Univerzita Pardubice

Fakulta ekonomicko-správní

Analýza občanské vybavenosti regionu s využitím fuzzy asociačních pravidel

Bc. Bohuslav Brenner

Diplomová práce

2009

Univerzita Pardubice
Fakulta ekonomicko-správní
Ústav systémového inženýrství a informatiky
Akademický rok: **2008/2009**

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení:	**Bohuslav BRENNER**
Studijní program:	**N6209 Systémové inženýrství a informatika**
Studijní obor:	**Informatika ve veřejné správě**

Název tématu:	**Analýza občanské vybavenosti regionu s využitím fuzzy asociačních pravidel.**

Z á s a d y p r o v y p r a c o v á n í :

1) Základní pojmy z oblasti regionu a jeho vybavenosti.
2) Analýza základních metod pro generování asociačních pravidel.
3) Porovnání (komparace) vybraných regionů s využitím asociačních pravidel.
4) Navržení a verifikace modelu.

Rozsah grafických prací:
Rozsah pracovní zprávy:
Forma zpracování diplomové práce: **tištěná/elektronická**
Seznam odborné literatury:

Petr, P.: Data Mining: Díl I, Pardubice: Univerzita Pardubice, 2006.
Pyle D. Data Preparation for Data Mining. San Diego, Academic Press, 1999, 540 s.. San Diego, 1999.
Rud, O. L. Data Mining - Praktický průvodce dolováním dat pro efektivní prodej, cílený marketing a podporu zákazníků (CRM). Praha, Computer Press, 2001, 330 s.. , 2001.

Vedoucí diplomové práce: **doc. Ing. Pavel Petr, Ph.D.**
Ústav systémového inženýrství a informatiky

Datum zadání diplomové práce: **6. října 2008**
Termín odevzdání diplomové práce: **1. května 2009**

doc. Ing. Renáta Myšková, Ph.D.
děkanka

L.S.

doc. Ing. Jiří Křupka, Ph.D.
vedoucí ústavu

V Pardubicích dne 6. října 2008

Prohlašuji:

Tuto práci jsem vypracoval samostatně. Veškeré literární prameny a informace, které jsem v práci využil, jsou uvedeny v seznamu použité literatury.

Byl jsem seznámen s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně.

V Pardubicích dne 24. 8. 2009

Bc. Bohuslav Brenner

Anotace

Tato práce je rozdělena na čtyři hlavní části, kde část první je věnována popisu regionu a vlastní občanské vybavenosti regionu, část druhá pak asociačním pravidlům a metodám, které jsou s nimi spojené. Třetí část se zabývá fuzzy problematikou a fuzzy asociačními pravidly. Čtvrtá, poslední část je pak porovnáním použitých metod.

Cílem práce je osvětlit pojmy region, občanská vybavenost regionu, asociační pravidla a fuzzy asociační pravidla, dále pak použít vhodné metody k analýze občanské vybavenosti regionu jak pomocí asociačních, tak i fuzzy asociačních pravidel, a následně dané metody porovnat.

Klíčová slova

Region, občanská vybavenost regionu, asociační pravidla, fuzzy asociační pravidla.

Annotation

This theses are devided into four main parts where the first part is devoted to a description of a region and civic facilities of the region, the second part includes association rules and a methodology which are conected to them. The third part deals with fuzzy matters and fuzzy association rules. The last fourth part then compares used methods.

The goal of this theses is to elucidate terms a region, a civic facilities of a region, association rules and fuzzy association rules, furthemore to use proper methods to analyse the civic facilities of the region through the use of both association and fuzzy association rules and afterwars compare these methods.

Key words

Region, civic facilities of a region, association rules, fuzzy association rules.

OBSAH

Úvod

Tato diplomová práce se zabývá analýzou občanské vybavenosti regionu. Občanská vybavenost regionu a její rozbor hraje v dnešní společnosti nemalou roli. V podstatě lze říci, že region s malou nebo jen nízkou úrovní občanské vybavenosti je jak pro obyvatele, tak i pro potencionální investory, méně zajímavým (výjimkou však může být například absence obchodního centra v regionu, ve kterém se developerská společnost rozhodne právě toto centrum vystavět). Upustím-li od uvažování polohy regionu, pak jeho vybavenost může také dokonce velkou částí regulovat ceny nemovitostí či služeb v tomto regionu.

Cílem práce je tedy ozřejmit pojmy, jako jsou region, občanská vybavenost regionu, asociační pravidla a fuzzy asociační pravidla, dále zvolení a použití vhodné metody k analýze občanské vybavenosti regionu, a to jak pomocí asociačních, tak i fuzzy asociačních pravidel, a následně porovnání daných metod.

V první části práce je definován region, jeho hranice, struktura, řád i hierarchie, načež navazuje územně správní (administrativní) členění České republiky a Německa. Dále pak je zde uveden pojem občanská vybavenost, který je také názorně popsán na příkladu města Bíliny jak ve sféře veřejno-prospěšné, tak komerční. Část druhá je věnována asociačním pravidlům, jejich vlastnostem, zobecnění a výjimkám, stejně jako metodám prohledávání a dalším metodám sloužícím ke generování pravidel, jako je metoda GUHA, algoritmus apriori či analýza nákupního košíku. Pomocí posledně jmenované metody je také provedena vlastní analýza občanské vybavenosti zvolených regionů (krajských měst). V další části je již osvětlena fuzzy problematika, a to počínaje fuzzy množinami, přes výrokovou fuzzy logiku, funkce příslušnosti, Mukaidonův fuzzy rezoluční princip, až po samotná fuzzy asociační pravidla. Následuje opět analýza občanské vybavenosti, ale tentokrát právě za použití zmíněných fuzzy asociačních pravidel. Poslední částí je část čtvrtá, věnovaná porovnání obou použitých metod.

1. Region a občanská vybavenost regionu

V této části práce jsou vysvětleny základní pojmy z oblasti regionalistiky, jako je region, jeho struktura a hranice či občanská vybavenost, která je navíc názorně popsána na konkrétním příkladu občanské vybavenosti ve městě Bílina.

1.1. Region – základní pojem regionalistiky

Pojem region[1] je od počátku 20. století pojmem velmi frekventovaným, ale zároveň nejednoznačným a rozporuplně vyložitelným. V literatuře lze nalézt celou řadu definic regionu. V souvislosti s regionální historiografií uvádí J. Bartoš jednu definici, která je výhodná svou obecností a univerzálností [8]:

„Region je dnes i v mezinárodním měřítku chápán, i když s některými odlišnostmi podle zemí a oborů, jako obecný název pro menší územně ohraničené jednotky určitého většího prostorového nebo územního celku, tj. světa, světadílů a jednotlivých zemí. Toto nejširší pojetí je však ve společenských vědách většinou specifikováno a zúženo na označení tzv. středně velkých územně rozložených a vymezitelných jednotek uvnitř států, tedy na jednotky s rozlohou menší než stát a větší než jednotlivé lokality a základní sídelní a správní jednotky – obce. Záleží potom na vědním nebo praktickém oboru, zda je touto jednotkou nějaká přírodní krajina či ekosystém nebo je to určitá teritoriální skupina či lidská pospolitost...“.

Regiony jsou objekty regionalistických studií, jejich předmětem jsou prostorové vztahy, jevy a procesy. Prostorové vztahy, jevy a procesy jsou v současné společnosti složitě organizovány a zabývají se osídlením prostoru a strukturou tohoto osídlení. Úkolem regionalistiky je rozdělit osídlení a jeho strukturu do regionů tak, aby vytvářely charakteristické jednotky jako soubory prvků osídlení a jeho struktury, které jsou typické a zároveň od jiných jednotek odlišné. Struktura osídlení tedy znamená uspořádání prvků sídelní soustavy a jejich vazeb (hlavní dělení vazeb je na vnitřní a vnější, slabé a silné) a obsahuje jak přírodně prostorové prvky (např. geologické, geomorfologické, klimatické atd.), tak prvky demografické (obyvatele a jejich mobilitu), a také prvky vytvořené lidskou aktivitou (např. materiální výroba, technická infrastruktura, občanská vybavenost, přenos informací atd.). Předmětné zaměření regionalistiky je studováno na objektech [9]:

[1] Z lat. *regio,* hranice, světová strana, spravovaný kraj

prostorové organizace (struktura osídlení prostoru),

společnosti (společenství lidí přetváří přírodní prostor kulturou a v rámci toho uspořádává své vztahy k přírodě a obývanému prostoru,

regionech (typických, od sebe vzájemně odlišitelných jednotkách), na které lze rozdělit složitou prostorovou organizaci soudobé společnosti.

Regiony jsou také sociálně konstruované, protože v moderní společnosti lidé stále častěji ovlivňují prostředí svého žití. Díky tomu existuje tolik definic regionů vycházejících z řady kritérií, jež jsou brána v úvahu při typizaci a vymezování regionů. Jedná se hlavně o kritéria geografického, demografického, sociokulturního a politickosprávního charakteru. Podle toho se pak regiony třídí a vymezují např. na městské a venkovské, na hospodářsky slabé a strukturálně postižené na jedné straně a relativně hospodářky silné na druhé straně, z hlediska globálního měřítka je využívána klasifikace regionů na země prvního a třetího světa, atd.

Jedna z dalších definic regionu dle [9] je, že *„region je komplex vznikající regionální diferenciace krajiny"*. Regionální diferenciace krajiny je velmi starým historickým pojmem, ale též stále pokračujícím procesem, ve kterém lze rozpoznat základní tendence:

- osvobozovaní lidských společností od závislosti na přírodních faktorech jejich žití (jedná se o faktory mající přírodní ráz, jsou z hlediska činností lidí vnější, exogenní, působí „shora"),

- stále silnější uplatňování kulturních aktivit lidských společností (jedná se o faktory kulturní mající společenský ráz, jsou z hlediska činností lidí vnitřní, endogenní, působí „zdola", jde především o aktivity hospodářské, sociokulturní a občansko-politické).

Stoupenci tzv. systémových teorií obměňují tuto definici takto [9]: *„region je sociálně geografický subsystém a současné sociálně geografické systémy jsou výsledkem formování nové geografické organizace společnosti."* V této definici je zdůrazněno, že strukturu regionu určuje geografické umístění (přírodní faktor), na kterém se projevuje působení společenských (kulturní faktorů).

Když srovnáme obě výše uvedené definice regionů, vidíme, že obě vyjadřují základní tendenci - na prostorové organizaci společnosti se stále více uplatňují faktory, které nemají přírodní charakter, ale kulturní povahu, jsou vytvořeny lidmi. Druhá definice navíc obsahuje aspekt tzv. řádu a hierarchie regionu.

Pro lepší pochopení obou definic připojuji zjednodušený model faktorů a jejich ukazatelů pro popis regionů. Tabulka číslo 1 je pouhou ukázkou výběru druhu faktorů (kritéria pro rozlišení regionů) a k tomu příslušného jednoho ukazatele velikostního (V) a jednoho ukazatele strukturálního (S) typu.

Tabulka 1: Tabulka charakteru regionu, zdroj: [9].

Charakter regionu	Druh faktoru	Ukazatel faktoru
Exogenní (shora) ↓ Zásah Člověka ↓ Endogenní (zdola)	přírodní (geografický)	nadmořská výška (V), poměr nížiny k pahorkatině (S)
	demografický	průměrný věk (V), poměr dětské k produktivní populaci (S)
	ekonomický	počet podnikatelů (V), struktura pracovních příležitostí v zemědělství, průmyslu a službách
	správní	počet obcí (V), poměr obcí ke všem sídlům (S)
	sociálně kulturní	počet knihoven (V), vzdělaností struktura - poměr obyvatelstva se ZŠ, SŠ a VŠ vzděláním (S)

V současné době se pojem region nejčastěji používá v souvislosti s „regionálním rozvojem“ a se strukturální a regionální politikou Evropské unie. Hovoří se o regionech zaostalých, strukturálně postižených, o venkovských mikroregionech, o „regionech soudržnosti“ a podobně. Toto pojetí implikuje region jako jednotku menší než stát a rovinu regionálního rozhodování jako nižší než státní či vládní úroveň.

K základním dílčím charakteristikám regionů a jejich struktury patří hranice regionu, struktura regionu, regionální struktura a řád a hierarchie regionů. Veškeré údaje k těmto pojmům a tedy i kapitolám 1.1.1. až 1.1.4. byly čerpány z [9], [14] a [15].

1.1.1.Hranice regionu

Region je považován za vzájemně se odlišující území, což znamená, že každý region musí být nějak ohraničen. Lze rozlišit dva typy hranic[2], a to fyzickogeografickou hranici, která může být ostrá (také liniová), např. řeka, anebo méně ostrá (tzv. zóna, pásmo), např. pohoří, a dále hranice, která vznikla činností člověka. Ta má zpravidla ostřejší podobu, například ulice, politické a správní hranice. Ve skutečnosti se častěji objevují přechody mezi regiony, které mají spíš zonální charakter. Příklad typického zonálně rozhraničeného území je model používaný ve Francii v rámci venkovského prostoru. Nejtypičtější liniovou hranicí jsou hranice mezi správními okresy (counties) v americkém státě Missouri.

Hranice regionu jsou pak většinou určovány na základě přírodních, historických či správních hranic.

[2] Slovem hranice je obvykle myšlena pomyslná čára vedená na zemském povrchu, která odděluje dvě různé přírodní či společenské entity. [15]

1.1.2.Struktura regionu

Pod strukturou regionu rozumíme vnitřní strukturu regionu a podle této velice podstatné vlastnosti rozlišujeme dva základní typy regionů – homogenní a nehomogenní.

Vymezování regionů na základě principu homogenity vyhledává stejnost, podobnost. Můžeme vymezit region dle shody podle jednoho kritéria (např. převaha pěstování jedné plodiny), anebo dle shody ve více kritériích. Vždy ale platí, že do jedné třídy regionů se zahrnují jen ty, u nichž je nalezena shoda ve všech zvolených kritériích.

Nehomogenní regiony se označují také jako heterogenní, nodální, spádové, uzlové, funkční. Výrazy nodální a uzlové znamenají, že jsou tvořeny jádrem, uzlem, centrem a zázemím. Prostor centra/center a zázemí je funkčně propojen vazbami (slabými – silnými). Takový funkčně propojený celek je charakteristický přemísťováním obyvatel a zboží .

Rozdíl mezi heterogenním a homogenním regionem je především ve funkčnosti. Homogenní regiony se vyznačují stejnorodostí funkcí, heterogenní regiony jsou naopak typické svou různorodostí funkcí, které zajišťují.

1.1.3.Regionální struktura

Zpravidla ji popisujeme pomocí pojmů makroregion, mezoregion a mikroregion (v podstatě podle velikosti), přičemž uvnitř tohoto členění regionu se ještě rozlišují vyšší a nižší stupně.

Česká republika je považována za makroregion vyššího stupně a je tvořena dvěma základními makroregiony nižšího stupně (Čechy, Morava s „českým“ Slezkem). Mezoregiony jsou rozsáhlé územní jednotky, jejichž vnitřní integrita je již méně vázána na prostorové vztahy obyvatelstva a je celkově nižší. Mezoregionální centra jsou významnými prostory koncentrace socioekonomických aktivit. Praha je v tomto smyslu pak metropolí mezinárodního významu, Brno je regionální metropolí I. řádu, Ostrava regionální metropolí II. řádu. Dále následuje mezoregionální centrum I. řádu – Plzeň, a poté mezoregionální centra II. řádu, Olomouc, Liberec, Hradec Králové, Ústí nad Labem, České Budějovice, Pardubice, Zlín a Karlovy Vary. [9] Nejmenšími jednotkami regionální struktury jsou mikroregiony, které jsou považovány za územní celky, v jejichž rámci jsou relativně uzavřeny nejintenzivnější regionální procesy, především dojíždění za prací a základní druhy služeb. Z výše uvedeného lze tedy usoudit, že vnitřní integrita je relativně nejvyšší u mikroregionů. V naší republice je vyvinuta dvoustupňová mikroregionální organizace, což znamená, že vyšší stupeň mikroregionů je dán zpravidla jen správní funkcí. Pro Českou republiku také vyplývá, že nejméně je rozvinutá „střední vrstva“ regionální struktury, mezoregiony, a to proto, že oproti makroregionům i mikroregionům mají nižší míru vnitřní integrity.

1.1.4.Řád a hierarchie regionu

Určení řádu a hierarchie regionů vymezuje organizaci a vztah středisek mezi sebou na ose vyšší – nižší řád, anebo nadřazenost – podřazenost regionů v regionální struktuře z hlediska vybavenosti službami a zbožím. Vyšší řád služby a zboží poskytuje a nižší řád služby a zboží přijímá. Na základě toho rozlišujeme tedy regiony, které vyjadřují vyšší řád, a subregiony, které vyjadřují nižší řád. V odborné literatuře se pro znázornění nejčastěji používá klasifikace W. Christalera (čtyři stupně středisek – vyššího, nižšího, nejnižšího řádu a pomocná střediska), anebo klasifikace A. Lösche, která neobsahuje ostře ohraničené stupně, ale spíš plynulý sled středisek. Obě koncepce jsou založeny na modelu, jehož struktura je zobrazena jako soustava šestiúhelníků s centrálním místem ve středu a okolními vrstvami, v nichž se postupně nacházejí střediska nižšího řádu, nejnižšího řádu a pomocná střediska. [9]

Systém střediskových měst, řád a hierarchie regionů nejsou neměnné. Změny přinášejí jak rozvoj průmyslu, tak služeb, a v dnešní době především komunikačních služeb. Pokud je cílem regionalistických snah poznání za účelem řešení regionálních disparit, vždy se pracuje s modely prostorového rozvoje společnosti, které vedou k naplnění společenských požadavků, k tzv. koncepci prostorového uspořádání společnosti. V České republice se jedná o skupinu regionů se soustředěnou pomocí státu, která je vnitřně členěná na tzv. strukturálně postižené, hospodářsky slabé a venkovské regiony. Dále je to skupina ostatních regionů podporovaných státem z jiných důvodů, např. pohraniční území, bývalé vojenské prostory, regiony postižené živelnými pohromami a další.

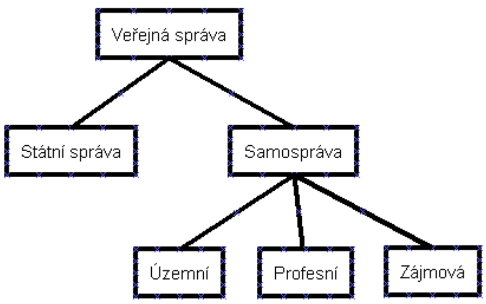
1.1.5.Územně správní uspořádání společnosti

Toto uspořádání společnosti vyjadřuje způsob, kterým je soudobý stát rozdělen na jednotky, v nichž je vykonávána veřejná správa prostřednictvím územních správních úřadů. Územně správní jednotky slouží k výkonu státní správy v samostatné a přenesené působnosti a zároveň to jsou územní společenství s vlastní samosprávou. Územně správní jednotky se významnou měrou podílejí na veřejné správě jako souboru záměrných a cílených činností v institucionálním a organizovaném rámci, který je pomocí specializovaných prostředků a nástrojů regulován, aby efektivně fungoval v zájmu veřejné správy.

Veřejná správa je složena ze dvou subsystémů – státní správy a samosprávy. Oba subsystémy pak můžeme charakterizovat [9]:

- organizačně - to znamená podle toho, **kdo** vykonává veřejnou správu:
  státní správu stát a jeho orgány,
  samosprávu vykonávají orgány uzemní , profesní a zájmové samosprávy,
- věcně (charakteristika činnostní) - to znamená podle toho, **co** je veřejnou správou vykonáváno.

Schématicky vyjádříme základní podobu veřejné správy takto na obrázku číslo 1:



Obrázek 1: Základní podoba veřejné správy, zdroj: [9].

Veřejná správa je vykonávána prostřednictvím územně správních úřadů obecních, krajských, a speciálních (zvláštních s jistou dílčí působností, mohou být tzv. jednostupňové či dvoustupňové / dekoncentrované).

Veřejná správa je vykonávána v územně správních jednotkách. Pro Českou republiku platí výkon v [9]:

základních jednotkách, tj. v obcích (se společným majetkem, rozpočtem, orgány a organizacemi, programem a legislativními nástroji, kterými jsou vyhlášky),

krajích (se společným majetkem, rozpočtem, orgány = zastupitelstvem / radou, hejtmanem / úřadem a jistými zvláštními orgány s právní pravomocí a předpisy).

Jako příklad je dále uvedeno územně správní (administrativní) členění České republiky a Německa.

1.1.6.Územně správní (administrativní) členění ČR

Česká republika se dělí na kraje, obce s rozšířenou působností, pověřené úřady a obce. S platností Zákona o vyšších územně správních celcích zaniká správní funkce okresů, které zůstávají pouze statistickou jednotkou.

Již na počátku sedmdesátých let zavedl Statistický úřad Evropského společenství (Eurostat) pro potřeby klasifikování jednotné unifikované struktury územních jednotek klasifikaci NUTS (La Nomenclature des Unités Territoriales Statistiques), tzv. nomenklatura územních statistických jednotek. Základní kameny systému územních statistických jednotek NUTS tvoří jednotky NUTS 1 až NUTS 3. Tyto tři základní jednotky jsou doplněny o jednu vyšší jednotku NUTS 0 pro ty státy, které výrazně velikostně překračují vymezení NUTS 1 a o dvě jednotky menší, tzv. LAU 1 a LAU 2 (Local Administration Unit), jejichž vymezení je v jednotlivých státech dobrovolné[3].[16], [14]

Přehled statistických a administrativních jednotek i s příkladem pro Českou republiku znázorňuje tabulka číslo 2,

Tabulka 2: Přehled statistických a administrativních jednotek, zdroj: [9].

Jednotka NUTS	Velikostní vymezení	Příklad - Česká republika
NUTS 0	Stát	není
NUTS 1	3 - 7 milionů obyvatel	Česká republika
NUTS 2	0,8 - 3 miliony obyvatel	Regiony soudržnosti (8)
NUTS 3	150 - 180 tisíc obyvatel	Kraje (14)
LAU 1		Bývalé okresy (77)
LAU 2	Obce	Obce (6249)

Z toho je ještě 387 pověřených úřadů a 205 obcí s rozšířenou působností.

1.1.7.Územně správní (administrativní) uspořádání Německa

Spolková republika Německo se člení na 16 spolkových zemí (NUTS 1), jejichž úplné oficiální názvy jsou následující (v závorce je uvedeno hlavní město země, české synonymum je použito v případech, kdy se běžně používá) [9]:

Badensko – Württembersko (Stuttgart),

[3] Jedná se však o klasifikační jednotky na úrovni nižší, než je kraj (NUTS 3).

Svobodný stát Bavorsko (Mnichov),
Spolkové hlavní město Berlín,
Braniborsko (Postupim),
Svobodné a hanzovní město Brémy,
Svobodně a hanzovní město Hamburk,
Hessensko (Wiesbaden),
Meclenbursko – Přední Pomořany (Schwerin),
Dolní Sasko (Hannover),
Severní Porýní – Vestfálsko (Düsseldorf),
Porýní – Falc (Mainz),
Sársko (Saarbrücken),
Svobodný stát Sasko (Drážďany),
Sasko – Anhaltsko (Magdeburg),
Šlesvicko – Holštýnsko (Kiel),
Svobodný stát Durynsko (Erfurt).

Administrativně se tedy Spolková republika Německo dělí na [9]:

16 spolkových zemí	NUTS 1,
22 vládních krajů (Regierungsbezirke)	NUTS 2,
439 okresů (Stadt- und Landkreise),	

z toho na 116 městských a 323 venkovských okresů NUTS 3 a 13176 obcí LAU 2.

1.1.8.Mezinárodní politické regiony

Pojmem region však mohou být označovány i oblasti mnohem větší než stát, např. celé kontinenty či subkontinenty. S takovými pojmy pak pracuje obor zvaný geopolitika, která zpravidla rozlišuje tři základní typy mezinárodních politických regionů: panregiony, transnacionální regiony a regiony mezihraniční.

Panregiony jsou největšími mezinárodními regiony. Někdy bývají také označovány jako makroregiony a jedná se o rozsáhlé oblasti s „kontinentálními“ rozměry, které mohou zahrnovat i velké množství zemí. Počet zemí ovšem není rozhodující. Existují i panregiony, jež jsou tvořeny velkými státními útvary, jako například Severní Amerika či Dálný východ. Kromě příslušných parametrů musí mít panregion i určité jednotící kulturní a civilizační prvky, jimiž se odlišuje od ostatních panregionů. Ve většině panregionů probíhá jak integrace ekonomická a kulturní, tak integrace politická. V každém případě jde o hledání vzájemné rovnováhy a konsenzu. Nejsložitější strukturu má evropský panregion. Uvnitř panregionů tvořených větším počtem států mohou vzniknout menší transnacionální politické regiony, které zahrnují několik států, jež mají určité společné rysy a zájmy, někdy i společnou historii. Transnacionální politický region se může další integrací přeměnit ve státní útvar (např. Velká Británie), ale může také dojít k rozpadu velkého státu na menší státy, které pak vytvářejí transnacionální politický region. [9]

Pro českou politiku má obrovský význam středoevropský transnacionální region, kterým se podrobněji zabývá B. Hnízdo.

Dále existují mezihraniční politické regiony, které nezahrnují celé státy, jen jejich příhraniční oblasti, jež se spojují s částí území sousedního státu či států. Při integraci pohraničních oblastí do mezihraničního regionu mohou působit různé faktory – etnické, národnostní, náboženské, jazykové apod. V České republice lze k historickým mezihraničním regionům řadit zejména Těšínsko.

1.2. Občanská vybavenost regionu

V období centrálně plánované ekonomiky bylo řešení počtů, druhů a kapacit zařízení občanského vybavení velmi zjednodušeno na direktivně přidělované prostředky, za které byly realizovány objekty a zařízení bez ohledu na to, zda budou sloužit svému účelu po celou dobu své životnosti. V minulém období bylo v rámci občanské vybavenosti příznačné, že veškerá zařízení byla utvářena podle přesně stanovených pravidel, která v řadě případů neodpovídala potřebám, nebo byla naopak poddimenzována. Byl též stanoven standard vybavení podle velikosti sídla, což představoval počet obyvatel bez ohledu na to, zda-li je či není objekt optimálně využit. [10]

V rámci tržní ekonomiky získává tato problematika zcela odlišný charakter. V průběhu doby se stávající zařízení občanské vybavenosti rozčlenila do dvou hlavních kategorií, a to zařízení veřejně prospěšná a zařízení komerční [10]:

Zařízení veřejně prospěšná – jedná se o nekomerčně využívaná zařízení, která si nejsou schopná na sebe a na svůj provoz vydělávat, přesto je ale potřeba, aby fungovala. Provoz těchto zařízení je v péči obce, jelikož o ně není ze strany podnikatelů zájem. Do této kategorie se řadí téměř veškerá zařízení **školská, kulturní, zdravotnická** a dále také **sociální péče**. Tato zařízení jsou ve městě co do kvantity zastoupena dostatečně, obtížnější je však zajišťovat jejich provoz z rozpočtu města.

Zařízení komerční – jde v podstatě o celou síť obchodů, ubytovacích a stravovacích zařízení a zařízení služeb. V rámci této kategorie již začínají působit mechanismy nabídky a poptávky. Soubor těchto subjektů je v současné době již zcela v privátním držení, který nelze direktivně nařizovat.

1.2.1.Občanská vybavenost města Bílina

V této kapitole bude dále popsán konkrétní příklad občanské vybavenosti města Bílina. Veškeré zde uvedené údaje jsem čerpal z oficiálních www stránek města Bílina [10].

1.2.1.1. Zařízení veřejně prospěšná

Zařízení školství

Objekty škol všech stupňů jsou ve městě zastoupeny v dostatečné míře, jež svou kvantitou a kapacitou vyhovují potřebám města. Vzhledem k očekávanému demografickému vývoji se bude v následujících letech počet žáků snižovat, tím se nejspíš objeví problémy s optimálním využití zejména základních škol, které byly dimenzovány značně velkoryse, avšak velmi jednoúčelově. O využití budov k jiným účelům se prozatím neuvažuje, jelikož by to bylo problematické a zřejmě i nákladné.

Hodnocení podle druhu (stupně) školských zařízení

Současný počet mateřských škol vyhovuje potřebám města a je jich plně kapacitně využito. Optimální je i rozložení těchto zařízení v území vzhledem k docházkovým vzdálenostem. V roce 1997 byla celková kapacita pro 565 dětí, zapsáno bylo 80,4 % kapacity MŠ. Provoz MŠ zajišťuje 39 pedagogických pracovníků a 38 ostatních pracovníků.

Základní školy jsou ve městě čtyři (Základní škola Lidická, Základní škola Za Chlumem, Základní škola praktická, Základní škola Aléská) a jsou v současném období kapacitně vyhovující. Jak již bylo naznačeno výše, neustále se diskutuje, jak prostor těchto objektů lépe využít. Jako schůdné řešení se jeví využití pro různé druhy škol středních. Způsob využití se operativně mění, např. Základní škola Teplické předměstí je nyní Střední odborné učiliště stavební. Ve školním roce 1996/97 bylo ve městě celkem 1778 žáků (číslo zahrnuje i žáky z okolních obcí). O výchovu se ve školách nyní stará 173 pracovníků, z toho 105 pedagogů.

Na území města se objevují i specifická školská zařízení jako Základní umělecká škola a Zvláštní škola internátní. Základní umělecká škola je nyní na Mírovém náměstí ve Vrchlického ulici a zaměřuje se na hudbu, tanec a výtvarnou výchovu. Budova je po rekonstrukci. Tuto školu nyní navštěvuje 250 žáků a pracuje zde 15 pedagogů. Nynějším největším problémem je vyřešení nedostatku prostor pro pohybovou výchovu, zejména potřeba sálu se záchody. Zvláštní škola internátní sídlí v Kmochově ulici a je řízena školským úřadem Teplice. Navštěvuje ji 130 žáků a má 24 pracovníků, z toho 16 pedagogů. Internát má kapacitu 42 lůžek, zvýšení kapacity se neplánuje, pouze stavební rozšíření o tělocvičnu.

V Bílině je, kromě nově připravovaných středních škol, které budou umístěny v objektech škol základních, pouze Gymnázium, které navštěvuje 337 studentů a je zde zaměstnáno 39 pedagogů a 7 dalších pracovníků. Výhledově se nepočítá s dalším rozšířením.

Dále je ve městě Dům dětí a mládeže, který se nachází v Havířské ulici. Toto zařízení pro volný čas dětí mělo zapsáno 550 žáků, zaměstnanců je zde 9, k tomu 35 externích spolupracovníků. S rozšířením se v budoucnu nepočítá.

Zařízení kulturní

Kulturní zařízení v městě Bílina má z části na starosti město, z části kulturní centrum Kaskáda, z části soukromí sektor. Provoz těchto zařízení je v nynější době velmi nákladný a z ekonomického hlediska značně neefektivní. Tento pohled z ekonomického hlediska je pouze jedním z mnoha, ze kterých je třeba na činnosti kulturních zařízení města pohlížet, i když řada z nich je z hlediska návštěvnosti v hluboké krizi. I přes všechny nedostatky provozní povahy kulturní zařízení města fungují.

Město zajišťuje chod knihoven, které se v současnosti nacházejí ve čtyřech lokalitách. Knihovna, která sídlila na Vrchlického ulici, byla přesunuta na Mírové náměstí místo ZŠ. V části budovy se počítá se zřízením muzea. Knihovna stačí kapacitně pokrýt zájem čtenářů i s minimálním personálem, který tvoří celkem šest zaměstnanců.

Kulturní centrum Kaskáda spravuje tyto objekty – **divadlo**[4]**, kino** Hvězda, přírodní divadlo v Kyselce a kulturní dům Za Chlumem. Divadlo a kino Hvězda potřebují podle sdělení vedení kulturního centra Kaskáda rekonstrukci a obnovu zařízení a technologií, aby vyhovovaly dnešnímu standardu. Přírodní amfiteátr je ve velice špatném stavu. Rekonstrukce bude velmi nákladná a výsledek a využití tohoto areálu budou ztrátové, jelikož se jedná o sezónní aktivitu. Kulturní dům je v současnosti využíván pouze k pronájmům. Náplň činností je předmětem jednání. Předpokládá se, že činnosti, které budou v kulturní domě provozovány, si budou muset na svůj chod vydělávat, což může znamenat, že se budou od

[4] Tučně zvýrazněné zařízení jsou použita k následné analýze občanské vybavenosti v kapitolách 2 a 3.

kultury nejspíše vzdalovat. Předpokládá se však, že půjde pouze o jev přechodný, který pomůže zvládnout chod zařízení, které je sice pro Bílinu jako celek přitažlivé, ale svojí polohou potřebám města ne zcela vyhovující.

Výhledově by se také měly začít využívat ke kulturním účelům prostory zámku, jako například výstavní síň, koncertní sál. Zámek přešel do soukromého držení, což ovlivní i výběr kulturních činností, které se zde budou provozovat.

Zařízení sportovní a tělovýchovná

Sportovní a tělovýchovná zařízení jsou ve srovnání s jinými městy stejné velikostní kategorie zastoupena poměrně bohatě, čili toto zastoupení je dostačující. Otázkou je vhodnost či nevhodnost jejich umístění z hlediska jejich další perspektivy. Jedná se především o sportovní plochy na Kyselce, které jsou neperspektivní svým umístěním mezi tratí a řekou, což neumožňuje další rozvoj. Navíc je tento areál umístěn v těsné blízkosti na **lázně** Kyselka, které jsou dnes již také v soukromých rukou. Některá ze sportovišť tohoto území jsou již také privátní, například minigolf, tenisové kurty. Naopak areál koupaliště s restaurací přešel v roce 1996 do majetku města.

Pro potřeby města se vytvořil návrh na vybudování nového sportovního areálu, který bude zahrnovat lehkoatletický stadion, 2 fotbalová hřiště, tenisové a volejbalové kurty, sportovní halu včetně parkoviště. Tento sportovní areál bude postaven v severní části města, aby navazoval na nové plochy pro výstavbu rodinných domů a areálů podnikatelských aktivit Za Chlumem. Tento areál by měl nahradit výše uvedená sportoviště v Kyselce, která by výhledově sloužila jako sportovní a rekreační zázemí tamních lázní. Návrh nového sportovního zařízení je připraven ve dvou alternativách.

Dále je zde plavecká hala se dvěma letními bazény, která do budoucna vyhovuje daným potřebám. Dalším významným objektem je zimní stadión s umělou ledovou plochou. V současné době je nezastřešený, což bohužel snižuje jeho kvalitní využití. Jako velice perspektivní se jeví jeho zastřešení i se zázemím (tribuny, šatny apod.), které by umožňovalo využití objektu i k jiným činnostem než sporty, které potřebují ledovou plochu. Plně budou využívány také sportovní objekty v areálech jednotlivých základních škol.

Velice specifickým areálem je bikrosová dráha, která se nachází u bývalých kasáren a jejíž využití je závislé pouze na aktivitě současných a budoucích uživatelů.

Zařízení zdravotnická

Hlavním zdravotnickým areálem města Bílina je Hornická **nemocnice** s poliklinikou na Pražské ulici. Objekt je v pronájmu Hornické zdravotní zaměstnanecké pojišťovny Most. Je zde lůžková část s kapacitou 70 lůžek (interní a neurorehabilitační oddělení). Celkový počet všech zaměstnanců je 260, z toho 32 zaměstnanců jsou lékaři. Pozemek areálu je rozlehlý, jelikož v době jeho vzniku se počítalo s podstatně větším rozsahem. Umožňuje tedy další využití. Jednou z možností je výstavba penzionu pro osoby důchodového věku, pro které by byla blízkost přilehlé polikliniky výhodou. Dále se nabízí stavba domu s pečovatelskou službou.

Kromě Hornické nemocnice s poliklinikou se zde nacházel také stacionář Za Chlumem, který byl vybudován z bývalých jeslí. Jeho provoz byl velice nákladný, proto byl stacionář zrušen a děti byly převedeny do jeslí v ulici M. Švabinského a do Čapkovy mateřské školy. V objektu jsou zřízeny ordinace lékařů.

Mezi zdravotní zařízení jsou započítávány i **jesle**. Tato zařízení jsou však velice málo využívána, proto jsou ve městě pouze jedny. Do budoucna se plánuje jejich zrušení. Ostatní objekty bývalých jeslí jsou využívány především kulturně, nachází se zde například klub důchodců, knihovna, anebo komerčně (restaurační zařízení).

Dalším objektem, který se řadí mezi zdravotnické, jsou **lázně** Kyselka. Lázně jsou již privátní, takže jejich rozvoj je závislý výhradně na majiteli, který vzhledem k poměrně malému počtu lázeňských hostů zrekonstruoval starou lázeňskou budovu na hotel s kapacitou šedesáti lůžek, kde je možno také využít lázeňských služeb. V současnosti majitel lázní uvažuje o prodeji. Hlavní zájem o lázně má pochopitelně město, avšak finanční požadavky majitelů jsou vzhledem k možnostem městské pokladny příliš vysoké. Prozatím se hledá nový majitel lázní, který snad bude mít zájem zachovat lázeňský provoz i do budoucna, protože i když jsou lázně svým rozsahem malé, poskytují kvalitní péči a prostředí. Město Bílina má díky lázním vydán Statut lázeňského místa, který vydal KNV v Ústí nad Labem podle ustanovení § 45 odst. 1 zákona č.20/1966, usnesením rady KNV č. 22, ze dne 20.10.1970.

1.2.1.2. Zařízení komerční

Druhou hlavní skupinu občanského vybavení města tvoří zařízení zakládající se na komerční bázi, to znamená, že jejich činnost a existence vyplývá z tržní ekonomiky. Podle statistické nomenklatury jde o zařízení obchodu, stravování, ubytování a služeb. Většina těchto zařízení je v současné době již zprivatizována, což znamená, že jejich činnost je plně závislá na možnostech jejich majitelů. Proto zde není možné hodnotit, zda momentálně v síti obchodů chybí ten či onen druh prodeje, a v rámci řešení územního plánu navrhovat jeho doplnění. V rámci územního plánu je možno stanovit pro lokalizaci jednotlivých druhů vybavení určitá pravidla, která by jejich existenci na území města regulovala prostřednictvím správce území, tj. městského zastupitelstva.

Pro většinu zařízení občanského vybavení fungujícího na komerčních základech jsou využívány především prostory nacházející se ve stávajícím stavebním fondu, které byly nebo jsou využity jiným způsobem, než bylo původně zamýšleno. Jedná se zejména o původní centrální části města. Urbanistická koncepce předpokládá povýšení tohoto centrálního území, které je pro svou kulturní a historickou hodnotu i jádrem městské památkové zóny, na vyšší kvalitativní úroveň, na tzv. faktického reprezentanta města.

V této oblasti jsou velice důležitá pravidla regulace, z nichž nejvýznamnější jsou tato [10]:

- pro uspokojení potřeb vybavenosti a služeb využít veškerých rezerv ve stávajícím stavebním fondu,
- realizací navržených změn v městské dopravní síti dosáhnout podstatného zklidnění této části města,
- koncentrovat zařízení občanského vybavení na vysoké úrovni kvality právě do centra města,
- zamezit dalšímu vylidňování centra města, umístit zejména na Mírové náměstí nadstandardní byty, např. bydlení pro významné městské představitele, hosty, apod.,
- vhodnou dopravní organizací ve středu města (šikany, snížení rychlosti, atd.) zamezit průjezdné dopravě (nikoli dopravu osobními auty zrušit) a vytvořit podmínky pro realizaci pěší zóny, včetně jejího rozšíření o část na Mosteckém předměstí,
- pokusit se ve spolupráci s majiteli jednotlivých objektů, zejména objektů historicky a architektonicky cenných, nalézt takovou funkční náplň, která by odpovídala jejich kvalitě, umístění a reprezentačním potřebám města.

Návrh konceptu územního plánu počítá s několika novými plochami v centrální části města, které by byly využitelné právě pro vyšší vybavenost a služby. Jde především o plochu v blízkosti Žižkova a Pivovarského náměstí, kam byla umístěna velkoprodejna s potravinářským zbožím – Diskont Plus, která je již v provozu, dále dvě záchytná parkoviště a plocha pro hromadné garáže. V jižní části centrální zóny je navržena plocha pro umístění zařízení občanského vybavení nadměstského typu – nové centrum, prozatím bez konkrétní náplně.

Dalším problémem v rámci komerční vybavenosti je vytvoření prostorových podmínek, které by sloužily podnikatelským aktivitám další skupiny tohoto občanského vybavení. Jedná se o činnosti, které v dnešním prostředí již existují, ale plochy pro ně chybí, jelikož jsou náročné na území. Jde například o plošné zajištění nároků na přechodné umístění různých zábavných podniků, jako je např. cirkus, lunaparky, nebo plošných nároků na výstavbu prodejních ploch marketů, kam se za nákupy jezdí. Jsou-li lokalizovány u komunikací, musí být vybaveny parkovištěm přiměřené velikosti, apod. Současný zájem vypovídá o tom, že toto zařízení bude ve městě pouze jedno. Tržní poměry však mohou tento zájem změnit a bude třeba v území lokalizovat další tato zařízení . Proto je potřebné vytvoření určitých územních rezerv tak, aby se tyto možné aktivity nemusely odmítat. Návrh počítá s takovouto rezervní plochou v severní části města.

Kromě toho jsou při výjezdu z města směrem na Louny navrženy také plochy pro novou oboustrannou čerpací stanici pohonných hmot, včetně ostatních potřebných služeb, a u stávající benzínové stanice na Teplickém předměstí je navrženo její plošné rozšíření o myčku aut. Navíc je uvažováno s rezervní plochou pro služby pro motoristy v návaznosti na ČOV při výjezdu z města ve směru na Teplice.

Otázka rozložení a řešení služeb je pojata v komplexu s rozvojem jak obytného území, tak ploch tzv. areálů pro podnikatelské aktivity. V určených lokalitách se kromě výstavby provozoven pro nerušící výrobu (u mnohých včetně bydlení pro majitele), počítá i s umístěním drobných provozoven služeb, které nejsou prostorově náročné a na druhé straně je nelze provozovat ve standardním rodinném domě. Kombinace bydlení a drobné řemeslné činnosti nebo služeb se navrhuje v pěti lokalitách v severní a severovýchodní části města.

Občanské vybavení v ostatních obcích řešeného území nebude z hlediska územních nároků kvantitativně růst. Ve všech obcích je možnost využití všech prostor existujících i dočasně nefungujících k rozvoji řady činností. Podstatné však bude, aby se časem do sídel, která lze považovat za venkovská, vrátila jejich dříve tak samozřejmá relativně nízká závislost na velkých sídlech (městech), hlavně z pohledu zabezpečování a krytí základních potřeb bydlících obyvatel. Zde půjde o proces zdlouhavý, který bude realizován postupně, v plné závislosti na celkové ekonomické úrovni celého řešeného území.

2. Asociační pravidla

Klasická asociační pravidla, tedy IF-THEN pravidla, lze najít v mnoha programovacích jazycích, ale používají se i v běžné řeči. Například IF I buy something to eat in a shop THEN I must not go to the restaurant today, tedy KDYŽ si koupím něco k jídlu v obchodě, POTOM nemusím jít dnes do restaurace. Je tedy zřejmé, že takováto pravidla jsou společně s rozhodovacími stromy nejčastějším prostředkem, jak vyjádřit určitou znalost, a to ať již se jedná o znalost expertů, nebo pouhý automatizovaný extrakt z dat.

Asociační pravidla přišla ve známost počátkem 90. let, a to ve spojitosti s Analýzou nákupního košíku. Použitím této analýzy se povětšinou zjišťuje, jaké druhy zboží kupuje zákazník dohromady - například současná koupě těstovin a kečupu v potravinách. Nemusí se však nutně jednat pouze o analýzu zboží v supermarketech. V dalším textu se budu snažit použít tuto metodu právě na analýzu občanské vybavenosti v jednotlivých územních celcích. V podstatě to znamená, že se bude jednat o vyhledávání vzájemných vazeb (asociací) mezi objekty občanské vybavenosti, přitom však nebude upřednostňován žádný z těchto objektů jako závěr pravidla.

2.1. Vlastnosti asociačních pravidel

Každé pravidlo se skládá z předpokladu, tedy levé strany pravidla, a ze závěru, ten je naopak na straně pravé. Pro tuto práci, a to v celém rozsahu, je použit pro název předpoklad také anglický ekvivalent antecedent (zkráceně *Ant*) a pro název závěr zase conseqvent (*Con*). Obecně má každé asociační pravidlo následující tvar [1]:

$$Ant \Rightarrow Con. \qquad (1)$$

Nutno poznamenat, že u pravidel, která jsou vytvořena z dat, je kladen důraz na to, kolik příkladů splňuje předpoklad, kolik závěr, kolik současně předpoklad i závěr či kolik jich splňuje předpoklad, ale již ne závěr a tak dále. To se dá nejpřehledněji vyjádřit v tzv. kontingenční tabulce. Tabulka číslo 3 je příkladem kontingenční tabulky, kde

$n(Ant \wedge Con) = a$ je množství příkladů, které jsou pokryty jak předpokladem, tak závěrem,

$n(\neg Ant \wedge Con) = c$ je množství příkladů, které nejsou pokryty předpokladem, ale jsou pokryty závěrem, n($Ant \wedge \neg Con$) = b je množství příkladů, které jsou pokryty předpokladem, ale nejsou pokryty závěrem,

n($\neg Ant \wedge \neg Con$) = d je množství příkladů, které nejsou pokryty ani předpokladem, ani závěrem,

Tabulka 3: Kontingenční tabulka, zdroj: [1].

	Con	¬Con	Σ
Ant	a	b	r
¬Ant	c	d	s
Σ	k	l	n

Dále pak platí následující vztahy [1]:

n(Ant) = a + b = r;
n(Con) = a + c = k;
n($\neg Ant$) = c + d = s;
n($\neg Con$) = b + d = l;
n = a + b + c + d.

Výše uvedené údaje se také v některých zdrojích označují jako četnost nebo i frekvence kombinace. Mohou také sloužit k výpočtu různých charakteristik pravidel, a tak kvantifikovat a následně hodnotit znalosti, které byly nalezeny.

V tzv. Agrawalově pojetí jsou základní charakteristiky asociačních pravidel definovány jako podpora (anglicky support) a spolehlivost (anglicky confidence). Podpora je množství příkladů, které splňují jak předpoklad, tak závěr, tedy výše uvedená proměnná „a“, může být dále členěna jako absolutní nebo relativní. Relativní četnosti kombinace (dále jen Comb) budu považovat za odhad pravděpodobnosti výskytu této kombinace v datech P(Comb). Formálně tedy zapsáno následovně [1]:

$$P(Ant \wedge Con) = \frac{a}{a+b+c+d}, \tag{2a}$$

spolehlivost je podmíněná pravděpodobnost závěru za předpokladu, že platí předpoklad tedy Antecedent. Spolehlivost se také často nazývá platnost, konsistence či správnost (anglické překlady po řadě jsou validity, consistenci, accuracy)

$$P(Con \mid Ant) = \frac{a}{a+b}, \tag{2b}$$

další charakteristiky, které zde nemohou být opomenuty jsou například

absolutní a relativní množství příkladů, které splňují předpoklad

$$\text{absolutní a + b, a relativní } P(Ant) = \frac{a+b}{a+b+c+d}, \tag{2c}$$

množství příkladů (opět absolutní i relativní), které splňují závěr

$$\text{absolutní a + c, a relativní } P(Con) = \frac{a+c}{a+b+c+d}, \tag{2d}$$

pokrytí (anglicky coverage) nebo také úplnost, což je opět podmíněná pravděpodobnost, tentokráte předpokladu, pokud platí závěr pravidla

$$P(Con \mid Ant) = \frac{a}{a+c}, \tag{2e}$$

kvalita, ta je váženým součtem spolehlivosti a pokrytí

$$Kvalita = w_1 \frac{a}{a+b} + w_2 \frac{a}{a+c}, \tag{2f}$$

kde $w_1 + w_2 = 1$, tedy například ($w_1 = 0{,}5$ a $w_2 = 0{,}5$ nebo $w_1 = 0{,}6$ a $w_2 = 0{,}4$).

Dále je možno na základě spolehlivosti a úplnosti (pokrytí) rozdělit implikace do následujících skupin [1]:

Konzistenční pravidla – tato pravidla mají spolehlivost rovno 1, tedy strana levá (v implikaci) je dostatečnou podmínkou pro splnění strany pravé.
Úplná pravidla – tato pravidla mají pro změnu pokrytí rovno 1, tedy strana levá je v tomto případě nutnou podmínkou pro splnění strany pravé.

Deterministická pravidla – tato pravidla mají jak spolehlivost, tak úplnost rovno 1, tedy strana levá je podmínkou nutnou a zároveň postačující pro splnění strany pravé.

V některých zdrojích jsou popsány i další doplňující charakteristiky asociačních pravidel. Příkladem budiž další text, kde krom již zmíněných vlastností jako podpora a spolehlivost, jež se zde mimo jiné nazývají popisná podpora a popisná spolehlivost, uvádí autor níže uvedené [4]:

kauzální podpora (Causal support)

$$P(Ant \wedge Con) + P(\neg Ant \wedge \neg Con) = \frac{a+d}{a+b+c+d}, \tag{3a}$$

kauzální spolehlivost (Causal confidence)

$$\frac{1}{2}P(Con \mid Ant) + \frac{1}{2}P(\neg Ant \mid \neg Con) = \frac{1}{2}\frac{a}{a+b} + \frac{1}{2}\frac{d}{b+d}, \tag{3b}$$

deskriptivní potvrzení (Descriptive confirmation)

$$P(Ant \wedge Con) - P(Ant \wedge \neg Con) = \frac{a-d}{a+b+c+d}, \tag{3c}$$

kauzální potvrzení (Causal confirmation)

$$P(Ant \wedge Con) + P(\neg Ant \wedge \neg Con) - 2P(Ant \wedge \neg Con) = \frac{a+d-2b}{a+b+c+d}, \tag{3d}$$

ujištění (Conviction)

$$\frac{P(Ant)P(\neg Con)}{P(Ant \wedge \neg Con)} = \frac{(a+b)(b+d)}{d(a+b+c+d)}, \tag{3e}$$

zajímavost (interestingness)

$$\frac{P(Ant \wedge Con)}{P(Ant)P(Con)} = \frac{a(a+b+c+d)}{(a+b)(a+c)}, \tag{3f}$$

závislost (dependency)

$$P(Con\,|\,\mathrm{Ant}) - P(Suc) = \frac{a}{a+b} - \frac{a+c}{a+b+c+d}\,. \tag{3g}$$

Z výše uvedeného textu je tedy vidět, že z kontingenční tabulky lze spočítat různé charakteristiky asociačních pravidel. Další vztahy, které jsou rovněž tvořené z této tabulky, je možné nalézt v literatuře [3] věnované metodě GUHA.

2.2. Vytváření kombinací

Vytváření kombinací, nebo také generování kombinací či konjunkcí (v konjunkcích se neopakují atributy) hodnot atributů, je základem všech algoritmů, které se využívají pro vlastní hledání asociačních pravidel. Vytváření kombinací hodnot atributů je prohledáváním prostoru všech těchto kombinací. K těmto účelům jsou všeobecně k dispozici tři základní metody, a to [5]:

- Prohledávání do šířky
- Prohledávání do hloubky
- Heuristická metoda prohledávání.

První dvě metody patří do skupiny takzvaného neinformovaného (slepého) prohledávání. U tohoto typu prohledávání se vyskytuje nedeterministický prvek. Není zde totiž striktně určeno, v jakém pořadí či jakým konkrétním způsobem budou voleny uzly, které jsou určené na expandování. Je-li takovéto pořadí nezávislé na dané úloze, pak se jedná o neinformované prohledávání. Třetí z metod je naopak prohledáváním informovaným. Využívá-li strategie volby uzlů informace z dané úlohy, pak jde o tento typ prohledávání, tedy o prohledávání informované. Konkrétně lze říci, že heuristická metoda umožňuje nalezení hledaného cíle rychleji než metody slepého prohledávání, avšak je také možné, že se tato metoda ocitne ve „slepé uličce“. [5]

2.2.1. Prohledávání do šířky

Tento typ prohledávání (generování kombinací) je založen na principu vytváření kombinací podle délek. Základem je tabulka vstupních dat, ze které se kombinace generují.

Tabulka vstupních dat je v následujícím příkladě znázorněna jako tabulka číslo 4,

Tabulka 4: Tabulka dat, zdroj: [vlastní].

I	A	a
I	B	b
II	A	b

V prvním kroku se vytvoří veškeré kombinace délky jedna. Ve druhém kroku veškeré kombinace délky dvě a tak dále. Pro názornost je zde uvedena tabulka 5, která obsahuje již samotné kombinace a jejich četnosti,

Tabulka 5: Prohledávání do šířky, zdroj: [1].

Číslo	Četnost	Kombinace	Číslo	Četnost	Kombinace
1	2	I	14	1	II b
2	1	II	15	1	A a
3	2	A	16	1	A b
4	1	B	17	0	B a
5	1	a	18	1	B b
6	2	b	19	1	I A a
7	1	I A	20	0	I A b
8	1	I B	21	0	I B a
9	1	I a	22	1	I B b
10	1	I b	23	0	II A a
11	1	II A	24	1	II A b
12	0	II B	25	0	II B a
13	0	II a	26	0	II B b

2.2.2.Prohledávání do hloubky

Při tomto typu generování kombinací[5] se vychází od první kombinace s délkou jedna, která se v druhém kroku prodlouží o první kategorii následujícího atributu. Takto se prodlužuje až do té doby, kdy dosáhne maximální zadané délky kombinace či při vyčerpání všech atributů a jejich hodnot. Ve chvíli, kdy již další prodloužení není možné, dojde ke změně kategorie posledního z atributů. Jsou-li již všechny kategorie tohoto posledního

[5] Podrobný popis generování kombinací je možné najít například ve zdrojích [1] a [5].

atributu již také vyčerpané, tedy nelze je již změnit, pak následuje zkrácení kombinace a současná změna poslední kategorie. [1]

V následujícím příkladě je čerpáno ze stejných vstupních dat jako v kapitole 2.2.1. a výsledek tohoto postupu viz. tabulka číslo 6,

Tabulka 6: Prohledávání do hloubky, zdroj: [1].

Číslo	Četnost	Kombinace	Číslo	Četnost	Kombinace
1	2	I	14	0	II B
2	1	I A	15	0	II B a
3	1	I A a	16	0	II B b
4	0	I A b	17	0	II a
5	1	I B	18	1	II b
6	0	I B a	19	2	A
7	1	I B b	20	1	A a
8	1	I a	21	1	A b
9	1	I b	22	1	B
10	1	II	23	0	B a
11	1	II A	24	1	B b
12	0	II A a	25	1	a
13	1	II A b	26	2	b

Jak již bylo uvedeno dříve, jedná se o tzv. slepé prohledávání. To v tomto konkrétním případě znamená, že se oba způsoby provádějí pouze s ohledem na seznam hodnot atributů a nikoli s ohledem na vlastní data. Důsledkem toho je, že se vygenerují kombinace, které nejsou vůbec v datech obsaženy. Takové kombinace pak mají četnost 0 v obou výše uvedených tabulkách.

2.2.3.Heuristická metoda prohledávání

Třetím druhem prohledávání dle četností je heuristické prohledávání. Heuristické prohledávání generuje kombinace seřazené v závislosti na jejich výskytu v datech, a to od největšího počtu k nejmenšímu - tj. kombinace, které mají nulovou četnost, budou na konci seznamu. Na tomto místě je nutné podotknout, že generování kombinací je při velkém počtu atributů výpočtově velmi náročný proces. Je však možné odhadnout počet kombinací předem. Nechť K_{Ai} je počet kategorií atributu $A_{1,2,\ldots,n.}$ Počet kombinací délky jedna pak je možno zapsat jako [1]:

$$\sum_{i=1}^{n} K_{Ai}\,, \tag{4a}$$

a celkový počet kombinací pak

$$\prod_{j=1}^{n}(1+K_{Aj})-1. \tag{4b}$$

V následujícím příkladě opět vycházím ze stejné tabulky dat jako v kapitolách 2.2.1. a 2.2.2.. V tebulce číslo 7 je pak znázorněn výsledek této metody,

Tabulka 7: Heuristické prohledávání, zdroj: [1].

Číslo	Četnost	Kombinace	Číslo	Četnost	Kombinace
1	2	I	14	1	A b
2	2	A	15	1	B b
3	2	b	16	1	I A a
4	1	II	17	1	I B b
5	1	B	18	1	II A b
6	1	a	19	0	II B
7	1	I A	20	0	II a
8	1	I B	21	0	B a
9	1	I a	22	0	I A b
10	1	I b	23	0	I B a
11	1	II A	24	0	II A a
12	1	II b	25	0	II B a
13	1	A a	26	0	II B b

V prvních dvou uvedených příkladech se jedná o celkový počet kombinací 26. Tento počet je však sumarizován bez ohledu na vstupní data a řídící parametry (jsou zde započteny i kombinace s nulovou četností). Počet kombinací ve třetím případě tedy klesne, a to právě o součet jednotlivých kombinací, které mají četnost nula. Konkrétně v tomto případě pak počet kombinací klesne z původních 26 na 18. V obou variantách však lze tvrdit, že množství vygenerovaných konjunkcí je exponenciálně závislé právě na počtu daných atributů.

2.3. Algoritmus apriori

Algoritmus zvaný apriori je vůbec nejznámějším a nejvíce používaným algoritmem hledání asociačních pravidel. Slovo A priori či Apriori je spojení z latiny a doslovný překlad je „z předchozího“. Toto spojení lze tedy interpretovat jako „předem“ nebo „předchůdný“, což již naznačuje základní vlastnost tohoto algoritmu. Tvůrcem tohoto algoritmu je Rakesh Agrawal, který jej vyvinul při analýze jedné z metod spojených s nakupováním. Tato metoda se nazývá Analýza nákupního košíku, kterou se budu zabývat v jedné z dalších kapitol této práce a pokusím se ji i použít a přehledně interpretovat na konkrétním případu. [15]

Hlavním úkolem zmíněného algoritmu je vyhledávání množin položek, které se často opakují. V anglické literatuře se často používá také název frequent itemsets. Jedná se, podle [1] o „kombinace (konjunkce) kategorií, které dosahují předem zadané četnosti (podpory minsup) v datech.“ K tomu ještě autor dodává, že „V případě analýzy nákupního košíku jsou kategorie typu máslo(ano), chleba(ano) apod., které můžeme stručněji zapisovat máslo, chleba a chápat je jako položky zboží.“

Je-li hledána kombinace kategorií s vysokou četností o délce k, pak je využíváno skutečnosti, že je již známa kombinace délky k-1. Právě spojování těchto kombinací o délce k-1 je pak využíváno pro vytvoření kombinace délky k. Z předešlého textu je tedy jasné, že se jedná o prohledávání (generování) kombinací do šířky. Základním požadavkem pro vytvoření kombinace délky k je splnění četnosti u všech jejích podkombinací délky k-1. Příkladem mohou být následující tříčlenné kombinace čísel 1,2,3,4 a 5

{1,2,3; 2,3,4; 1,2,4; 1,3,5; 1,3,4}

Tyto kombinace splňují požadavek na četnost a dá se z nich vytvořit jedna čtyřčlenná kombinace (1,2,3,4).

Pseudokód algoritmu apriori[6] je uveden níže [11]:

„for (k=2; ;k++) begin

    L_{k-1}={ kandidáti z C_{k-1}, kteří mají podporu vyšší než minimální}

    if (množina L_{k-1} je prázdná) break

    C_k=AprioriGen(L_{k-1});

    for each (transakce t) begin

        C_t= subset(C_k, t);

[6] Algoritmus je v mnoha formách a podobách dostupný na internetu. Tento konkrétní byl nalezen a převzat z materiálů k výuce problematiky *Dolování asociačních pravidel*, viz. [11].

for each (kandidáti c ∈ C_t) zvyš o 1 podporu kandidáta c

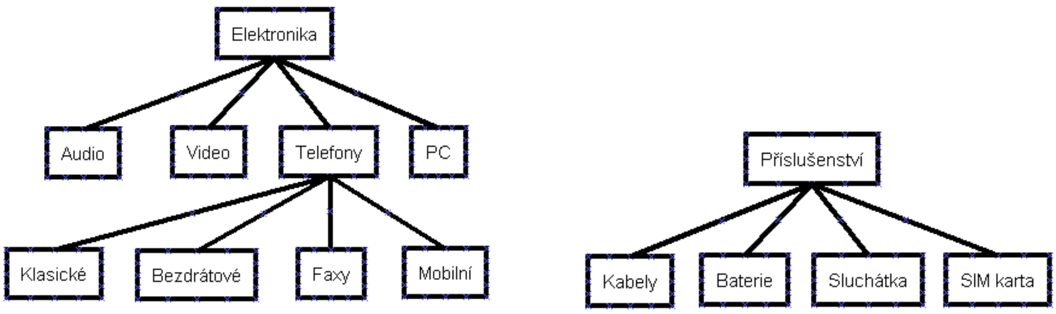
end;

end;

Zde metoda AprioriGen() vytváří veškeré k-množiny, tzv. kandidáty z množin frekventovaných, a následně vyřazuje takové množiny, které obsahují podmnožinu, jež není frekventovaná.“

2.4. Asociační pravidla ve zobecněné podobě

Tento problém je nejlépe definovatelný na konkrétním případě. Představme si tedy nákupní vozík, tak jak je znám z obchodu s potravinami. Pak veškeré zboží, které je zákazníkem ukládáno do tohoto nákupního košíku, je součástí tzv. přirozené taxonomie. V obchodě je obvykle možné zakoupit mnoho druhů zboží - například elektroniku, masné výrobky či pečivo viz. obrázek číslo 2.[7]



Obrázek 2: Elektronika – taxonomie, zdroj: [vlastní]

Taxonomie tohoto typu se pak využívá k vyhledávání asociačních pravidel ve zobecněné podobě. Takto vytvořené pravidla vyjadřují vzájemné asociace položek na rozlišných úrovních obecnosti. Z tohoto lze tedy vyvodit, že nejsou zajímavá pouze pravidla umístěná na nejnižší úrovni výše uvedené hierarchie, jako například

Klasické telefony => Kabely,

ale také pravidla na vyšší úrovni obecnosti, která mohou mít mnohdy při využití této taxonomie větší vypovídací hodnotu

[7] Obrázek je vytvořen v prostředí Dia 0.95-1 dostupném na adrese http://www.gnome.org/projects/dia/ [12].

Telefony => Příslušenství.

Je potřeba zde také uvést, že takovéto pravidlo je pak ve tvaru *Ant* => *Con*, jak je definováno v kapitole 2.1. Vlastnosti asociačních pravidel. Přičemž vždy platí, že žádná položka v *Con* není vlastním předchůdcem žádné z položek uvedených v *Ant*. Tedy pravidlo nemůže být ve tvaru Kabely => Příslušenství. Mohou se však v pravidlech vyskytovat i položky z různých hierarchických vrstev.

Problémem vyhledávání zobecněných asociačních pravidel může být vhodnost volby minimální hranice pro požadovanou podporu. Při zvolení příliš vysoké hodnoty není možné najít pravidla na spodní (nejníže položené) úrovni, ale také může nastat situace, kdy nebude nalezeno ani příslušné obecnější pravidlo (pro nízkou spolehlivost pravidla). Dalším důsledkem příliš vysoké minimální hranice podpory je kombinatorická exploze a přítomnost položek vyšší úrovně hierarchie ve všech možných kombinacích v pravidlech. Zde platí, že má-li podporu na požadované úrovni kombinace

$$Ant \wedge Con \,, \qquad (5)$$

pak i předchůdci *Ant* a *Con* budou mít požadovanou podporu. Pokud však bude pravidlo ve tvaru

$$Ant \Rightarrow Con \,, \qquad (1)$$

a taktéž bude splňovat podporu a spolehlivost na požadované úrovni, pak je však zaručeno pouze pro pravidlo

$$Ant \Rightarrow \text{Předchůdce}\,(Con)\,, \qquad (6)$$

že splní oba dané parametry. Výše uvedené se pokusím ilustrovat na názorném příkladu. V první tabulce číslo 8 je výběr nákupů elektroniky v obchodním domě. Transakce jsou pak jednotlivé nákupy a položky zase zakoupené zboží v transakci

Tabulka 8: Transakce, zdroj: [vlastní].

Transakce	Položky
1	Video
2	Klasický telefon
3	Mobilní telefon + Baterie
4	Klasický telefon + Kabely

V tabulce číslo 9 je ke každé položce (zakoupenému zboží) vypočtena její četnost ve všech uvedených transakcích

Tabulka 9: Tabulka četností, zdroj: [vlastní].

Položky	Četnost
Klasický telefon	2
Příslušenství	2
Telefony	3
Elektronika	4

Ve třetí tabulce číslo 10 jsou po řadě uvedeny položky s alespoň 50% podporou a položky s téže podporou a navíc i spolehlivostí vyšší nebo rovno 60%,

Tabulka 10: Asociační pravidla ve zobecněném tvaru, zdroj: [vlastní].

Pravidlo	Podpora v %	Spolehlivost v %
Telefon => Příslušenství	50	66
Příslušenství => Telefon	50	100
Příslušenství => Elektronika	50	100

Výše uvedené hodnoty podpory a spolehlivosti byly vypočteny dle vztahů (2a) a (2b).

Za povšimnutí stojí fakt, že v tabulce 10 chybí pravidlo Elektronika => Příslušenství. Toto pravidlo totiž nedosáhlo hranice spolehlivosti 60%. K vypořádání se s otázkou rozlišné podpory v hierarchické struktuře je potřeba dynamicky zaměňovat minimální hranici požadované podpory, a to právě v závislosti na hierarchické úrovni pro danou kombinaci.

2.5. Pravidla a výjimky

Každý algoritmus, který je používán pro vyhledávání asociačních pravidel, generuje určité množství pravidel (seznam), které je potřeba nějakým způsobem dále interpretovat. Jednou z nejčastějších možností je vyhledání nějaké zajímavosti v seznamu pravidel. Takovéto pravidlo je pak formálně definováno na základě následujících tří pravidel [1]:

$$Comb_A \Rightarrow Con\,, \tag{7a}$$

zde pravidlo odpovídá tzv. ustáleným představám, tedy pravidlo má vysokou míru podpory i spolehlivosti. Dále pak pravidlo

$$Comb_A \wedge Comb_B \Rightarrow \neg Con\,, \qquad (7b)$$

je právě hledanou výjimkou. tj. má nízkou míru podpory, ale vysokou míru spolehlivosti. A konečně poslední z trojice pravidel je tzv. referenční pravidlo

$$Comb_B \Rightarrow \neg Con\,, \qquad (7c)$$

a to má buď nízkou míru podpory, nebo nízkou míru spolehlivosti. Možná je i kombinace obojího, tedy nízká míra podpory i spolehlivosti. Velmi známý příklad je taktéž uveden v [1]:

„1. použité bezpečnostní pásy => přežití automobilové havárie
(obecně uznávané pravidlo o účinnosti bezpečnostních pásů),
2. použité bezpečnostní pásy $\wedge$ věk = předškolní => úmrtí při havárii
(překvapivá výjimka, pro malé děti nejsou pásy vhodné),
3. věk = předškolní => úmrtí při havárii
(referenční pravidlo, při haváriích umírá málo předškolních dětí)“

Tato pravidla je však možné vyhledávat právě až ve vygenerovaném seznamu asociačních pravidel. Existují však také alternativní metody přímého generování.

2.6. Dvojité implikace a ekvivalence

V předešlém textu byl doposud pouze předpoklad asociačních pravidel v tzv. Agrawalově pojetí, tedy byla vyjadřována jako implikace

$$Ant => Con. \qquad (1)$$

Znaky, které charakterizují pravidla tohoto druhu, jsou podpora a spolehlivost, tedy

$$\frac{a}{a+b} \text{ a } \frac{a}{a+b+c+d}. \qquad \text{(2a) a (2b)}$$

Další trochu obecnější, avšak velmi zajímavý pohled na vlastní typy asociačních pravidel, tedy vztahy mezi *Ant* a *Con*, které vycházejí z numerické charakteristiky kontingenční tabulky, nabízí různé práce zabývající se metodou GUHA. Tato metoda je ryze českého původu, pracuje s vlastní terminologií, která však vychází z predikátové logiky. Jedním z nejznámějších autorů je Petr Hájek. Pro předpoklad tedy *Ant* je zde opět použit výraz antecedent, pro závěr však *Suc*, tedy sukcekvent, a ne výraz *Con*, conseqvent[8]. Dále pak je zde zaveden i termín kvantifikátor pro typ pravidla.

Kvantifikátor je všeobecně známý termín hojně využívaný v matematice. Pracuje se především s následující dvojicí kvantifikátorů [2]:

$$\text{obecný } (\forall) \text{ a existenční } (\exists)$$

Příkladem může být formule $\forall x Y(x)$, která platí pouze v případě, že právě všechna *x* splňují uvedené tvrzení *Y*. Naopak formule $\exists x Y(x)$ platí, když alespoň jediné *x* splňuje uvedené tvrzení *Y*.

Další text této kapitoly je již věnovaný typologiím vztahů mezi *Ant* a *Suc*. Proměnnou F je zde označena funkce, která definuje kvantifikátor. Pokud tato funkce F přímo závisí pouze na hodnotách parametrů *a* a *b* z kontingenční tabulky, pak je zápis ve tvaru $\sim(a,b)$. Funkce F, jež je závislá na parametrech *a*, *b*, *c*, má obdobný zápis $\sim(a,b,c)$ a konečně i funkce F závislá na všech parametrech má zápis $\sim(a,b,c,d)$. Nutno doplnit, že všechny výše uvedené zápisy funkce F jsou v rozsahu [0,1]. [1]

Základní (fundované) typy pravidel jsou dle [1]:

základní implikace $Ant \Rightarrow Suc$, kde

$$\Rightarrow (a,b) = \frac{a}{a+b}, \qquad \text{(8a)}$$

základní dvojitá implikace $Ant \Leftrightarrow Suc$, kde

8 V tomto případě přijímám pro závěr označení sukcekvent, pro dodržení „čistoty“ vzniklých pravidel. Je ale stejně tak možné používat nadále výraz conseqvent.

$$\Leftrightarrow (a, b, c) = \frac{a}{a+b+c}, \qquad (8b)$$

základní ekvivalence $Ant \equiv Suc$, kde

$$\equiv (a, b, c, d) = \frac{a+d}{a+b+c+d}, \qquad (8c)$$

První z uvedených typů, tedy implikace je vztah asymetrický, ostatní pak jsou vztahy symetrické. Lze tedy psát, že $Ant \Leftrightarrow Suc$ je to samé jako $Suc \Leftrightarrow Ant$ a že $Ant \equiv Suc$ je taktéž to samé jako $Suc \equiv Ant$. Dále pak pravidlo dvojité implikace $Ant \Leftrightarrow Suc$ je ekvivalentním zápisem pro $(Ant \Rightarrow Suc) \wedge (Suc \Rightarrow Ant)$ a $Ant \equiv Suc$ je vzájemnou závislostí *Ant* a *Suc*. [1]

2.7. Obecný úvod do metody GUHA

S návrhem konceptu asociačních pravidel přišla skupina vědců kolem Petra Hájka již asi třicet let před Agrawalem. Hlavní myšlenkou metody General Unary Hypotheses Automaton, zkráceně jen GUHA, kterou tato skupina vyvinula, je nalézt v zadaných datech veškeré zajímavé souvislosti, a poté je uživateli nabídnout k dalšímu zpracování. Tato metoda se v období svého vzniku řadila mezi metody extrapolační analýzy dat. Tento druh analýzy si klade za cíl nejen ověřit (testovat) platnost zvolené statistické hypotézy, jak tomu je u konfirmační analýzy, ale i tyto hypotézy daným způsobem vytvářet. Výstižné je v tomto případě přirovnání konfirmační analýzy k chytání ryb v rybníce na udici, oproti přirovnání metody GUHA k výlovu tohoto rybníku.

Vlastní podstatou metody GUHA je sloučení metod, které hypotézy vytváří, s metodami, které naopak hypotézy testují. Hypotéz, a tím i algoritmů pro jejich vytváření, které popisují vztah mezi kombinacemi binárních atributů či hledají v nominálních datech zdroje jejich závislosti, byla definována jen omezená množina. Například původní GUHA-procedura ASSOC či od ní odvozená GUHA-procedura 4FT-Miner. Hypotézy, které jsou vytvářené a ověřované procedurou 4FT-Miner jsou tvaru

$$Ant \sim Suc / Cond, \qquad (9)$$

kde opět *Ant* – antecedent, *Suc*-sukcedent, *Cond*-condition jsou konjunkce literálů a ~ je zobecněný kvantifikátor, který charakterizuje typ vztahu antecedentu a sukcedentu na množině objektů splňujících podmínku (condition). Základem pro vytváření hypotéz je pozitivní nebo negativní literál. V případě pozitivního literálu je definován jako atribut (koeficient) a v případě literálu negativního zase jako ¬atribut (koeficient). Seznam hodnot tohoto atributu (koeficientu) může být podmnožina, interval či řez. [1]

Kombinace vzniklé pomocí *Ant*, *Suc* a *Cond* jsou mnohem obsáhlejší svou strukturou, než je zvykem u ostatních systémů hledajících asociační pravidla. Uvedená procedura 4FT-Miner navíc nabízí ještě větší množství kvantifikátorů. Krom kvantifikátorů implikačních, dvojitě implikačních a ekvivalenčních ještě kvantifikátory fisherovské a asociační. Přesný popis uvedených kvantifikátorů lze vyhledat v literatuře zabývající se metodou GUHA. Jedna z těchto prací je uvedena v seznamu literatury pod číslem [3].

2.8. Analýza nákupního košíku

Analýza nákupního košíku, zkráceně jen MBA, z anglického jazyka Market Basket Analysis, je jak již sám název napovídá spojená s nakupováním. Jedná se v podstatě o analýzu prodeje, kde je zkoumáno, které položky nákupu jsou v nákupním košíku uloženy dohromady, tedy jaké zboží nebo druhy zboží zákazník kupuje současně. Použití této metody nalezne uplatnění například při rozhodování, jaké výrobky na jakém konkrétním místě a jakým způsobem budou umístěny v prodejně tak, aby to zákazníkovi co nejvíce vyhovovalo a utratil tak v obchodě co nejvíce finančních prostředků. Také však vidím využití například při školení zaměstnanců, kteří výrobky přímo prodávají zákazníkům, a při koupi jednoho výrobku mohou okamžitě dle výsledků analýzy doporučit výrobek jiný, a to ten, který se nejčastěji prodává společně s výrobkem již zákazníkem vybraným. Příkladem budiž nákup plovoucí podlahy, kdy prodávající nabídne ke koupi kupujícímu další produkt, např. leštidlo na podlahy, okrajové lišty, řezačku nebo izolační vrstvu, související s produktem, který si již kupující vybral. Dále pak může prodejce nabídnou zákazníkovi při prodeji jednoho výrobku slevu na výrobek další a tak podobně.

Výsledky analýzy MBA jsou vyjádřeny formou asociačních pravidel, která by měla být snadno pochopitelná, ověřitelná a použitelná. S tím je spojena i potřeba bezprostředního použití těchto pravidel, Ta musí obsahovat natolik užitečné informace, aby pomocí nich mohlo dojít k dalším intervencím. Naproti tomu však získaná asociační pravidla nesmějí být triviální, tj. pravidla, která jsou již všeobecně známá a nepřinášejí tedy žádnou další užitnou hodnotu a nevysvětlitelná, to jsou ta pravidla, k nimž neexistuje žádné logické vysvětlení a ve svém důsledku tedy nevedou k žádné další akci. [13]

Před vlastním osvětlením této metody MBA je nutné definovat základní pojmy [13]:

Položka – je konkrétní zboží, produkt či nabídka služeb,
Transakce – je složena z několika jednotlivých položek,
Tabulka četností – obsahuje údaje o počtu výskytů libovolných kombinací položek právě v některé z výše uvedených již provedených transakcí.

To znamená, že v této tabulce se vyskytují údaje o tom, kolikrát byly dva určité produkty zakoupeny dohromady. V tabulce četností se na diagonále píší hodnoty, které odpovídají počtu jednotlivých transakcí, které obsahují příslušnou položku.
Pravidlo – má tvar IF podmínka THEN výsledek, tedy Pravidlo R: IF položka x THEN položka y.

Ukazateli určujícími vlastnosti asociačních pravidel jsou stejně jako v předchozím textu podpora, spolehlivost a zlepšení [13]:
Podpora – vyjadřuje, jak často lze pravidlo využít

$$Podpora(Pravidlo\_R) = \frac{Pocet\_transakci\_obsahujicich\_polozky\_i\_j}{Pocet\_v\sec h\_transakcí} * 100\%, \qquad (10a)$$

spolehlivost – vyjadřuje, jak se lze na výsledky pravidla spolehnout

$$Splehlivost(Pravidlo\_R) = \frac{Pocet\_transakci\_obsahujicich\_polozky\_i\_j}{Pocet\_transakci\_obsahujicich\_polozku\_i} * 100\%, \qquad (10b)$$

zlepšení – vyjadřuje, o kolik je lepší při predikci dané pravidlo použít, než výsledek pouze předpokládat

$$Zlepseni(Pravidlo\_R) = \frac{p(i\_a\_j)}{p(i) * p(j)}. \qquad (10c)$$

Zde také platí, že pokud je zlepšení < 1 , pak je dané pravidlo při predikci horší, než vlastní náhodná volba. Z toho vyplývá, že negace výsledku vede k lepšímu pravidlu. tj. IF podmínka THEN NOT výsledek.

Kroky, ve kterých se analýza nákupního košíku provádí, jsou v podstatě tři základní. Za prvé je to volba odpovídající (sledované) položky na adekvátní úrovni. Za druhé vlastní vytvoření pravidla, a to na základě dat z tabulky četností, tj. výpočet podmíněných pravděpodobností výskytu položek a samozřejmě i jejich kombinací v transakcích. Při omezení prohledávání prahovou hodnotou podpory. Třetím krokem je určení nejlepšího pravidla, a to analýzou vypočtených pravděpodobností. Jak již bylo řečeno v kapitole věnované algoritmu apriori, hledají se zajímavosti v pravidlech. Tato kapitola je v některých částech dosti podobná kapitole předešlé, avšak analýza nákupního košíku je v této práci natolik důležitý pojem, že jsem se rozhodl věnovat se mu ještě podrobněji a definovat jej poněkud srozumitelněji bez ohledu na jakýkoli jiný algoritmus či metodu. [13]

Ta tomto místě je také uveden jednoduchý příklad analýzy nákupního košíku, a to se všemi potřebnými výpočty, aby byl vysvětlen celý princip této analýzy a následně se k těmto výpočtům již nemuselo znovu vracet a podrobně je popisovat. Dá se totiž předpokládat, že to velmi pomůže k úplnému pochopení následného příkladu, který je jedním z pilířů této práce.

Princip této metody je nejlépe zřejmý právě na příkladu nakupování potravin v obchodě, jak je uvedeno v mnohých materiálech dostupných na internetu. Nyní jde však již o pokus mírně proniknout do občanské vybavenosti regionu a na tomto velmi jednoduchém případě budou zaměněny transakce, tedy nákup uložený v košíku za město, respektive hned několik měst, které obsahují místo jednotlivých položek nákupu vlastní entity občanské vybavenosti, ale také více či méně komerčních firem v daném městě.

Příklad: Pomocí analýzy nákupního košíku jsou zde sledována města (transakce) a jejich bezprostřední okolí v Ústeckém kraji, tedy Chomutov, Most, Litvínov, Ústí nad Labem a Litoměřice. V těchto městech jsou sledovány a zaznamenány výskyty následujících organizací (položek): chemička, uhelné doly, strojírenský průmysl a elektrárna. Data v příkladu jsou jen velmi mírně modifikována, aby byl výpočet názornější. Vše je nejprve shrnuto do přehledné transakční tabulky (tabulka číslo 11),

Tabulka 11: Transakční tabulka, zdroj: [vlastní].

Město	**Organizace**
Chomutov	Doly + Strojírenský průmysl
Litoměřice	Chemička + Strojírenský průmysl
Litvínov	Doly + Elektrárna + Chemička
Most	Doly + Elektrárna
Ústí nad Labem	Chemička + Strojírenský průmysl

Nyní je možno přistoupit k sestrojení tabulky četností, kde jsou jak ve sloupcích, tak v řádcích, uvedeny všechny sledované organizace. Tabulka číslo 12 tedy vychází z tabulky 11.

Tabulka 12: Tabulka četností, zdroj: [vlastní].

X	**Doly**	**Elektrárna**	**Chemička**	**Strojírenský průmysl**
Doly	3	2	0	1
Elektrárna	2	2	1	0
Chemička	0	1	3	2
Strojírenský průmysl	1	0	2	3

Na první pohled jsou již některé pravidla jasná, např. kde je elektrárna, jsou i uhelné doly, nebo kde je chemička, naopak žádné uhelné doly nejsou. V tabulce číslo 13 jsou pak uvedena veškerá pravidla i s příslušnými hodnotami podpory, spolehlivosti a zlepšení, které jsou vypočítávány ze vztahů (10a), (10b) a (10c)

Tabulka 13: Tabulka asociačních pravidel, zdroj: [vlastní].

Pravidlo	**Podpora (%)**	**Spolehlivost (%)**	**Zlepšení**
IF Doly THEN Elektrárna	40	66,67	1,67
IF Doly THEN Chemička	0	0	0
IF Doly THEN Strojírenský průmysl	20	33,33	0,56
IF Elektrárna THEN Doly	40	100	1,67
IF Elektrárna THEN Chemička	20	50	0,83
IF Elektrárna THEN Strojírenský průmysl	0	0	0
IF Chemička THEN Doly	0	0	0
IF Chemička THEN Elektrárna	20	33,33	0,83
IF Chemička THEN Strojírenský průmysl	40	66,67	1,11
IF Strojírenský průmysl THEN Doly	20	33,33	0,56
IF Strojírenský průmysl THEN Elektrárna	0	0	0
IF Strojírenský průmysl THEN Chemička	40	66,67	1,11

Při větším počtu asociačních pravidel je nutností tyto pravidla nějakým způsobem redukovat, a to tak, aby výsledná pravidla měla nějakou vypovídací hodnotu. Jak již bylo zmíněno, tak by zlepšení pravidla mělo mít hodnotu větší než 1. Toto také bude jedním z předpokladů při sestavování redukované tabulky pravidel. Druhým předpokladem může být i stanovení podpory pravidla na minimální hodnotě 30% a taktéž stanovení minimální hladiny pro spolehlivost. V tomto konkrétním případě je zvolena hladina spolehlivosti na 50%.

Při respektování všech výše uvedených předpokladů je sestrojena následující tabulka číslo 14, která obsahuje pouze taková asociační pravidla, která vyhovují právě výše uvedeným předpokladům,

Tabulka 14: Redukovaná tabulka asociačních pravidel, zdroj: [vlastní].

Pravidlo	Podpora (%)	Spolehlivost (%)	Zlepšení
IF Doly THEN Elektrárna	40	66,67	1,67
IF Elektrárna THEN Doly	40	100	1,67
IF Chemička THEN Strojírenský průmysl	40	66,67	1,11
IF Strojírenský průmysl THEN Chemička	40	66,67	1,11

V tabulce 14 jsou tedy již jen pravidla splňující kritéria: Podpora > 30%, Spolehlivost > 50 a Zlepšení > 1. Tyto pravidla se pak dají považovat za použitelná a je možné s nimi dále pracovat.

2.9. Analýza občanské vybavenosti regionu pomocí analýzy nákupního košíku

Pro analýzu občanské vybavenosti regionu prostřednictvím MBA jsou zvoleny za zkoumané regiony veškerá Česká krajská města s výjimkou hlavního města Prahy. Důvodem pro tuto volbu je fakt, že krajská města jsou všeobecně považována za kulturní centra jednotlivých krajů a dá se tedy předpokládat, že občanská vybavenost v těchto obcích bude na vyšší, v zásadě vzájemně porovnatelné úrovni, než ve městech menších. Zmíněnou výjimku tvoří hlavní město Praha, které se jak svou polohou, počtem obyvatel tak i přítomností mnoha ojedinělých institucí (například vládních), liší od ostatních krajských měst natolik, že bylo z analýzy vypuštěno. Města do analýzy zahrnutá jsou tedy, dle abecedního pořadí následující: Brno, České Budějovice, Hradec Králové, Jihlava, Karlovy Vary, Liberec, Olomouc, Ostrava, Pardubice, Plzeň, Ústí nad Labem, Zlín.

Jako ukazatele občanské vybavenosti výše uvedených regionů jsem zvolil organizace z různých odvětví tak, aby tato analýza měla širší rozsah a tak i obecnější informační hodnotu. Jedná se pak zejména o následující organizace[9] (opět v abecedním pořadí): Divadlo (Div.), Jesle (Jes.), Krematorium (Krem.), Lázeňské léčebny (Láz.), Letecká a kosmická doprava (Let.), Multikino (Mul.), Nemocnice (Nem.), Ústavy sociální péče pro dospělé (USP), Vodní doprava (Vod.), Vysoká škola (VS), Zoologická zahrada (ZOO). Dané organizace i jejich

[9] Tyto organizace, jakož i veškeré údaje k nim uvedené jsou čerpány z oficiálních stránek Českého statistického úřadu (ČSÚ). [14]

počet, případně nepřítomnost, jsou uvedeny v následující tabulce číslo 15, která má stejnou podobu pro všech dvanáct sledovaných krajských měst[10]

Tabulka 15: Tabulka vstupních dat pro město Brno, zdroj: [vlastní].

	Organizace	Počet
Brno	Divadlo	14
	Jesle	3
	Krematorium	1
	Lázeňské léčebny	0
	Letecká a kosmická doprava	2
	Multikino	1
	Nemocnice	11
	Ústavy sociální péče pro dospělé	2
	Vodní doprava	0
	Vysoká škola	9
	Zoologická zahrada	1

Vynásobíme-li počet měst a počet organizací, tedy 12 * 11 dostaneme sumu všech údajů, které budou do analýzy zahrnuty, tj. 132 údajů. Tyto tabulky vstupních dat jsou sice přehledné a pro reprezentaci občanské vybavenosti jednotlivých krajských měst vhodné, ale pro samotnou analýzu je výhodnější je transformovat do tabulky číslo 16,

Tabulka 16: Souhrnná tabulka vstupních dat, zdroj: [vlastní].

Město	**Organizace**										
	Divadlo	Jesle	Krematorium	Lázně	Letecká dop.	Multikino	Nemocnice	Ústav soc. péče	Vodní dop.	Vysoká škola	ZOO
Brno	14	3	1	0	2	1	11	2	0	9	1
České Budějovice	5	1	1	0	3	1	1	0	3	2	0
Hradec Králové	2	1	0	0	0	1	2	0	0	1	0
Jihlava	2	1	1	0	0	0	1	0	0	1	1
Karlovy Vary	9	0	1	40	0	0	1	0	1	1	0
Liberec	4	0	1	0	0	0	1	1	2	1	1
Olomouc	2	1	1	0	1	1	2	1	0	2	1
Ostrava	6	0	1	0	0	1	3	1	1	3	1
Pardubice	4	1	1	0	1	1	3	0	4	1	0
Plzeň	10	2	1	0	2	1	4	3	0	2	1
Ústí nad Labem	2	3	2	0	1	0	1	0	7	2	1
Zlín	2	2	1	1	0	0	2	1	0	1	1
SUMA	62	15	12	41	10	7	32	9	18	26	8

[10] Kompletní přehled všech tabulek vstupních dat viz. přílohy 1 a 2, strana I a II.

Pro analýzu v rámci MBA však není potřeba uvažovat počety uvedených organizací ve všech městech, ale postačí pouze údaje o tom, zda se daná organizace v tom či onom městě nachází či nikoliv. Je proto vhodné souhrnnou tabulku vstupních dat dále transformovat na tabulku obsahující, v případě přítomnosti organizace v obci, na daném místě číslo 1 a naopak, v případě nepřítomnosti organizace, číslo 0. Tato tabulka je zde uvedena jako tabulka číslo 17,

Tabulka 17: Tabulka přítomnosti organizací v obcích, zdroj: [vlastní].

Město	**Organizace**										
	Divadlo	Jesle	Krematorium	Lázně	Letecká dop.	Multikino	Nemocnice	Ústav soc. péče	Vodní dop.	Vysoká škola	ZOO
Brno	1	1	1	0	1	1	1	1	0	1	1
České Budějovice	1	1	1	0	1	1	1	0	1	1	0
Hradec Králové	1	1	0	0	0	1	1	0	0	1	0
Jihlava	1	1	1	0	0	0	1	0	0	1	1
Karlovy Vary	1	0	1	1	0	0	1	0	1	1	0
Liberec	1	0	1	0	0	0	1	1	1	1	1
Olomouc	1	1	1	0	1	1	1	1	0	1	1
Ostrava	1	0	1	0	0	1	1	1	1	1	1
Pardubice	1	1	1	0	1	1	1	0	1	1	0
Plzeň	1	1	1	0	1	1	1	1	0	1	1
Ústí nad Labem	1	1	1	0	1	0	1	0	1	1	1
Zlín	1	1	1	1	0	0	1	1	0	1	1

Nyní je možné přistoupit k vlastní analýze. Na základě získaných údajů, lze již poměrně snadno vytvořit tabulku jednotlivých transakcí, tedy tabulku číslo 18. V tomto konkrétním případě je jako transakce uvažováno město a položky této transakce jsou jednotlivé organizace, které jsou v daném městě obsaženy,

Tabulka 18: Transakční tabulka, zdroj: [vlastní].

Město	**Organizace**
Brno	Divadlo + Jesle + Krematorium + Letecká dop. + Multikino + Nemocnice + Ústav soc. péče + Vysoká škola + ZOO
České Budějovice	Divadlo + Jesle + Krematorium + Letecká dop. + Multikino + Nemocnice + Vodní dop. + Vysoká škola
Hradec Králové	Divadlo + Jesle + Multikino + Nemocnice + Vysoká škola
Jihlava	Divadlo + Jesle + Krematorium + Nemocnice + Vysoká škola + ZOO
Karlovy Vary	Divadlo + Krematorium + Lázně + Nemocnice + Vodní dop. + Vysoká škola
Liberec	Divadlo + Krematorium + Nemocnice + Ústav soc. péče + Vodní dop. + Vysoká škola + ZOO
Olomouc	Divadlo + Jesle + Krematorium + Letecká dop. + Multikino + Nemocnice + Ústav soc. péče + Vysoká škola + ZOO
Ostrava	Divadlo + Krematorium + Multikino + Nemocnice + Ústav soc. péče + Vodní dop. + Vysoká škola + ZOO
Pardubice	Divadlo + Jesle + Krematorium + Letecká dop. + Multikino + Nemocnice + Vodní dop. + Vysoká škola
Plzeň	Divadlo + Jesle + Krematorium + Letecká dop. + Multikino + Nemocnice + Ústav soc. péče + Vysoká škola + ZOO
Ústí nad Labem	Divadlo + Jesle + Krematorium + Letecká dop. + Nemocnice + Vodní dop. + Vysoká škola + ZOO
Zlín	Divadlo + Jesle + Krematorium + Lázně + Nemocnice + Ústav soc. péče + Vysoká škola + ZOO

Dále pak, je již možno vytvořit tabulku četností, tabulku číslo 19, která jak již bylo řečeno, je základem dalších výpočtů. V tomto případě má tabulka následující tvar,

Tabulka 19: Tabulka četností, zdroj: [vlastní].

	Divadlo	**Jesle**	**Krematorium**	**Lázně**	**Letecká dop.**	**Multikino**	**Nemocnice**	**Ústav soc. péče**	**Vodní dop.**	**Vysoká škola**	**ZOO**
Divadlo	12	9	11	2	6	7	12	6	6	12	8
Jesle	9	9	8	1	6	6	9	4	3	9	6
Krematorium	11	8	11	2	6	6	11	6	6	11	8
Lázně	2	1	2	2	0	0	2	1	1	2	1
Letecká dop.	6	6	6	0	6	5	6	3	3	6	4
Multikino	7	6	6	0	5	7	7	4	3	7	4
Nemocnice	12	9	11	2	6	7	12	6	6	12	8
Ústav soc. péče	6	4	6	1	3	4	6	6	2	6	6
Vodní dop.	6	3	6	1	3	3	6	2	6	6	3
Vysoká škola	12	9	11	2	6	7	12	6	6	12	8
ZOO	8	6	8	1	4	4	8	6	3	8	8

Tabulka četností obsahuje data, která jsou kombinacemi jednotlivých organizací ve všech uvažovaných městech dohromady. Například letecká doprava a ústav sociální péče pro dospělé se vyskytuje zároveň ve třech městech (Brno, Olomouc a Plzeň). Diagonální prvky této tabulky jsou absolutní četnosti daných organizací ve všech obcích. To znamená, že např. prvek umístěný v pravém dolním rohu tabulky, na průsečíku organizace zoologická zahrada a zoologická zahrada s hodnotou 8 vyjadřuje skutečnost, že v osmi z dvanácti uvažovaných měst je právě tato organizace obsažena.

Na základě dosud uvedených informací je v dalším kroku již možné generovat vlastní asociační pravidla. Část těchto pravidel je uvedena v tabulce číslo 20,[11]

[11] Vygenerovaných asociačních pravidel je v tomto případě 110.

Tabulka 20: Tabulka asociačních pravidel, zdroj: [vlastní].

Pravidlo	Podpora (%)	Spolehlivost (%)	Zlepšení
IF Divadlo THEN Jesle	75,00	75,00	1,00
IF Divadlo THEN Krematorium	91,67	91,67	1,00
IF Divadlo THEN Lázně	16,67	16,67	1,00
IF Divadlo THEN Letecká dop.	50,00	50,00	1,00
IF Divadlo THEN Multikino	58,33	58,33	1,00
IF Divadlo THEN Nemocnice	100,00	100,00	1,00
IF Divadlo THEN Ústav soc. péče	50,00	50,00	1,00
IF Divadlo THEN Vodní dop.	50,00	50,00	1,00
IF Divadlo THEN Vysoká škola	100,00	100,00	1,00
IF Divadlo THEN ZOO	66,67	66,67	1,00
IF Jesle THEN Divadlo	75,00	100,00	1,00
IF Jesle THEN Krematorium	66,67	88,89	0,97
IF Jesle THEN Lázně	8,33	11,11	0,67
IF Jesle THEN Letecká dop.	50,00	66,67	1,33
IF Jesle THEN Multikino	50,00	66,67	1,14
IF Jesle THEN Nemocnice	75,00	100,00	1,00
IF Jesle THEN Ústav soc. péče	33,33	44,44	0,89
IF Jesle THEN Vodní dop.	25,00	33,33	0,67
IF Jesle THEN Vysoká škola	75,00	100,00	1,00
IF Jesle THEN ZOO	50,00	66,67	1,00
IF Krematorium THEN Divadlo	91,67	100,00	1,00
IF Krematorium THEN Jesle	66,67	72,73	0,97
IF Krematorium THEN Lázně	16,67	18,18	1,09
IF Krematorium THEN Letecká dop.	50,00	54,55	1,09
IF Krematorium THEN Multikino	50,00	54,55	0,94
IF Krematorium THEN Nemocnice	91,67	100,00	1,00
IF Krematorium THEN Ústav soc. péče	50,00	54,55	1,09
IF Krematorium THEN Vodní dop.	50,00	54,55	1,09
IF Krematorium THEN Vysoká škola	91,67	100,00	1,00
IF Krematorium THEN ZOO	66,67	72,73	1,09

Tato tabulka obsahuje čtyři sloupce, kde v prvním z nich jsou jednotlivá asociační pravidla, v druhém procentuální míra podpory, v třetím spolehlivost, opět v procentech a v posledním čtvrtém sloupci je uvedeno zlepšení pravidla. Například pravidlo „IF Divadlo THEN Jesle“ má podporu 75%, spolehlivost taktéž 75% a zlepšení rovno 1. Uvedené hodnoty jsou opět vypočítány dle vztahů (10a), (10b) a (10c).

Je zřejmé, že v tomto, a mnohdy i ještě větším množství, některá z asociačních pravidel v podstatě nelze v praktických úlohách využít, nebo, že mají malou vypovídající hodnotu. Právě proto dochází v posledním kroku této analýzy k redukci asociačních pravidel na takovou míru, která je dle zadání požadována, či dává uspokojivé výsledky. V tomto případě je redukce pravidel uskutečněna stanovením minimální hranice pro všechny z uvedených vlastností asociačních pravidel následovně:

$$\text{Podpora} \geq 50\% \qquad \text{Spolehlivost} \geq 70\% \quad \text{Zlepšení} \geq 1$$

Výsledná tabulka číslo 21 obsahuje jen ta asociační pravidla, která splňují veškeré výše uvedené podmínky,

Tabulka 21: Redukovaná tabulka asociačních pravidel, zdroj: [vlastní].

Pravidlo	Podpora (%)	Spolehlivost (%)	Zlepšení
IF Divadlo THEN Jesle	75,00	75,00	1,00
IF Divadlo THEN Krematorium	91,67	91,67	1,00
IF Divadlo THEN Nemocnice	100,00	100,00	1,00
IF Divadlo THEN Vysoká škola	100,00	100,00	1,00
IF Jesle THEN Divadlo	75,00	100,00	1,00
IF Jesle THEN Nemocnice	75,00	100,00	1,00
IF Jesle THEN Vysoká škola	75,00	100,00	1,00
IF Krematorium THEN Divadlo	91,67	100,00	1,00
IF Krematorium THEN Nemocnice	91,67	100,00	1,00
IF Krematorium THEN Vysoká škola	91,67	100,00	1,00
IF Krematorium THEN ZOO	66,67	72,73	1,09
IF Letecká dop. THEN Divadlo	50,00	100,00	1,00
IF Letecká dop. THEN Jesle	50,00	100,00	1,33
IF Letecká dop. THEN Krematorium	50,00	100,00	1,09
IF Letecká dop. THEN Nemocnice	50,00	100,00	1,00
IF Letecká dop. THEN Vysoká škola	50,00	100,00	1,00
IF Multikino THEN Divadlo	58,33	100,00	1,00
IF Multikino THEN Jesle	50,00	85,71	1,14
IF Multikino THEN Nemocnice	58,33	100,00	1,00
IF Multikino THEN Vysoká škola	58,33	100,00	1,00
IF Nemocnice THEN Divadlo	100,00	100,00	1,00
IF Nemocnice THEN Jesle	75,00	75,00	1,00
IF Nemocnice THEN Krematorium	91,67	91,67	1,00
IF Nemocnice THEN Vysoká škola	100,00	100,00	1,00
IF Ústav soc. péče THEN Divadlo	50,00	100,00	1,00
IF Ústav soc. péče THEN Krematorium	50,00	100,00	1,09
IF Ústav soc. péče THEN Nemocnice	50,00	100,00	1,00
IF Ústav soc. péče THEN Vysoká škola	50,00	100,00	1,00
IF Ústav soc. péče THEN ZOO	50,00	100,00	1,50
IF Vodní dop. THEN Divadlo	50,00	100,00	1,00
IF Vodní dop. THEN Krematorium	50,00	100,00	1,09
IF Vodní dop. THEN Nemocnice	50,00	100,00	1,00
IF Vodní dop. THEN Vysoká škola	50,00	100,00	1,00
IF Vysoká škola THEN Divadlo	100,00	100,00	1,00
IF Vysoká škola THEN Jesle	75,00	75,00	1,00
IF Vysoká škola THEN Krematorium	91,67	91,67	1,00
IF Vysoká škola THEN Nemocnice	100,00	100,00	1,00
IF ZOO THEN Divadlo	66,67	100,00	1,00
IF ZOO THEN Jesle	50,00	75,00	1,00
IF ZOO THEN Krematorium	66,67	100,00	1,09
IF ZOO THEN Nemocnice	66,67	100,00	1,00
IF ZOO THEN Ústav soc. péče	50,00	75,00	1,50
IF ZOO THEN Vysoká škola	66,67	100,00	1,00

3. Fuzzy teorie

V této části práce jsou popsány fuzzy množiny, rozdíl mezi fuzzy a klasickou množinou, výroková fuzzy logika, Mukaidonův fuzzy rezoluční princip a samozřejmě i fuzzy asociační pravidla. V závěru kapitoly je pak také vlastní příklad analýzy občanské vybavenosti pomocí fuzzy asociačních pravidel.

3.1. Fuzzy množiny

V mnoha odvětvích je běžné, že jsou i zásadní problémy nejprve definované pomocí přirozeného jazyka. Tento jazyk je reprodukován jedincem, který může formulovat daný problém poněkud vágním, či neurčitým způsobem. Fuzzy množiny se pak většinou používají jako prostředek, umožňující definovat existující veličiny libovolného objektu, které nemusí mít nutně numerický charakter, ve tvaru numerickém. V tomto případě, lze tedy tvrdit, že fuzzy množiny umožňují zpracování sémantiky slov daného přirozeného jazyka a to právě v numerickém tvaru.

Americký profesor Íránského původu Lotfali Askar-Zadeh který zavedl v roce 1965 pojem fuzzy množin, definoval na příkladu neurčitých a složitých informací o systémech následující [5]:

„*Tak, jak roste složitost nějakého systému, klesá naše schopnost uskutečňovat precizní a přitom stále ještě použitelná tvrzení o jeho chování, pokud není dosažený práh (hranice), za kterým se stávají preciznost a použitelnost téměř vzájemně se vylučujícími charakteristikami. Jinak řečeno, studovaná oblast reálných systémů vykazuje mnoho nejasného a neurčitého.*“

Tomuto výstižnému popisu problému, lze rozumět i tak, že s rostoucí precizností zápisu tvrzení velmi složitého problému, klesá jeho srozumitelnost a tím aplikovatelnost (použitelnost). Při určité, vysoké, míře preciznosti, se pak mohou preciznost a použitelnost vzájemně negovat.

Jedním z výše uvedených pojmů je také neurčitost, která hraje v teorii fuzzy množin velkou roli. Neurčitost je většinou dělena na dvě základní skupiny a to na neurčitost stochastickou a sémantickou. Stochastická neurčitost znemožňuje predikci budoucího stavu či vývoje objektu z důvodu neúplnosti informací o tomto objektu či jeho okolí. Naopak neurčitost sémantická tkví v nepřesném nebo nejednoznačném popisu objektu, či jeho stavů.

Tato sémantická neurčitost se projevuje například při snaze o popis objektu v přirozeném jazyce, který nemá striktně definované, či dvousmyslné významy jednotlivých slov. [6]

3.2. Fuzzy x Klasické množiny

U klasických množin, tak jak jsou notoricky známé z oblasti matematiky, je vlastnost určující přítomnost prvku v dané množině jednoznačná. Tj. je přesně určeno, do jaké množiny (množin) zvolený prvek patří a do jaké množiny (množin) nikoli. V binárním vyjádření se potom dá tato vlastnost reprezentovat jako čísla 0 a 1, kde 1 znamená, že prvek do množiny patří a 0, že prvek do množiny naopak nepatří.

V teorii fuzzy množin, se však uvažuje i o možnosti pouze částečné příslušnosti prvku k fuzzy množině. Charakteristická funkce fuzzy množiny pak může pracovat s hodnotami v intervalu [0,1], kde má horní i dolní prvek intervalu stejný význam jako u klasických množin. Ostatní hodnoty intervalu pak určují velikost stupně příslušnosti prvku k dané fuzzy množině. Následující definice, převzaty z [5].

„Nechť X je proměnná (ve smyslu klasické matematiky) a U množina (ve smyslu klasické teorie množin) hodnot, které může proměnná X dosahovat. Proměnná X se nazývá bázová proměnná a množina U univerzem nebo referenční množinou. Referenční množina je reprezentovaná osou reálných čísel nebo její podmnožinou. Nechť každému prvku u $\in$ U je přiřazeno reálné číslo $A(u) \in [0,1]$. Číslo A(u) udává na intervalu [0,1] míru (stupeň) možnosti toho, že bázová proměnná X dosahuje právě hodnotu u.“

„V teorii fuzzy množin je tímto způsobem, tj. pomocí funkce příslušnosti A(u) definována fuzzy množina na příslušném univerzu. Hodnota A(u) funkce příslušnosti v bodě $u \in U$ udává stupeň příslušnosti prvku do fuzzy množiny.“

Pro názornost zde uvádím příklad, kdy jsou porovnávány výsledky z oblasti klasických a fuzzy množin. Jde o příklad velikosti sklenice podle jejího obsahu v litrech, tj. vyjádření jazykové proměnném která může nabývat hodnot malá, střední a velká. Pokud je přijat předpoklad, že středně velká sklenice je o obsahu jeden litr, pak je pro tuto hodnotu zcela jistě jasné, že hodnota jazykové proměnné nabude jak v oblasti fuzzy množin, tak i v oblasti klasických množin právě hodnoty „*střední*“. Rozdíl však nastává, pokud je obsah sklenice jiný, něž jeden litr.

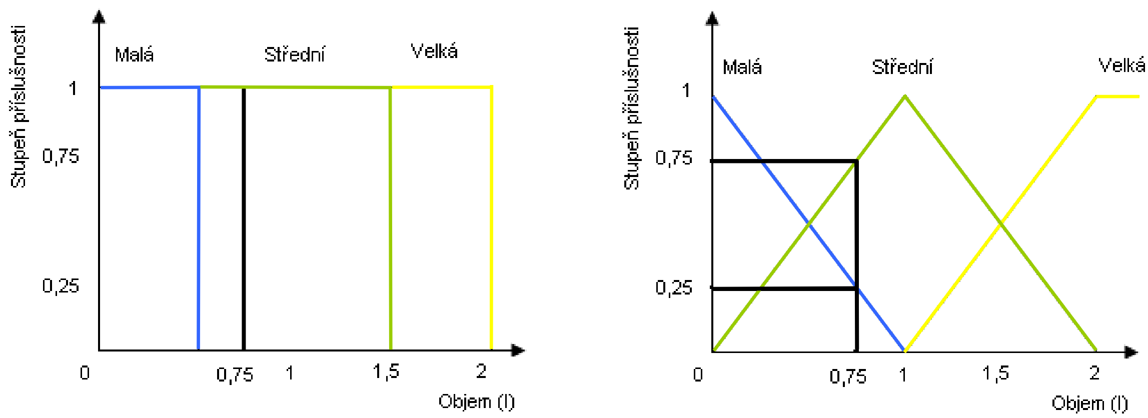
Otázka tedy může být položena následovně: Jak velká je sklenice o obsahu 0,75 litru? Je střední? V oblasti fuzzy množin, je možné v tomto případě na univerzu v rozsahu 0 až 2 litry definovat funkci příslušnosti A(u), která říká, že tato sklenice je střední, ale s hodnotou 0,75. Stejně tak to funguje i pro ostatní hodnoty, viz. tabulka číslo 22, kde jsou pro každou hodnotu (malá, střední, velká) vždy uvedeny hodnoty funkce příslušnosti,

Tabulka 22: Tabulka hodnot funkce příslušnosti, zdroj: [vlastní].

	Objem sklenice (l)								
Velikost	**0 l**	**0,25 l**	**0,5 l**	**0,75 l**	**1 l**	**1,25 l**	**1,5 l**	**1,75 l**	**2 l**
Malá	1	0,75	0,5	0,25	0	0	0	0	0
Střední	0	0,25	0,5	0,75	1	0,75	0,5	0,25	0
Velká	0	0	0	0	0	0,25	0,5	0,75	1

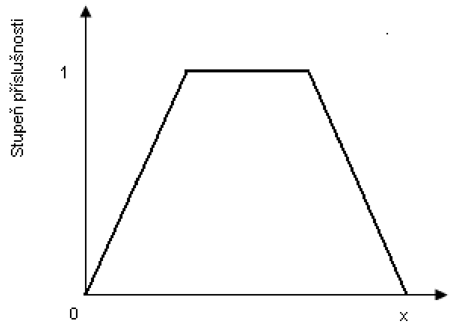
Z tabulky je patrné, že pokud je obsah sklenice 0,75 litru, lze dle funkce příslušnosti určit, že tato sklenice je střední s hodnotou 0,75 a malá s hodnotou 0,25.

Na obrázku číslo 3 jsou pak data z tabulky zakreslena pomocí grafického vyjádření klasických (vlevo) a fuzzy množin (vpravo),



Obrázek 3: Klasické a fuzzy množiny, zdroj: [6].

Fuzzy množiny (funkce příslušnosti) bývají často znázorněny právě pomocí tvaru připomínajícího překrývající se trojúhelníky, toto znázornění vzniklo z tzv. Bell function popotažením. Jedno z dalších možných a také velmi známých grafických vyjádření fuzzy množin je ve tvaru lichoběžníku, viz. následující obrázek číslo 4,



Obrázek 4: Lichoběžníkové grafické vyjádření fuzzy množiny, zdroj: [6].

3.3. Výroková fuzzy logika

Pro další text, je přijat předpoklad, že A, B a C jsou fuzzy množiny. Prázdná množina bude značena symbolem Ø. Nyní lze definovat tři základní funkce příslušnosti (μ), a to průniku, sjednocení a doplňku následovně [6]:

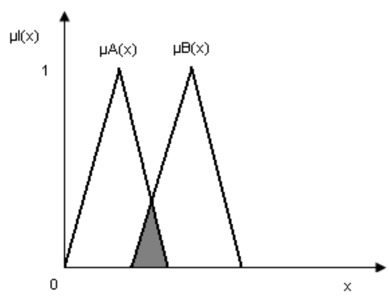
Funkce příslušnosti průniku $\mu_I(x)$ kde $I = A \cap B$ má tvar

$$\mu_I(x) = MIN\{\mu_A(x), \mu_B(x)\}, \tag{11a}$$

následující vlastnosti

$$A \cap X = A,\ A \cap \emptyset = \emptyset,\ A \cap B = B \cap A,\ A \cap (B \cap C) = (A \cap B) \cap C\ ,$$

grafické znázornění je pak



Obrázek 5: Funkce příslušnosti průniku $\mu_I(x)$, zdroj: [6],

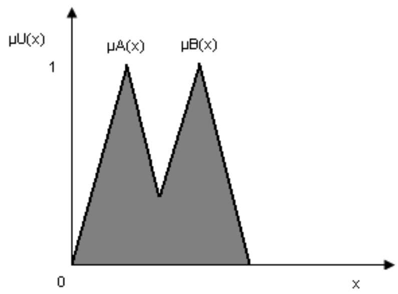
funkce příslušnosti sjednocení $\mu_U(x)$ kde $U = A \cup B$ má tvar

$$\mu_U(x) = MAX\{\mu_A(x), \mu_B(x)\}, \tag{11b}$$

následující vlastnosti

$$A \cup X = X,\ A \cup \emptyset = A,\ A \cup B = B \cup A,\ A \cup (B \cup C) = (A \cup B) \cup C\ ,$$

a grafické znázornění

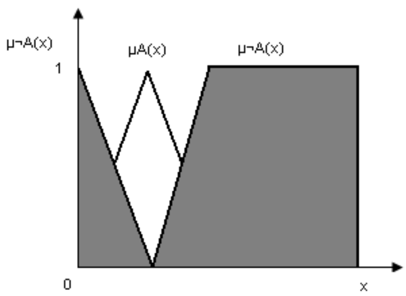


Obrázek 6: Funkce příslušnosti sjednocení $\mu_U(x)$, zdroj: [6],

funkce příslušnosti doplňku $\mu_{\neg A}(x)$ kde $\neg A = 1 - A$ má tvar

$$\mu_{\neg A}(x) = 1 - \mu_A(x), \tag{11c}$$

a grafické znázornění



Obrázek 7: Funkce příslušnosti doplňku $\mu_{\neg A}(x)$, zdroj: [6].

Prvně zmíněnému vztahu $\mu_I(x) = MIN\{\mu_A(x), \mu_B(x)\}$ (tvar funkce příslušnosti průniku) vyhovuje i jistá třída funkcí, která se jmenuje trojúhelníková norma nebo také t-norma a je možné ji vyjádřit následovně [6]:

$$I = A \cap B \Leftrightarrow \forall x \in X : \mu_I(x) = \mu_A(x) t \mu_B(x) \leq MIN\{\mu_A(x), \mu_B(x)\}$$

Když je uvažována t-norma jako zobrazení t: [0,1] × [0,1] →[0,1] pro každé w, x, y, z $\in [0,1]$, pak je tato t-norma buď [6]:

monotónní	$t(x, y) \leq t(w, z), \forall x \leq w, y \leq z$,
komutativní	$t(x, y) = t(y, x)$,
asociativní	$t(t(x, y), z) = t(x, t(y, z))$,
nebo ohraničená	$t(x, 1) = x, t(0, 1) = 0$.

Také vztahu $\mu_U(x) = MAX\{\mu_A(x), \mu_B(x)\}$ (funkce příslušnosti doplňku) vyhovuje jistá třída funkcí, která se naopak jmenuje s-norma, nebo také t-conorma a je vyjádřena následovně [6]:

$$U = A \cup B \Leftrightarrow \forall x \in X : \mu_U(x) = MAX\{\mu_A(x), \mu_B(x)\} \leq \mu_A(x) s \mu_B(x)$$

Platí, že pokud je s-norma (t-conorma) uvažována jako zobrazení s: [0,1] × [0,1] → [0,1] pro každé w, x, y, z $\in [0,1]$, pak je opět buď [6]:

monotónní	$s(x, y) \leq s(w, z), \forall x \leq w, y \leq z$,
komutativní	$s(x, y) = s(y, x)$,
asociativní	$s(s(x, y), z) = s(x, s(y, z))$,
nebo ohraničená	$s(x, 0) = x, s(1, 0) = 1$.

Dále je také možné nahlížet na výrokovou fuzzy logiku (*VFL*) jako na algebraický systém, který lze podle [6] zapsat ve tvaru:

$$VFL = \langle [0, 1], \wedge, \vee, \neg \rangle$$

Kde *VFL* značí výrokovou fuzzy logiku, [0, 1] je uzavřený interval obsahující pravdivostní hodnoty fuzzy výroku a $\wedge, \vee, \neg$ jsou operátory konjunkce, disjunkce a negace.

Pravdivostní hodnoty fuzzy výroku jsou tedy, jak již bylo uvedeno výše, v rozsahu intervalu od 0 do 1. Pokud se tato hodnota konkrétního fuzzy výroku rovná nebo blíží k hranici intervalu, tedy k hodnotám 0 a 1, potom je možné o takovém fuzzy výroku (s velkou pravděpodobností) prohlásit, že je to výrok pravdivý nebo naopak nepravdivý. Problémovým bodem na intervalu [0, 1] je bod 0,5, kdy o fuzzy výroku nelze s jistotou říci, zda je pravdivý či nikoli. Je-li přijat fakt, že fuzzy výrok má označení A a jeho pravdivostní hodnota je stanovena jako $a \in [0, 1]$, pak je navíc možné definovat tzv. koeficient určitosti fuzzy výroku c_a, a to jako

$$c_a = (a - 0{,}5) * 2\,, \tag{12}$$

z tohoto vztahu je evidentní, že pokud se hodnota a, tedy pravdivostní hodnota fuzzy výroku, bude rovnat jedné, pak se koeficient určitosti tohoto fuzzy výroku bude také rovnat 1. Dále, pokud se a bude rovnat 0, pak se c_a bude rovnat -1. A konečně pro hodnotu a=0,5 se bude c_a rovnat 0. [6]

3.4. Mukaidonův fuzzy rezoluční princip

Nejprve je potřeba definovat základní operace a pojmy [5]:

- základními operacemi jsou opět $\wedge, \vee, \neg$;
- výrokové proměnné jsou dány konečnou množinou symbolů X_1, X_2, …,X_n;
- literál je pak název pro jednotlivé X_i a jejich případné negace;
- dalším pojmem je klauzule (C), pro kterou platí, že každý literál X je klauzule C a že i výraz $C \vee X$ je klauzule;
- formule (F) je pak definována následovně – každá klauzule C je formule F, $\neg F$ je též formule a v neposlední řadě také platí, že konjunkce či disjunkce dvou formují je také formule.

Mukaidonův fuzzy rezoluční princip se využívá na výpočet koeficientu určitosti (KU) klauzulí. Jeho následná definice je doslovně převzata z [5]:

„Nechť $C_1 = X \vee L_1$ a $C_2 = \overline{X} \vee L_2$ jsou ukazatele, kde L_1 a L_2 neobsahují literály X a $\overline{X}$ a nemají žádné navzájem komplementární proměnné. Potom klauzuli $L_1 \vee L_2$ možno nazvat rezolventou klauzulí C_1 a C_2 s klíčovým slovem X. Rezolventa klauzulí C_1 a C_2 se označuje $R(C_1, C_2)c$, kde

$$c = |c_x| = [MAX(x, \overline{x}) - 5] * 2$$

je absolutní hodnotou KU klíčového slova X a nazývá se KU rezolventy $R(C_1, C_2)$.

Nechť S je libovolná množina klauzulí. Potom množinu všech rezolvent odvozených ze všech dvojic klauzulí S a všech klauzulí S možno označit následovně $R^1(S)c^1$ a nazvat rezoluční množinou S první třídy, kde $c^1 = c1$, přičemž c1 je minimální KU z rezolvent R^1.

Rezoluční množina S n-té třídy označená $R^n(S)c^n$ je definována následujícím způsobem:

$$R^0(S)c^0 = S,\ R^n(S)c^n = R^1(R^{n-1}(S)c^{n-1})c^1$$

Kde $c^0 = 1$, $c^n = MIN(c1, c2, \ldots, cn)$, přičemž ci je minimální KU z rezolvent R^i."

Mukaidonům fuzzy rezoluční princip se také často používá na vytváření či odvozování nových formulí ze stávajících axiomů a taktéž při výpočtech koeficientů určitosti rezolventy těchto formulí. Rezoluční množina pak obsahuje dané axiomy a samotnou negovanou formuli.

Jako příklad odvozovacích pravidel v binární logice se uvádí v [5] následující:

modus ponenc

$$\text{IF } A \text{ and } A \Rightarrow B \text{ THEN } B, \tag{13a}$$

modus tollens

$$\text{IF } \neg B \text{ and } A \Rightarrow B \text{ THEN } \neg A, \tag{13b}$$

disjunktivní sylogizmus

$$\text{IF } A \vee B \text{ and } \neg A \text{ THEN } B, \tag{13c}$$

hypotetický sylogismus

$$\text{IF } A \Rightarrow B \text{ } \textit{and} \text{ } B \Rightarrow C \text{ THEN } A \Rightarrow C, \tag{13d}$$

konstruktivní dilema

$$\text{IF } (A \Rightarrow B) \wedge (C \Rightarrow D) \text{ } \textit{and} \text{ } A \vee C \text{ THEN } B \vee D, \tag{13e}$$

destruktivní dilema

$$\text{IF } (A \Rightarrow B) \wedge (C \Rightarrow D) \text{ } \textit{and} \text{ } \neg B \vee \neg D \text{ THEN } \neg A \vee \neg C. \tag{13f}$$

Tyto inferenční pravidla je možné použít k odvozování, na základě fuzzy interpretace.

Výše uvedený popis fuzzy množin a výrokové fuzzy logiky je zde uveden z důvodu ulehčení pochopení a dalšího popisu problému, kterým je rovněž analýza občanské vybavenosti regionu, tentokráte však s využitím metody zakládající se právě na fuzzy problematice a fuzzy asociačních pravidlech.

3.5. Fuzzy asociační pravidla

Fuzzy asociační pravidla jsou generována na základě fuzzy množin a fuzzy logiky, které byly osvětleny v předcházejících kapitolách. Nejprve vždy dochází k fuzzifikačnímu procesu, kde jsou ostré hodnoty převedeny na hodnoty fuzzy. Poté, dle použitého algoritmu, jsou generována vlastní asociační pravidla, či jsou provedeny další výpočty, obsahující například i výpočet fuzzy support value (hodnota fuzzy podpory) nebo fuzzy confidence value (hodnota fuzzy spolehlivosti). Opět je zde možné, nebo spíš nutné, vygenerovaná pravidla pomocí stanovených kritérií filtrovat a získat tak pravidla, které jsou použitelná.

V dalším textu bude použita jako příklad jedné z metod získávání fuzzy asociačních pravidel metoda publikovaná ve článku Mining fuzzy association rules for web access case adaptation[12]. Tento článek se zabývá problematikou webu a přístupu na něj, což má podle autorů zásadní vliv například na internetový obchod. Metoda se nazývá Fuzzy Association Rule Discovery, což je možno volně přeložit jako objevování fuzzy asociačních pravidel. Vlastní postup metody je tedy převzat z [7], avšak zvolené téma týkající se občanské vybavenosti i veškerá vstupní data jsou původní a v souladu s předešlým příkladem na „klasická“ asociační pravidla. Záměrně byla použita stejná vstupní data z důvodu následného porovnání uvedených metod.

[12] Autoři jsou Cody Wong, Simon Shiu (oba z Department of Computing, Hong Kong Polytechnic University) a Ankar Pal (Distinguished Scientist & Head Maschine inteligence Unit, Indian Statistical Institute) [7].

3.6. Analýza občanské vybavenosti regionu pomocí fuzzy asociačních pravidel

První ze zde uvedených metod generování fuzzy asociačních pravidel je, jak jsem již uvedl výše, metoda Fuzzy Association Rule Discovery. Prvním krokem je tedy fuzzifikační proces.

3.6.1.Fuzzifikační proces

Fuzzifikační proces je převod ostrých hodnot na fuzzy hodnoty. Vlastní princip tohoto procesu byl již popsán v kapitole číslo 3.2. na příkladu malé, střední, či velké sklenice. V celém tomto příkladě se opět bude vycházet ze vstupních dat, která byla použita již v prvním příkladě analýzy občanské vybavenosti v kapitole 2.9.. Tato data měla stejný tvar pro každé z uvažovaných krajských měst (Brno, České Budějovice, …), jednotlivé položky jsou zřejmé z tabulky číslo 23,

Tabulka 23: Tabulka vstupních dat pro město Brno, zdroj: [vlastní].

	Organizace	**Počet**
Brno	Divadlo	14
	Jesle	3
	Krematorium	1
	Lázeňské léčebny	0
	Letecká a kosmická doprava	2
	Multikino	1
	Nemocnice	11
	Ústavy sociální péče pro dospělé	2
	Vodní doprava	0
	Vysoká škola	9
	Zoologická zahrada	1

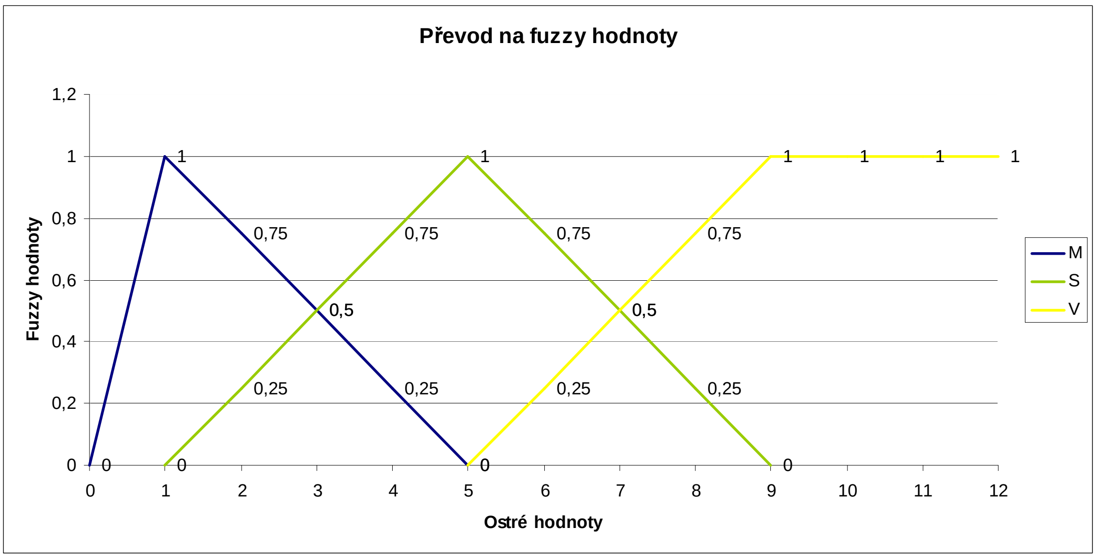
V tomto případě však budou použity nejen údaje o přítomnosti či nepřítomnosti dané organizace v krajském městě, ale i jejich konkrétní počet, který bude převeden na fuzzy hodnotu. Přehledná tabulka s počtem všech organizací v daných městech viz. tabulka číslo 24,

Tabulka 24: Souhrnná tabulka vstupních dat, zdroj: [vlastní].

Město	**Organizace**										
	Divadlo	Jesle	Krematorium	Lázně	Letecká dop.	Multikino	Nemocnice	Ústav soc. péče	Vodní dop.	Vysoká škola	ZOO
Brno	14	3	1	0	2	1	11	2	0	9	1
České Budějovice	5	1	1	0	3	1	1	0	3	2	0
Hradec Králové	2	1	0	0	0	1	2	0	0	1	0
Jihlava	2	1	1	0	0	0	1	0	0	1	1
Karlovy Vary	9	0	1	40	0	0	1	0	1	1	0
Liberec	4	0	1	0	0	0	1	1	2	1	1
Olomouc	2	1	1	0	1	1	2	1	0	2	1
Ostrava	6	0	1	0	0	1	3	1	1	3	1
Pardubice	4	1	1	0	1	1	3	0	4	1	0
Plzeň	10	2	1	0	2	1	4	3	0	2	1
Ústí nad Labem	2	3	2	0	1	0	1	0	7	2	1
Zlín	2	2	1	1	0	0	2	1	0	1	1
SUMA	62	15	12	41	10	7	32	9	18	26	8

V tabulce jsou uvedena v řádcích veškerá krajská města a ve sloupcích jednotlivé organizace. Údaje zde uvedené tedy vyjadřují počet výskytu organizace v daném městě. Příkladem tedy může být, že v Brně je 14 divadel. Těmto hodnotám se pak říká ostré hodnoty, tj. hodnoty z klasické teorie množin.

Tyto ostré hodnoty je nyní třeba převést na fuzzy hodnoty následujícím způsobem. Nejprve je nutné určit počet a název fuzzy proměnných. V tomto případě se bude jednat o hodnoty určující počet organizací v daném městě. Tedy počet malý, zkráceně jen M, počet střední, zkráceně S, a počet velký, zkráceně V. Vše je pak zřejmé z následujícího grafu číslo 1



Graf 1: Fuzzy hodnoty, zdroj: [vlastní].

Z tohoto grafu je tedy vidět, jaké hodnoty funkce příslušnosti (fuzzy hodnoty) přísluší ostrým hodnotám počtu jednotlivých organizací. Jsou zde znázorněny tři fuzzy množiny, potažmo funkce příslušnosti: Množina M, která je znázorněna modrou barvou, množina S, zelenou barvou a množina V, která je znázorněna barvou žlutou a její neklesající tendence od bodu 9 na ose y znamená, že každá větší hodnota než 9 spadá do této množiny. Hodnoty z grafu jsou také přehledně viditelné v tabulce číslo 25,

Tabulka 25: Převodová tabulka fuzzy hodnot, zdroj: [vlastní],

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12 - ∞**
M	0	1	0,75	0,5	0,25	0	0	0	0	0	0	0	0
S	0	0	0,25	0,5	0,75	1	0,75	0,5	0,25	0	0	0	0
V	0	0	0	0	0	0	0,25	0,5	0,75	1	1	1	1

která obsahuje opět tři uvedené množiny M, S a V a to s hodnotami funkce příslušnosti pro jednotlivé ostré hodnoty (uvedené v prvním řádku tabulky). Jen pro úplnost je ještě jednou nutné podotknout, že zmíněné ostré hodnoty jsou počty výskytů jednotlivých organizací ve městech.

Jako příklad je možné uvést počet divadel v Liberci (Nyní již budou používány i zkratky pro organizace uvedené v kapitole 2.9. tj. například Divadlo (Div.), Jesle (Jes.), či Krematorium (Krem.)). Tento počet, tedy ostrá hodnota, je čtyři, což dle uvedeného grafu číslo 1 a tabulky číslo 25 (společně korespondující) znamená následující fuzzy hodnotu.

Divadlo Liberec = 4 => Div.M 0,25 a Div.S 0,75

V zásadě to vypovídá o tom, že počet divadel v Liberci je Malý s hodnotou funkce příslušnosti 0,25 a střední s hodnotou 0,75. Tímto způsobem je potřeba převést z ostrých na fuzzy hodnoty veškeré původní údaje. Tyto hodnoty jsou pak znázorněny v následující tabulce číslo 26.[13]

[13] Kompletní tabulku se všemi údaji je možno nalézt v příloze číslo 3 na straně III.

Tabulka 26: Tabulka fuzzy hodnot jednotlivých organizací ve městech, zdroj: [vlastní].

Město	**Organizace**								
	Div	Div	Div	Jes	Jes	Jes	Krem	Krem	Krem
	M	S	V	M	S	V	M	S	V
Brno	0	0	1	0,5	0,5	0	1	0	0
České Budějovice	0	1	0	1	0	0	1	0	0
Hradec Králové	0,75	0,25	0	1	0	0	0	0	0
Jihlava	0,75	0,25	0	1	0	0	1	0	0
Karlovy Vary	0	0	1	0	0	0	1	0	0
Liberec	0,25	0,75	0	0	0	0	1	0	0
Olomouc	0,75	0,25	0	1	0	0	1	0	0
Ostrava	0	0,75	0,25	0	0	0	1	0	0
Pardubice	0,25	0,75	0	1	0	0	1	0	0
Plzeň	0	0	1	0,75	0,25	0	1	0	0
Ústí nad Labem	0,75	0,25	0	0,5	0,5	0	0,75	0,25	0
Zlín	0,75	0,25	0	0,75	0,25	0	1	0	0

V této tabulce jsou uvedeny pouze tři první organizace a to divadlo, jesle a krematorium. Ke každé organizaci jsou zde pak uvedeny fuzzy hodnoty pro M, S a V počet.

3.6.2.Vytváření fuzzy asociačních pravidel

Jsou-li ostré hodnoty převedeny na fuzzy hodnoty, pak je již možné generovat fuzzy asociační pravidla. Tato pravidla mají opět známý IF-THEN tvar:

IF X THEN Y,

kde X a Y jsou jednotlivé položky transakce, v tomto případě konkrétně prvky organizací, například Div.M, Div.S, Div.V, Jes.M, … Tato pravidla vyjadřují přítomnost jedné organizace Y za předpokladu přítomnosti jiné organizace, a to obecně ve všech městech, tj. ve všech transakcích. IF-THEN pravidla je možné generovat z tabulky 26, a to kombinacemi daných organizací a jejich velikostí, respektive kombinacemi jednotlivých sloupců tabulky. Názorný příklad fuzzy asociačních pravidel viz tabulka číslo 27,

Tabulka 27: Příklady fuzzy asociačních pravidel, zdroj: [vlastní].

Fuzzy asociační pravidla
IF Div.M THEN Jes.M
IF Div.M THEN Jes.S
IF Div.M THEN Jes.V
IF Div.M THEN Krem.M
IF Div.M THEN Krem.S
IF Div.M THEN Krem.V
IF Div.M THEN Láz.M
IF Div.M THEN Láz.S
IF Div.M THEN Láz.V
IF Div.M THEN Let.M
IF Div.M THEN Let.S
IF Div.M THEN Let.V
IF Div.M THEN Mul.M
IF Div.M THEN Mul.S
IF Div.M THEN Mul.V
IF Div.M THEN Nem.M
IF Div.M THEN Nem.S
IF Div.M THEN Nem.V
IF Div.M THEN USP.M
IF Div.M THEN USP.S
IF Div.M THEN USP.V
IF Div.M THEN Vod.M
IF Div.M THEN Vod.S
IF Div.M THEN Vod.V
IF Div.M THEN VS.M
IF Div.M THEN VS.S
IF Div.M THEN VS.V
IF Div.M THEN ZOO.M
IF Div.M THEN ZOO.S
IF Div.M THEN ZOO.V

Tabulka číslo 27 obsahuje veškerá fuzzy asociační pravidla pro antecedent Div.M. Pro ostatní prvky organizací jsou pravidla tvořena analogicky. Je zřejmé, že tato pravidla nejsou v plné míře dost dobře použitelná a proto je vždy nutné pravidla nějakým způsobem redukovat tak, aby zbyla pouze taková pravidla, která mají co možná nejvyšší informační hodnotu, tedy taková pravidla, která jsou dále použitelná a mají nějakým způsobem zajímavý charakter. Je proto vhodné stanovit nějaká kritéria, podle kterých budou pravidla redukována.

V tomto případě bude použita tzv. Fuzzy Support Value (Fuzzy Podpora) a Fuzzy Confidence Value (Fuzzy Spolehlivost).

3.6.3.Fuzzy podpora

K výběru použitelných fuzzy asociačních pravidel je nejprve nutný výpočet hodnoty fuzzy podpory těchto pravidel. Fuzzy podpora je jednou ze dvou omezujících podmínek pro výběr redukovaných fuzzy asociačních pravidel. Je tedy nutné tuto hodnotu nejprve vypočítat pro všechny kombinace prvků (items) ze kterých se fuzzy asociační pravidla skládají a to ve všech transakcích, tedy ve všech krajských městech. V tomto případě se pak jedná o prvky „organizace“, tj. například Div.M, Jes.S či Krem.V.

Fuzzy support value se podle autorů článku, Mining fuzzy association rules for web access case adaptation, vypočítává následujícím způsobem [7]:

$$FS_{\langle X,A \rangle} = \frac{\sum_{ui \in \cup \pi xj \in x} (a_j \in A, t_i * x_j)}{|C_s|}.$$

kde v čitateli je skalární součin všech hodnot jednotlivých dvojic items (organizací) a ve jmenovateli je pak počet těchto dvojic hodnot. Jako příklad výpočtu je možno uvést příklad na výpočet fuzzy podpory následujících dvou organizací Div.M a Jes.M. V tabulce číslo 28 jsou uvedené hodnoty, ze kterých se bude fuzzy podpora dále vypočítávat,

Tabulka 28: Tabulka hodnot pro výpočet fuzzy podpory, zdroj:[vlastní].[14]

Div	**Jes**
M	**M**
0	0,5
0	1
0,75	1
0,75	1
0	0
0,25	0
0,75	1
0	0
0,25	1
0	0,75
0,75	0,5
0,75	0,75

[14] Tyto hodnoty je také možno nalézt v tabulce číslo 26 (Tabulka fuzzy hodnot jednotlivých organizací ve městech), odkud jsou zkopírovány.

FS (Fuzzy support) = (0*0,5 + 0*1 + 0,75*1 + 0,75*1 + 0*0 + 0,25*0 + 0,75*1 + 0*0 + + 0,25*1 + 0*0,75 + 0,75*0,5 + 0,75*0,75) / 12 = **0,29**

Hodnota fuzzy podpory pro dvojici Div.M a Jes.M, neboli fuzzy asociačního pravidla

IF Div.M THEN Jes.M

se tedy rovná 0,29. Tato hodnota fuzzy podpory je na základě komutativnosti prvků skalárního součinu v čitateli použitelná i pro tuto dvojici v obráceném pořadí, tj. Jes.M a Div.M. Tuto skutečnost zde uvádím proto, protože fuzzy asociační pravidlo

IF Div.M THEN Jes.M a pravidlo IF Jes.M THEN Div.M

nejsou totožná.

Přehled některých hodnot fuzzy podpory je uveden v následující tabulce s číslem 29,

Tabulka 29: Tabulka vybraných hodnot fuzzy podpory, zdroj: [vlastní].

		Div	Div	Div	Jes	Jes	Jes	Krem	Krem	Krem	Láz	Láz	Láz	Let	Let	Let
		M	S	V	M	S	V	M	S	V	M	S	V	M	S	V
Div	M				0,29	0,05	0,00	0,28	0,02	0,00	0,06	0,00	0,00	0,15	0,00	0,00
Div	S				0,23	0,02	0,00	0,35	0,01	0,00	0,02	0,00	0,00	0,15	0,04	0,00
Div	V				0,10	0,06	0,00	0,27	0,00	0,00	0,00	0,00	0,08	0,13	0,04	0,00
Jes	M							0,53	0,01	0,00	0,06	0,00	0,00	0,33	0,07	0,00
Jes	S							0,11	0,01	0,00	0,02	0,00	0,00	0,09	0,02	0,00
Jes	V							0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Krem	M										0,08	0,00	0,08	0,40	0,08	0,00
Krem	S										0,00	0,00	0,00	0,02	0,00	0,00
Krem	V										0,00	0,00	0,00	0,00	0,00	0,00
Láz	M													0,00	0,00	0,00
Láz	S													0,00	0,00	0,00
Láz	V													0,00	0,00	0,00
Let	M															
Let	S															
Let	V															

Po samotném výpočtu již stačí jen stanovit prahovou hodnotu fuzzy podpory, která byla v tomto případě zvolena na hladině **0,3**. Tento práh pak bude dále sloužit k redukci fuzzy asociačních pravidel.

3.6.4.Fuzzy spolehlivost

Dalším a v tomto případě i posledním prvkem, který slouží k redukci fuzzy asociačních pravidel je tzv. Hodnota fuzzy spolehlivosti (Fuzzy Confidence Value). Vychází z výpočtu skalárního součinu všech hodnot jednotlivých dvojic items (organizací), tak jak tomu bylo v předchozím výpočtu fuzzy podpory, avšak ve jmenovateli se tentokrát nachází prostý součet všech hodnot prvního prvku dvojice (antecedentu pravidla). Její výpočet opět pro dvojici Div.M a Jes.M neboli fuzzy asociační pravidlo

IF Div.M THEN Jes.M

je následující.

FC (Fuzzy confidence) = (0*0,5 + 0*1 + 0,75*1 + 0,75*1 + 0*0 + 0,25*0 + 0,75*1 + 0*0 + + 0,25*1 + 0*0,75 + 0,75*0,5 + 0,75*0,75) / (0 + 0 + 0,75 + 0,75 + 0 + 0,25 + 0,75 + 0 + + 0,25 + 0 + 0,75 + 0,75) = **0,81**.

Pro pravidlo IF Jes.M THEN Div.M je výpočet prováděn analogicky. Tzn., že čitatel zůstane nezměněn a jmenovatel bude obsahovat součet všech hodnot prvního prvku dvojice tj. Jes.M. Proto se bude výsledek ve většině případů lišit a je proto nutné tuto hodnotu fuzzy spolehlivosti na rozdíl od fuzzy podpory počítat pro „obrácená“ pravidla ještě jednou. Pro dvojici Jes.M, Div.M je pak výsledkem hodnota **0,46**. V podstatě to znamená, že fuzzy asociační pravidlo IF Div.M THEN Jes.M má větší spolehlivost, než pravidlo IF Jes.M THEN Div.M.

Část již vypočítaných hodnot je možné vidět v tabulce 30,

Tabulka 30: Tabulka vybraných hodnot fuzzy spolehlivosti, zdroj: [vlastní].

		Div	Div	Div	Jes	Jes	Jes	Krem	Krem	Krem	Láz	Láz	Láz	Let	Let	Let
		M	S	V	M	S	V	M	S	V	M	S	V	M	S	V
Div	M				0,81	0,13	0,00	0,78	0,04	0,00	0,18	0,00	0,00	0,41	0,00	0,00
Div	S				0,63	0,04	0,00	0,93	0,01	0,00	0,06	0,00	0,00	0,39	0,11	0,00
Div	V				0,38	0,23	0,00	1,00	0,00	0,00	0,00	0,00	0,31	0,46	0,15	0,00
Jes	M							0,85	0,02	0,00	0,10	0,00	0,00	0,53	0,11	0,00
Jes	S							0,92	0,08	0,00	0,17	0,00	0,00	0,71	0,13	0,00
Jes	V							0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Krem	M										0,09	0,00	0,09	0,44	0,09	0,00
Krem	S										0,00	0,00	0,00	1,00	0,00	0,00
Krem	V										0,00	0,00	0,00	0,00	0,00	0,00
Láz	M													0,00	0,00	0,00
Láz	S													0,00	0,00	0,00
Láz	V													0,00	0,00	0,00
Let	M															
Let	S															
Let	V															

Opět je zde nutné, ze stejného důvodu jako tomu bylo u hodnoty fuzzy podpory, stanovit prahovou hodnotu fuzzy spolehlivosti. Ta je stanovena na hladině **0,6**.

3.6.5.Výsledná fuzzy asociační pravidla

Posledním krokem této metody je vlastní výběr, či redukce počtu fuzzy asociačních pravidel. Vyhledají se taková pravidla z množiny všech pravidel, která splňují obě výše popsané podmínky. tj. hladina fuzzy podpory musí být vyšší než 0,3 a hladina fuzzy spolehlivosti zase vyšší než hodnota 0,6. Výčet pravidel splňujících uvedené podmínky je možné vidět v tabulce číslo 31.

Tabulka 31: Tabulka fuzzy asociačních pravidel, zdroj: [vlastní].

FUZZY ASOCIAČNÍ PRAVIDLA	Podpora	Spolehlivost
IF Div.M THEN VS.M	0,32	0,91
IF Div.S THEN Krem.M	0,35	0,93
IF Div.S THEN VS.M	0,31	0,83
IF Jes.M THEN Krem.M	0,53	0,85
IF Jes.M THEN Mul.M	0,44	0,70
IF Jes.M THEN Nem.M	0,44	0,70
IF Jes.M THEN VS.M	0,52	0,83
IF Krem.M THEN Nem.M	0,63	0,70
IF Krem.M THEN VS.M	0,69	0,77
IF Krem.M THEN ZOO.M	0,65	0,72
IF Let.M THEN Mul.M	0,33	0,80
IF Mul.M THEN VS.M	0,40	0,68
IF Nem.M THEN VS.M	0,63	0,88
IF Nem.M THEN ZOO.M	0,44	0,62
IF USP.M THEN VS.M	0,30	0,69
IF USP.M THEN ZOO.M	0,44	1,00
IF VS.M THEN ZOO.M	0,48	0,61
IF Let.M THEN Jes.M	0,33	0,79
IF Mul.M THEN Jes.M	0,44	0,75
IF Nem.M THEN Jes.M	0,44	0,62
IF VS.M THEN Jes.M	0,52	0,65
IF Let.M THEN Krem.M	0,40	0,95
IF Mul.M THEN Krem.M	0,50	0,86
IF Nem.M THEN Krem.M	0,63	0,88
IF USP.M THEN Krem.M	0,44	1,00
IF VS.M THEN Krem.M	0,69	0,88
IF ZOO.M THEN Krem.M	0,65	0,97
IF VS.M THEN Nem.M	0,63	0,79
IF ZOO.M THEN Nem.M	0,44	0,66
IF ZOO.M THEN USP.M	0,44	0,66
IF ZOO.M THEN VS.M	0,48	0,72

Konkrétně je tedy možné popsat například první z fuzzy asociačních pravidel, tedy pravidlo IF Div.M THEN VS.M. Toto pravidlo má fuzzy podporu 0,32 a fuzzy spolehlivost 0,91. To znamená že je splněn požadavek na minimální hladinu obou hodnot, která je u fuzzy podpory 0,3 a u fuzzy spolehlivosti 0,6 a pravidlo je tak na této hladině použitelné (má tedy význam ho využít při dalším průzkumu či analýze). Rozhodne-li se však řešitel problému zvednout hranici podpory na 0,4, pak se toto pravidlo stává pro další použití nevhodným.

4. Porovnání (komparace) uvedených metod

Postupy výše uvedených metod získávání asociačních a fuzzy asociačních pravidel jsou zřejmé z vlastních příkladů, které jsou obsažené v příslušných kapitolách. Jednou z nejvíce odlišných vlastností je u metody získávání fuzzy asociačních pravidel, oproti klasickým asociačním pravidlům, převod ostrých čísel na fuzzy hodnoty (hodnoty funkce příslušnosti), což také ovlivňuje výsledky daných metod. V této části práce budou tedy porovnány zmíněné výsledky obou metod a to jak na hladině, v příkladech uvedených kritérií, tak na hladině stejných hodnot kritérií.

4.1. Porovnání na stejné hladině hodnot kritérií

V prvním příkladě v kapitole 2.9. byla vygenerována veškerá asociační pravidla, která byla následně omezena kritérii:

$$\text{Podpora} \geq 50\% \qquad \text{Spolehlivost} \geq 70\%$$

Výsledný efekt byl pak snížení počtu asociačních pravidel na cca 40. Tato pravidla je možné vidět v tabulce číslo 21 v příkladě analýzy občanské vybavenosti regionu pomocí MBA.

Pokud se však shodné hodnoty kritérií, tedy

$$\text{Fuzzy\_Podpora} \geq 50\% \qquad \text{Fuzzy\_Spolehlivost} \geq 70\%$$

použijí při analýze stejného problému a se stejnými vstupními daty pomocí metody generování fuzzy asociačních pravidel, pak počet pravidel, které odpovídají těmto kritériím, bude pouze cca 10, tedy zhruba jen jedna čtvrtina.

V následující tabulce číslo 32 je přehled právě těch fuzzy asociačních pravidel, která daným kritériím vyhovují. Za povšimnutí stojí také fakt, že všechny hodnoty, které jsou obsaženy ve výsledných pravidlech, jsou s postfixem M, tedy malý výskyt organizace v regionu,

Tabulka 32: Tabulka výsledných fuzzy asociačních pravidel, zdroj: [vlastní].

FUZZY ASOCIAČNÍ PRAVIDA	Fuzzy podpora	Fuzzy spolehlivost
IF Jes.M THEN VS.M	0,52	0,83
IF Krem.M THEN VS.M	0,69	0,77
IF Krem.M THEN ZOO.M	0,65	0,72
IF Nem.M THEN Krem.M	0,63	0,88
IF Nem.M THEN VS.M	0,63	0,88
IF VS.M THEN Krem.M	0,69	0,88
IF VS.M THEN Nem.M	0,63	0,79
IF ZOO.M THEN Krem.M	0,65	0,97
IF Jes.M THEN Krem.M	0,53	0,85

Porovnáním výsledných pravidel v obou metodách, se dá tedy zjistit, která pravidla jsou uvedena jak v prvním, tak v druhém případě. Odpovídající asociační pravidla jsou uvedena v tabulce číslo 33,

Tabulka 33: Tabulka odpovídajících asociačních pravidel, zdroj: [vlastní].

ASOCIAČNÍ PRAVIDLA	Podpora (%)	Spolehlivost (%)
IF Jes THEN VS	75,00	100,00
IF Krem THEN VS	91,67	100,00
IF Krem THEN ZOO	66,67	72,73
IF Nem THEN Krem	91,67	91,67
IF Nem THEN VS	100,00	100,00
IF VS THEN Krem	91,67	91,67
IF VS THEN Nem	100,00	100,00
IF ZOO THEN Krem	66,67	100,00

Tato tabulka obsahuje hodnoty podpory a spolehlivosti v procentech (%) a proto i tabulka fuzzy asociačních pravidel bude na procenta převedena. V tabulce číslo 34 jsou tedy vygenerovaná fuzzy acociační pravidla s hodnotami fuzzy podpory a fuzzy spolehlivosti v procentech (%),

Tabulka 34: Tabulka odpovídajících fuzzy asociačních pravidel, zdroj: [vlastní].

FUZZY ASOCIAČNÍ PRAVIDLA	Fuzzy Podpora (%)	Fuzzy Spolehlivost (%)
IF Jes.M THEN VS.M	51,56	82,50
IF Krem.M THEN VS.M	69,27	77,33
IF Krem.M THEN ZOO.M	64,58	72,09
IF Nem.M THEN Krem.M	62,50	88,24
IF Nem.M THEN VS.M	62,50	88,24
IF VS.M THEN Krem.M	69,27	87,50
IF VS.M THEN Nem.M	62,50	78,95
IF ZOO.M THEN Krem.M	64,58	96,88

Z výše uvedených tabulek číslo 33 a 34 je zřejmé, že pravidla, která se zde vyskytují, jsou si nějakým způsobem „podobná“. Například první z pravidel

IF Jes THEN VS a IF Jes.M THEN VS.M

vyjadřují fakt, že pokud ve zvolených regionech (transakcích) jsou obsaženy jesle, pak je zde obsažena i vysoká škola, a to s uvedenou podporou a spolehlivostí. Pravidlo však již neplatí opačně, tedy pokud je v regionech obsažena vysoká škola, nemusí zde nutně být jesle.

Je-li porovnán počet asociačních a fuzzy asociačních pravidel, pak je možné dojít k závěru, že pouze čtvrtinový počet fuzzy asociačních pravidel vede k větší přesnosti těchto pravidel. Je to způsobeno tím faktorem, že fuzzy asociační pravidla (fuzzy množiny) mají v tomto případě vlastnost rozlišovat velikosti daných položek. To také znamená, že klasická asociační pravidla mají zde větší míru obecnosti a neberou tak v potaz rozdělení položky (organizace) do dalších dílčích skupin[15]. Tento fakt je založen na rozdílu mezi klasickou a fuzzy teorií množin.

Pravidla, která nejsou v tomto porovnání obsažena, avšak jsou výsledkem jedné z metod získávání pravidel, jsou samozřejmě také platná. V tomto případě lze také tvrdit, že výsledná fuzzy asociační pravidla jsou na jiné úrovni obecnosti také obsažena ve výsledných asociačních pravidlech. To je také důvodem, proč je zde asociačních pravidel více.

[15] Neobsahují tedy funkce příslušnosti a jednotlivé organizace nejsou děleny na M, S a V

4.2. Porovnání na různé hladině hodnot kritérií

V předchozím případě (kapitola 4.1.) byla zvolena stejná hladina hodnot kritérií. Důsledkem toho bylo zjištění, že počet fuzzy asociačních pravidel je čtvrtinový oproti klasickým asociačním pravidlům. Pro zvýšení počtu fuzzy asociačních pravidel je nutné snížit hodnoty fuzzy podpory a fuzzy spolehlivosti, tak, jak tomu bylo v příkladě analýzy občanské vybavenosti pomocí fuzzy asociačních pravidel v kapitole 3.6.. Nyní bude tedy opět zvolen práh obou podmínek tak, aby hodnota fuzzy podpory byla vyšší než 0,3 a hladina fuzzy spolehlivosti zase vyšší než hodnota 0,6. Protože tyto hodnoty jsou totožné s výše uvedeným příkladem, je možné vycházet právě z výsledků uvedených v tabulce číslo 31, kde je již počet nalezených fuzzy asociačních pravidel cca. 30. Došlo tedy k nárůstu pravidel řádově o dvě desítky při snížení hodnot fuzzy podpory z hodnoty 0,5 na 0,3 a fuzzy spolehlivosti z hodnoty 0,7 na 0,6.

Tento počet se přeci jen o něco více blíží počtu klasických asociačních pravidel, při zachování hladiny podpory i spolehlivosti na úrovni 0,5 a 0,7, jak tomu bylo v původním příkladu kapitoly 2.9.. Při uvedených hladinách hodnot kritérií je tedy možné vycházet nadále z výsledných tabulek obou příkladů, tedy z tabulky číslo 21 a 31.

V následující tabulce číslo 35 jsou uvedena taková asociační pravidla, která jsou v tomto případě v obou metodách srovnatelná.

Tabulka 35: Tabulka odpovídajících asociačních pravidel 2, zdroj: [vlastní].

ASOCIAČNÍ PRAVIDLA	Podpora (%)	Spolehlivost (%)
IF Div THEN Krem	91,67	91,67
IF Div THEN VS	100,00	100,00
IF Jes THEN Nem	75,00	100,00
IF Jes THEN VS	75,00	100,00
IF Krem THEN Nem	91,67	100,00
IF Krem THEN VS	91,67	100,00
IF Krem THEN ZOO	66,67	72,73
IF Let THEN Jes	50,00	100,00
IF Let THEN Krem	50,00	100,00
IF Mul THEN Jes	50,00	85,71
IF Mul THEN VS	58,33	100,00
IF Nem THEN Jes	75,00	75,00
IF Nem THEN Krem	91,67	91,67
IF Nem THEN VS	100,00	100,00
IF USP THEN Krem	50,00	100,00
IF USP THEN VS	50,00	100,00
IF USP THEN ZOO	50,00	100,00
IF VS THEN Jes	75,00	75,00
IF VS THEN Krem	91,67	91,67
IF VS THEN Nem	100,00	100,00
IF ZOO THEN Krem	66,67	100,00
IF ZOO THEN Nem	66,67	100,00
IF ZOO THEN USP	50,00	75,00
IF ZOO THEN VS	66,67	100,00

V tabulce číslo 35 jsou tedy taková asociační pravidla, která je v jiné formě možné nalézt i v tabulce číslo 31, tedy ve výsledné tabulce fuzzy asociačních pravidel. Následuje tabulka odpovídajících fuzzy asociačních pravidel,[16]

[16] V tabulce jsou opět převedeny hodnoty fuzzy podpory i spolehlivosti na procenta.

Tabulka 36: Tabulka odpovídajících fuzzy asociačních pravidel 2, zdroj: [vlastní].

FUZZY ASOCIAČNÍ PRAVIDLA	**Fuzzy Podpora (%)**	**Fuzzy Spolehlivost (%)**
IF Div.M THEN VS.M	32,29	91,18
IF Div.S THEN Krem.M	34,90	93,06
IF Jes.M THEN Nem.M	43,75	70,00
IF Jes.M THEN VS.M	51,56	82,50
IF Krem.M THEN Nem.M	62,50	69,77
IF Krem.M THEN VS.M	69,27	77,33
IF Krem.M THEN ZOO.M	64,58	72,09
IF Let.M THEN Jes.M	32,81	78,75
IF Let.M THEN Krem.M	39,58	95,00
IF Mul.M THEN Jes.M	43,75	75,00
IF Mul.M THEN VS.M	39,58	67,86
IF Nem.M THEN Jes.M	43,75	61,76
IF Nem.M THEN Krem.M	62,50	88,24
IF Nem.M THEN VS.M	62,50	88,24
IF USP.M THEN Krem.M	43,75	100,00
IF USP.M THEN VS.M	30,21	69,05
IF USP.M THEN ZOO.M	43,75	100,00
IF VS.M THEN Jes.M	51,56	65,13
IF VS.M THEN Krem.M	69,27	87,50
IF VS.M THEN Nem.M	62,50	78,95
IF ZOO.M THEN Krem.M	64,58	96,88
IF ZOO.M THEN Nem.M	43,75	65,63
IF ZOO.M THEN USP.M	43,75	65,63
IF ZOO.M THEN VS.M	47,92	71,88

Porovnáním výsledků obou metod, je tedy možné dojít k závěru, že snížením hodnot kritérií fuzzy podpory a fuzzy spolehlivosti je dosaženo nejen zvýšení počtu fuzzy asociačních pravidel, ale také zvýšení počtu pravidel (trojnásobně), která mají podobný význam, a to jak u klasických, tak u fuzzy asociačních pravidel.

Výše uvedené taktéž vede k závěru, že veškeré hodnoty fuzzy podpory a fuzzy spolehlivosti jsou vždy nižší než tytéž hodnoty u odpovídajících asociačních pravidel.

Závěr

Text této diplomové práce je zaměřen na v každé době aktuální a neustále se vyvíjející problém analýzy občanské vybavenosti regionu. Tato problematika je zde pak řešena jak pomocí asociačních, tak fuzzy asociačních pravidel.

Úvodní kapitola je věnována pojmům jako je region, hranice regionu, jeho struktura či charakter. Volně také navazuje územně správní (administrativní) členění ČR a Německa jako názorný příklad a konečně také vlastní občanská vybavenost, kde je použit popis konkrétní občanské vybavenosti ve městě Bílina.

Druhá kapitola obsahuje již nejen definici asociačních pravidel a jejich vlastností, ale také vytváření kombinací, zobecnění asociačních pravidel, různé druhy vyhledávání pravidel či popis některých metod a algoritmů, které jsou ke generování pravidel často používané. Završením této kapitoly je samozřejmě vlastní příklad analýzy občanské vybavenosti pomocí metody MBA (analýza nákupního košíku), která přináší již konkrétní výsledky v podobě vygenerovaných asociačních pravidel.

Následující část osvětluje samotnou fuzzy problematiku, zejména pak fuzzy množiny, porovnání klasických a fuzzy množin, výrokovou fuzzy logiku a v neposlední řadě také fuzzy asociační pravidla. Na základě získaných informací z této oblasti je pak v závěru kapitoly opět konkrétní příklad na analýzu občanské vybavenosti se stejnými vstupními daty, ale tentokrát pomocí fuzzy asociačních pravidel. Protože obě výše uvedené metody nemají stejné výsledky, je v závěrečné kapitole zpracováno jejich porovnání (komparace), a to nejprve na stejné hladině hodnot omezujících podmínek, tedy podpory a spolehlivosti, respektive fuzzy podpory a fuzzy spolehlivosti, a poté následuje porovnání na snížené hladině fuzzy omezujících podmínek.

Pevně věřím, že se v textu podařilo splnit veškeré v úvodu vytyčené cíle. Po shlédnutí této práce by měl mít čtenář tedy již konkrétní představu o pojmu region, jeho vlastnostech a občanské vybavenosti. Dále pak o asociačních pravidlech a metodách generování těchto pravidel, a také samozřejmě o fuzzy problematice a fuzzy asociačních pravidlech. Oba zde uvedené příklady by pak měly podtrhnout rozdíl mezi asociačními pravidly, fuzzy asociačními pravidly a metodami, kterými byla tato pravidla generována.

Seznam literatury

[1] Berka, P., *Dobývání znalostí z databází*. Praha: Academia, 2003. ISBN 80-200-1062-9

[2] Chytil, M., *Automaty a gramatiky*. 1. vydání. Praha: SNTL, 1984. Typové číslo L11-E1-IV-41F/11 878

[3] Hájek, P., Havránek, T., *Mechanizing Hypothesis Formation: Mathematical Foundations for a General Theory*. Berlin, Heidelberg, New York: Springer-Verlag, 1976. ISBN 3-540-08738-9

[4] Kodratoff, Y., *Knowledge extraction form texts: Knowledge Discovery and Data Mining*. PAKDD2000. Kjoto, 2000

[5] Olej, V., Petr, P., *Expertní systémy*. Pardubice: Univerzita Pardubice, 1997

[6] Olej, V., Petr, P., *Umělá a výpočetní inteligence: část: Fuzzy množiny – distanční opora*. Pardubice: Univerzita Pardubice, 2004

[7] Wong, C., Shiu, S., Pal, S., *Mining fuzzy association rules for web access case adaptation*. Department of Computing, Hong Kong Polytechnic University

[8] Čmejrek, J., Bubeníček V., Luhanová, M. *Politika v regionálním rozvoji : Úvod do studia.* 1. vydání. Praha: PEF ČZU ve vydavetelství CREDIT, 2004. ISBN 80-213-1157-6

[9] Hudečková, H., Lošťák, M., Ševčíková A. *Regionalistka, regionální rozvoj a rozvoj venkova.* 1. vydání. Praha: PEF ČZU, 2006. ISBN 80-213-1413-3

[10] Město Bílina. *Http://www.bilina.cz* [online]. c2005 , © 2009 Město Bílina [cit. 2009-08-15]. Dostupný z WWW: <http://www.bilina.cz/clanek.asp?idc=162>

[11] Trávníček, M., Dolování asociačních pravidel. *Dolování asociačních pravidel* [online]. 2009 [cit. 2009-08-10]. Dostupný z WWW: http://www.fit.vutbr.cz/study/courses/ZZD/public/seminar0405/travnicek.pdf

[12] *Http://projects.gnome.org/dia/* [online]. c2005 [cit. 2009-08-10]. Dostupný z WWW: http://live.gnome.org/Dia/Download

[13] Mrázová, I., Analýza nákupního košíku. *Analýza nákupního košíku* [online]. 2009 [cit. 2009-08-10]. Dostupný z WWW: <http://ksvi.mff.cuni.cz/~mraz/datamining/lecture/dmw10.pdf>

[14] Český statistický úřad. *Http://www.czso.cz/* [online]. c2009 , 19.8. 2009 [cit. 2009-08-10]. Dostupný z WWW: http://vdb.czso.cz/xml/mos.html

[15] *Http://cs.wikipedia.org* [online]. c2009 , 24. 7. 2009 [cit. 2009-08-05]. Dostupný z WWW: http://cs.wikipedia.org

[16] *Http://www.nuov.cz* [online]. c2008 , 2008 [cit. 2009-08-05]. Dostupný z WWW: <http://www.nuov.cz/klasifikace-jednotek-nuts>

Seznam obrázků

Seznam tabulek

Seznam grafů

Seznam příloh

Seznam použitých zkratek

Ant	Antecedent
Comb	Combination
Con	Conseqvent
Cond	Condition
ČSÚ	Český statistický úřad
GUHA	General Unary Hypotheses Automaton
LAU	Local Administration Unit
MBA	Market Basket Analysis
NUTS	Nomenclature des Unités Territoriales Statistiques
Suc	Sukcedent

Přílohy

Příloha 1: Tabulky vstupních dat 1, zdroj: [vlastní].

Brno	Divadlo	14
	Jesle	3
	Krematorium	1
	Lázeňske léčebny	0
	Letecká doprava	2
	Multikino	1
	Nemocnice	11
	ÚSP pro dospělé	2
	Vodní doprava	0
	Vysoká škola	9
	Zoologická zahrada	1

Jihlava	Divadlo	2
	Jesle	1
	Krematorium	1
	Lázeňske léčebny	0
	Letecká doprava	0
	Multikino	0
	Nemocnice	1
	ÚSP pro dospělé	0
	Vodní doprava	0
	Vysoká škola	1
	Zoologická zahrada	1

České Budějovice	Divadlo	5
	Jesle	1
	Krematorium	1
	Lázeňske léčebny	0
	Letecká doprava	3
	Multikino	1
	Nemocnice	1
	ÚSP pro dospělé	0
	Vodní doprava	3
	Vysoká škola	2
	Zoologická zahrada	0

Karlovy Vary	Divadlo	9
	Jesle	0
	Krematorium	1
	Lázeňske léčebny	40
	Letecká doprava	0
	Multikino	0
	Nemocnice	1
	ÚSP pro dospělé	0
	Vodní doprava	1
	Vysoká škola	1
	Zoologická zahrada	0

Hradec Králové	Divadlo	2
	Jesle	1
	Krematorium	0
	Lázeňske léčebny	0
	Letecká doprava	0
	Multikino	1
	Nemocnice	2
	ÚSP pro dospělé	0
	Vodní doprava	0
	Vysoká škola	1
	Zoologická zahrada	0

Liberec	Divadlo	4
	Jesle	0
	Krematorium	1
	Lázeňske léčebny	0
	Letecká doprava	0
	Multikino	0
	Nemocnice	1
	ÚSP pro dospělé	1
	Vodní doprava	2
	Vysoká škola	1
	Zoologická zahrada	1

Příloha 2: Tabulky vstupních dat 2, zdroj: [vlastní].

Olomouc	Divadlo	2
	Jesle	1
	Krematorium	1
	Lázeňske léčebny	0
	Letecká doprava	1
	Multikino	1
	Nemocnice	2
	ÚSP pro dospělé	1
	Vodní doprava	0
	Vysoká škola	2
	Zoologická zahrada	1

Plzeň	Divadlo	10
	Jesle	2
	Krematorium	1
	Lázeňske léčebny	0
	Letecká doprava	2
	Multikino	1
	Nemocnice	4
	ÚSP pro dospělé	3
	Vodní doprava	0
	Vysoká škola	2
	Zoologická zahrada	1

Ostrava	Divadlo	6
	Jesle	0
	Krematorium	1
	Lázeňske léčebny	0
	Letecká doprava	0
	Multikino	1
	Nemocnice	3
	ÚSP pro dospělé	1
	Vodní doprava	1
	Vysoká škola	3
	Zoologická zahrada	1

Ústí nad Labem	Divadlo	2
	Jesle	3
	Krematorium	2
	Lázeňske léčebny	0
	Letecká doprava	1
	Multikino	0
	Nemocnice	1
	ÚSP pro dospělé	0
	Vodní doprava	7
	Vysoká škola	2
	Zoologická zahrada	1

Pardubice	Divadlo	4
	Jesle	1
	Krematorium	1
	Lázeňske léčebny	0
	Letecká doprava	1
	Multikino	1
	Nemocnice	3
	ÚSP pro dospělé	0
	Vodní doprava	4
	Vysoká škola	1
	Zoologická zahrada	0

Zlín	Divadlo	2
	Jesle	2
	Krematorium	1
	Lázeňske léčebny	1
	Letecká doprava	0
	Multikino	0
	Nemocnice	2
	ÚSP pro dospělé	1
	Vodní doprava	0
	Vysoká škola	1
	Zoologická zahrada	1

Příloha 3: Tabulka fuzzy hodnot jednotlivých organizací ve městech, zdroj: [vlastní].

Město	Organizace																																			
	Div	Div	Div	Jes	Jes	Jes	Krem	Krem	Krem	Láz	Láz	Láz	Let	Let	Let	Mul	Mul	Mul	Nem	Nem	Nem	USP	USP	USP	Vod	Vod	Vod	VS	VS	VS	ZOO	ZOO	ZOO
	M	S	V	M	S	V	M	S	V	M	S	V	M	S	V	M	S	V	M	S	V	M	S	V	M	S	V	M	S	V	M	S	V
Brno	0	0	1	0,5	0,5	0	1	0	0	0	0	0	0,75	0,25	0	1	0	0	0	0	1	0,75	0,25	0	0	0	0	0	0	1	1	0	0
České Budějovice	0	1	0	1	0	0	1	0	0	0	0	0	0,5	0,5	0	1	0	0	1	0	0	0	0	0	0,5	0,5	0	0,75	0,25	0	0	0	0
Hradec Králové	0,75	0,25	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0,75	0,25	0	0	0	0	0	0	0	1	0	0	0	0	0
Jihlava	0,75	0,25	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0
Karlovy Vary	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0
Liberec	0,25	0,75	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0,75	0,25	0	1	0	0	1	0	0
Olomouc	0,75	0,25	0	1	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0,75	0,25	0	1	0	0	0	0	0	0,75	0,25	0	1	0	0
Ostrava	0	0,75	0,25	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0,5	0,5	0	1	0	0	1	0	0	0,5	0,5	0	1	0	0
Pardubice	0,25	0,75	0	1	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0,5	0,5	0	0	0	0	0,25	0,75	0	1	0	0	0	0	0
Plzeň	0	0	1	0,75	0,25	0	1	0	0	0	0	0	0,75	0,25	0	1	0	0	0,25	0,75	0	0,5	0,5	0	0	0	0	0,75	0,25	0	1	0	0
Ústí nad Labem	0,75	0,25	0	0,5	0,5	0	0,75	0,25	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0,5	0,5	0,75	0,25	0	1	0	0
Zlín	0,75	0,25	0	0,75	0,25	0	1	0	0	1	0	0	0	0	0	0	0	0	0,75	0,25	0	1	0	0	0	0	0	1	0	0	1	0	0