

Interakce člověk–počítač v přirozeném jazyce (ICP)

LS 2013 — Dynamic Time Warping (DTW)

Tino Haderlein

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

Dynamické borcení časové osy (DTW)

- Srovnání šablon není v podstatě pro klasifikaci možné kvůli počtu možných vzorů.
- Je možné pro klasifikační problémy, kde existuje jen málo vzorů.
- Problém při aplikaci na rozpoznávání jednotlivých slov: různé způsoby výslovnosti a délky slov
- Délky ploviv se mění jen málo, ale např. délky samohlásek a nazál mají širokou variaci.
- Řešení: dynamické borcení časové osy (dynamic time warping, DTW), poprvé používáno pro analýzu řeči (Sakoe & Chiba 1978; Rabiner & Juang 1993)

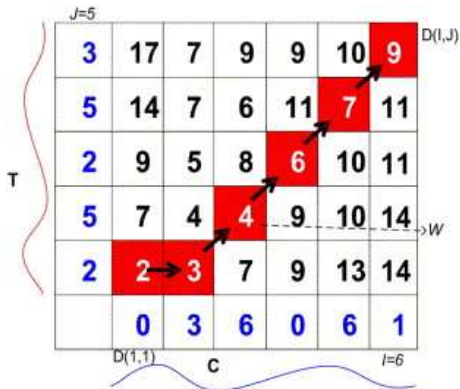
Dynamické borcení časové osy (DTW)

- Rozdělíme referenční a srovnávaný signál do I popř. J časových bodů, mezi kterými měříme vzdálenost signálů.
- Je možné časové osy „zbortit“, aby se odstranila závislost na délce signálů.
- Pokud jsou signály posloupností čísel, lze použít euklidovskou metriku (diferenci) jako míru vzdálenosti.
- Složitější signály (např. vektory spektrálních nebo cepstrálních koeficientů) vyžadují složitější míry vzdálenosti.
- V čase t ($t=1, \dots, T$) se podíváme na prvek $i(t)$ referenčního signálu a na prvek $j(t)$ srovnávaného signálu.
- rekurzivní definice vzdálenosti $D_{ij} = D(i(t), j(t))$:

$$D_{ij} = \begin{cases} D_{1,1}, & \text{když } i = j = 1 \\ \min\{D_{i-1,j-1}, D_{i-1,j}, D_{i,j-1}\}, & \text{když } i > 1, j > 1 \\ \infty & \text{jinak} \end{cases}$$

Znázornění DTW

Matice obsahuje celkové vzdálenosti k odpovídajícímu bodu:



<http://ars.els-cdn.com/content/image/1-s2.0-S1566253506000674-gr1.jpg>; C = class, reference; T = test, srovnávaný signál

Podmínky počítání vzdáleností DTW

- monotonicita (neklesající funkce):
 $i(t-1) \leq i(t), j(t-1) \leq j(t)$
- kontinuita (každý bod musí být zohledněn, cesta je spojitá):
 $i(t) - i(t-1) \leq 1, j(t) - j(t-1) \leq 1$
- hraniční podmínky (začátky a konce sekvencí si odpovídají):
 $i(1) = 1, j(1) = 1, i(T) = I, j(T) = J$

Popis semestrální práce 1. části LS 2013 (1/3)

Naprogramujte pomocí DTW jednoduchý klasifikátor pro rozpoznávání posloupností číslic (0-9). Program má rozčlenit nahrávku do číslic pomocí prahu energie, počítat pro každý úsek spektrální anebo cepstrální příznaky a srovnat je prostřednictvím DTW s příznaky uložených referenčních signálů. Pak má vrátit rozpoznanou posloupnost jako text (a pro kontrolu i měřené vzdálenosti mezi signály).
Poznámka: Rozpoznávač bude ovšem závislý na mluvčím.

Popis semestrální práce 1. části LS 2013 (2/3)

- Pro počítání příznaků můžete užívat existující programové knihovny (pro DFT etc.).
- Algoritmus DTW musí být Vaše vlastní práce, i míra vzdálenosti, se kterou srovnáváte příznaky.
- Zpracování nahrávek může být provedena offline.
- Musí být možné snadno rozšiřovat klasifikátor na jiná slova, tj. přidat do seznamu referencí další soubory s referenčními příznaky spolu s jejich významem.
- Použijte libovolný programovací jazyk.
- Musí být možné Váš program spustit na jiných počítačích bez velkých potíží.
- Pro zlepšení výsledků můžete např. průběh příznaků vyhladit, tj. počítat průměrné hodnoty v krátkém okně.

Popis semestrální práce 1. části LS 2013 (3/3)

Demonstrace semestrální práce:

- Odevzdejte psaný popis své práce: technické detaily a požadavky, možnosti programu, návod k obsluze
- Program osobně předved'te.