelací. Resp. H_A : chyby jsou AR(1) proces

- ALT+T, D resp.: Tests > Durbin-Watson p-value „vyšlo“ významně na 10 %, takže zamítáme H_0 . MLR 2 ☒

- MLR3 – plníme, resp. GRETl žádnou proměnnou nevyhodil, ale zkontrolujeme tak, že se vrátíme do okna, kde jsou proměnné, obě vybereme a pravím myšítkem correlation matrix.
- MLR.4a – konstanta je signifikantní, takže to vypadá dobře, ale otestujeme, zda je model dobře specifikovaný, použijeme tzv. RESET test. Tests > RESET. Ten testuje, zda náhodou není parabola (přítomnost \hat{y}^2), nebo křivka s kubickým (\hat{y}^3) členem lepší než současný odhad.

RESET - H_0 : koeficienty u dodatečných kvadratických a kubických regresorů jsou nevýznamné

- Vidíme, že všechny členy jsou významné, test „vyšel“, p-value blízké nule, zamítáme H_0 .
- MLR 4a
- MLR.4b – endogenita v MLR (potřebujeme více proměnných x_i). Princip testování je u průřezových dat poměrně jednoduchý, pokud předpokládáme, že naše odhadnutá rezidua \hat{u} (uhat) jsou blízka ideálu teoretické chybové složky u (která je neznámá a nikdo neví, jak vypadá), která je nekorelovaná se všemi regresory (x_1, x_2, \dots, x_i) a to najednou, ve skupinkách i jednotlivě a ještě i s vysvětlovanou proměnnou. Z odhadnutých reziduí \hat{u} (uhat) uděláme proměnnou a použijeme ji v „obrácené“ regresi (prohodíme vysvětlovanou proměnnou za podezřelou proměnnou z x_i , odebereme jeden x regresor a přidáme uložená rezidua z původní regrese).
 - Přepneme se do hlavního okna a znovu zadáme OLS model, dáme více proměnných. Dependent bude cpiuacns a do Regressors dáme zbytek (u2ratensa a currcir)
 - Z aktivní regrese si uložíme rezidua – Save, Residuals (uhat1).
 - Přepneme se do hlavního okna a znovu zadáme OLS model, tentokát obráceně a přidáme proměnnou uhat1. Model > OLS... > u2ratensa jako dependent a Regressors bude uhat1 a cpiuacns.
 - Pokud je náš model dobrý (víme, že není, není to překvapení, proměnné jsou jasně endogenní), pak rezidua by neměla být signifikantní (p-value = 0.0299), což jasně jsou na alfa = 5 %, což indikuje endogenitu. MLR 4b

- SLR 5: přítomnost heteroskedasticity → nevyžaduje se test, pokud jde o MLR a používáte metody, které jsou robustní na heteroskedasticitu (tj. minimálně robustní standardní chyby s popisem, co heteroskedasticitu v populaci způsobuje).
 - Tests > heteroscedasticity > White's test

White's - H_0 : heteroskedasticity není přítomna.

- p-value „vyšlo“ významně na 10 %, takže zamítáme H_0
- SLR 5
- SLR 6: normalita chybové složky
 - Zde máme možnost otestovat jen odhadnutou náhodnou složku.
 - Tests > Normality of residuals

Doornik-Hansen test - H_0 : normalita

- p-value „vyšlo“ významně na 10 %, takže zamítáme H_0 SLR 6

Závěrem: model je ekonometricky špatně specifikovaný, nekonzistentní a vychýlený.

Zadání samostatné práce (úkolů)

Pomocí regrese analyzujte vztah mezi mírou nezaměstnanosti (u2ratensa) a mírou inflace v České republice. Využijte dat z Českého statistického úřadu a programu Gretl.

Studijní literatura

Povinná literatura

HAMPEL, D., V. BLAŠKOVÁ a L. STŘELEČEK, 2016. *Ekonometrie* 2. 3. přeprac. vyd. Brno: Mendelova univerzita. ISBN 978-80-7509-427-8. (s. 49-76)

3.4 Modelování časových řad pomocí regresní analýzy

Klíčová slova

Jednotkový kořen, stacionarity, ADF, deterministická řada, difference

Cíle kapitoly

Cílem kapitoly je procvičení modelování časových řad pomocí regresní analýzy, zejména pak proces testování stacionarity časové řady.

Výstupy z učení

- 16.6 modeluje časové řady pomocí fuzzy logiky
- 16.7 modeluje časové řady pomocí umělých neuronových struktur
- 16.8 modeluje časové řady pomocí genetických algoritmů

Příklad, uvedení vzorového úkolu

1. Testování jednotkového kořene – Vývoj populace na území ČR 1785 – 2010.

Tabulka 6: přehledové statistiky populace na území ČR 1785 – 2010

Proměnná	Počet pozorování	Průměr	Směrodatná odchylka	Minimum	Maximum
Populace	226	8276268	2131372	4250000	1.12e+07

Zdroj dat: Vlastní zpracování

Nejprve si ukažme ADF testovací strategii na prvním příkladu pro typicky rostoucí (nebo klesající) časovou řadu, jako je třeba roční vývoj populace na území České republiky (dále to může být produkce, spotřeba, objem širšího peněžního agregátu apod.). Zde nejčastěji předpokládáme, že časová řada má jednotkový kořen, je bez signifikantního trendu a se signifikantní konstantou. ADF test provedeme v podobě rovnice (3) s trendem a přidáme odpovídající počet zpoždění (např. 3). První krok je tedy testování jednotkového kořenu $(\beta_2 - 1) = 0$ v rovnici (3) a sledování, zda zde existuje signifikantní $(\beta_1$ trend) a nenulová konstanta.

Tabulka 7: ADF test časové řady Populace na území ČR 1785-2010

D.populace		Coef.	Std. Err.	t	P>t
Populace	L1.	-0.011	0.007	-1.680	0.094
	L1D.	0.928	0.065	14.330	0.000
	L2D.	-0.601	0.080	-7.500	0.000
	L3D.	0.303	0.065	4.640	0.000
Trend		252.901	215.893	1.170	0.243
Konstanta		72769.240	33302.250	2.190	0.030

Zdroj dat: Vlastní zpracování

Z provedeného ADF testu plyne, že trend není signifikantní, konstanta je signifikantní, a pozorujeme koeficient u L1.Populace (jedno období zpožděná proměnná $y_{(t-1)}$), který je blízký nule (-0,011). To vše jsou indicie, které ukazují na přítomnost jednotkového kořenu. Pro otestování H_0 o přítomnosti jednotkového kořenu máme přesnější informaci v hodnotě MacKinnon p-value (= 0.7597). Ta podává informaci o tom, zda nám t statistika u proměnné L1.populace (= -1.680) spadá do širšího intervalu (než u t rozdělní) spolehlivosti od 1 % do 10 % (-3,999; -3,134) pro zamítnutí nulové hypotézy o jednotkovém kořenu. Díky této hodnotě se nám nepodařilo zamítnout nulovou hypotézu o jednotkovém kořenu v populaci. Populace na území ČR 1785 – 2010 je časová řada obsahující jednotkový kořen a je řadou nestacionární.

Pokud by $(\beta_2 - 1) < 0$ a $(\beta_1 \text{trend}) \neq 0$, tj. řada by měla signifikantní trend a jednotkový kořen by byl nejasný, pak je tu šance, že může jít o trendově stacionární časovou řadu H_1 , nebo časovou řadu s jednotkovým kořenem a driftem H_0 . Drift a jednotkový kořen odhalí společný F test koeficientů L1.populace a trendu. Pokud jsou společně signifikantní, zamítneme H_0 o jednotkovém kořenu a časová řada je díky tomu, že je typicky rostoucí, nebo klesající velmi pravděpodobně trendově stacionární.

Zadání samostatné práce (úkolů)

Otestujte přítomnost jednotkového kořenu v časové řadě přirozeného přírůstku obyvatelstva na území ČR.

Studijní literatura

Povinná literatura

HAMPEL, D., V. BLAŠKOVÁ a L. STŘELEČ, 2016. *Ekonomie 2*. 3. přeprac. vyd. Brno: Mendelova univerzita. ISBN 978-80-7509-427-8. (s. 77-93)

3.5 Pokročilé modelování časových řad pomocí regresní analýzy

Klíčová slova

Panel, série dat, autokorelace, pevné efekty, náhodný efekt

Cíle kapitoly

Cílem je pokročilé modelování časových řad pomocí regresní analýzy. Cílem je pokročilé modelování panelových dat pomocí regresní analýzy

Výstupy z učení

- 16.6 modeluje časové řady pomocí fuzzy logiky
- 16.7 modeluje časové řady pomocí umělých neuronových struktur
- 16.8 modeluje časové řady pomocí genetických algoritmů

Příklad, uvedení vzorového úkolu

1. Analyzujte panelová data – EU letecký průmysl 2003-2009

Tabulka 8: Přehled proměnných pro panelovou analýzu dat, EU letecký průmysl 2003-2009

Proměnná	Počet pozorování	Průměr	Std. Odchylka	Minimum	Maximum
Obrat (Y)	1021	671254.000	2377247	388	2.70E+07
Fixní aktiva (K)	1021	367185.800	1616156	5	1.68E+07
Zaměstnanců (L)	1021	2674.583	10248	50	110882
Veřejný podnik	1021	0.100	0.300	0	1
Strategie 1: Nízkonákladový podnik	1021	0.086	0.281	0	1
Strategie 2: Podniky v alianci dopravců	1021	0.085	0.279	0	1

Zdroj dat: Vlastní zpracování

Tradiční pojetí mikroekonometrické produkční funkce v ekonometrických modelech vychází z klasického teoretického podkladu Cobb-Douglésovi (C-D) produkční funkce. Tu definujeme jako produkční funkci vyjadřující závislost produktu na kapitálu a práci. Vyznačuje se mnoha nereálnými předpoklady a zjednodušením, které lze v ekonometrickém modelu do určité míry zohlednit. Jsou to např. konstantní výnosy z rozsahu a konstantní podíly kapitálových důchodů a pracovních důchodů na vytvořeném produktu. V práci použijeme v ekonometrickém modelu rozšířenou C-D produkční funkci. Odvození funkce vyjadřující produktivitu práce Q/L na kapitálové intenzitě K/L vychází ze standardní analýzy produkční funkce a je následující:

$$Q/L = Y = A \cdot (K/L)^\alpha (L)^{\alpha+\beta-1} \quad (1)$$

$$\ln(Y/L) = \alpha \cdot \ln(K/L) + (\alpha+\beta-1) \cdot \ln(L) + \ln(A) \quad (2)$$

Panelový odhad s fixním efektem, kde „ a_i “ reprezentuje individuální efekty:

$$\ln(Y_{it}/L_{it}) = a_i + \text{konstanta} + \beta_1 \cdot \ln(K/L)_{it} + \beta_2 \ln(L)_{it} + \beta' x'_{it} + u_{it} \quad (3)$$

Panelový odhad s náhodným efektem s chybou u_i , model bude obsahovat části individuálních efektů společných mezi sledovanými jednotkami.

$$\ln(Y_{it}/L_{it}) = \mu_i + \text{konstanta} + \beta_1 \cdot \ln(K/L)_{it} + \beta_2 \ln(L)_{it} + \beta' x'_{it} + u_{it} \quad (4)$$

Protože předpokládáme, že sledovaná strategie podniků bude společně určovat jejich strukturu a produktivitu, tj. efekt bude společný i mezi jednotkami v čase, otestujeme vhodnost použití efektivnějšího modelu s náhodným efektem oproti modelu s pevnými efekty Hausmanovým testem.

Tabulka 9: Odhad s fixním efektem

$\ln(Y/L)$	Koef.	Robustní Std. Ch.	t	P>t
$\ln(K/L)$	0.18	0.04	4.31	0.000
$\ln(L)$	-0.12	0.09	-1.28	0.200
konstanta	5.45	0.57	9.57	0.000
R-sq: within = 0.1358		F(2,211)	=	9.61***

Zdroj dat: Vlastní zpracování

Panelový odhad s fixními efekty vysvětluje cca 14 % variability modelu a neposkytuje příliš uspokojivé výsledky pro ověření stanovených hypotéz, protože veškeré časově stálé proměnné z modelu odstranil. Pro použití náhodného efektu musíme mít odůvodnění, samotné potvrzení Hausmanovým testem nestačí, v našem případě se domníváme, že podniky v leteckém průmyslu jsou si v určitých prvcích, které jsou časově stálé podobné. Provedený Hausmanův test ($\text{Prob} > \chi^2 = 0.9315$) nedokázal zamítnout nulovou hypotézu o nesystematičnosti rozdílů v koeficientech. Za konzistentní odhad v testu považujeme odhad s fixními efekty oproti pravděpodobně efektivnějšímu odhadu s náhodným efektem, který ovšem nemusí být konzistentní. Hausmanův test potvrdil naši domněnku o vhodnosti použití modelu s náhodným efektem a nyní je možné zapojit další proměnné definující vlastnictví a strukturu.

Tabulka 10: Odhad s náhodným efektem

ln(Y/L)	Koef.	Robustní Std. Ch.	t	P>t
ln(K/L)	0.185	0.038	4.80	0.000
ln(L)	-0.119	0.050	-2.39	0.017
Veřejný podnik	-0.604	0.177	-3.41	0.001
Nízkonákladový podnik	0.456	0.152	3.00	0.003
Podnik v alianci	0.563	0.170	3.31	0.001
Konstanta	5.423	0.323	16.81	0.000
R-sq: between = 0.1898		Wald $\chi^2(5) = 50.23***$		

Zdroj dat: Vlastní zpracování

Tato analýza se liší od původní autorčiny práce (Šubrtová, 2012) ve vynechání malých společností. Dospěli jsme k podobným výsledkům a také k potvrzení hypotézy o neefektivním veřejném vlastnictví v daném sledovaném období a výhodnosti strategického chování. Výhodnější je strategie aliančního sdružování, což v exponenciálním vyjádření (díky semilogaritmické podobě modelu v případě dummy proměnných) je přibližně rozdíl produktivity o 1,1 % od geometrického průměru produktivity nízkonákladové společnosti (357,45 tisíc eur). Na příkladu jsme si ukázali použití Hausmanova testu, který je standardní součástí panelového odhadu, pokud zvažujeme model s náhodným efektem.

Zadání samostatné práce (úkolů)

V podnikové databázi vyberte dostupné podniky v období 2012-2015 z Jihočeského a Středočeského kraje. Identifikujte proměnné Id, Rok účetní závěrky, Jihočeský kraj (0,1) tržby, dlouhodobý kapitál, počet zaměstnanců. Pomocí produkční funkce proveďte analýzu firem a rozhodněte o uplatnění modelu s náhodným efektem při zapojení časově stálé proměnné kraj.

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 6-33)

3.6 Průběžný test

Klíčová slova

Regresní model, časové řady, panelová data, logika

Cíle kapitoly

Cílem kapitoly je procvičit si látku z předchozích kapitol a připravit se na průběžný test.

Výstupy z učení

- 16.6 modeluje časové řady pomocí fuzzy logiky
- 16.7 modeluje časové řady pomocí umělých neuronových struktur
- 16.8 modeluje časové řady pomocí genetických algoritmů

Příklad, uvedení vzorového úkolu

1. Jaký je rozdíl mezi odhadem panelových dat s fixními efekty a náhodným efektem? Panelová data jsou ideální pro územní, mezioborové a meziodvětvové srovnání, kdy pro sledovanou jednotku (i) území, jednotlivce, domácnost, či firmu máme více pozorování v čase (t). Panelová struktura umožňuje sledovat zákonitosti a variabilitu uvnitř těchto jednotek, dále prozkoumat zákonitosti mezi jednotkami a mezi samotnými obdobími, pokud tyto jednotky mají např. nějaké společné prvky (forma vlastnictví, stejné odvětví, úroveň bohatství, stupeň tolerance).

Fixní efekty v modelu jsou výchozí strategií pro panelový odhad. Předpokládáme, že zde existují pevně dané nenáhodné a časově (t) stálé individuální charakteristiky jednotek (i), které nevstupují do našeho odhadu a zkoumání. Jinak řečeno z modelu vyjmeme to, co by mohlo porušit MLR 2 podmínky. Sledujeme tedy pouze odhad časově (t) proměnných zákonitosti uvnitř sledovaných jednotek (i). Avšak ne vždy fixní efekty odpovídají realitě, přesto je tato strategie odhadu důležitá pro další testování efektivnosti odhadu.

Odhad je postaven na OLS, se zahrnutím dummy proměnných pro každou firmu. Necht' panel má následující strukturu.

Tabulka 11: Struktura dat

Id (Panel identifikátor)	Rok (Panel identifikátor)	Y Tržby v mil. Kč	X1 Počet zaměstnanců	X2 Objem kapitálu v mil. Kč
1	2004	145	123	17
1	2005	151	122	18
2	2004	144	105	5
2	2005	143	108	6
3	2004	1211	412	48
3	2005	1212	428	49

Zdroj dat: Vlastní zpracování

Abychom mohli v OLS modelu udělat fixní efekty, přidáme 3 dummy proměnné, počet dummy proměnných odpovídá počtu unikátních id v datovém vzorku. Pokud bychom měli pro 78 firem po 3 pozorování v letech 2004, 2005 a 2006, pak vytvoříme 78 dummy proměnných. Zpět k našemu příkladu, kde máme jen 3 firmy.

Tabulka 12: Struktura dat s dummy proměnnou pro fixní efekty

Id (Panel identifikátor)	Rok (Panel identifikátor)	Y	X1	X2	D_id_1 1 když Id==1 Jinak 0	D_id_2 1 když Id==2 Jinak 0	D_id_3 1 když Id==3 Jinak 0
1	2004	145	123	17	1	0	0
1	2005	151	122	18	1	0	0
2	2004	144	105	5	0	1	0
2	2005	143	108	6	0	1	0
3	2004	1211	412	48	0	0	1
3	2005	1212	428	49	0	0	1

Zdroj dat: Vlastní zpracování

Zahrnutím dummy proměnných do modelu zajistíme „odfiltrování“ individuálních v čase stejných charakteristik. Z předchozího výkladu o dummy proměnných si vzpomeňte, že pomocí dummy proměnné vytváříme a definujeme nové konstanty pro všechny sledované skupiny, které určují do modelu vložené dummy proměnné.

$$Y_{it} = a_i + \text{konstanta} + \beta_1 X_{1it} + \beta_2 X_{2it} + u_{it} \quad (1)$$

Rovnice 1 pak ukazuje, jak bude vypadat odhad. Výraz „ a_i “ lze rozepsat následovně:

$$a_i = \alpha_1 D_{id_1} + \alpha_2 D_{id_2} + \alpha_3 D_{id_3} \quad (2)$$

Výraz a_i zahrnuje veškeré „ID dummy“ proměnné a abychom nespadli do pasti dummy proměnných (dummy trap), jednu proměnnou (např. D_{id_3}) odebereme, ta je definována konstantou v modelu (1). Konstanta v modelu popisuje průměrné tržby pro podnik s $id = 3$ při nula zaměstnancích a nulovém kapitálu.

Ale co když některé časově stále charakteristiky (zda je to firma v daném regionu, zda je to technologicky náročné odvětví, zda je tam vysoký podíl manažerek...) jsou zajímavé? Model s fixními efekty nám je „vyhodí“ a nelze je ani moc dobře interpretovat, víme např., že firmy se adaptují na nové technologie, mají nové zaměstnance, nové manažery a jen velmi málo věcí zůstává časově neměnných. Nechť náš příklad se rozroste o takovou X_3 dummy proměnnou (ženy ve vedení > 50 %), která je v čase stálá.

Tabulka 13: Struktura dat s dummy proměnnou pro fixní efekty a časově stálou proměnnou

Id (Panel identifikátor)	Rok (Panel identifikátor)	Y	X1	X2	D_id_1 1 když Id==1 Jinak 0	D_id_2 1 když Id==2 Jinak 0	D_id_3 1 když Id==3 Jinak 0	X3 1 když ženy jinak 0
1	2004	145	123	17	1	0	0	1
1	2005	151	122	18	1	0	0	1
2	2004	144	105	5	0	1	0	0
2	2005	143	108	6	0	1	0	0
3	2004	1211	412	48	0	0	1	0
3	2005	1212	428	49	0	0	1	0

Zdroj dat: Vlastní zpracování

Lze vidět, že mezi lety 2004 a 2005 se X_3 nemění, je to časově stálé a navíc v perfektní kolinearitě se skupinou všech $D_{id_}$ proměnných. OLS odhad s fixními efekty proto proměnnou vyhodí. Co když je ale X_3 proměnná, která je náhodná a neporušuje MLR 2 a MLR 4 podmínky? Pokud tomu tak je – lze použít odhad s náhodným efektem (RE model).

$$Y_{it} = \mu_i + \text{konstanta} + \beta_1 K_{it} + \beta_2 L_{it} + \gamma X_{3t} + u_{it} \quad (3)$$

Náhodný efekt (RE) v modelu je alternativní metodou odhadu s fixními efekty (rovnice 3). Fixní efekty, které byly dány stranou lze do určité míry „vpustit do modelu“. Aneb X_3 proměnou vytáhneme z členu a_i a ponecháme ji v odhadu. Člen μ_i jakoby rozdrobíme na:

$$\mu_i = a_i + \gamma X_3 \quad (4)$$

Tyto individuální efekty mohou být náhodné, dvě pozorování o dvou ID jednotkách (firma 2 a firma 3) si budou více podobná (manažerky zvyšují tržby), než dvě pozorování téže jednotky (ve firmě 2, teoretická změna vedení z manažerek na manažery). Budou zde existovat efekty, původně individuální, ale do určité míry vpuštěné do modelu, budou s určitým vlivem na sledovanou vysvětlovanou proměnou a nezávislé na ostatních vysvětlovaných proměnných a chybovém členu, nechceme aby γX_3 bylo korelované s u_{it} .

Tento model je pochopitelně mnohem efektivnější vysvětlenou variabilitou a měl by být vždy upřednostněn před fixními efekty. Nicméně, my víme, že tento model je postaven na možnosti porušit MLR 2 (náhodnost výběru), proto musíme být opatrní. V případě, že „vpuštěný náhodný efekt“ koreluje s ostatními regresory (porušení MLR 4) a jsou si více podobná dvě pozorování v čase o téže jednotce než mezi dvěma jednotkami ve stejném období, model s náhodným efektem je nepoužitelný a zkreslený, porušující G-M podmínky.

Pro test takové hypotézy slouží Sargan Hansenův, případně Hausmanův test, který pomůže otestovat vhodnost použití efektivnějšího modelu s náhodným efektem, který využívá tzv. GLS metodu odhadu – generalizované nejmenší čtverce, oproti modelu s fixními efekty. Resp. Hausmanův test nám řekne, zda model s náhodným efektem lze použít. Co tento test dělá? Postaví proti sobě odhadnuté koeficienty konzistentního modelu s fixními efekty, oproti modelu s náhodným efektem (pravděpodobně porušení MLR 2 i MLR 4). Pokud se odhadnuté koeficienty RE a FE modelu výrazně odlišují, bude Hausmanův test ($p\text{-value} < 0.05$) signifikantní, což podporuje tvrzení o nekonzistenci modelu s náhodným efektem a musí být použit model s fixními efekty. Jinak řečeno, je nutné ponechat individuální efekty „mimo“, protože mají vliv na konzistenci odhadu. Problémem Hausmanova testu však je, že není silným testem. Nesignifikantní Hausmanův test by měl být doplněn vždy testem silnějším. Hausmanův test vyžaduje splnění SLR 5 podmínky (homoskedasticity), jinak řečeno nelze jej použít pro odhady s přítomnou heteroskedasticitou. Silnější test je Sargan-Hansenův test, který nevyžaduje striktní homoskedasticitu a otestuje heteroskedasticitně robustní (robustní standardní chyba) RE oproti FE.

Zadání samostatné práce (úkolů)

1. Proveďte analýzu panelu dat firem, které školily své zaměstnance

Stáhněte si soubor JTRAIN.GDT pro program Gretl:
<https://www.dropbox.com/s/cwxt4s27w59m83d/jtrain.gdt?dl=0>

Ten popisuje panel dat firem, které školily své zaměstnance. Vysvětlovaná proměnná „hrsemp“ je počet hodin školení na zaměstnance. Ostatní proměnné jsou vysvětlující (počet zaměstnanců employ, logaritmovaný počet zaměstnanců lemploy, poskytnutý grant, hodnota grantu v předchozím roce grant_1...). Odhadněte model s fixními efekty, robustní standardní chybou a se zahrnutím časových indikátorových proměnných (stačí to zaškrtnout ve specifikaci modelu a proměnné dt_2 a dt_3 se vytvoří samy):

$$\text{Hrsemp}_{it} = \alpha_i + \text{konstanta} + \beta_1 \text{grant}_{it} + \beta_2 \text{grant_1}_{it} + \beta_3 \text{lemploy}_{it} + \gamma_4 \text{dt_2} + \gamma_5 \text{dt_3} + u_{it}$$

- Kolik firem bylo zahrnuto do odhadu?
- Kolik bylo celkem pozorování?
- Kolik by bylo celkem pozorování, pokud by byl panel balancovaný?
- Jaký je efekt grantu na počet hodin školení na zaměstnance?
- Jaké může být reálné vysvětlení toho, že proměnná grant_1 (grant v předchozím roce) není statisticky signifikantní?
- Je efekt velikosti formy přítomný? Pokud má firma o 10 % více zaměstnanců, jaký je efekt na počet hodin školení na zaměstnance?
- Plní model podmínku MLR 4? Jaká je střední hodnota reziduí? Mají rozdělení alespoň podobné normálnímu?

Studijní literatura

Povinná literatura

HAMPEL, D., V. BLAŠKOVÁ a L. STŘELEČ, 2016. *Ekonometrie 2*. 3. přeprac. vyd. Brno: Mendelova univerzita. ISBN 978-80-7509-427-8. (s. 1-93)

3.7 Procvičování základních principů teorie fuzzy množin a fuzzy logiky

Klíčová slova

Množina, sjednocení, funkce, náležitost

Cíle kapitoly

Cílem je procvičování základních principů teorie fuzzy množin a fuzzy logiky.

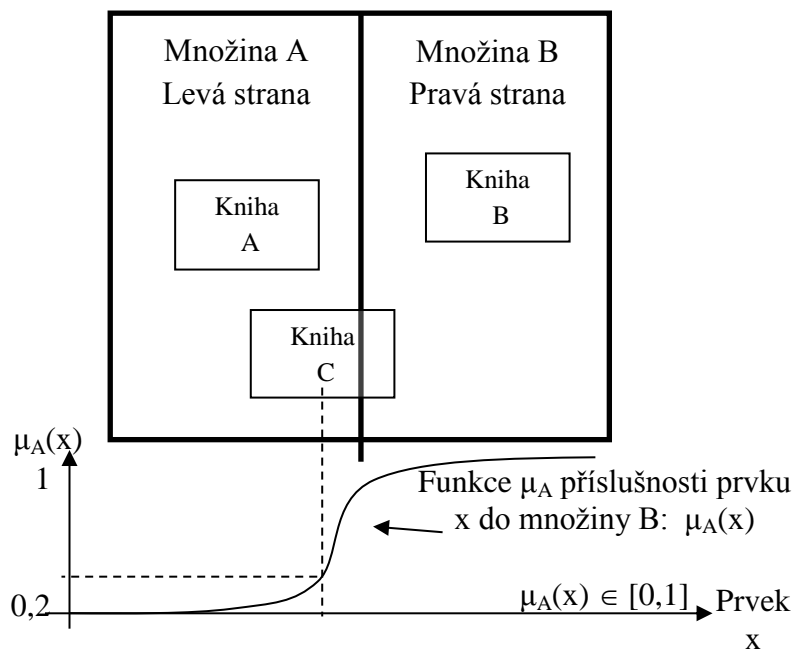
Výstupy z učení

- 16.1 zná základní pojmy a pravidla fuzzy logiky
- 16.6 modeluje časové řady pomocí fuzzy logiky

Příklad, uvedení vzorového úkolu

1. Definujte vhodnou funkci příslušnosti prvku x do množiny B

Obrázek 12: Fuzzy množina a charakteristická fuzzy funkce náležitosti do množiny A a B



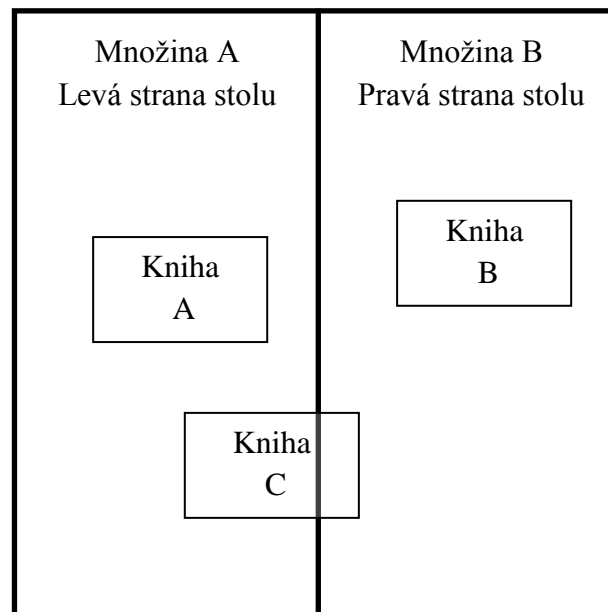
Zdroj: Vlastní zpracování

Knihy C náleží do množiny B jen částečně, hodnota $\mu_B(x) = 0,2$, tj. 20% náležitost do množiny B.

Zadání samostatné práce (úkol)

Definujte vhodnou funkci příslušnosti Knihy B do množiny B.

Obrázek 13: Fuzzy množina a charakteristická fuzzy funkce



Zdroj: Vlastní zpracování

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 6-54)

3.8 Modelování časových řad pomocí fuzzy logiky

Klíčová slova

Funkce příslušnosti, množinové operace, jádro, nosič

Cíle kapitoly

Cílem je procvičit modelování časových řad pomocí fuzzy logiky.

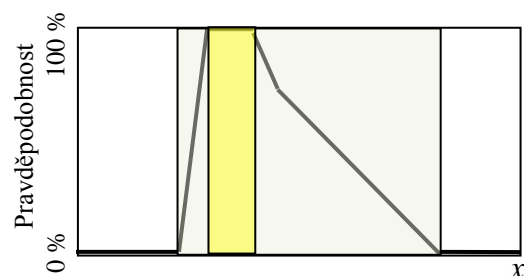
Výstupy z učení

- 16.1 zná základní pojmy a pravidla fuzzy logiky
- 16.6 modeluje časové řady pomocí fuzzy logiky

Příklad, uvedení vzorového úkolu

1. Pro danou funkci příslušnosti definujte nosič fuzzy množiny a jádro fuzzy množiny.

Obrázek 14: Funkce příslušnosti pro hodnocení odměny zákaznické podpory



Zdroj: Vlastní zpracování

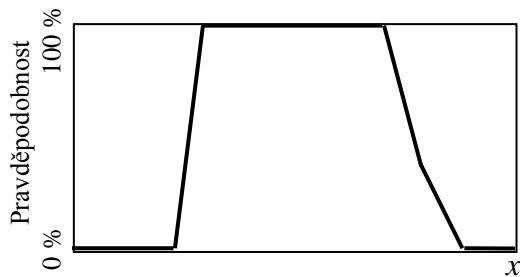
Nosič fuzzy množiny A definované na univerzu X s prvky x je ostrá množina všech prvků s nenulovou příslušností do této fuzzy množiny. V obrázku je vyznačeno světle zelenou barvou.

Jádro fuzzy množiny A definované na univerzu X s prvky x je ostrá množina všech prvků, jejichž příslušnost se rovná hodnotě 1, $\alpha = 1$. V obrázku je vyznačeno žlutou barvou.

Zadání samostatné práce (úkolů)

1. Pro danou funkci příslušnosti definujte nosič fuzzy množiny a jádro fuzzy množiny.

Obrázek 15: Funkce příslušnosti



Zdroj: Vlastní zpracování

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 55-110)

3.9 Procvičování základních principů umělých neuronových sítí

Klíčová slova

Skrytá, vstupní, výstupní vrstva, učení, neuron, perceptron

Cíle kapitoly

Cílem je procvičování základních principů umělých neuronových sítí.

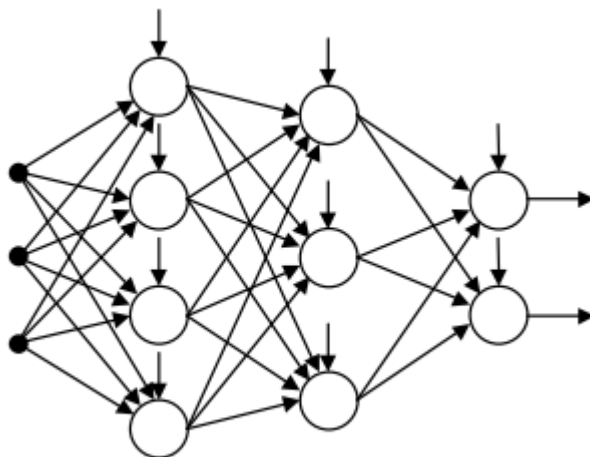
Výstupy z učení

- 16.2 rozumí fungování umělých neuronových struktur
- 16.7 modeluje časové řady pomocí umělých neuronových struktur

Příklad, uvedení vzorového úkolu

1. Popište danou strukturu na obrázku:

Obrázek 16: Struktura sítě



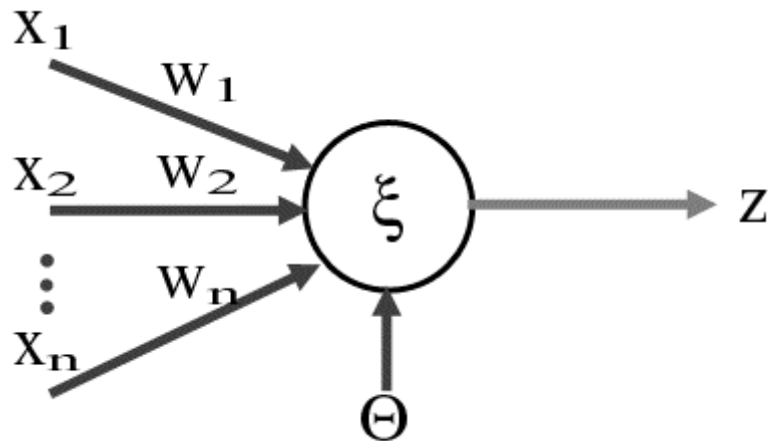
Zdroj: Vlastní zpracování

Jde o jednoduchou třívrstvou neuronovou síť se třemi vstupy a dvěma výstupy. Uprostřed je skrytá vrstva. Jednotlivé neurony jsou vzájemně propojeny spoji ohodnocenými vahami. Základem této sítě je perceptron s prahem citlivosti neuronu.

Zadání samostatné práce (úkolů)

1. Popište daný obrázek

Obrázek 17: Struktura UNS



Zdroj: Vlastní zpracování

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 111-139)

VOCHOZKA, M. et al., 2017. *Využití neuronových sítí při komplexním hodnocení podniků*. Praha: C. H. Beck. ISBN 978-80-7400-642-5. (s. 67-80)

3.10 Modelování časových řad pomocí UNS

Klíčová slova

Statistica, programová výbava, vzorkování, zpoždění

Cíle kapitoly

Cílem kapitoly je procvičit modelování časových řad pomocí UNS.

Výstupy z učení

- 16.2 rozumí fungování umělých neuronových struktur
- 16.7 modeluje časové řady pomocí umělých neuronových struktur

Příklad, uvedení vzorového úkolu

Jak probíhá práce s časovými řadami v programu STATISTICA?

K modulu *Automatizované neuronové sítě* v programu STATISTICA se lze dostat přes menu Statistiky. V úvodním dialogu neuronových sítí pro možnost Časové řady jsou zvoleny „Proměnné“ a vybrána spojitá „Cílová proměnná“. Strategii pro vytváření prediktivních modelů lze ponechat na defaultní možnosti „Vyhledávání automatiz. sítí (ANS)“. Tato volba ulehčuje práci s neuronovými sítěmi, lze zvolit i Vlastní nastavení. Dále v dialogovém okně „SANS: Výběr dat“ je třeba přepnout na záložku „Vzorkování (VNS a ANS)“, kde je možnost výběru vzorkovacích metod.

Vstupní soubor je třeba rozdělit na 3 části. Trénovací množina – náhodně vybraná část dat, která slouží pro učení sítě, Testovací množina – další část dat sloužící k zastavení trénování, aby nedošlo k přeučení sítě, Validační množina – zbytek dat, na kterém se ověří konečná kvalita modelu. Jde o data, která model dosud k dispozici neměl. Obvykle se rozdělení datového souboru na tyto množiny doporučuje v poměru 50-25-25, případně 70-15-15. V záložce Časové řady v úvodním dialogu SANS je dále k dispozici možnost definovat počet období (např. dva minulé roky) zpožděných hodnot časové řady a také počet predikovaných hodnot (pět budoucích let). V následujícím dialogu je vhodné vybrat i síť RBF a poté již program vytvoří modely neuronových sítí. Po trénování sítí naběhne

výsledkový dialog „SANS: Výsledky“. Pro následné zkoumání kvality se doporučuje vybrat tu síť, která je jednak nejlepší a jednak stabilní ve všech sledovaných výkonech. Výsledky vybraného modelu lze vizualizovat pomocí tlačítka Graf projekce.

Zadání samostatné práce (úkolů)

Proveďte predikci časové řady indexu PX50 (měsíční průměry) pomocí 5 zpoždění a predikcí 2 následujících období.

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 140-182)

VOCHOZKA, M. et al., 2017. *Využití neuronových sítí při komplexním hodnocení podniků*. Praha: C. H. Beck. ISBN 978-80-7400-642-5. (s. 80-106)

3.11 Pokročilé modelování časových řad pomocí UNS

Klíčová slova

Neuronové sítě, modelování, časové řady

Cíle kapitoly

Procvičit znalosti o fungování umělých neuronových struktur a modelování časových řad pomocí umělých neuronových struktur.

Výstupy z učení

- 16.2 rozumí fungování umělých neuronových struktur
- 16.7 modeluje časové řady pomocí umělých neuronových struktur

Příklad, uvedení vzorového úkolu

Jaké jsou nejčastěji používané modely reálných vztahů v oblasti predikce ekonomických proměnných? Jaké metody byste využili při modelování vývoje hrubého domácího produktu pomocí dopředných neuronových sítí? Najděte v databázi Proquest tři další studie aplikující neuronové sítě.

Řešení: metoda extrapolace časových řad, ekonometrické modely, metoda scénářů, reflexivní a intuitivní metody, expertní metody.

Popis metod například: ŠEDA, I. a V. OLEJ, 2007. Modelování vývoje hrubého domácího produktu pomocí dopředných neuronových sítí. *Scientific papers of the University of Pardubice. Series D, Faculty of Economics and Administration*. **11**(-), 118-125. ISSN 1211-555X.

Tři studie:

GJYLAPI, D. a V. DURMISHI, 2014. Artificial neural networks in forecasting tourists' flow, an intelligent technique to help the economic development of tourism in Albania. *Academicus International Scientific Journal*. -(10), 202-211. ISSN 2079-3715.

LAI, S. L. et al, 2014. Energy consumption forecasting in Hong Kong using ARIMA and artificial neural networks models. *Applied Mechanics and Materials*. **672-674**(-), 2085-2097. ISSN 1662-7482.

AYYOUB, M. a A. RIAZ, 2017. The Artificial Neural Network Method: A Practical Guide for Business Research. *Journal of Business Strategies*. **11**(1), 113-132, 171-172. ISSN 2085-2097.

Zadání samostatné práce (úkolů)

Práce s vědeckým textem – replikace modelu HDP

Seznamte se s tímto textem: ŠEDA, I. a V. OLEJ, 2007. Modelování vývoje hrubého domácího produktu pomocí dopředných neuronových sítí. *Scientific papers of the University of Pardubice. Series D, Faculty of Economics and Administration*. **11**(-), 118-125. ISSN: 1211-555X.

Zjistěte, jaký navrhli model na predikci hrubého domácího produktu.

Řešení:

ŠEDA, I. a V. OLEJ, 2007. Modelování vývoje hrubého domácího produktu pomocí dopředných neuronových sítí. *Scientific papers of the University of Pardubice. Series D, Faculty of Economics and Administration*. **11**(-), 120-125. ISSN 1211-555X.

Studijní literatura

Povinná literatura

VOCHOZKA, M. et al., 2017. *Využití neuronových sítí při komplexním hodnocení podniků*. Praha: C. H. Beck. ISBN 978-80-7400-642-5. (s. 107-166)

3.12 Procvičování základních principů genetických algoritmů

Klíčová slova

Genetické algoritmy, modelování, časové řady

Cíle kapitoly

Procvičit znalosti o genetických algoritmech, procvičit modelování časových řad pomocí genetických algoritmů.

Výstupy z učení

- 16.3 rozumí genetickým algoritmům
- 16.8 modeluje časové řady pomocí genetických algoritmů

Příklad, uvedení vzorového úkolu

Zjistěte, jaké postupy byly využity – práce s vědeckým textem.

Zjistěte, jaké postupy byly využity při vytváření přehledu a srovnání existujících knihoven pro genetické algoritmy v C# .NET v práci: HUNĚK, M., 2016. *Knihovna pro genetické algoritmy v C#.NET*. Diplomová práce. Masarykova univerzita. Fakulta informatiky.

Dostupné

z:

https://is.muni.cz/th/393640/fi_m/Knihovna_pro_geneticke_algoritmy_C_.pdf

Řešení:

HUNĚK, M., 2016. *Knihovna pro genetické algoritmy v C#.NET*. Diplomová práce. Masarykova univerzita. Fakulta informatiky. Dostupné z: https://is.muni.cz/th/393640/fi_m/Knihovna_pro_geneticke_algoritmy_C_.pdf, s. 28-65.

Zadání samostatné práce (úkolu)

Srovnejte mezi sebou tři základní operace genetických algoritmů, tj. selekce, křížení a mutaci.

Řešení:

STUDNIČKA, V., 2010. *Genetické algoritmy – Multi-core CPU implementace*. Diplomová práce. Vysoké učení technické v Brně. Fakulta strojního inženýrství. Ústav automatizace a informatiky. Dostupné z: https://www.vutbr.cz/www_base/zav_prace_soubor_verejne.php?file_id=33586

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 183-217)

3.13 Modelování časových řad pomocí GA

Klíčová slova

Genetické algoritmy, modelování, časové řady

Cíle kapitoly

Procvičit znalosti o genetických algoritmech, procvičit modelování časových řad pomocí genetických algoritmů.

Výstupy z učení

- 16.3 rozumí genetickým algoritmům
- 16.8 modeluje časové řady pomocí genetických algoritmů

Příklad, uvedení vzorového úkolu

Seznamte se se zhodnocením provedených experimentů a metodickými doporučeními v práci: ŠTENCL, M. 2010. *Predikce a optimalizace reálných dat pomocí algoritmů umělé inteligence*. Disertační práce. Mendelova univerzita v Brně. Provozně ekonomická fakulta. Dostupné z: <https://is.mendelu.cz/lide/clovek.pl?zalozka=13;id=19296;studium=34172>.

Jak byste mohli dané metody využít při zpracování Vaší diplomové práce?

Řešení:

ŠTENCL, M. 2010. *Predikce a optimalizace reálných dat pomocí algoritmů umělé inteligence*.

Disertační práce. Mendelova univerzita v Brně. Provozně ekonomická fakulta. Dostupné z: <https://is.mendelu.cz/lide/clovek.pl?zalozka=13;id=19296;studium=34172>. (s. 91-99)

Podstatné jsou autorova doporučení pro aplikace v rámci systémů podpory rozhodování a pro využití neuronových sítí v rámci analytických nástrojů na úrovni operačního managementu při paralelním využívání statistických metod. V závěru uvedl: „*Obecně tak lze doporučit použití neuronových sítí v rámci jednotlivých fází rozhodovacího procesu. Pro nasazení metod neuronových sítí je třeba mít na paměti nutnost přeučování sítě.*

Celkový proces nového natrénování sítě je však možné kombinovat s různými dalšími technikami. Ve vlastní práci byla pro odvození vhodné topologie použita metoda cross-validace. Se správně připravenými datovými soubory by bylo možné metodu naimplementovat automatizovaně. To je i oblast dalšího pokračování vlastní práce. Jako další uváděnou metodu pro stanovení počtu neuronů lze zmínit genetické algoritmy. Právě při pohledu na provedené experimenty ve vlastní práci je možné identifikovat hlavní problém při použití neuronových sítí z hlediska automatizovaného zpracování. Zcela jistě se jedná o proces trénování sítě, pomineme-li mohutnost použitých datových souborů. Problém s mohutností datových souborů lze označit za obecný problém i při použití metod statistiky. Problém trénování je spojen se dvěma rovinami. První jsou znalosti a zkušenosti nutné pro kvalitní natrénování (včetně volby dalších typů sítí) neuronové sítě s požadovanou přesností výstupu. Druhým problémem je výpočetní náročnost procesu trénování. Ani jeden z těchto problémů nelze zatím účinně eliminovat. Pracovat však lze s naučenou neuronovou sítí. Z pohledu uživatele jsou natrénované neuronové sítě černou skříňkou, která mu vrací požadované údaje. Mnohdy tak ani neví, že pracuje s neuronovou sítí (Štencl 2010, s. 98-99).

Zadání samostatné práce (úkol)

Zjistěte studie, které se zabývají využitím neuronových sítí pro modelování na burze cenných papírů. Porovnejte je, zda autoři doporučují nahradit statistické metody využitím neuronových sítí.

Řešení:

GAVUROVÁ, B. et al, 2017. Analysis of impact of using the trend variables on bankruptcy prediction models performance. *Ekonomický časopis*. **65**(4), 370-383. ISSN 0013-3035.

NORADDIN, M. A., A. A. MOHAMMAD a M. F. MAHDI, 2014. An integrated system based on fuzzy genetic algorithm and neural networks for stock price forecasting. *The International Journal of Quality & Reliability Management*. **31**(3), 281-292. ISSN 0265-671X.

RAJIHY, Y., K. NERMEND a A. ALSAKAA, 2017. Back-propagation artificial neural networks in stock market forecasting. an application to the warsaw stock exchange WIG20. *Aestimatio*. -(15), 88-99. ISSN 2173-0164.

ZÁKOUTSKÝ, V., 2013. *Modelování na burze cenných papírů metodami umělé inteligence*. Diplomová práce. Univerzita Pardubice. Fakulta ekonomicko-správní. Ústav systémového inženýrství a informatiky. Dostupné z: https://dk.upce.cz/bitstream/handle/10195/51913/ZakoutskyV_ModelovaniNaBurze_VO_2013.pdf?sequence=3&isAllowed=y

Studijní literatura

Povinná literatura

DOSTÁL, P., 2015. *Soft computing v podnikatelství a veřejné správě*. Brno: CERM Akademické nakladatelství. ISBN 978-80-7204-898-4. (s. 218-243)