

Aplikace fuzzy a pravděpodobnostních automatů

Martin Jašek

12. září 2016 — ??

Obsah

1	Definice a značení	2
2	Fuzzy automaty, gramatiky a jazyky	4
2.1	Jazyk rozpoznávaný fuzzy automatem	5
2.2	Fuzzy a bivalentní regulérní jazyky	5
2.3	Fuzzy regulární výrazy	5
3	Další varianty fuzzy automatů	5
3.1	Fuzzy zásobníkové automaty	5
3.2	Fuzzy tree automaty	5
3.3	Buněčné fuzzy automaty	7
4	Rozpoznávání textových vzorů	7
4.1	Formální zavedení problému	8
4.2	Motivace k použití fuzzy automatů	8
4.3	Automat rozpoznávající ω	9
4.4	Podobnost symbolů	10
4.5	Fuzzy symboly	12
4.6	Editační operace	14
4.7	Deformovaný automat	18
4.8	Shrnutí	19
5	Konkrétní příklady	19
5.1	Rozpoznání ručně psaného textu	19
5.2	Detekce překlepů	20

1 Definice a značení

Tato kapitola zatím poslouží jako „skladiště“ pro definice a zavedení značení pro ostatní kapitoly.

Abecedy, řetězce, jazyky

Abecedy budou značeny standardně, tedy velkými řeckými písmeny (typicky Σ). Řetězce pak malými písmeny (ω, α, \dots). Jazyky velkými kaligrafickými písmeny. Jazyk přijímaný automatem A bude značen $\mathcal{L}(A)$.

Fuzzy teorie

Fuzzy množiny a relace budou po vzoru [7] nejčastěji malými řeckými písmeny (obdobně jako jejich členské (angl. „membership“) funkce). Množinu všech fuzzy podmnožin množiny S budeme značit $\mathcal{F}(S)$.

Deterministický bivalentní automat

(zde bude doplněno: zdroj: Eilenberg S.: Automata, Languages and Machines, Vol. A, Academic Press, New York, 1974. Pokud ji někde seženu (odkazuje se na ni Bel v [1])
(zde bude doplněno: co citování definic? půlku jsem si vymyslel ...)

Definice 1.1. Konečný deterministický (bivalentní) automat je pětice $A = (Q, \Sigma, \delta, q_0, F)$, kde Q je konečná množina stavů, Σ je vstupní abeceda, $\delta : Q \times \Sigma \rightarrow Q$ je přechodová funkce, $q_0 \in Q$ je počáteční stav a $F \subseteq Q$ je množina koncových stavů.

Nedeterministický bivalentní automat

(zde bude doplněno: Značení převzato z FJAA, dohledat zdroj)

Definice 1.2. Konečný nedeterministický (bivalentní) automat je pětice $A = (Q, \Sigma, \delta, I, F)$, kde Q je konečná množina stavů, Σ je vstupní abeceda, $\delta : Q \times \Sigma \rightarrow 2^Q$ je přechodová funkce, $I \subseteq Q$ je množina počátečních stavů a $F \subseteq Q$ je množina koncových stavů.

Základní definice nedeterministického fuzzy automatu

Značení je převzato z [7] a lehce upraveno.

Definice 1.3 (Nedeterministický fuzzy automat). Nedeterministický fuzzy automat A je pětice $(Q, \Sigma, \mu, \sigma, \eta)$, kde Q je konečná množina stavů, Σ je abeceda, μ je fuzzy přechodová funkce (fuzzy relace $Q \times \Sigma \times Q \rightarrow [0, 1]$) a σ a η jsou po řadě fuzzy množiny nad Q počátečních, resp. koncových stavů.

{def-ZaklDefNedFuzzAut}

Definice 1.4 (Fuzzy stav). Mějme nedeterministický fuzzy automat A . Pak jako fuzzy stav označujeme fuzzy podmnožinu jeho stavů, tj. $V \in \mathcal{F}(Q)$.

{def-FuzzStav}

Definice 1.5 (Aplikace fuzzy relace na fuzzy stav). Mějme nedeterministický fuzzy automat A a fuzzy symbol V . Pak aplikací binární fuzzy relace $R : Q \times Q \rightarrow [0, 1]$ na fuzzy stav V obdržíme fuzzy symbol $V \circ R$ splňující pro každé $p \in Q$: $(V \circ R)(p) = \max_{q \in Q} (V(q) \otimes R(q, p))$.

Definice 1.6 (Přechodová funkce fuzzy stavů). *Mějme nedeterministický fuzzy automat A . Pak přechodová funkce fuzzy stavů je fuzzy relace $\hat{\mu} : \mathcal{F}(F) \times \Sigma \rightarrow \mathcal{F}(F)$ taková, že pro každý fuzzy stav $V \in \mathcal{F}(Q)$ a symbol $x \in \Sigma$ je $\hat{\mu}(V, x) = V \circ \mu[x]$.*

{def-PreFunFuzzStav}

Poznámka 1.1. *Označení $\mu[x]$ je fuzzy relace, pro kterou platí: $\mu[x](p, q) = \mu(p, x, q)$ pro všechna $x \in \Sigma$ a $p, q \in Q$.*

{def-PreFunFuzzStav}

Definice 1.7 (Rozšířená přechodová funkce). *Mějme nedeterministický fuzzy automat A . Pak rozšířená přechodová funkce (fuzzy stavů) je fuzzy relace $\mu^* : \mathcal{F}(F) \times \Sigma^* \rightarrow \mathcal{F}(F)$ (zde bude doplněno: co je to F ? Nemá to být Q ?!)) daná následujícím předpisem:*

1. $\mu^*(V, \epsilon) = V$ pro všechna $V \in \mathcal{F}(Q)$
2. $\mu^*(V, \alpha x) = \hat{\mu}(\mu^*(V, \alpha), x)$ pro všechna $V \in \mathcal{F}(Q), \alpha \in \Sigma^*, x \in \Sigma$

{def-RetPriAut}

Definice 1.8 (Řetězec přijímaný automatem). *Mějme nedeterministický fuzzy automat A . Pak řetězec $\alpha \in \Sigma^*$ je automatem A přijat ve stupni*

$$A(\alpha) = \max_{q \in Q} (\mu^*(\sigma, \alpha)(q) \otimes \eta(q))$$

(zde bude doplněno: ověřit, dohledat, ozdrojovat)

{def-JazRozpAut}

Definice 1.9 (Jazyk rozpoznávaný automatem). *Mějme nedeterministický fuzzy automat A . Pak fuzzy množinu $\mathcal{L}(A)(\alpha) = A(\alpha)$ nad univerzem Σ^* nazýváme fuzzy jazyk rozpoznávaný automatem A .*

(zde bude doplněno: ověřit, dohledat, ozdrojovat)

Nedeterministický fuzzy automat s ϵ přechody

{def-NedFuzzAutEpsPre}

Definice 1.10 (Nedeterministický fuzzy automat s ϵ přechody). *(zde bude doplněno: dohledat přesně, zkontrolovat a ozdrojovat) Nedeterministický fuzzy automat A je pětice $(Q, \Sigma, \mu, \sigma, \eta)$, kde Q je konečná množina stavů, Σ je abeceda, μ je fuzzy přechodová funkce (fuzzy relace $Q \times (\Sigma \cup \{\epsilon\}) \times Q \rightarrow [0, 1]$) a σ a η jsou po řadě fuzzy množiny nad Q počátečních, resp. koncových stavů.*

(zde bude doplněno: Tady by asi bylo vhodné rozebrat ϵ -uzávěry)

Reprezentace fuzzy automatu

(zde bude doplněno: dohledat zdroje)

(zde bude doplněno: Značení fuzzy množiny $\sigma = \{x/0.5\}$ vs. $\sigma(x) = 0.5$)

Přechodový diagram: Notace s lomítky např. zde: [9].

Tabulka: stav \times symbol nebo stav \times stav [10]?

Konstrukce fuzzy automatu z konečného automatu

V praxi se často setkáme z problémem, kdy máme k dispozici konečný bivalentní automat avšak my potřebujeme pro naši práci fuzzy automat. Je tedy třeba zkonstruovat takový fuzzy automat, který rozpoznává odpovídající jazyk odpovídající jazyku rozpoznávaném naším bivalentním automatem.

Důležité je zmínit, že nelze zkonstruovat fuzzy automat, rozpoznávající stejný jazyk neboť fuzzy automat rozpoznává fuzzy jazyk, zatímco bivaletní automat klasický „bivaletní“ jazyk. Můžeme však sestavit automat takový, který přijímá řetězce ve stupni 0 nebo 1 podle toho, jestli je přijímal bivaletní automat.

Formálně řečeno, pro konečný (nedeterministický) bivaletní automat A budeme konstruovat nedeterministický fuzzy automat A' takový, že bude pro všechna $x \in \Sigma^*$ splněna následující rovnost:

$$\mathcal{L}(A')(\omega) = \begin{cases} 1 & \text{pokud } \omega \in \mathcal{L}(A) \\ 0 & \text{pokud } \omega \notin \mathcal{L}(A) \end{cases}$$

Poznámka 1.2. *Postup budeme provádět pro nedeterministické automaty. To jednak proto, že nedeterministické automaty jsou obecnější, než deterministické, a navíc, protože jsou v praxi využívány častěji. (zde bude doplněno: ozdrojovat, klidně někde, kde rozebírám determinismus vs. nedeterminismus)*

Nyní se podíváme na to, jak výsledný fuzzy automat bude vypadat. Abeceda i množina stavů automatu zůstanou zachovány, lišit se tedy bude pouze množina počátečních a koncových stavů a přechodová funkce. (zde bude doplněno: fuzzy subset I , F a δ ? nebo tak něco, z teorie fuzzy množin?)

Definice 1.11 (Fuzzy automat bivaletního automatu). *Mějme řetězec konečný nedeterministický automat $A = (Q, \Sigma, \delta, I, F)$. Pak nedeterministický fuzzy automat přijímající korespondující jazyk je automat $A' = (Q, \Sigma, \mu, \sigma, \epsilon)$ kde pro všechna $q_i, q_j \in Q$ a $x \in \Sigma$:*

{def-FuzzAutBivAut}

- $\sigma(q_i) = \begin{cases} 1 & \text{pokud } q_i \in I \\ 0 & \text{pokud } q_i \notin I \end{cases}$
- $\eta(q_i) = \begin{cases} 1 & \text{pokud } q_i \in F \\ 0 & \text{pokud } q_i \notin F \end{cases}$
- $\mu(q_i, x, q_j) = \begin{cases} 1 & \text{pokud } q_j \in \delta(q_i, x) \\ 0 & \text{pokud } q_j \notin \delta(q_i, x) \end{cases}$

(zde bude doplněno: rozebrat, jestli tento automat skutečně dělá to, co má? Asi by to chtělo)

(zde bude doplněno: vymyslet nějaký fakt pěkný příklad)

2 Fuzzy automaty, gramatiky a jazyky

(zde bude doplněno: nějak to uvést. Budou pojmy jako regulérní jazyk a gramatika popsány v nějaké předchozí kapitole?)

(zde bude doplněno: pojem „Lattice language“)

(zde bude doplněno: značení: „Fuzzy množina ϕ “ vs. „ L -množina $\phi : X \rightarrow L$ “; „fuzzy podmnožina“ vs. „fuzzy množina nad“)

2.1 Jazyk rozpoznávaný fuzzy automatem

Věta 6.3 [14] (pro lattice monoid, není to někde jen pro $[0, 1]$?).

Dle definice 4 [9] je fuzzy regulární jazyk fuzzy podmnožina bivalentního.

Automat s bivalentní μ and η (a fuzzy σ) taky rozpoznává fuzzy regulární jazyk [5]. Neřešil něco takového i Bel? Jinak řečeno, support konečný automat [9].

2.2 Fuzzy a bivalentní regulární jazyky

Univerzum fuzzy jazyka je regulární jazyk, pozorování 6.1 [14].

Stejně tak, zaříznutý jazyk (a -řez jazyka) je také regulární, věta 2.2 [5].

Pumping lemma pro fuzzy regulární jazyky, lemma 4-7 pro různé typy automatů [9].

Uzávěrové vlastnosti fuzzy regulárních jazyků, např. [5].

2.3 Fuzzy regulární výrazy

LiPed-FuzzFinAutFuzzRegExMembValLattOrdMon, definice 5.1, 5.2 (+ opsat důkaz, že $[0, 1]$ je lattice monoid) [14].

Algoritmus převodu reg na aut, [13]. Ale zdá se mi to až moc složité.

3 Další varianty fuzzy automatů

V této kapitole budou rozebrány další výpočetní modely, které jsou založeny na fuzzy automatech. Tyto výpočetní modely sice přichází o jednoduchost „klasických“ fuzzy automatů, nicméně jsou blíže k reálným aplikacím a proto je na místě se jimi zabývat.

3.1 Fuzzy zásobníkové automaty

(zde bude doplněno: tohle vzít asi jen teoreticky, v aplikacích se vlastně ani moc nepoužívají)

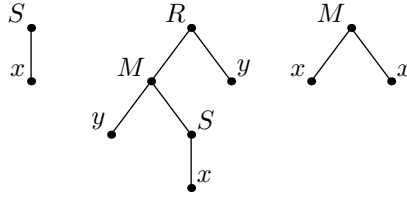
3.2 Fuzzy tree automaty

Fuzzy tree automaty jsou speciální třídou automatů, které jsou navrženy pro rozpoznávání řetězců, které mají v sobě obsaženu určitou stromovou strukturu. Jak bude ukázáno, fuzzy tree automaty tak mohou rozpoznávat vybrané jazyky popsané bezkontextovými gramatikami.

Fuzzy tree automaty vznikly fuzzyfikací „klasických“ tree automatů. O „klasických“ tree automatech je možné se dočíst více informací např. v [?], popř. [6] a [?]. Problematicke fuzzy tree automatů se věnuje například [?], [11], [12], [?] a [?]. V této kapitole vycházeno z [?].

Fuzzy automaty pracují se speciálními strukturami symbolů, tzv. stromy. Následuje tedy definice stromů.

Definice 3.1 (Doména stromu). *Mějme abacedu Σ uspořádanou pomocí \leq . Pak konečnou množinu $U \subseteq \Sigma^*$ nazvěme doména konečného stromu, pokud splňuje následující podmínky:*



Obrázek 1: Kořenové stromy termů z příkladu 3.2

{img:RootTree}

- jestliže $w \in U$ a $w = uv$ pak $u \in U$ pro všechna $u, v, w \in \Sigma^+$ (tj. množina je prefixově uzavřena)
- $wn \in U$ a $m \leq n$ implikuje $wm \in U$, pro všechna $w \in \Sigma^+$ a $m, n \in \Sigma$

Doména stromu je vlastně množina řetězců, které formují prefixový strom. Množinu U tak rozložit na množinu \bar{U} listových uzlů jako

$$\bar{U} = \{w \in U \mid U \text{ neobsahuje žádný řetězec } w' \neq w \text{ s prefixem } w\}$$

a množninu vnitřních uzlů jako $U \setminus \bar{U}$.

Definice 3.2 (Částečně spořádaná abeceda). Částečně spořádaná abeceda je dvojice (N, T) , kde N a T jsou dvě disjunktí konečné abecedy (tj. $N \cap T = \emptyset$).

Definice 3.3 (Strom). Strom t nad částečně spořádanou abecedou (N, T) je zobrazení z domény U stromu do $(N \cup T)$ (psáno $t : U \rightarrow (N, T)$) takové, že

- $t(w) \in N$ pokud $w \in U \setminus \bar{U}$
- $t(w) \in T$ pokud $w \in \bar{U}$

Takto nadefinovaný strom lze snadno graficky ztvárnit. Stačí si uvědomit, jak se vyobrazují prefixové stromy (tím se zobrazí doména stromu t) a následně se přeznačí všechny uzly w hodnotou $t(w)$.

{ex:trees}

Příklad 3.1. Příklady stromů nad částečně spořádanou abecedou (N, T) , kde $N = \{m, r, s\}$ a $T = \{x, y\}$ jsou na obrázku 1.

Automaty jsou však modely, které pracují s řetězci, tedy (lineárními) posloupnostmi symbolů. Je proto třeba zavést způsob, jak zakódovat stromy do řetězců. Řetězcům, které kódují stromy budeme říkat pseudotermy.

Definice 3.4 (Pseudoterm). Označme $D_{(N,T)}^p$ nejmenší podmnožinu $(N \cup T \cup \{(\cdot, \cdot)\})^*$ splňující následující podmínky:

- $T \subset D_{(N,T)}^p$
- pokud $n > 0$, $A \in N$ a $t_1, \dots, t_n \in D_{(N,T)}^p$, pak $A(t_1 \dots t_n) \in D_{(N,T)}^p$

Prvky množiny $D_{(N,T)}^p$ nazýváme pseudotermy.

Příklad 3.2. Pro částečně spořádanou abecedu s předchozího příkladu mohou být pseudotermy například: y , $S(x)$, $M(xx)$, $R(M(yS(x))y)$

{ex:terms}

Nyní je třeba ukázat, že mezi stromy a pseudotermy platí vzájemně jednoznačný vztah.

Věta 3.1. *Ke každému pseudotermu $p(t) \in D_{(N,T)}^p$ existuje odpovídající strom t .*

Důkaz. Existenci stromu dokážeme přes strukturu pseudotermu:

- je-li pseudoterm atomický, tj. $p(t) = a$, kde $a \in T$, pak doménou stromu t je množina $\{\epsilon\}$ a $t(\epsilon) = a$
- pokud je pseudoterm ve tvaru $p(t) = A(t_1^p \dots t_m^p)$, pak doménou stromu t je množina $\bigcup_{i \leq m} \{iw \mid w \in \text{domain}(t_i)\} \cup \{\epsilon\}$ a

$$t(w) = \begin{cases} A & \text{pokud } w = \epsilon \\ t_i(w') & \text{pokud je } w = iw' \text{ a } w \text{ je v doméně } t \end{cases}$$

□

Značení. Množinu všech stromů nad (N, T) značme $D_{(N,T)}$ a jednotlivé stromy $t \in D_{(N,T)}$. Množinu všech pseudotermů značme $D_{(N,T)}^p$ a pseudoterm korespondující stromu t značme $p(t)$.

Definice 3.5 (Fuzzy jazyk stromů). Fuzzy množinu T nad $D_{(N,T)}$ nazvěme fuzzy jazyk stromů.

3.3 Buněčné fuzzy automaty

4 Rozpoznávání textových vzorů

Rozpoznávání vzorů obecně je jednou z nejvýznamějších aplikací informatiky. V běžném životě se často setkáváme se situacemi, kdy je třeba v datech najít výskyt učitěho vzoru, popř. jeho další vlastnosti. Případně určit podobnost ke vzoru, nebo nejpodobnější vzor.

Typickým příkladem je např. detekce obličeje na fotografii, tedy rozpoznávání vzorů v obrazových datech. Vzory je však možné rozpoznávat v téměř jakýchkoliv datech, například textech, zvukových záznamech či výsedcích měření nebo pozorování.

Z pohledu teoretické informatiky je však základem vyhledávání vzorů v textových datech. Textová data, tedy řetězce, mají jednoduchou strukturu a lze s nimi snadno manipulovat. Na druhou stranu, jsou schopna reprezentovat nebo kódovat široké spektrum dat. Právě z tohoto důvodu je studium rozpoznávání textových vzorů klíčové pro zpracovávání jakýchkoliv dalších typů dat.

Poznámka 4.1. *Pokud nebude uvedeno jinak, pojem „rozpoznávání textových vzorů“ bude v této kapitole zkracován jen na „rozpoznávání vzorů“.*

4.1 Formální zavedení problému

Stejně tak, jak se mohou různit aplikace rozpoznávání vzorů, i samotný pojem „rozpoznávání vzorů“ bývá chápán různě. V nejzákladnější podobě se jedná o problém určení, zda-li pozorovaný řetězec odpovídá předem stanovenému vzoru. Vzorem bývá obvykle také řetězec, ale může jím být například regulérní výraz. Také - může nás zajímat buď exaktní shoda pozorovaného řetězce se vzorem, nebo jen nějaká forma podobnosti.

V rozšířeném smyslu může být problém chápán jako klasifikace. Tedy, určení třídy, do které by měl pozorovaný řetězec spadat, typicky na základě podobnosti s vybranými reprezentanty jednotlivých tříd.

V této kapitole se však budeme zabývat pouze určováním podobnosti vzorového a pozorovaného řetězce. U každé instance problému budeme znát abecedu se kterou pracujeme a také vzor. Vzorem bude libovolný řetězec nad touto abecedou. Řešením tohoto problému pro nějaký, tzv. pozorovaný, vstupní řetězec bude úroveň podobnosti tohoto řetězce s vzorovým. Jako podobnost zde budeme uvažovat reálné číslo z intervalu $[0, 1]$, kde 0 znamená úplnou rozdílnost a 1 úplnou shodu.

Poznámka 4.2. *Vzorový řetězec budeme v této kapitole vždy značit ω , pozorovaný pak α .*

Nyní máme zadefinován problém samotný, nicméně je třeba zdůraznit, že v jeho definici se používá vágní pojem „podobnost řetězců“. Podobnost řetězců je totiž pojem, který souvisí s konkrétní instancí problému a nelze jej nějak přesně, ale současně dostatečně obecně popsat. Jediné, co o podobnosti řetězců můžeme říct, je, že čím vyšší toto číslo je, tím by si měly být řetězce podobnější.

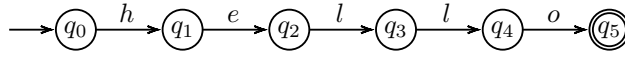
Například, budeme-li porovnávat vstup zadaný z klávesnice počítače oproti nějakému vzoru, je možné, že uživatel udělá překlep. V takovém případě bude vzorovému řetězci určitě více podobný řetězec obsahující dva překlepy (záměna symbolu za některý sousedící na klávesnici) než jiný, který se sice bude lišit jen v jednom symbolu, ale to takovém, který je na opačné straně klávesnice.

Obdobně, pokud budeme pracovat s abecedou malých a velkých písmen (majuskule a minuskule). Uvažujme vzorový řetězec `hello`. Řetězec `HELLO` se s ním neshoduje v ani jednom symbolu, ale přesto jejich podobnost může být blízka jedné.

4.2 Motivace k použití fuzzy automatů

Klasická teorie automatů vznikla jako nástroj pro zpracování textových řetězců. Z tohoto důvodu je rozpoznávání textových vzorů jejím základním výsledkem. Automaty obecně jsou nástroje sloužící pro rozhodování, zda-li řetězec odpovídá vzoru automatem reprezentovanému. Použití pro rozpoznávání řetězcového vzoru tak bude jen speciálním případem jejich užití.

V předchozí podkapitole jsme si stanovili, že řešením našeho problému je číslo z intervalu $[0, 1]$. Z tohoto důvodu nebude možné využít klasické bivalentní automaty. Fuzzy automaty pracují se stupněm pravdivosti, který by mohl s hodnotou podobnosti řetězců korespondovat. Navíc, v praxi se často setkáme s texty, které jsou nepřesné a nedokonalé. Fuzzy přístup by nám tak mohl pomoci na tyto nepřesnosti adekvátně reagovat.



Obrázek 2: Automat rozpoznávající hello

{diag-AutRozpHell}

(zde bude doplněno: a co pravděpodobnostní?) (zde bude doplněno: Protože například: „pozorovaný řetězec se se vzorovým shoduje ve stupni x “ ale „je pravděpodobnost y , že uživatel zadal požadovaný řetězec“)

4.3 Automat rozpoznávající ω

Klíčovým pro rozpoznávání vzorů (chceme-li využívat fuzzy automaty) je bivalentní automat rozpoznávající vzorový řetězec. Tedy automat takový, který přijímá jediný řetězec ω a všechny ostatní zamítá. Nyní si takovýto automat zkonstruujeme.

Uvažujme, že máme k dispozici vzorový řetězec ω nad abecedou Σ . Označme $\mathcal{L}(\omega)$ jako jednoprvkový jazyk obsahující pouze řetězec ω . Vzhledem k tomu, že jazyk $\mathcal{L}(\omega)$ je konečný, je také regulérní a existuje tak konečný deterministický automat, který jej rozpoznává.

Automat bude v každém kroku konzumovat symboly ze vstupního řetězce a porovnávat je se symboly vzorového řetězce na odpovídajících pozicích. Pokud dojde ke shodě na všech pozicích, automat dojde do koncového stavu a sledovaný řetězec přijme. Pokud se symboly shodovat nebudou, automat nebude mít definován žádný odpovídající přechod, kterým by pokračoval ve výpočtu, a řetězec tak zamítne.

Takovýto automat označme jako *automat rozpoznávající ω* .

Definice 4.1 (Automat rozpoznávající ω (deterministický)). *Mějme řetězec ω délky n nad abecedou Σ . Automat rozpoznávající ω je pak konečný automat $A(\omega) = (Q, \Sigma, \delta, q_0, F)$ takový, že jeho množina stavů Q se sestává z právě n stavů q_0, \dots, q_n , q_0 je počáteční stav, $F = \{q_n\}$ množina koncových stavů a δ je přechodová funkce definována pro všechna $0 \leq k < n$ následovně:*

$$\delta(q_k, a_k) = q_{k+1} \text{ kde } a_k \text{ je } k\text{-tý symbol řetězce } \omega$$

Tato definice automatu je vcelku intuitivní. K stejnému výsledku bychom došli, pokud bychom automat zkonstruovali konverzí gramatiky nebo regulérního výrazu.

Příklad 4.1. *Příklad automatu rozpoznávající řetězec $\omega = \text{hello}$ se nachází na obrázku 2.*

My však budeme potřebovat fuzzy automat rozpoznávající ω . To znamená, že musíme nejdříve automat z předchozí definice převést na nedeterministický a poté dle definice 1.11 k němu zkonstruovat odpovídající fuzzy automat.

{def-AutRozpOme}

Definice 4.2 (Automat rozpoznávající ω (nedeterministický)). *Mějme řetězec ω nad abecedou Σ z předchozí definice. Nedeterministický automat rozpoznávající ω je pak konečný automat $A'(\omega) = (Q, \Sigma, \delta, I, F)$ takový, že jeho množina stavů Q je stejná jako v předchozí definici, dále $I = \{q_0\}$ je množina počátečních a*

$F = \{q_n\}$ množina koncových stavů a δ je přechodová funkce definována pro všechna $0 \leq k < n$ následovně:

$$\delta(q_k, a_k) = \begin{cases} \{q_{k+1}\} & \text{pokud je } a_k \text{ } k\text{-tý symbol řetězce } \omega \\ \emptyset & \text{jinak} \end{cases}$$

Následuje vytvoření fuzzy automatu.

{def-FuzzAutRozp0me}

Definice 4.3 (Fuzzy automat rozpoznávající ω). Mějme řetězec ω nad abecedou Σ délky n . Fuzzy automat rozpoznávající ω je pak automat $A''(\omega)$ vytvořený z nedeterministického automatu rozpoznávající ω (definice 4.2) dle definice 1.11. Bude to tedy automat $A''(\omega) = (Q, \Sigma, \mu, \sigma, \epsilon)$ kde

- $\sigma(q_0) = 1$ a $\sigma(q_i) = 0$ pro všechna $i > 0$
- $\epsilon(q_n) = 1$ a $\epsilon(q_i) = 0$ pro všechna $i < n$
- $\mu(q_k, a_k, q_{k+1}) = \begin{cases} 1 & \text{pokud je } a_k \text{ } k\text{-tý symbol řetězce } \omega \\ 0 & \text{jinak} \end{cases}$

Nyní máme k dispozici fuzzy automat, který ostře rozpoznává vzorový řetězec. V následujících podkapitolách následuje výčet několika technik, které tuto ostrost (pomocí dalších informací) odstraňují a nahrazují podobností.

4.4 Podobnost symbolů

Nezákladnější technika pro zanesení neostrého (stupňovitého) rozpoznávání je s využitím podobnostní relace symbolů. Tato technika byla přejata z [2]. Myšlenkou této techniky je, že symbol v pozorovaném řetězci může být snadno zaměněn za jiný, podobný, jemu odpovídající v řetězci vzorovém.

Pro realizaci této techniky je potřeba mít k dispozici fuzzy relaci $g_s : \Sigma \times \Sigma \rightarrow [0, 1]$. Tato relace popisuje podobnost dvojice symbolů. Tedy, je-li pro nějakou dvojici symbolů $x, y \in \Sigma$ $g_s(x, y) = 0$, pak se jedná o naprosto rozdílné symboly. Naopak, pokud bude $g_s(x, y) = 1$, pak se jedná o shodné symboly. Je zjevné, že by relace g_s měla být symetrickou a reflexivní. (zde bude doplněno: v článku to nepíše, ale měla by to být relace ekvivalence (Sym, Ref, Tra). Existuje něco, jako fuzzy relace ekvivalence?)

Příklad 4.2. Jako příklad podobnostní relace (nad abecedou písmen anglické abecedy) může posloužit například vzdálenost patřičných kláves na klávesnici. V takovém případě by určitě platilo kupříkladu $g_s(a, s) > g_s(a, d) > g_s(a, l)$. Protože klávesy *A* a *S* jsou si blíže (a tudíž symboly *a* a *s* jsou si „podobnější“) než například *A* a *D* či *A* a *L*.

Jiným příkladem může být například vizuální podobnost napsaných (malých psacích) písmen. V takovém případě by zřejmě platilo $g_s(a, o) > g_s(m, t)$, protože malá psací písmena *a* a *o* jsou si vizuálně podobnější než *m* a *t*, která vypadají úplně rozdílně.

Máme-li k dispozici relaci g_s , je nutné ji zakomponovat do automatu. Jak autoři uvádějí, tato technika může pracovat s libovolným konečným automatem. Podíváme se proto nejdříve, jak využít relaci g_s obecně. Následně ji aplikujeme na automat rozpoznávající ω , čímž získáme nástroj pro podobnostní rozpoznávání textového vzoru.

{def-AutPracGS}

Definice 4.4 (Automat pracující s g_s). *Uvažujme, že máme nedeterministický automat A a relaci podobnosti symbolů g_s . Pak k automatu A můžeme zkonstruovat fuzzy automat A' , který navíc pracuje s g_s . Takový automat bude zkonstruován dle definice 1.11 s tím rozdílem, že přechodová funkce μ bude definována pro všechna $q_i, q_j \in Q$ a $x \in \Sigma$ následovně:*

$$\mu(q_i, x, q_j) = \bigvee_{y \in \Sigma} (g_s(x, y) \wedge \delta_y(q_i, q_j))$$

$$\text{kde } \delta_x(q_i, q_j) = \begin{cases} 1 & \text{pokud } q_j \in \delta(q_i, x) \\ 0 & \text{pokud } q_j \notin \delta(q_i, x) \end{cases} \text{ pro všechna } q_i, q_j \in Q \text{ a } x \in \Sigma.$$

Definice je vcelku přímočará. Pro každý přechod ze stavu q_i do stavu q_j přes symbol x , procházíme přechody původního automatu. Obsahovala-li přechodová funkce původního automatu přechod ze stavu q_i přes symbol y do stavu q_j , pak je $g_s(x, y) \wedge \delta_y(q_i, q_j)$ rovno podobnosti x a y . V opačném případě je roven nule. Hodnota tohoto výrazu je díky spojení přes všechny symboly maximalizována.

Nyní aplikujeme tento způsob konstrukce fuzzy automatu na automat rozpoznávající ω .

Definice 4.5 (Automat rozpoznávající ω pracující s g_s). *Mějme abecedu Σ , řetězec ω nad touto abecedou a fuzzy relaci g_s nad touto abecedou. Dle definice 4.2 můžeme zkonstruovat nedeterministický bivalentní automat $A(\omega)$ rozpoznávající ω . Jako automat rozpoznávající ω pracující s g_s označme nedeterministický fuzzy automat $A'(\omega)$, který byl z automatu $A(\omega)$ vytvořen podle definice 4.4.*

(zde bude doplněno: neměl by se takovýto automat místo $A'(\omega)$ značit třeba $A_{g_s}(\omega)$?)

Následuje jednoduchý příklad takového automatu.

{ex-AutRozpOmePodSym}

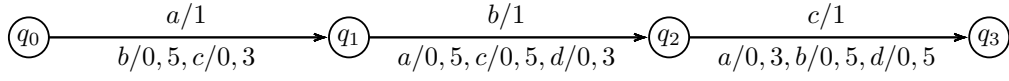
Příklad 4.3. *Mějme abecedu $\Sigma = \{a, b, c, d\}$. Dále uvažujme relaci podobnosti symbolů g_s takovou, že*

- každý symbol je podobný sám sobě ve stupni 1
- každý symbol je podobný symbolu ve stupni 0,5 jedná-li se o symboly reprezentující sousedící písmena abecedy
- každý symbol je podobný symbolu ve stupni 0,3 jedná-li se o symboly reprezentující ob-jedno písmeno sousedící písmena abecedy
- všechny ostatní dvojice symbolů jsou si podobny ve stupni 0

Tuto relaci můžeme zapsat do matice (sloupce i řádky odpovídají po řadě symbolům a, b, c, d):

$$g_s = \begin{pmatrix} 1,0 & 0,5 & 0,3 & 0,0 \\ 0,5 & 1,0 & 0,5 & 0,3 \\ 0,3 & 0,5 & 1,0 & 0,5 \\ 0,0 & 0,3 & 0,5 & 1,0 \end{pmatrix}$$

Nyní mějme vzorový řetězec $\omega = abc$. Pak můžeme podle předchozí definice sestavit automat $A(\omega)$ rozpoznávající ω pracující s g_s . Přechodový diagram takového automatu je na obrázku 3.



Obrázek 3: Automat rozpoznávající abc pracující s g_s

{diag-AutRozpABCPracGS}

Tento automat evidentně rozpoznává řetězec abc ve stupni 1. Pokud v pozorovaném řetězci nahradíme symbol a za b , bude jej automat přijímat ve stupni 0, 5. Pokud nahradíme b za d , bude jej automat přijímat ve stupni 0, 3.

Pokud na začátek pozorovaného řetězce vložíme symbol a (tedy $\alpha = aabc$), automat jej přijme ve stupni 0. Stejnětak, pokud odebereme symbol c z konce vzorového řetězce (tedy $\alpha = ab$). Pokud vložíme symbol a na začátek a současně odebereme c z konce vzorového řetězce, obdržíme pozorovaný řetězec $\alpha = aab$. Tento řetězec bude přijat ve stupni 0, 5 (zde bude doplněno: $1 \otimes 0, 5 \otimes 0, 5$, záleží tedy na \otimes).

Z příkladu jasně vyplývá, že automat pracující s g_s je schopen akceptovat pouze náhradu symbolu jiným symbolem. Bude-li pozorovaný řetězec oproti vzorovému obsahovat vložený symbol nebo naopak z něj bude symbol odebrán, tento typ automatu selže. Na druhou stranu jeho princip i konstrukce jsou jednoduché a snadno se s nimi pracuje.

4.5 Fuzzy symboly

Fuzzy symbol je technika využívající podobnosti symbolů. Ve své podstatě se jedná o téže techniku jak v předchozí podkapitole, jen je na ni nahlíženo jinak. Oproti podobnosti symbolů je použití fuzzy symbolů komplikovanější, avšak umožňuje jednoduše tuto techniku kombinovat s jinými. Princip fuzzy symbolů byl přejat z [8].

Mějme abecedu Σ a relaci p podobnosti symbolů (stejně jako relace g_s v předchozí podkapitole). Fuzzy symbolem symbolu $x \in \Sigma$ označujeme fuzzy množinu symbolů takových, které jsou podle relace p symbolu x „podobné“.

Definice 4.6 (Fuzzy symbol). *Mějme abecedu Σ a fuzzy relaci $p \subseteq \Sigma \times \Sigma$. Pak pro každý symbol $y \in \Sigma$ definujeme fuzzy symbol \tilde{y} symbolu y jako fuzzy množinu nad Σ takovou, že pro všechna $x \in \Sigma$ platí*

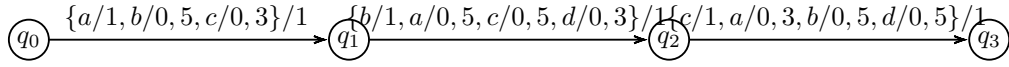
$$\tilde{y}(x) = p(y, x)$$

Vzhledem k tomu, že fuzzy symbol máme definován pro všechny $y \in \Sigma$, můžeme množinu všech takových fuzzy symbolů nazvat abecedou fuzzy symbolů.

Definice 4.7 (Abeceda fuzzy symbolů). *Mějme abecedu Σ a fuzzy symboly \tilde{y} pro všechna $y \in \Sigma$. Pak množinu všech těchto fuzzy symbolů nazvěme abeceda fuzzy symbolů abecedy Σ a označme $\tilde{\Sigma}$. Tedy $\tilde{\Sigma} = \{\tilde{y} \mid y \in \Sigma\}$.*

Máme-li abecedu fuzzy symbolů $\tilde{\Sigma}$, můžeme pracovat s řetězci $\tilde{\alpha} \in \tilde{\Sigma}^*$ nad touto abecedou. Ještě si však doplníme, jak vytvořit k řetězci $\alpha \in \Sigma^*$ jemu odpovídající řetězec fuzzy symbolů $\tilde{\alpha} \in \tilde{\Sigma}^*$.

(zde bude doplněno: sjednotit značení ω vs. α , když se používá jen jeden obecný řetězec)



Obrázek 4: Automat rozpoznávající abc pracující s g_s

{diag-AutRozpOmePraFuzSym}

Definice 4.8 (Řetězec fuzzy symbolů). *Mějme abecedu Σ a nějaký řetězec $a_1 \dots a_n = \alpha \in \Sigma^*$. Pak definujme $\tilde{\alpha} = \tilde{a}_1 \dots \tilde{a}_n$ jako řetězec fuzzy symbolů řetězce α .*

V této fázi jsme schopni plnohodnotně pracovat s řetězcí fuzzy symbolů a konstruovat je z řetězců nad abecedou Σ . Nyní přejdeme k návrhu fuzzy automatu, který bude s fuzzy symboly pracovat. Stejně jako u podobnosti symbolů i fuzzy symboly mohou být aplikovány na libovolný typ automatu. Vytvoříme proto automat pracující s fuzzy symboly nejdříve obecně, pro libovolný automat A .

{def-AutPracFuzzSym}

Definice 4.9 (Automat pracující s fuzzy symboly). *Mějme nedeterministický fuzzy automat A . Pak fuzzy automat \tilde{A} pracující s fuzzy symboly vytvoříme tak, že v definici automatu A nahradíme Σ za $\tilde{\Sigma}$.*

(zde bude doplněno: může to tak být? a co související pojmy)

Formální zavedení automatu pracujícího s fuzzy symboly je intuitivní, jedná se jen o formalitu. Abychom však využili potenciál fuzzy symbolů, je třeba pozměnit výpočet automatu. Proces jeho výpočtu se změní ve fázi výpočtu přechodové funkce fuzzy stavů. Připomeňme, že ta je definována (definice 1.7) jako fuzzy relace $\hat{\mu}$ přiřazující každému fuzzy stavu V a fuzzy symbolu x fuzzy stav dle předpisu

$$\hat{\mu}(V, x) = V \circ \mu[x]$$

Zde je zjevně nutné nahradit $\mu[x]$ spojením přes všechny fuzzy symboly. Bude tedy vypadat následovně:

$$\hat{\mu}(V, x) = V \circ \bigvee_{y \in \Sigma} (\mu[x] \wedge \tilde{x}(y))$$

Tím, že je změna zakořeněna ve výpočtu automatu, nám umožňuje další práci se samotným automatem. Můžeme tedy bez problémů zkonstruovat automat rozpoznávající ω pracující s fuzzy symboly.

Definice 4.10 (Automat rozpoznávající ω pracující s fuzzy symboly). *Mějme abecedu Σ , řetězec ω nad touto abecedou a abecedu fuzzy symbolů $\tilde{\Sigma}$. Dle definice 4.3 můžeme zkonstruovat fuzzy automat $A(\omega)$ rozpoznávající ω . Následně pak podle definice 4.9 automat $\tilde{A}(\omega)$ rozpoznávající ω pracující s fuzzy symboly.*

Postup konstrukce takového automatu je opět vcelku intuitivní. Následuje demonstrace na příkladu.

Příklad 4.4. *Mějme abecedu Σ , vzorový řetězec a podobnostní relaci $p \equiv g_s$ stejné jako v příkladu 4.3. Na obrázku 4 je zobrazen diagram automatu $\tilde{A}(\omega)$ rozpoznávající ω pracující s fuzzy symboly.*

Co se týče vlastností automatů (rozpознаvajících ω) pracujících s fuzzy symboly, jejich charakteristika je vesměs stejná jako u automatů pracujících s podobností symbolů. Pouze, jak již bylo zmíněno v úvodu, nezasahují do struktury automatu jako takového.

4.6 Editační operace

Další technikou pro podobnostní porovnávání pozorovaného a vzorového řetězce je s využitím editačních operací. Tato technika byla přejata z [7]. Základní idea této techniky spočívá v trojici jednoduchých editačních operací, jejíž složením jsme schopni popsat transformaci pozorovaného řetězce na vzorový. Množství transformace pak udává podobnost pozorovaného a vzorového řetězce.

Následuje formální definice editačních operací a pojmů s nimi souvisejících. Následně přejdeme ke konstrukci automatu, který s nimi bude schopen pracovat.

Definice 4.11 (Editační operace). *Mějme abecedu Σ , uvažujme množinu $E = (\Sigma \cup \{\epsilon\}) \times (\Sigma \cup \{\epsilon\}) \setminus \{(\epsilon, \epsilon)\}$. Pak každou dvojici $(x, y) = z \in E$ nazvěme editační operace. Speciálně pak, pro všechna $x, y \in \Sigma$, $(x, y) \in E$ znamená nahrazení symbolu x symbolem y , $(x, \epsilon) \in E$ znamená odebrání symbolu x a naopak $(\epsilon, y) \in E$ pak vložení symbolu y . Navíc jako editační operaci uvažujeme i všechny dvojice $(x, x) \in E$ (pro každé $x \in \Sigma$) symbolizující „žádnou editaci“.*

Máme-li editační operaci $(x, y) = z \in E$, pak označme $x = z^\downarrow$ a $y = z^\uparrow$.

Editační operace jsou tedy tři a to náhrada symbolu, vložení symbolu a odebrání symbolu. Například řetězec **hallo** vznikl záměnou **e** za **a** v řetězci **hello**. Obdobně, řetězec **hellow** vznikl přidáním **w** na konec a řetězec **helo** odebráním (prvního nebo druhého) symbolu **l**.

My však obvykle očekáváme, že došlo k více, než jedné jednoduché editaci. Je proto vhodné zavést koncept mnohanásobné editace. Jednotlivé editace za sebe seřadíme do posloupnosti v pořadí, v jakém mají být postupně aplikovány, a takovouto posloupnost nazvěme vyrovnaním řetězce α na řetězec ω .

Uvažujme nyní množinu E editačních operací jako abecedu. Pak každé vyrovnaní ζ řetězce α na řetězec ω (posloupnost $z_1 z_2 \dots z_n$ symbolů $z_1, z_2, \dots, z_n \in E$), tak můžeme považovat za řetězec nad abecedou E .

(zde bude doplněno: fakt E považovat za abecedu a G za jazyk? není to zbytečná komplikace? je to tam nutné?)

Například všechny tři následující řetězce jsou vyrovnaním řetězce **ahoj** na řetězec **hello**:

$$\begin{aligned}\zeta_1 &= (a, \epsilon)(h, h)(o, e)(j, l)(\epsilon, l)(\epsilon, o) \\ \zeta_2 &= (a, \epsilon)(h, \epsilon)(o, \epsilon)(j, \epsilon)(\epsilon, h)(\epsilon, e)(\epsilon, l)(\epsilon, l)(\epsilon, o) \\ \zeta_3 &= (a, h)(h, e)(o, l)(j, l)(\epsilon, o)\end{aligned}$$

Na tomto příkladu je vhodné si povšimnout, že obecně může existovat více než 1 vyrovnaní mezi libovolnou dvojicí řetězců. Bude proto vhodné neuvažovat vyrovnaní jednotlivá, ale množinu všech možných vyrovnaní mezi dvojicí řetězců.

Podíváme-li se nyní jen na levé části editačních operací ve vyrovnaní ζ_1 z předchozího příkladu, zjistíme, že jejich zřetězením získáme řetězec α :

$$(a, \epsilon)^\downarrow (h, h)^\downarrow (o, e)^\downarrow (j, l)^\downarrow (\epsilon, l)^\downarrow (\epsilon, o)^\downarrow = ahoj$$

Stejně tak, zřetězením pravých částí editačních operací v ζ_1 získáme řetězec ω :

$$(a, \epsilon)^\uparrow (h, h)^\uparrow (o, e)^\uparrow (j, l)^\uparrow (\epsilon, l)^\uparrow (\epsilon, o)^\uparrow = hello$$

Tato vlastnost nám udává, v jakém pořadí mají být editační operace aplikovány. Stejně tak nám odstraňuje nadbytečné editační operace (např. opakované

přidávání a odebírání téže znaku, které by mohlo vést až k nekonečné posloupnosti editací). Proto nám tato vlastnost poslouží jako definiční pro formání zavedení vyrovnání řetězců.

Definice 4.12 (Vyrovnání řetězců [7]). *Jako množinu všech vyrovnání $G(\alpha, \omega)$ řetězce α na řetězec ω (kde $(\epsilon, \epsilon) \neq \alpha, \omega \in \Sigma^*$) označme takovou množinu $\{\zeta \in E^+ \mid \zeta \text{ splňuje vlastnosti 1., 2. i 3.}\}$*

1. $\zeta = z_1 z_2 \dots z_r, z_i \neq (\epsilon, \epsilon)$ pro všechna $i \in 1, \dots, r$,
2. $z_1^\downarrow z_2^\downarrow \dots z_r^\downarrow = \alpha$
3. $z_1^\uparrow z_2^\uparrow \dots z_r^\uparrow = \omega$

V tento okamžik máme formálně zavedena vyrovnání řetězců. Můžeme tedy přejít k práci s nimi. Ukážeme si způsob, jak pomocí vyrovnání řetězců spočítat podobnost dvojice řetězců. Na základě tohoto výpočtu pak sestavíme automat, který tento výpočet bude realizovat.

Pro určení podobnosti na základě vyrovnání řetězců budeme potřebovat znát míry pravdivosti editačních operací. Vstupem pro výpočet podobnosti řetězců tak bude navíc binární fuzzy relace R nad množinou všech editačních operací (E), udávající stupeň akceptovatelnosti každé z možných editačních operací. S touto znalostí můžeme nadefinovat relaci podobnosti řetězců, tzv. fuzzy míru dvojice řetězců α a ω .

(zde bude doplněno: Musí být R reflexivní a symetrická (def. automatu to vyžaduje, ale je to nutné?). A co T -tranzitivita?) (zde bude doplněno: Takové relaci se říká relace podobnosti (proximity relation))

Definice 4.13 (Fuzzy míra [7]). *Mějme binární fuzzy relaci R nad $\Sigma \cup \{\epsilon\}$. Pak jako fuzzy míru mezi řetězci $\alpha, \omega \in \Sigma^*$ (značenou $S_{\Sigma, R, \otimes}$) označme fuzzy relaci danou následujícím předpisem:*

{def-FuzzMir}

$$S_{\Sigma, R, \otimes} = \begin{cases} 1 & \text{pokud } (\alpha, \omega) = (\epsilon, \epsilon) \\ \max_{\zeta \in G(\alpha, \omega)} (\bigotimes_{i=1}^{|\zeta|} R(\zeta_i)) & \text{pokud } (\alpha, \omega) \neq (\epsilon, \epsilon) \end{cases}$$

Definice fuzzy míry je vcelku intuitivní. Počítá se míra všech možných vyrovnání z nichž se vybírá ta největší. Míra vyrovnání se určuje jako t-norma ze všech $R(\zeta_i)$, tedy stupňů akceptovatelnosti jednotlivých editačních operací. Navíc, míra mezi dvojicí prázdných řetězců je dodefinována jako 1.

Označme $\mathcal{L}(\omega)$ jako fuzzy jazyk řetězců „podobných“ řetězci ω s podobností danou relací R . Takový jazyk pak můžeme nadefinovat pro všechna $\alpha \in \Sigma^*$ následujícím předpisem

$$\mathcal{L}(\omega)(\alpha) = S_{\Sigma, R, \otimes}(\alpha, \omega)$$

Nyní zkonstruujeme nedeterministický fuzzy automat s ϵ -přechody, který jazyk \mathcal{L} rozpoznává. (zde bude doplněno: rozpoznává vs. přijímá, pozor na to)

{def-AutRozpCall}

Definice 4.14 (Automat rozpoznávající \mathcal{L} [7]). *Mějme binární fuzzy relaci R nad $\Sigma \cup \{\epsilon\}$ (stejná jako v definici 4.13). Pak pro vzorový řetězec $a_1 a_2 \dots a_n = \omega \in \Sigma^*$ označme $M_{\Sigma, R, \otimes}(\omega)$ automat rozpoznávající jazyk $\mathcal{L}(\omega)$ dle definice 4.10, takový, že*

1. množina stavů $Q = \{q_0, q_1, \dots, q_n\}$
2. fuzzy přechodová funkce μ pro všechny $x \in \Sigma$:
 - (a) $\mu(q_i, q_i, x) = R(x, \epsilon)$ pro všechny $q_i \in Q$ taková, že $i = 0, \dots, n$
 - (b) $\mu(q_i, q_{i+1}, x) = R(x, a_{i+1})$ pro všechny $q_i, q_{i+1} \in Q$ taková, že $i = 0, \dots, n-1$
 - (c) $\mu(q, q', x) = 0$ pro všechny $q, q' \in Q$ nesplňující předchozí dva body
 - (d) $\mu(q_i, q_i, \epsilon) = 1$ pro všechny $q_i \in Q$ taková, že $i = 0, \dots, n$
 - (e) $\mu(q_i, q_{i+1}, \epsilon) = R(\epsilon, a_{i+1})$ pro všechny $q_i \in Q$ taková, že $i = 0, \dots, n-1$
 - (f) $\mu(q, q', \epsilon) = 0$ pro všechny $q, q' \in Q$ nesplňující předchozí dva body
3. množina počátečních stavů $\sigma: \sigma(q_0) = 1$ a pro všechny ostatní $q_0 \neq q' \in Q$: $\sigma(q') = 0$
4. množina koncových stavů $\eta: \eta(q_n) = 1$ a pro všechny ostatní $q_n \neq q' \in Q$: $\eta(q') = 0$

Máme nadefinován fuzzy automat rozpoznávající jazyk $\mathcal{L}(\omega)$. Bylo by však vhodné dokázat, že jazyk $\mathcal{L}(\omega)$, který tento automat rozpoznává je skutečně jazykem řetězců podobných řetězci ω s podobností danou relací R . Vzhledem ke složitosti důkazu tohoto tvrzení se v této práci spokojíme pouze s ilustrací na příkladu.

Věta 4.1. *Mějme binární fuzzy relaci R nad $\Sigma \cup \{\epsilon\}$ (stejná jako v definici 4.13) a vzorový řetězec $\omega \in \Sigma^*$. Pak pro automat $M_{\Sigma, R, \otimes}(\omega)$ sestavený dle předcházející definice a fuzzy míru $S_{\Sigma, R, \otimes}$ platí následující rovnost*

$$\mathcal{L}(M_{\Sigma, R, \otimes}) = \mathcal{L}(\omega)$$

Důkaz. Kompletní důkaz je k nalezení v [7]. □

(zde bude doplněno: sazba symbolů v matematickém módu vs. verbatim řetězce v textovém)

Příklad 4.5. *Uvažujme abecedu $\Sigma = \{a, b, c\}$ a vzorový řetězec $\omega = abc$. Zkonstruujeme automat, který bude akceptovat ve stupni 0.5 náhradu symbolu x symbolem s s ním v abecedě sousedícím. Navíc uvažujme vložení symbolu a ve stupni 0.2 a odebrání symbolu c ve stupni 0.1. Tedy, relace R bude vypadat následovně:*

$$R = \{(a, a)/1, (b, b)/1, (c, c)/1, \\ (b, a)/0.5, (a, b)/0.5, (c, b)/0.5, (b, c)/0.5, \\ (a, \epsilon)/0.2, (\epsilon, c)/0.1\}$$

Dle definice 4.14 můžeme sestavit automat $M_{\Sigma, R, \otimes}$. Jako \otimes použijme produktovou t -normu. Získáme tak automat $M_{\Sigma, R, \otimes} = (Q, \Sigma, \mu, \sigma, \epsilon)$ takový, že:

1. $Q = \{q_0, q_1, q_2, q_3\}$
2. $\mu = \{$
 - (a) $(q_0, q_0, a)/0.2, (q_1, q_1, a)/0.2, (q_2, q_2, a)/0.2, (q_3, q_3, a)/0.2,$

- (b) $(q_0, q_1, a)/1, (q_1, q_2, a)/0.5,$
 $(q_0, q_1, b)/0.5, (q_1, q_2, b)/1, (q_2, q_3, b)/0.5,$
 $(q_1, q_2, c)/0.5, (q_2, q_3, c)/1,$
(d) $(q_0, q_0, \epsilon)/1, (q_1, q_1, \epsilon)/1, (q_2, q_2, \epsilon)/1, (q_3, q_3, \epsilon)/1,$
(e) $(q_2, q_3, \epsilon)/0.1$

} (přechody dle bodů (b) a (d) v definici jsou s nulovým stupněm a ve výpisu jsou vynechány)

3. $\sigma = \{q_0/1, q_1/0, q_2/0, q_3/0\}$

4. $\eta = \{q_0/0, q_1/0, q_2/0, q_3/1\}$

(zde bude doplněno: $\sigma = \{x/y, \dots\}$ by se mělo přepsat na $\sigma(x) = y, \dots$, ne?)

Přechodový diagram tohoto automatu je k nalezení na obrázku 5. V přechodovém diagramu jsou červeně zvýrazněny pravidla pro rozpoznávání ω , ostatní pravidla (doplněna dle definice) jsou černá.

Nyní si na pár řetězcích zkusíme ukázat platnost věty 4.1. Dle definice 1.8 spočítáme stupeň, v jakém automat náš testovací rozpoznává řetězec α :

$$\begin{aligned} M_{\Sigma, R, \otimes}(\alpha) &= \max_{q \in Q} (\mu^*(\sigma, \alpha)(q) \otimes \eta(q)) \\ &= \max\{\mu^*(\sigma, \alpha)(q_0) \otimes \eta(q_0), \mu^*(\sigma, \alpha)(q_1) \otimes \eta(q_1), \\ &\quad \mu^*(\sigma, \alpha)(q_2) \otimes \eta(q_2), \mu^*(\sigma, \alpha)(q_3) \otimes \eta(q_3)\} \\ &= \max\{\mu^*(\sigma, \alpha)(q_0) \otimes 0, \mu^*(\sigma, \alpha)(q_1) \otimes 0, \\ &\quad \mu^*(\sigma, \alpha)(q_2) \otimes 0, \mu^*(\sigma, \alpha)(q_3) \otimes 1\} \\ &= \mu^*(\sigma, \alpha)(q_3) \end{aligned}$$

- řetězec $\alpha = abc$: Určíme stupeň akceptance automatem:

$$M_{\Sigma, R, \otimes}(\alpha) = \mu^*(\sigma, abc)(q_3) = \hat{\mu}(\hat{\mu}(\hat{\mu}(\hat{\mu}(\sigma, \epsilon), a), b), c)(q_3) = \{q_3/1\}(q_3) = 1$$

A následně ověříme fuzzy míru. Množina všech vyrovnaní bude obsahovat například $\zeta_1 = (a, a), (b, b), (c, c)$ či $\zeta_2 = (a, \epsilon)(\epsilon, a)(b, \epsilon)(\epsilon, b)(c, \epsilon)(\epsilon, c)$. Snadno zjistíme, že $\bigotimes_{i=1}^{|\zeta|} R(\zeta_i)$ je maximální právě pro $\zeta = \zeta_1$ a nabývá stupně 1. A tedy $S_{\Sigma, R, \otimes}(\alpha) = 1$.

- řetězec $\alpha = ab$:

$$M_{\Sigma, R, \otimes}(\alpha) = \mu^*(\sigma, ab)(q_3) = \hat{\mu}(\hat{\mu}(\hat{\mu}(\sigma, \epsilon), a), b)(q_3) = \{q_2/1, q_3/0.1\}(q_3) = 0, 1$$

$$S_{\Sigma, R, \otimes}(\alpha) = R(a, a) \otimes R(b, b) \otimes R(\epsilon, c) = 1 \otimes 1 \otimes 0.1 = 0, 1$$

- řetězec $\alpha = bbb$:

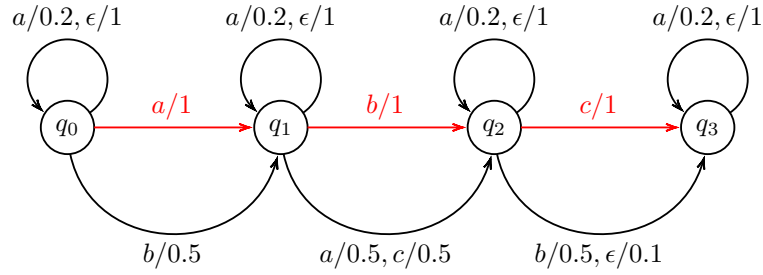
$$M_{\Sigma, R, \otimes}(\alpha) = \mu^*(\sigma, bbb)(q_3) = \dots = \{q_3/0, 25\}(q_3) = 0, 25$$

$$S_{\Sigma, R, \otimes}(\alpha) = R(b, a) \otimes R(b, b) \otimes R(b, c) = 0, 5 \otimes 1 \otimes 0, 5 = 0, 25$$

- řetězec $\alpha = abaca$:

$$M_{\Sigma, R, \otimes}(\alpha) = \mu^*(\sigma, abaca)(q_3) = \dots = \{q_3/0, 04\}(q_3) = 0, 04$$

$$S_{\Sigma, R, \otimes}(\alpha) = R(a, a) \otimes R(b, b) \otimes R(a, \epsilon) \otimes R(c, c) \otimes R(a, \epsilon) = 1 \otimes 1 \otimes 0, 2 \otimes 1 \otimes 0, 2 = 0, 04$$



Obrázek 5: Přechodový diagram automatu z příkladu 4.5

{img-AutRozpCaLL}

- řetězec $\alpha = cba$:

$$M_{\Sigma, R, \otimes}(\alpha) = \mu^*(\sigma, cba)(q_3) = \dots = (\emptyset)(q_3) = 0$$

$$S_{\Sigma, R, \otimes}(\alpha) = R(c, a) \otimes R(b, b) \otimes R(a, c) = 0 \otimes 1 \otimes 0 = 0$$

...

Je tedy zřejmé, že výpočet automatu $M_{\Sigma, R, \otimes}$ je v korespondenci s fuzzy mírou $S_{\Sigma, R, \otimes}$.

Je vidět, že automat zkonstruován dle editačních operací je značně silný nástroj. Umožňuje nám velmi pohodlně popsat, jak moc mohou být konkrétní editační operace akceptovány. Editací operace vložení symbolu, náhrada symbolu a odebrání symbolu jsou pro popis modifikace vzorového řetězce přirozené.

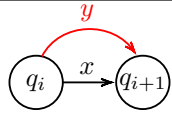
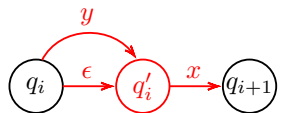
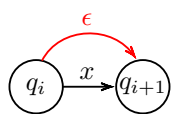
Nevýhodou této techniky je, že jednotlivé editační operace jsou akceptovány bez ohledu na jejich výskyt v řetězci. Automat akceptuje nastanuvší editační operaci pokaždé, kde může nastat, ve stejném stupni. Často je však třeba v jiném stupni stejnou editační operaci přijímat např. na začátku a na konci řetězce pokaždé však v jiném stupni. Tento požadavek však automat zkonstruovaný pomocí editačních operací neumí zpracovat. Řešením může být například použití následující techniky.

4.7 Deformovaný automat

Dalším ze způsobů, jak přijímat řetězec podobný vzorovému je s využitím deformovaného (fuzzy) automatu. Tato technika využívá tzv. deformovaného automatu, neboli automatu který byl stanoveným způsobem upraven, neboli deformován. Tato technika byla přejata z [3].

Jako deformace může být použita prakticky jakákoliv úprava automatu. Mějme fuzzy automat A . Provedením deformace x získáme deformovaný automat A' , který rozpoznává jiný jazyk, než automat A . Mezi nejzákladnější tři deformace patří náhrada symbolu, vložení symbolu před symbol a odebrání symbolu. Možných deformací existuje nekonečně mnoho. Mezi další deformace může patřit například (pro nějaké $x, y, z \in \Sigma$ a $i \geq 0$): „náhrada symbolu x na i -té pozici symbolu yz “, „odebrání všech výskytů symbolu x , které se nachází před symbolem y “ nebo „vložení sudého počtu symbolů y mezi symboly x a z “.

Provedeme-li deformaci fuzzy automatu rozpoznávající ω , můžeme se na trojici základních deformací podívat konkrétně. Ukázka toho, jak by vypadal

Deformace	Význam deformace
<p>NÁHRADA symbolu na i-té pozici (symbolu x) symbolem y $\delta' = \delta \cup \{(q_i, y, q_{i+1})\}, Q' = Q$</p>	
<p>VLOŽENÍ symbolu y na i-tou pozici (před symbol x) $\delta' = \delta \setminus \{(q_i, x, q_{i+1})\} \cup \{(q_i, \epsilon, q'_i), (q_i, y, q'_i), (q'_i, x, q_{i+1})\}, Q' = Q \cup \{q'_i\}$</p>	
<p>ODEBRÁNÍ symbolu z i-té pozice (symbolu x) $\delta' = \delta \cup (q_i, \epsilon, q_{i+1}), Q' = Q$</p>	

Tabulka 1: Deformace deformovaného automatu
(zde bude doplněno: pozor, toto je pro konečné automaty, ne pro fuzzy automaty!)

{tbl-DefAutDef}

deformovaný automat po provedení jedné ze základních deformací je vyobrazeno v tabulce 1.

Je vidět, že deformace mohou být účinným nástrojem pro rozpoznávání modifikovaných pozorovaných řetězců. Na druhou stranu, deformování automatu vyžaduje znalost fungování automatů. Často také může nastat situace, kdy výsledný zdeformovaný automat bude více, než deformovaný automat rozpoznávající ω , automatem reprezentující samostatný netriviální vzor.

4.8 Shrnutí

V této kapitole byl zaveden pojem rozpoznávání textových vzorů. Bylo ukázáno, že pro klasickou teorii automatů je to triviální problém, který však kvůli nepřesnostem reálných dat vyžaduje nasazení fuzzy automatů. Bylo představeno několik technik, které pomocí fuzzy automatů umožňují přijímání řetězců podobných vzorovému.

Nejjednodušší z nich, využívající relaci podobnosti symbolů, je vhodná na prosté nahrazování podobných symbolů. Technika fuzzy symbolů funguje na stejném principu, jen se liší ve formálním zavedení. Technika s využitím editačních operací umožňuje specifikovat stupeň akceptance základních editačních operací (vlození symbolu, odebrání symbolu, náhrada symbolu). Poslední technika, deformovaný automat, umožňuje libovolnou transformaci automatu, vedoucí až k libovolnému vzoru.

5 Konkrétní příklady

5.1 Rozpoznávání ručně psaného textu

...

5.2 Detekce překlepů

[4] až na konci, Fuzzy Aho-Corasick algorithm (popř. dohledat jiný zdroj).

Reference

- [1] Radim Bělohlávek. *Fuzzy Relational Systems: Foundations and Principles*. Kluwer, New York, 2002.
- [2] H. A. Girijamma Dr. V. Ramaswamy. Conversion of finite automata to fuzzy automata for string comparison. *International Journal of Computer Applications*, 2012.
- [3] J. J. Astrain; J. R. Garitagoitia; J. R. Gonzalez De Mendivil; J. Villadangos; F. Farina. Approximate string matching using deformed fuzzy automata: A learning experience. *Springer Science & Business Media*, 2004.
- [4] Abdulwahed Almarimi Gabriela Andrejková and Asmaa Mahmoud. Approximate pattern matching using fuzzy logic. *CEUR Workshop Proceedings*, 2003.
- [5] S.S. Yau G.F. DePalma. Fractionally fuzzy grammars with application to pattern recognition. *US–Japan Seminar on Fuzzy Sets and their Applications*, 1974. článek jako e-book: <http://bit.ly/2cumxjz>, výcuc v MorMal-FuzzAutAndLangs 10.7.
- [6] Kou-Yuan Huang. *Syntactic Pattern Recognition for Seismic Oil Exploration*. World Scientific, 2002.
- [7] J.R. Garitagoitia J. Astrain, J.R. González de Mendivil. Fuzzy automata with ϵ -moves compute fuzzy measures between strings. *Fuzzy Sets and Systems* 157, 2005.
- [8] José R. Garitagoitia Carlos F. Alastrueya Javier Echanobe, José R. Gonzáles Mendivil.
- [9] José R. Garitagoitia José R. González de Mendivil. Fuzzy languages with infinite range accepted by fuzzy automata: Pumping lemma and determination procedure. *Fuzzy sets and Systems*, 2014.
- [10] S. C. Kremer M. Doostfateme. New directions in fuzzy automata. *International Journal of Approximate Reasoning*, 2004.
- [11] M. M. ZAHEDI S. MOGHARI and R. AMERI. New direction in fuzzy tree automata. *Iranian Journal of Fuzzy Systems*, 2011.
- [12] Mukta N. Joshi S. R. Chaudhari. A note on fuzzy tree automata. *International Journal of Computer Applications*, 2012.
- [13] Miroslav Stamenkovic, Aleksandar; Ciric. Construction of fuzzy automata from fuzzy regular expressions. *Fuzzy Sets and Systems*, 2012.
- [14] Witold Pedrycz Yongming Li. Fuzzy finite automata and fuzzy regular expressions with membership values in lattice-ordered monoids. *Fuzzy Sets and Systems*, 2005.