How to create dialog system

Jan Balata
Dept. of computer graphics and interaction
Czech Technical University in Prague
Outline

- Intro
- Architecture
- Types of control
- Designing dialog system
- IBM Bluemix
- How to start
Why go for speech UI

- **Good**
  - Speech is fast (large lists, dates, times)
  - Speech is natural and intuitive
  - Speech input device is small
  - Capturing emotional state
  - Determining speaker identity

- **Bad**
  - Speech is transient (no history on the screen)
  - Speech is “serial”
  - Limited short term memory of the user
  - Real time apps (speech is slow for games)
  - Problems with noisy environment
  - Other modalities more effective in some cases
  - Privacy

https://www.youtube.com/watch?v=5FFRoYhTJQQ
Application areas

- Large list selections, dates and times
- Hands busy situations
- Embedded systems with no keyboard or screen
- Telephony
- Pervasive systems – Car, Home environments
- Accessibility
Dialog systems are trending

- Bots are here, they’re learning — and in 2016, they might eat the web
  - Text based dialog systems
    - Big trend around 2000 and now again
    - Apps are dead - The average person spends 80 percent of their time on mobile devices using just 3 apps (ComScore 2015)
  - Facebook M
  - Slackbot
  - Nikabot
    - [http://www.nikabot.com](http://www.nikabot.com)
  - Magic
    - [https://getmagicnow.com/](https://getmagicnow.com/)
  - Luka
    - [https://luka.ai/](https://luka.ai/)
  - Lark
  - Penny
    - [https://www.pennyapp.io/](https://www.pennyapp.io/)
Architecture
Architecture

Dialog System

- Automatic Speech Recognition
- Natural Language Understanding
- Natural Language Generation
- Dialog Management
- Text To Speech
Architecture

- ASR (automatic speech recognition)
- NLU (natural language understanding)
- DM (dialog management)
- NLG (natural language generation)
- TTS (text to speech)
Almost 65 years of research and development
- 1952 Bell Labs: single words ~ 10 words
- 1960 Standford: continuous speech ~ 200 words
- 1971 DARPA: continuous speech ~ 1,000 words
- 1970s Institute for Defense Analysis: Hidden Markov Model
- 1980s IBM ~ 20,000 words (statistical models, HMM)
- 1990s more words than average human
- 2000s Dragon Systems
- 2010s University of Toronto, Microsoft, Google, and IBM (deep neural networks, deep learning)

Best accuracy (for English)
- 92% in optimal conditions
NLU

- Parsing the input text, determine the meaning
- Voice commands vs. full comprehension of news article
  - Real application somewhere between (text classification for the automatic analysis of emails)
- Lexicon, parser, grammar, ontology ... many person-year of effort
DM

- State variables
- Unanswered questions
- Error handling
- Initiative control
- Decisions
  - Scripted (defined by human)
  vs.
  reinforced learning (Markov Decision Process – action selected based on state and reward function)
Simple but hard to do correctly

- Content determination: Deciding what information to mention in the text
- Document structuring: Overall organisation of the information to convey
- Aggregation: Merging of similar sentences to improve readability and naturalness
- Lexical choice: Putting words to the concepts
- Realisation: Creating actual text with syntax, grammar correct
TTS

- Artificial production of human speech from text
  - *Text normalization*
  - *Phonetic transcription*

- **Approaches**
  - **Concatenation synthesis**
    - Segments of recorded speech put together
  - **Formant synthesis**
    - Additive synthesis and acoustic model (most text-to-speech synthesizers)
  - **Articulatory synthesis**
    - Based on models of human vocal tract
Control of dialog
Finite-state based system

user is taken through a dialog consisting of a sequence of pre-determined steps

System: What is your destination?
User: London
System: Is that London?
User: Yes.
System: What day do you want to travel?
User: Friday.
System: Was that Sunday?
User: No.
System: What day do you want to travel?
Frame based system

dialog flow is not pre-determined but depends on the content of the user’s response and pieces of information that the system recognize from it

System: *What is your destination?*
User: *London*
System: *What day do you want to travel?*
User: *Friday.*
System: *When do you want to travel?*
User: *Around 10 in the morning.*
System: *You want to travel from London on Friday around 10 in the morning?*
User: *Yes.*
System: *I have the following connection ...*
user can take control of the dialogue and use spontaneous and unconstrained speech to interaction with the system

User: I’m looking for a job in the Calais area. Are there any servers?

System: No, there aren’t any employment servers for Calais. However, there is and employment server for Pas-de-Calais and an employment server for Lille. Are you interested in one of these?
Designing
Dialog from User perspective

- **Input/output**
  - Text, voice, gui, multimodal
Dialog from User perspective

- Course of dialogue
  - Directed dialog, mixed initiative dialog, turn taking, Believe state modeling, Deep learning, Anaphora resolution, turn taking, POMDP (Partially Observable Markov Decision Processes)
Dialog from User perspective

- **When to speak?**
  - Pust to talk, silence detection, always speak mode, trigger words

- **What can user say?**
  - List of phrases, dictation
Creating DS

- Start with a scripted dialog system
  - Rapidly assembled with the expert of the chief stakeholder
  - Use in-house dialog modeling languages
- Transition to a POMDP-based approaches as data becomes available
How to write the speech application

- Indicate that user speaks to the machine
- Keep in mind short term memory of the user
- Provide “what can I say” option through the app
- Provide “go back” option throughout the app
- Build in an error correction mechanism
Input verification

- **Explicit**
  - Ask right away

- **Implicit**
  - Interpret and ask the next question

  **User:** I want to travel from Milano to Roma.
  **System:** At what time do you want to leave from Milano to Roma?
  **User:** No I want to leave from Milano in the evening.
IBM Bluemix
Many cloud services

- **“Cognitive apps”**
  - Emotion extraction, ASR, TTS, dialog, personality insights, image recognition, ...
IBM Bluemix

- Free 30 day trial
- Redeem code for 6 months (ask me)
Hacking

- Responses from server can include snippets of js, html, etc.
  - You can put images into dialog
  - You can load external content from different services

![Walk through the tourniquets and continue straight until you reach a glass building](image-url)
Semestral project

- Jednoduchý domácí společník
  - domácí robot
  - “…zejména pak konverzační způsob interakce (pomocí například tabletu nebo mobilního telefonu). Robot by měl budít dojem společníka, který komunikuje pomocí přirozeného jazyka a využívá kontextové informace (denní doba, zvuky, profil uživatele, léčebná terapie, atd.)…”
How to start

- NUR semestral projects

- Quick start to IBM Bluemix
For dialog practice

- Account on Bluemix (use quickstart from previous slide)
- Send me Credentials (next week!)
- Bring your own laptop
- Select topic of your project
Thank you

Jan Balata
Czech Technical University in Prague
jan.balata@fel.cvut.cz