

Dialogové systémy

Luděk Bártek

Laboratoř vyhledávání a dialogu, Fakulta Informatiky Masarykovy Univerzity,
Brno

jaro 2021

- Rozpoznávání plynulé řeči – převádí souvislou promluvu na psaný text.
- Rozpoznávání izolovaných slov/příkazů.
- Princip rozpoznávání:
 - 1 získání vektoru příznaků pomocí metod krátkodobé analýzy signálu,
 - 2 klasifikace na základě vektoru příznaku získaného v předchozím kroku.

Rozpoznávání plynulé řeči

Dialogové
systémy

Luděk Bártek

Rozpoznávání
řeči

Rozpoznávání
plynulé řeči

Jazykový model

Gramatiky pro
podporu
roznávání řeči

Sémantická
interpretace
promluvy

- Hlavní rozdíly oproti rozpoznávání slov:
 - nelze vytvořit databázi vzorů
 - nutno brát zřetel na prozodické faktory
 - nutno určovat hranice mezi slovy
 - vypořádání se s výplňkovými zvuky a chybami řeči.
- Řešení – statistický přístup:
 - jazykový model
 - model uživatele.
- Příklad: HMM vrátí stejnou pravděpodobnost např. pro slova „máma“ a „nána“ – nejspíše se použije máma – je častější.

Rozpoznávání plynulé řeči

Jazykové modely

Dialogové
systémy

Luděk Bártek

Rozpoznávání
řeči

Rozpoznávání
plynulé řeči

Jazykový model

Gramatiky pro
podporu
roznávání řeči

Sémantická
interpretace
promluvy

- Máme:
 - posloupnost slov (promluva) $W = (w_1, \dots, w_n)$
 - posloupnost akustických vektorů $O = (o_1, \dots, o_t)$.
- Chceme nalézt W^* (množinu všech promluv), která maximalizuje $P(W|O)$.
- Dle Bayesova pravidla platí:

$$P(W^*|O) = \max P(W|O) = \max \frac{P(W) * P(O|W)}{P(O)}$$

Rozpoznávání plynulé řeči

Jazykové modely – pokračování

Dialogové
systémy

Luděk Bártek

Rozpoznávání
řeči

Rozpoznávání
plynulé řeči

Jazykový model

Gramatiky pro
podporu
roznávání řeči

Sémantická
interpretace
promluvy

- Pro nalezení maxima potřebujeme znát:
 - model řečníka – $P(O|W)$
 - jazykový model – $P(W)$.
- Model řečníka lze nahradit pravděpodobností generování W odpovídajícím Markovovým modelem.
- Trigramový model:
 - Experimentálně ověřeno, že platí:

$$P(w_n | w_1 \dots w_{n-1}) \cong P(w_n | w_{n-2} w_{n-1})$$

Rozpoznávání plynulé řeči

Rozpoznávání tématu

Dialogové
systémy

Luděk Bártek

Rozpoznávání
řeči

Rozpoznávání
plynulé řeči

Jazykový model

Gramatiky pro
podporu
roznávání řeči

Sémantická
interpretace
promluvy

- Úspěšnost rozpoznávání řeči se pohybuje cca 50 % — 99 % v závislosti na úkolu, jazyku, . . .
- Úspěšnost rozpoznávání lze zvýšit omezením domény rozpoznávání:
 - rozpoznání tématu
 - použitím gramatik pro rozpoznávání řeči.
- Známé téma:
 - změna stavového prostoru a pravděpodobnosti trigramů:
 - např. burzovní zprávy – rozpoznáno „honey“ nebo „money“?
 - možnost vytvoření přesnějšího jazykového modelu.

Gramatiky pro podporu rozpoznávání řeči

Dialogové
systémy

Luděk Bártek

Rozpoznávání
řeči

Rozpoznávání
plynulé řeči

Jazykový model

Gramatiky pro
podporu
rozpoznávání řeči

Sémantická
interpretace
promluvy

- Úspěšnost obecného rozpoznávání plynulé řeči může klesnout až na cca 50 %.
- Zvýšení lze dosáhnout omezením domény – např. specifikováním přípustných vstupů.
- Lze použít gramatiky pro podporu rozpoznávání řeči:
 - bezkontextové gramatiky
- Způsoby zápisů gramatik:
 - prostředky logického programování
 - proprietární řešení
 - otevřené standardy – JSGF, W3C SRGS, ...

Gramatiky pro podporu rozpoznávání řeči

Java Speech Grammar Specification (JSGF)

Dialogové
systémy

Luděk Bártek

Rozpoznávání
řeči

Rozpoznávání
plynulé řeči

Jazykový model

Gramatiky pro
podporu
rozpoznávání řeči

Sémantická
interpretace
promluvy

- Textový zápis gramatiky nezávislý na platformě a prodejci.
- Určen pro použití při rozpoznávání řeči.
- Součást Java Speech API.
- Používá styl a konvence jazyka Java.
- Aktuální verze 1.0 (říjen 1998).
- Použit např. v rozpoznávači Sphinx-4, VoiceXML interpretru VoiceGlue, ...
- Podrobněji v 2. polovině semestru při probírání tvorby dialogových rozhraní.

Gramatiky pro podporu rozpoznávání řeči

Ukázka JSGF

Dialogové
systémy

Luděk Bártek

Rozpoznávání
řeči

Rozpoznávání
plynulé řeči

Jazykový model

Gramatiky pro
podporu
rozpoznávání řeči

Sémantická
interpretace
promluvy

#JSGF

<koren> = Chci jet <cim> .|

Chci jet <cim> z <odkud> do <kam> .|

Chci jet <cim> z <odkud> do <kam> v <kdy> .;

<cim> = vlakem | autobusem;

<odkud> = <czMesto>;

<kam> = <czMesto>;

<kdy> = <czCas>;

Gramatiky pro podporu rozpoznávání řeči

W3C Speech Recognition Grammar Specification (SRGS)

Dialogové
systémy

Luděk Bártek

Rozpoznávání
řeči

Rozpoznávání
plynulé řeči

Jazykový model

Gramatiky pro
podporu
rozpoznávání řeči

Sémantická
interpretace
promluvy

- Standard W3C.
- Aktuální verze 1.0 (březen 2004).
- Definuje způsob zápisu pravidel a jejich odkazování.
- Dva způsoby zápisu:
 - XML
 - ABNF (Augmented BNF).
- Podrobněji v 2. polovině semestru při probírání tvorby dialogových rozhraní.

Ukázka W3C SRGS

Dialogové
systémy

Luděk Bártek

Rozpoznávání
řeči

Rozpoznávání
plynulé řeči

Jazykový model

Gramatiky pro
podporu
roznávání řeči

Sémantická
interpretace
promluvy

```
#ABNF 1.0 UTF-8
```

```
root $pozdrav;
```

```
language cs-CZ;
```

```
mode voice;
```

```
$pozdrav = ahoj
```

```
<?xml version="1.0" encoding="utf-8" ? >
```

```
<grammar root="pozdrav" xml:lang="cs-CZ" version="1.0" >
```

```
<rule id="pozdrav" >
```

```
ahoj
```

```
< /rule>
```

```
< /grammar>
```

Sémantická interpretace promluvy

Dialogové
systémy

Luděk Bártek

Rozpoznávání
řeči

Rozpoznávání
plynulé řeči

Jazykový model

Gramatiky pro
podporu
roznávání řeči

Sémantická
interpretace
promluvy

- Cíl – počítači srozumitelná interpretace informací zadaných uživatelem.
- Příklad:
Chtěl bych si koupit Zkrocení zlé ženy od Shakespeara.
 - akce = nákup
 - titul = Zkrocení zlé ženy
 - autor = Shakespeare
- Reprezentace – dvojice (atribut, hodnota).
- Obecné kroky sémantické analýzy:
 - 1 zjištění struktury rozpoznané promluvy
 - 2 zjištění významů jednotlivých částí rozpoznané promluvy
 - 3 případně odvození významu celé promluvy z významů jednotlivých částí promluvy.
- Sémantická interpretace \neq smyslu promluvy.

Sémantická interpretace promluvy

Implementace

Dialogové
systémy

Luděk Bártek

Rozpoznávání
řeči

Rozpoznávání
plynulé řeči

Jazykový model

Gramatiky pro
podporu
roznávání řeči

Sémantická
interpretace
promluvy

- Pravidlům gramatiky pro rozpoznávání řeči přidáme atributy, do kterých ukládáme jejich sémantickou reprezentaci.
- Na attributech lze provádět operace sloužící k sestavení sémantické interpretace celé promluvy z interpretací jednotlivých pravidel.
 - např jazykem ECMAScript (viz standard Sematic Interpretation for Speech Recognition)
- Ke zjištění smyslu sdělení je případně nutno jej zasadit do kontextu.
 - Kontext lze popsat pomocí konečného automatu s výstupem (Mealyho automatu – viz některá z dalších přednášek).

Popis Sémantické Interpretace

Dialogové
systémy

Luděk Bártek

Rozpoznávání
řeči

Rozpoznávání
plynulé řeči

Jazykový model

Gramatiky pro
podporu
roznávání řeči

Sémantická
interpretace
promluvy

■ JSGF:

- přiřazena pomocí značek (tags)
- zápis – {sémantická interpretace}

< *sentence* > = < *intro* > < *titul* > od < *autor* >

< *titul* > = Pejska a kočičku

{Povídání o pejskovi a kočičce}|

(Zlou ženu|Zkrocení zlé ženy) {Zkrocení zlé ženy}|...

■ SRGS – standard SISR :

- standard W3C Voice Browser Activity.
- Je postaven na jazyce ECMAScript.
- K pravidlům se přidává pomocí značky nebo atributu *tag*.
- Do dialogu je interpretace vracena ve formátu JSON.

■ ...