Dialog System Personalization

Vicky Zayats

work done during summer 2015

April 24, 2018

Dialog System Description

Goal: to build a demo of a simple dialog system

* all applications and examples in this presentation are fictitious

Dialog System Description

Characteristics:

- ASR/TTS based on existing tools (Alexa, OK Google, Cortana, etc.)
- Intent classification + some span detection
- Rule-based dialog management

Challenges

- Lack of training and testing data

Data Collection

Solution: Using Mechanical Turk to collect simulated data

"Pretend that you are in a situation ______. How would you communicate it with the system?"

Challenges

- Lack of training and testing data
  - simulated data is limited
  - simulated data is not real (different from our target data)
  - small vocabulary size

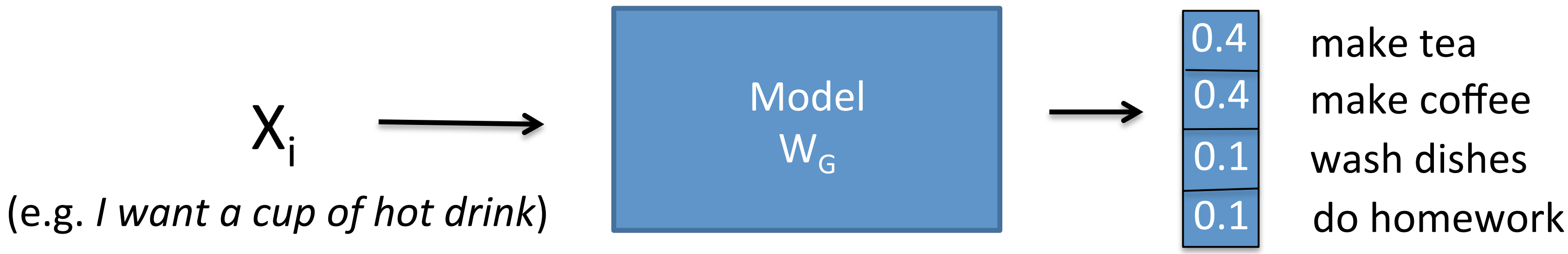
Challenges

- Lack of training and testing data
  - simulated data is limited
  - simulated data is not real (different from our target data)
  - small vocabulary size
- ASR errors
  - ASR tools are trained on irrelevant data
  - out-of-vocabulary lexicon
  - non-native speech is hard
  - far-field ASR accounts for additional errors

Solution

- Personalized Dialog System that would learn from the user generated data on-the-go
- Obtaining true labels for user generated data:
  - confirmation from the user on the action, otherwise ask to rephrase
  - at the end, ask if the intent was identified correctly
  - if so, add the sentence(s) to a user-generated pool of samples for training a personalized model

Dialog System



Model: Logistic Regression (Maximum Entropy)

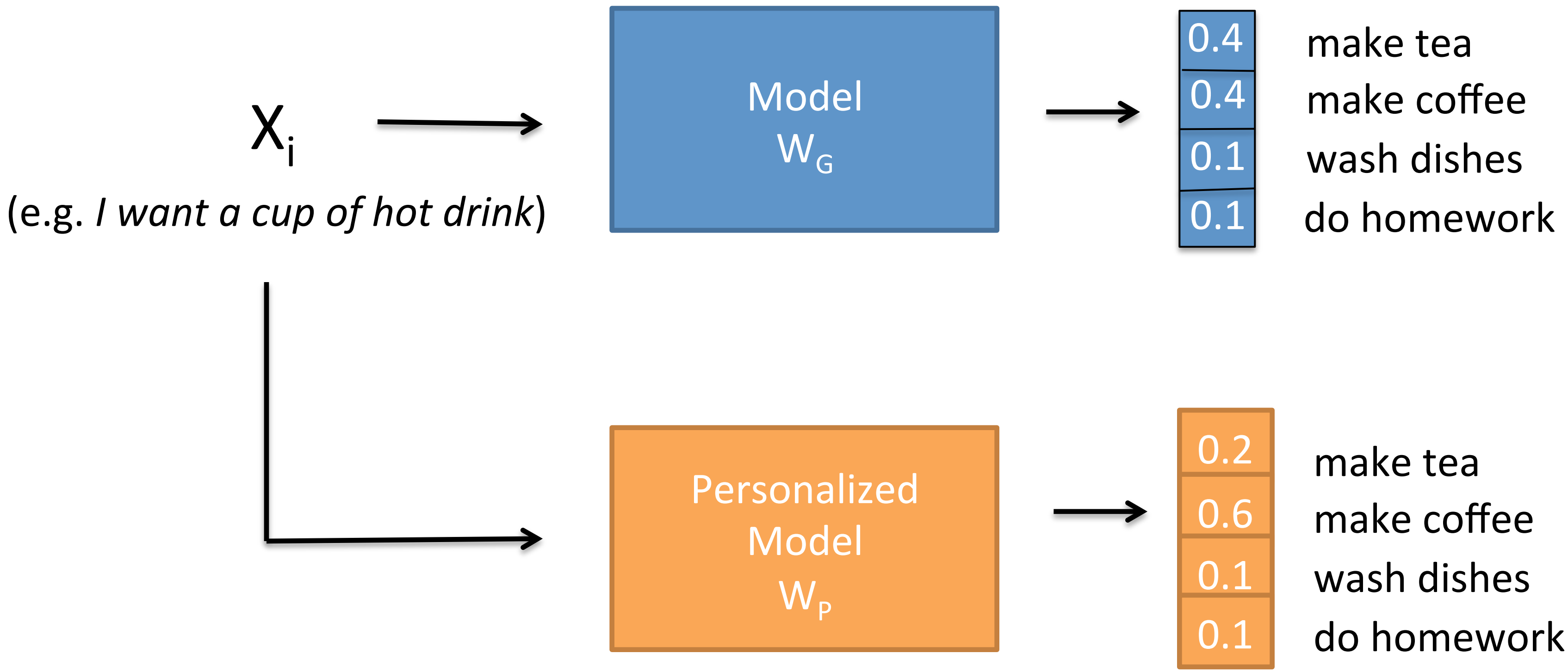
for *k*<*R*

$$P(Y = y_k|X) = \frac{\exp(w_{k0} + \sum_{i=1}^{n} w_{ki}X_i)}{1 + \sum_{j=1}^{R-1} \exp(w_{j0} + \sum_{i=1}^{n} w_{ji}X_i)}$$

for *k*=*R* (normalization, so no weights for this class)

$$P(Y = y_R|X) = \frac{1}{1 + \sum_{j=1}^{R-1} \exp(w_{j0} + \sum_{i=1}^{n} w_{ji}X_i)}$$

Dialog System Personalization



- Separate models allow a better tuning of personalized model, as well as erasing personalization when it is required

Models fusion

$$P(Y=k|X,W_G,W_P)=\alpha_t \underbrace{P(Y=k|X,W_G)}_{\text{General model}}+(1-\alpha_t)\underbrace{P(Y=k|X,W_P)}_{\text{Personalized model}}$$

$$\alpha_t \in [0,\ 1]$$

$\alpha_t \approx 1$ when $t \rightarrow 0,\ D_G \gg D_P$

$\alpha_t \approx 0$ when $t \rightarrow \infty,\ D_P \gg D_G$

D_G - # training sentences

D_P - # user generated sentences

Online learning

- Logistic loss as a stochastic function:

$$E_{\mathbf{x}}\left[\ell(\mathbf{w},\mathbf{x})\right] = E_{\mathbf{x}}\left[\ln P(y|\mathbf{x},\mathbf{w}) - \frac{\lambda}{2}||\mathbf{w}||_2^2\right]$$

- Batch gradient ascent updates:

$$\boxed{w_i^{(t+1)} \leftarrow w_i^{(t)} + \eta\left\{-\lambda w_i^{(t)} + \frac{1}{N}\sum_{j=1}^{N} x_i^{(j)}[y^{(j)} - P(Y=1|\mathbf{x}^{(j)},\mathbf{w}^{(t)})]\right\}}$$

- Stochastic gradient ascent updates:
  - Online setting:

$$\boxed{w_i^{(t+1)} \leftarrow w_i^{(t)} + \eta_t\left\{-\lambda w_i^{(t)} + x_i^{(t)}[y^{(t)} - P(Y=1|\mathbf{x}^{(t)},\mathbf{w}^{(t)})]\right\}}$$

*This slide is taken from lectures by Emily Fox, CSE 547

Conclusion

- Fuse general and personalized models to overcome noisy and inaccurate data:
  - Mismatch between training and test data
  - ASR errors
  - Small vocabulary size
- Update personalized model using online learning algorithm

Questions?