

Interakce člověk–počítač v přirozeném jazyce (ICP)

LS 2013 — Úvod

Tino Haderlein, Elmar Nöth

Katedra informatiky a výpočetní techniky (KIV)
Západočeská univerzita v Plzni

Lehrstuhl für Mustererkennung (LME)
Friedrich-Alexander-Universität Erlangen-Nürnberg

Co bylo řečeno?

- signál 1
- signál 2

Co bylo řečeno?

- *It is not hard to recognize speech*
- *It is not hard to wreck a nice beach*

signál 2

signál 1

Znalost o světě: Toto je přednáška o
rozpoznávání řeči,
ne o
cestovním ruchu a životním prostředí.

Statistické rozpoznávání vzorů

- Klasifikátor, který minimalizuje riziko chybné klasifikace, využívá Bayesovo pravidlo:

$$p(\Omega|\mathbf{O}) = \frac{p(\Omega) \cdot p(\mathbf{O}|\Omega)}{p(\mathbf{O})}$$

- Aplikace na rozpoznávání řeči:

\mathbf{O} = signál (nahrávka), Ω = množina všech sekvencí slov

→ rozhodni se pro sekvenci slov \mathbf{w} , která maximalizuje $p(\mathbf{w}|\mathbf{O})$

- Problém:

$p(\mathbf{O}|\text{It is not hard to recognize speech}) <$

$p(\mathbf{O}|\text{It is not hard to wreck a nice beach})$

ale

$p(\text{It is not hard to recognize speech}) \gg$

$p(\text{It is not hard to wreck a nice beach})$

Rozpoznávání vzorů



Obsah

- úvod
- třídy, fonetické základy
- příznaky
- klasifikace
- teorie skrytých Markovových modelů
- Markovové modely, modelace fonů
- jazykové modely
- dekodování
- dialogové systémy

Rozpoznávání vzorů

rozpoznávání vzorů:

automatická transformace sensorového signálu do symbolického popisu, který je specifický pro konkrétní úlohu

One of the most interesting aspects of the world is that it can be considered to be made up of patterns. A pattern is essentially an arrangement. It is characterized by the order of the elements of which it is made rather than by the intrinsic nature of these elements.

Norbert Wiener

porozumění vzorů:

automatické vypočítání symbolického popisu, který je optimálně kompatibilní s uloženými znalostmi a který se optimálně hodí k signálu.

Pattern Recognition

Pattern Recognition:

The automatic transformation of a sensor signal into an application specific symbolic description. The result of this transformation can automatically be processed or be processed by a human expert.

Examples:

1 an address reader for a postal sorting machine

automatic

2 a system that analyzes x-ray pictures w.r.t. tuberculosis, displays them for the doctor and prepares a medical expertise, which the doctor can accept or reject

manual

Both systems can also be semi-automatic:

(only) present the doubtful cases to an expert

(introduction of a rejection class)

Definice

- **Okolí** je celek hodnot, které jsou měřitelné fyzikálními stroji. Je reprezentováno množinou

$$U = \{ {}^\rho b(x) \mid \rho = 1, 2, \dots \}$$

funkcí ${}^\rho b(x)$.

- **Problémový okruh** Ω obsahuje jen objekty nebo funkce, které patří k přísně omezené aplikaci nebo výseku okolí. Je definován množinou

$$\Omega = \{ {}^\rho f(x) \mid \rho = 1, 2, \dots \} \subset U$$

funkcí ${}^\rho f(x)$ a je podmnožinou okolí U .

Definice

- Prvky množiny Ω (problémový okruh) se jmenují **vzory**. Proto je vektor funkce

$$f(\mathbf{x}) = \begin{pmatrix} f_1(x_1, x_2, \dots, x_n) \\ f_2(x_1, x_2, \dots, x_n) \\ \dots \\ f_m(x_1, x_2, \dots, x_n) \end{pmatrix}.$$

Definice

- **Rozpoznávání vzorů** se týká matematických a technických aspektů automatického pátrání po logických zobrazení faktů. To obsahuje
klasifikace jednoduchých vzorů a analýzu a porozumění komplexních vzorů.
- Při klasifikaci jednoduchých vzorů je každý vzor zpracován jako celek a je nezávisle od ostatních vzorů přiřazen jedné třídě Ω_k z k možných tříd $\Omega_\lambda, \lambda = 1, \dots, k$. Je také možné vzor odmítnout, tj. přiřadit jej ke $(k + 1)$. třídě Ω_0 , která je nazývána třída odmítnutých rozhodnutí (rejection class).

Třídy

Ω_{κ} se vznikne segmentací množiny Ω do k (nebo $k + 1$) podmnožin Ω_{κ} , $\kappa = 1, \dots, k$ nebo $\kappa = 0, 1, \dots, k$.

Musí platit:

$$\begin{aligned} \Omega_{\kappa} &\neq \phi, \kappa = 1, \dots, k, \\ \Omega_{\kappa} \cap \Omega_{\lambda} &= \phi, \kappa \neq \lambda, \\ \bigcup_{\kappa=1}^k \Omega_{\kappa} &= \Omega, \text{ anebo } \bigcup_{\kappa=0}^k \Omega_{\kappa} = \Omega \end{aligned}$$

Poznámky:

- Podle téhle definice jsou třídy *disjunkční*,
- *nedisjunkční* třídy mohou být vhodné pro několik aplikací,
- nahoře uvedené definice *nedají nám kritéria* jak najít třídy.

Třídy

Vzory jedné třídy

- by si měli být podobné,
- by se měli lišit od vzorů jiných tříd.

Když vzor nemůžeme spolehlivě klasifikovat, můžeme jej

- buď odmítnout, tj. klasifikovat do třídy Ω_0 ,
- anebo vydat několik alternativ, tj. klasifikovat do $\{\Omega_{\kappa,1}, \Omega_{\kappa,2}, \dots, \Omega_{\kappa,n_{\kappa}}\}$.

Když považujeme v přirozeném jazyce všechny předměty, které se nazýváme jedním slovem, za prvky jedné třídy:

Australští praobyvatelé

- mají **jedno** slovo pro „ženy“, „ohněň“ a „nebezpečné věci“,
- počítají „1, 2, 3, hodně“ a
- mají tucty slov pro klokany.

Postuláty

1 Testovací zkouška:

K dispozici je *reprezentativní* množina $\omega \subset \Omega$ vzorů $\mathbf{f}(\mathbf{x}) \in \Omega$.

There's no data like more data.

2 Příznaky:

(Jednoduchý) vzor má příznaky $c_\nu, \nu = 1, \dots, n$, které jsou charakteristické jeho příslušností k jedné třídě.

3 Kompaktnost:

Příznaky vzorů jedné třídy Ω_κ vyplňují kompaktní oblast v příznakovém prostoru.

4 Podobnost:

Dvě reprezentace (vzorů) jsou podobné, když vhodně definovaná míra vzdálenosti je malá.

Example Application 1 A

- access control to the computer system of a company: the user has to say his name to login (instead of typing a password)
 - only employees of the company (because of access control to the building)
 - cooperative user (wants easy access)
 - user says a password or a prompted text, i.e. system knows, what will be said
 - classes Ω_{K_i} : employees
 - Ω_0 : if the pattern is not similar to any of the known user, it is rejected (the user has to type a password)
 - the certainty is not important (except for the rejection threshold)
 - misclassifications are not critical (assuming that no employee wants to harm the company)
 - all errors are equally bad

Example Application 1 B

- access control to the computer system of a company: the user has to say his name to login (instead of typing a password)
 - some employees have higher system privileges
 - some employees might want to get unauthorized privileges
 - person could be intruder (because of no access control to the building)
 - cooperative user (wants easy or unauthorized access)
 - user says a password or a prompted text, i.e. system knows, what will be said
 - classes Ω_{κ} : employees
 - Ω_0 : if the pattern is not similar to any of the known user, it is rejected (the user has to type a password)
 - rejection class is really a union of two classes: intruder or employee, whose current pattern is not similar enough to his class model
 - the certainty is not important (except for the rejection threshold)
 - misclassifications are critical
 - **not** all errors are equally bad

Example Application 2 A

- surveillance of phone calls to a country to detect drug traffic
 - the surv. organization has a database of N known drug dealers
 - some dealers are inactive for a longer period of time
 - a suspicious phone call is recorded
 - the surv. organization wants to know if the two people are already in the database
 - the system doesn't know what is said (maybe not even the language)
 - the user is not cooperative (might suspect that he is observed)
 - if the person is in the database, it suffices that his identity shows up in the first M alternatives, to be verified by a human expert
 - if the person is not in the database, the similarity to most of the elements in the database should be bad enough that no human expert has to check the classification result
 - the certainty of classification is important (used to order the database)
 - missing the person in the database is more expensive than a false alarm

Example Application 2 B

- The evidence that the recorded voice of a criminal belongs to an accused person has to be weighed in court
 - the user is not or cannot be cooperative
 - a rightly accused person will refuse to repeat a recorded utterance
 - certain influencing factors like stress, fatigue, . . . cannot be simulated
 - the certainty of classification is VERY important
 - wrongly classifying the accused person is more expensive than wrongly rejecting the correct hypothesis

Classes

- all applications: an utterance has to be assigned to a person
- the utterance is longer than one sound
- the underlying words (or even the language) could be unknown
- one sound of one speaker can be more similar to the same sound of another speaker than to another sound of the same speaker
- if the classes are the speakers then postulate 3 is violated

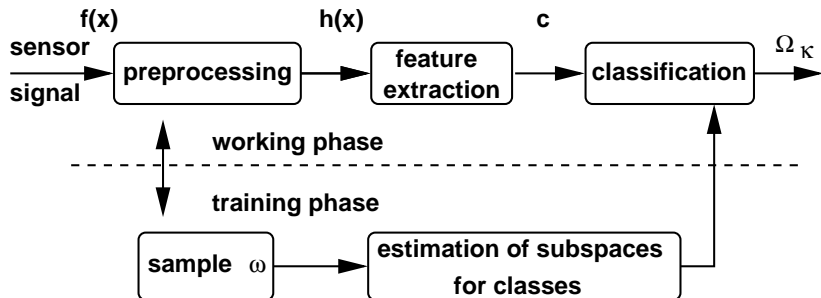
Postulate 3:

features of patterns from a class Ω_{κ} form a compact subspace

in the feature space \mathbb{R}^n .

- solution:
 - cut up the pattern into pieces that are smaller than one sound
 - classify these pieces w.r.t. the speakers
 - sum up the classification results for each speaker
 - decide for the one with the best score

A Classification System



Srovnání šablon

Pozorování: Zpracovává se vzory *digitálně*, tj. je jen *konečné číslo různých vzorů*.

Naivní nápad: Uložení všech vzorů; rozpoznávání vzoru je pak srovnání nového vzoru s uloženými.

⇒ **srovnání šablon**

Naivní nápad nejde kvůli počtu vzorů:

V rastru $18 \cdot 12$ (rozpoznávání znaků) je $2^{18 \cdot 12} \approx 10^{65}$ různých binárních rastrových obrazů.

Šedotónový obraz: $256^{512 \cdot 512} \approx 10^{630000}$

Srovnání šablon

Skutečnost, že je problém konečný, není dostatečná pro jeho praktickou řešitelnost. Výsledek by měl být dostupný co nejrychleji.

Princip řešení:

- Neuloží se všechny možnosti.
 - Uloží se algoritmus, který umí vypočítat řešení pro každou možnost.
- Příklad násobení: pro každou možnost
- Příklad rozpoznávání vzorů: pro skoro každou možnost

Spoken Dialogue System

