

14 Porovnání přístupů

Komplexní řešení jakékoliv úlohy, jakéhokoliv problému - ať již člověkem či technickým systémem (řešitelem) - má-li se jevit jako inteligentní, se musí nutně skládat ze dvou částí (z řešení dvou podúloh):

a) z části *analytické (klasifikační, diagnostické)*, během níž se řešitel seznamuje s potřebnými údaji a fakty, příslušným způsobem je analyzuje a vytváří si svoji představu, svůj model řešeného problému. V podstatě jde o úlohu vhodné interpretace výchozích dat.

b) z části *syntetizující (plánovací, terapeutické)*, v jejímž průběhu řešitel generuje plán řešení, tj. vhodnou posloupnost elementárních kroků tak, aby bylo dosaženo požadovaného cíle řešení.

Úlohy analytické jsou ve srovnání s úlohami plánovacími výrazně jednodušší. U úloh diagnostického charakteru totiž jde buď

a) o *úlohy ryze klasifikační*, tj. o úlohy prostého zařazování objektů (jevů, situací) do tříd na základě dat naměřených na objektech,

b) o *úlohy deskripčního vnímání*, kdy cílem není jen pouhá klasifikace, nýbrž i získání podrobnějšího popisu (strojového modelu) objektu, ev. zdůvodnění rozhodovacího postupu.

Umělá inteligence jako vědní disciplína dnes poskytuje k řešení úloh analytického charakteru v podstatě tři využitelné nástroje:

a) *příznakové metody rozpoznávání*, založené především na teorii statistického rozhodování,

b) *strukturální metody rozpoznávání* pro řešení úloh ryze klasifikačních i deskripčních, tato skupina metod vychází z nenumerického (symbolického) popisu objektů a metod zpracování takových popisů,

c) *expertní systémy diagnostického charakteru*, opírající se o explicitně vyjádřené znalosti expertů.

Tato publikace je věnována prvním dvěma přístupům, tj. příznakovým a strukturálním metodám rozpoznávání. Zabývá se základními teoretickými principy těchto metod (kapitoly 2 a 3), speciálními technikami jejich aplikace při zpracování optických obrazů (kap. 4) a řečových signálů (kap. 5) i příklady aplikací těchto metod v lékařské a technické diagnostice a robotice (kap. 6).

Expertní systémy vývojově navazují na strukturální metody rozpoznávání a představují velmi účinný a perspektivní nástroj řešení diagnostických úloh. Podrobné zpracování problematiky expertních systémů přesahuje rámec a cíle této publikace, nicméně stručná zmínka o těchto systémech nám umožní vymezit postavení těchto systémů při řešení diagnostických úloh.

Abychom mohli porovnat jednotlivé přístupy k řešení diagnostických úloh a jejich použitelnost, zmíníme se nyní o formalismu produkčních systémů.

14.1. Produkční systémy

Formalismus produkčních systémů vznikl jako určité zobecnění myšlenek použitých u systému STRIPS [7.3].

Produkční systém je definován [7.2] :

- bázi dat,
- souborem produkčních pravidel,
- řídící strategií, která určuje, jak pravidla na bázi dat aplikovat.

Báze dat popisuje okamžitý stav řešené úlohy (tj. model řešené úlohy, popř. data o řešené úloze). *Produkční pravidla* jsou pravidla tvaru

$$\{\text{situace } S\} \rightarrow \langle \text{akce } A \rangle ,$$

kteřá lze interpretovat takto: „Nastala-li v bázi dat situace S , vykonej akci A “. Akcí

je nějaká změna v bázi dat. Produkční systém pracuje v cyklech

rozpoznání situace -> vykonání akce.

Řídící mechanismus určuje, jak a v jakém pořadí aplikovat pravidla na bázi dat.

Z hlediska metodologického je důležité především to, že formalismus produkčních systémů umožňuje odlišovat *deklarativně* reprezentované poznatky (vyjadřují, co je, nebo má být poznáno) od *procedurálně* reprezentovaných znalostí (říkají, jak poznávat nebo odvozovat). V jistém zjednodušujícím pohledu deklarativně reprezentované znalosti popisují stavy řešení (tvoří bázi dat produkčního systému), procedurálně reprezentované znalosti pak přechody mezi stavy (z hlediska produkčního systému jde o produkční pravidla).

I na rozpoznávací systém založený na strukturálních metodách se lze dívat jako na produkční systém, byť systém s velkými omezeními na tvar produkčních pravidel. V nejběžnějším případě, kdy pracujeme se strukturálními popisy v podobě řetězců symbolů, lze rozpoznávací systém charakterizovat jako produkční systém tímto způsobem:

a) rozpoznávaný řetěz (tj. slovo) tvoří výchozí bázi dat řešeného konkrétního případu,

b) množina přepisovacích pravidel P příslušné gramatiky tvoří soubor produkčních pravidel, v nichž (situace) i <akce> jsou reprezentovány pouze řetězcy složenými z terminálních a neterminálních symbolů,

c) algoritmus syntaktické analýzy je řídicím mechanismem produkčního systému.

Právě prezentovaný pohled na systém strukturálního rozpoznávání má jak teoretický, tak i praktický význam. Z hlediska teoretického může posloužit k porovnání metod. V poslední době též narůstá praktický význam takového pohledu. Příčinou je existence speciálních vyšších programovacích jazyků typu „prázdného“ produkčního systému, kdy se začíná využívat těchto velmi obecných a silných nástrojů i k rychlé a levné realizaci systémů strukturálního rozpoznávání.

Všechny typy diagnostických systémů se vyznačují tím, že jsou v nich určitým způsobem reprezentovány znalosti o dané diagnostické úloze a že dodáním konkrétních dat (báze dat) systém dospívá k diagnostickému závěru. V následujícím porovnání metod rozpoznávání podle různých aspektů se budeme držet nejobecnější terminologie produkčních systémů (i když používání této terminologie v teorii rozpoznávacích systémů nemusí být vždy zcela obvyklé).

14.2. Porovnání metod

V tomto oddíle se pokusíme porovnat metody příznakového a strukturálního rozpoznávání.

V úvodu zdůrazněme zásadní rozdíl mezi metodami příznakového rozpoznávání na jedné straně a metodami strukturálního rozpoznávání na straně druhé. V příznakových metodách jsou využívány pouze numerické, kvantitativní informace, zatímco druhý přístup k řešení diagnostických úloh je založen především na symbolické kvalitativní reprezentaci vnitřní struktury diagnostikovaných (rozpoznávaných) objektů.

Reprezentace a využívání znalostí

V systémech příznakového rozpoznávání jsou znalosti o dané třídě zachyceny v podobě numerických parametrů ať již podmíněných hustot pravděpodobnosti, nebo diskriminačních funkcí, nebo rozdělovacích hyperploch. Numerická reprezentace je při řešení úlohy využívána přímočaře (bez návratů) jako součást pevných procedurálně reprezentovaných vzorců pro výpočet indikátoru třídy.

V systémech strukturálního rozpoznávání mohou být znalosti reprezentovány z hlediska teoretického nejen deklarativně (přepisovacími pravidly gramatiky), ale i ev. procedurálně (množinou podmínek a příkazů u atributových gramatik). Využívání znalostí je dáno použitou metodou syntaktické analýzy.

Data a jejich sémantika, vztah k vysvětlování

U systémů příznakového rozpoznávání je báze dat (údaje o konkrétním případě) tvořena vektorem numerických příznaků, a to vektorem konstantní dimenze N . Žádný příznak obvykle nesmí chybět. Každý příznak však může být zcela jiné fyzikální podstaty. Mohou to tedy být data z heterogenních zdrojů informací. V průběhu

rozpoznávání se ztrácí fyzikální interpretace dat (příznaků) i rozhodovacích pravidel. Nelze pracovat se sémantikou. Systémy příznakového rozpoznávání proto nemohou být vybaveny vysvětlovacími schopnostmi.

Symboly ve slovech strukturálního popisu obvykle reprezentují fakta (poznatky) stejného charakteru a původu. Informace z jiných zdrojů bývají kódovány jako doplňková číselná informace. V určitém rozsahu je tedy k dispozici možnost uvažovat sémantiku a kontext.

V průběhu zpracování dat metodami strukturálního rozpoznávání se neztrácí fyzikální interpretace, naopak je možné pracovat s hierarchicky organizovaným souborem symbolů. Vysvětlování procesu rozpoznávání je v omezeném rozsahu principiálně možné, ale není příliš ilustrativní, a proto se nepoužívá.

Úloha numerické informace

U příznakových metod rozpoznávání je numerická informace jediným typem reprezentace dat i znalostí. Data, tj. příznaky, lze chápat jako kvantitativní hodnoty veličin. Používají se k numerickým výpočtům, k určování maximální hodnoty diskriminačních funkcí (nebo adekvátních rozhodovacích formalismů).

U strukturálních metod má numerická informace charakter informace sémantické. Využívá se především k zavedení metriky v deformačních schématech a k fyzikální interpretaci výsledků syntaktického zpracování na úrovni sémantické analýzy, ale i k řízení procesu syntaktické analýzy. V dosavadních systémech je používána především ad hoc; účinný nástroj pro její jednotnou formalizaci a organické propojení se symbolickou (strukturální) reprezentací tvoří atributové gramatiky.

Počet tříd

Počet tříd není u příznakového klasifikátoru v principu omezen, praktické zkušenosti však ukazují, že přijatelných výsledků lze jediným klasifikátorem dosáhnout jen tehdy, je-li počet tříd maximálně 6 až 10, počet příznaků maximálně 10 až 20. Právě uvedené meze nelze brát zcela kategoricky, generalizace poznatků je velmi obtížná, neboť řešení konkrétní úlohy příznakového rozpoznávání silně závisí na charakteru dat. Při náročnějších klasifikacích (větší počet tříd, větší počet příznaků), mají-li být prováděny jen příznakově, je nezbytné vytvořit (hierarchické) soustavy elementárních klasifikátorů.

V případě strukturálních metod je klasifikovaný strukturální popis analyzován postupně (rozšířenými) gramatikami (nebo adekvátními automaty). Proces syntaktické analýzy je (s výjimkou jednoduchých regulárních gramatik) podstatně časově náročnější ve srovnání s příznakovými metodami. Slovo je zařazeno do té třídy, k jejíž gramatice leží nejbliže, a to ve smyslu deformačním schématem zavedené míry vzdálenosti (metriky). Zavedení míry vzdálenosti principiálně umožňuje též seřadit indikátory tříd pro dané slovo do relevantní posloupnosti. Této možnosti se však nevyužívá (nebývá k dispozici vysvětlovači mechanismus) s výjimkou těch situací, kdy je uvažována ev. oprava popisu ze sémantického hlediska či při hierarchickém řazení klasifikátorů se zpětnovazebním ovlivňováním.

Počet tříd (gramatik, automatů) v elementárním strukturálním klasifikátoru není v principu omezen, ale větší počet (více než 6 — 8) uvažovaných tříd nelze doporučit jak z časových důvodů, tak zejména z důvodů (obvykle) vysoké generativní schopnosti gramatik, a tedy rostoucí mohutnosti průníků jazyků generovaných gramatikami jednotlivých tříd. Při nezbytnosti strukturální klasifikace do většího počtu tříd lze opět použít hierarchické uspořádání klasifikátorů.

Učení

Nastavování příznakových klasifikátorů plně automatizovaným učením je nejrozsáhlejší metodou vytváření relevantní reprezentace znalostí u těchto systémů. Byla již vytvořena, ověřena a zdokumentována rozsáhlá škála těchto metod s vymezením jejich použitelnosti. Samostatnou, významnou skupinu tvoří metody samoučení příznakových klasifikátorů, tj. metody učení bez učitele, nazývané metodami shlukové analýzy.

V úloze nastavování příznakového klasifikátorů je trénovací množina chápána z hlediska statistického jako náhodný výběr; čím mohutnější je trénovací množina, tím kvalitnější nastavení klasifikátorů lze očekávat.

Pro systémy strukturálního rozpoznávání bylo vyvinuto v podstatě jenom několik, a to ještě velmi omezeně použitelných technik automatizovaného učení, tj. automatizované inference (regulárních) gramatik, popř. automatizované konstrukce (konečných deterministických) automatů. V praktických případech konstruktér buď vytváří vhodnou gramatiku, popř. automat, „ručně“ na základě podrobné analýzy úlohy, nebo jsou automatizovaně generovány celé třídy gramatik a konstruktér si z nich vybírá a upravuje je do finální podoby. Při stanovování numerických parametrů deformačního modelu lze však v hojně míře užít metod automatizovaného učení, zejména metod učení příznakových klasifikátorů.

Při inferenci gramatik není tak důležitá mohutnost trénovací množiny jako spíše splnění požadavku její reprezentativnosti. Při odvozování deformační gramatiky a jejích parametrů s využitím metod učení typických pro příznakové klasifikátory je však opět žádoucí co nejmohutnější náhodný výběr deformovaných obrazů.

Vklad konstruktéra

U příznakových metod konstruktér pouze volí příznaky a určuje způsob reprezentace tříd v klasifikátorů. Na základě (náhodně) vybrané trénovací množiny se již klasifikátor automaticky nastaví a je připraven k diagnostice.

U strukturálních metod musí konstruktér provádět obvykle náročnou strukturální analýzu s cílem vybrat elementy popisu (primitiva a relace). Dále musí vytvořit (obvykle „ručně“, méně často s pomocí automatického generátoru) gramatiky tříd pro nedeformované popisy a na základě dalšího podrobného rozboru množiny deformovaných popisů navrhuje konkrétní deformační schéma. Parametry deformačního schématu (tj. gramatiky deformací) lze určit jednoduchými metodami učení na základě mohutnější trénovací množiny. Gramatiku sémantické analýzy (např. na bázi atributové gramatiky) musí vždy vytvářet konstruktér.

14.3. Expertní systémy

Expertní systémy jsou programy simulující rozhodovací procesy špičkových specialistů (expertů) při řešení složitých úloh. Jsou založeny na znalostech převzatých od těchto expertů. Znalosti expertů bývají vyjádřeny nejčastěji v podobě pravidel, která však mívají (na rozdíl od pravidel v systémech strukturálního rozpoznávání) subjektivní charakter.

K využívání znalostí se aplikují algoritmy založené na metodách umělé inteligence (např. vycházející z teorie stavového prostoru a jeho prohledávání, z teorie reprezentace znalostí a zpracování nejisté informace apod.).

Cílem činnosti expertních systémů je dosáhnout kvality rozhodování na úrovni blízké se úrovni rozhodování experta.

Znalosti expertů jsou u expertních systémů vyjádřeny naprosto explicitně v podobě samostatného počítačového souboru *báze znalostí*, přičemž jsou uchovávány odděleně od algoritmů pro jejich využívání. Uchovávání znalostí v samostatných počítačových souborech zabezpečuje vyšší čitelnost znalostí, jednodušší přístup expertů ke znalostem i snadnou modifikovatelnost báze znalostí. Striktní oddělení báze znalostí od programového modulu pro jejich využívání (tzv. řídicího mechanismu) umožňuje využívat stejného řídicího mechanismu pro rozhodovací činnost v různých aplikačních oblastech. Problémově nezávislá část expertního systému se

obvykle nazývá prázdný expertní systém. Teprve připojením báze znalostí (naplněním) se vytváří použitelný expertní systém.

Báze znalostí obsahuje veškeré znalosti, kterých expert využívá při řešení úloh v příslušné problémové oblasti. Zachycuje pestrou paletu znalostí od nejobecnějších zákonitostí až po znalosti vysoce specializované, od „učebnicových“ znalostí až po znalosti „soukromé“ (tj. takové, které by si odborník ani nedovolil publikovat).

Expertní systémy jsou určeny k řešení zcela konkrétních případů. Data k danému případu poskytuje obvykle uživatel sekvenčně, v dialogovém režimu s počítačem. Dodávání dat lze chápat jako dosazování konkrétních údajů do obecných rozhodovacích struktur, zachycených v bázi znalostí. Dialog uživatele s počítačem má charakter dialogu laika či méně zkušeného odborníka s expertem. Otázky jsou expertním systémem voleny dynamicky. Systém předkládá v každém kroku konzultace tu otázku, od jejíhož zodpovězení očekává co nejrychlejší upřesnění svého vnitřního strojového modelu řešeného případu.

Množina všech údajů k danému případu se nazývá - na rozdíl od báze znalostí - bázi dat. Z dosavadního výkladu je patrné, že každý expertní systém může být chápán jako speciální produkční systém.

Expertní systémy bývají vybaveny schopností kombinovat nejisté znalosti s nejistými či nepřesnými daty v bázi dat. Nejisté znalosti bývají vyjádřeny jako znalosti s přidělenou mírou důvěry v jejich platnost, nejistá data jsou obsažena v nejistých odpovědích uživatele. Všimněme si, že na rozdíl od expertních systémů nevyjadřuje doplňková číselná informace u systémů strukturálního rozpoznávání nejistotu v platnost pravidla, nýbrž frekvenci či cenu použití tohoto pravidla jako podklad k řízení syntaktické analýzy.

Typickou vlastností expertních systémů je jejich schopnost vysvětlit i odůvodnit všechna hlediska rozhodovacích procesů.

U diagnostických expertních systémů se znalosti obvykle vyjadřují ve formě produkčních pravidel speciálního typu, tzv. pravidel $E \rightarrow H$:

JESTLIŽE (Předpoklad E), PAK (Závěr H) S (Váhou W),

kde předpoklad E i závěr H jsou tvrzení, váha W vyjadřuje neurčitost pravidla („sílu“ pravidla). Předpoklady některých pravidel jsou tvořeny obvykle závěry jiných pravidel atd. Vyjádříme-li každé pravidlo typu $E \rightarrow H$ graficky jako orientovanou hranu mezi uzlem E (reprezentujícím tvrzení E) a uzlem H , lze pak celou bázi znalostí vyjádřit grafem *inferenční sítě*. Cíle (cílové hypotézy) diagnostického procesu jsou představovány vrcholovými uzly, tj. těmi uzly, které vždy v pravidlech figurují pouze jako závěr. Inferenční síť lze u diagnostických systémů chápat jako pevnou inferenční strukturu umožňující [7.6]

a) vybírat vhodné dotazy uživateli (v režimu cílem řízeného prohledávání inferenční sítě),

b) šířit informaci od právě uživatelem zodpovězených uzlů k cílovým hypotézám (v režimu daty řízeného procesu, přičemž dochází k úpravě měr platnosti tvrzení v inferenční síti).

Inferenční síť slouží v diagnostických systémech současně i jako vnitřní (strojový) model řešeného případu, neboť obsahuje aktuální míry platnosti jednotlivých tvrzení.

14.4. Použitelnost jednotlivých přístupů

Z dosavadního stručného porovnání metod příznakového a strukturálního rozpoznávání a ze stručné charakteristiky expertních systémů lze usuzovat na to, v kterých případech řešení diagnostických úloh je výhodnější používat toho či onoho přístupu.

O nasazení systémů příznakového rozpoznávání lze uvažovat, pokud:

a) je nutné zpracovávat kvantitativní (numerická) data v zásadě heterogenního původu, u nichž nemá smysl hledat vnitřní logickou strukturu, jsou k dispozici

rozsáhlé soubory naměřených dat a neexistuje expert, schopný formulovat pravidla potřebná k rozhodování,

- b) není třeba explicitně vyjadřovat a uvažovat neurčitost v datech,
- c) nemá smysl v průběhu klasifikace uvažovat kontext a sémantiku,
- d) je vyžadován kategorický výsledek, bez vysvětlení průběhu

rozhodovacího procesu.

Je-li k dispozici soubor naměřených dat bez údajů o správném zařazení do třídy (tj. bez údajů od učitele), je jedinou použitelnou variantou příznakový systém na bázi metod shlukové analýzy.

O nasazení systémů strukturálního rozpoznávání lze uvažovat, pokud

a) je nutné zpracovávat data homogenního původu, u nichž lze určit výrazné strukturální elementy a relace mezi nimi,

b) jsou k dispozici soubory dat (slov) pro inferenci (ať již automatickou či „ruční“) pravidel,

c) slova strukturálního popisu nebývají poškozena nebo bývají poškozena jen ve velmi omezeném rozsahu,

d) u slov a jejich elementů není třeba explicitně vyjadřovat a zpracovávat jejich neurčitost,

e) má smysl uvažovat kontext a sémantiku jen v omezeném rozsahu,

f) je vyžadován kategorický výsledek, bez vysvětlení průběhu rozhodovacího

procesu.

Expertní systémy jsou vhodné k řešení diagnostických úloh, pokud:

a) je nutné zpracovávat nejruznější data (heterogenní, kvalitativní i kvantitativní),

b) existují expertní znalosti, přičemž pokud část expertních znalostí chybí, je pro tento úsek k dispozici soubor reálných dat,

c) je nutné vyhodnocovat i neúplná či velmi poškozená data,

d) část nebo všechna data mají neurčitý charakter, ;, ;,

e) existují silné kontextové a sémantické vazby mezi daty i pravidly,

f) jsou požadovány alternativní diagnózy s vysvětlením postupu, jak bylo závěrů dosaženo, tj. pokud si má uživatel zachovat právo konečného rozhodnutí.

Poznamenejme též, že při posuzování vhodnosti jednotlivých přístupů k řešení diagnostických úloh metodami umělé inteligence je vždy nutné brát do úvahy a individuálně posuzovat též hlediska paměťová a časová, zejména s ohledem na dostupnou výpočetní techniku. ~~

Z dosavadních úvah vyplývá, že systémy příznakového a strukturálního rozpoznávání a expertní systémy se hodí k řešení poněkud odlišných úloh.

Příznakové systémy nacházejí uplatnění při strojovém řešení úloh na té úrovni, kdy člověk nemůže dodat téměř žádnou informaci o řešeném problému (obvykle dodává jen trénovací množinu případů).

Systémy strukturálního rozpoznávání vyžadují větší vklad konstruktéra, uvažuje se v nich o vnitřní struktuře dat, pracuje se v omezeném rozsahu se sémantikou

a kontextem, lze počítat s jistou korekční schopností. Vytvoření strukturálního popisu však vyžaduje existenci (obvykle příznakového) systému pro extrakci primitiv a relací z obvyklých čistě numerických dat.

Expertní systémy jsou určeny pro řešení těch úloh, kdy dat je nedostatek nebo jsou neúplná, nejistá či poškozená, ale kdy člověk může poskytnout velmi hodnotné znalosti (v podobě báze znalostí) pro účinné zpracování takových dat.

Systémy příznakového rozpoznávání, strukturálního rozpoznávání a expertní systémy nejsou systémy konkurenční, nýbrž jsou určeny k řešení úloh na různých úrovních diagnostického procesu. Mnohdy se vzájemně doplňují (např. metod příznakového rozpoznávání se velmi často používá k vytváření těch částí báze znalostí, kde znalosti experta nejsou k dispozici; expertní systémy mohou naopak vyplňovat mezery v činnosti rozpoznávacích systémů).

Příznakové metody jsou obvykle aplikovány na nejnižší úrovni (např. při zpracování optické informace jsou určeny k získání elementů strukturálního popisu z rozsáhlé matice číselných hodnot), na střední úrovni rozhodování se uplatňují strukturální metody (např. detekce elementárních objektů a některých jejich vzájemných vazeb). Rozhodnutí pomocí expertního systému (rozhodnutí o stavu scény a potřebných zásazích) je pak zabezpečováno na hierarchicky nejvyšší úrovni.

Třídění nástrojů pro řešení úloh na různých úrovních diagnostického procesu má svůj přirozený původ, neboť příznakové metody vznikly jako analogie činnosti 1. signální soustavy, strukturální metody jako zjednodušená analogie formální stránky přirozeného jazyka, expertní systémy se snaží o napodobení „inteligentní“ mentální činnosti špičkového odborníka-experta.