# IN4080 exercises (autumn 2023)

## Week 11: Speech processing, dialogue management, system design and evaluation

(The two first questions are extracted from exams of previous years)

You wish to develop a (phone-based) spoken dialogue system that will call random U.S. citizens in order to collect opinion poll data for the next US election, with two candidates on the ballot: Kamala Harris and Ivanka Trump.

This system is framed as a *Markov Decision Process* (MDP) formalised as such:

- We have five possible states:
  - $s_1$ is the starting state
  - $s_2$ if the callee indicated their intention to vote for Kamala Harris
  - $s_3$ if the callee indicated their intention to vote for Ivanka Trump
  - $s_4$ if the callee expressed something else (that was not understood)
- The set of actions that can be taken by the dialogue system are as follows:
  - $a_1$ : Say ”Hi, I’m a automated bot developed to collect polling data. May I ask you for whom you plan to vote in the next election?”
  - $a_2$ : Say ”Sorry I did not understand. Who do you wish to vote for?”
  - $a_3$ : Say ”Ok, thank you for your help, and have a nice day!"
- The transition model is as follows:
  - In state $s_1$, only action $a_1$ is possible, with three possible transitions:
    $P(s'=s_2|s=s_1,a=a_1)=0.48, P(s'=s_3|s=s_1,a=a_1)=0.40, P(s'=s_4|s=s_1,a=a_1)=0.12$
  - In states $s_2$ and $s_3$, only $a_3$ is possible and terminates the dialogue.
  - In state $s_4$, only $a_2$ is possible, with the following transitions:
    $P(s'=s_2|s=s_4,a=a_2)=0.36, P(s'=s_3|s=s_4,a=a_2)=0.32, P(s'=s_4|s=s_4,a=a_2)=0.32$
- Finally, the reward model is defined as such:
  - $R(s=s_2,a=a_3)=R(s=s_3,a=a_3)=10$ (if the system manages to register the callee's political preference)
  - $R(s=s_4,a=a_2)=-1$ (to capture the annoyance of asking the callee to repeat)
  - Other actions have a reward of zero.

Based on this MDP model, calculate the expected cumulative reward of asking the callee to repeat when their answer was not properly understood, that is: $Q(s=s_4,a=a_2)$. You can assume a discount factor of 0.9.

## Question 2

Assume you wish to develop a talking robot that can respond to various requests. When a human is perceived around the robot, the first first step is to engage the human, for instance with a greeting, accompanied or not by gestures. We wish to apply reinforcement learning to determine the best engagement strategy among two options: *SayHi* and *SayHiWithGestures*.

Formally speaking, we can frame this problem as an MDP with

- 2 states: *HumanNotEngaged* (which is the starting state) and *HumanEngaged* (which is a final state).
- 3 actions: *SayHi*, *SayHiWithGestures*, and *AskHowCanIHelpYou*. The first two actions (*SayHi* and *SayHiWithGestures*) are only available for the starting state *HumanNotEngaged*, while the action *AskHowCanIHelpYou* is only available for the final state *HumanEngaged*.

The transition model $P(s'|s,a)$ for this MDP is as follows:

- If the robot executes action *SayHi* in the state *HumanNotEngaged*, we have a probability 0.5 of reaching the state *HumanEngaged*, and a probability 0.5 of staying in the state *HumanNotEngaged*.
- If the robot executes action *SayHiWithGestures* in the state *HumanNotEngaged*, we have a probability 0.7 of reaching the state *HumanEngaged*, and a probability 0.3 of staying in the state *HumanNotEngaged*.

The reward function $R(s,a)$ of this MDP is as follows:

- The reward of executing action *SayHi* in the state *HumanNotEngaged* is -1
- The reward of executing action *SayHiWithGestures* in the state *HumanNotEngaged* is -2, since the physical gestures "cost" more to the agent in terms of energy and mechanical wear.
- Finally, the reward of executing action *AskHowCanIHelpYou* in the state *HumanEngaged* is +10. You can consider this action as the final one in this MDP, which means that the expected cumulative reward Q(*HumanEngaged*, *AskHowCanIHelpYou*) = 10.

**Questions**:

1) Compute the expected cumulative rewards Q(*HumanNotEngaged*, *SayHi*) and Q(*HumanNotEngaged*, *SayHiWithGestures*) based on the MDP described above. To compute those Q-values, you need to use Bellman's equation and refine your estimates through several iterations. You can stop after 5 iterations and use a discount factor $\gamma$=0.9. For the first iteration, you can initialize the Q-values to zero. *(10 points)*

2) Based on those Q-values, which engagement strategy should the robot chose? *(2 points)*

**Tip**: Bellman's equation is:

$$Q(s,a) = R(s,a) + \gamma \sum_{s' \in S} P(s'|s,a)\ max_{a'} Q(s',a')$$

## Question 3

The PARADISE framework for the evaluation of dialogue systems operates by:

1. Defining a set of objective measures that can be extracted automatically from any given dialogue. Those measures may include both indicators of tasks success as well efficiency/quality costs (total elapsed time, number of clarification requests, etc.)

2. Inferring the weight of each quality measure by fitting a regression model that predicts the user satisfaction from those measures. This is done using a small dataset of dialogues where the user is asked to provide a rating for the dialogue that has just taken place.

Download the dataset at the following address:

https://home.nr.no/~plison/data/paradise_example.tsv

The file contains a tab-separated list of measures for 500 dialogues. For each dialogue, we provide:

- The total elapsed time (in seconds)
- The number of times the system had to ask the user to clarify their intent
- The number of times the system had to ask the user to repeat
- The total number of dialogue turns
- Whether the task was successfully completed at the end of the dialogue
- Whether the user uttered "thank you" at the end of the dialogue
- Finally, a rating provided by the user at the end of the dialogue, on a scale from 1 to 5 (where 1 means least satisfied and 5 most satisfied)

Based on the data above:

1) Train a linear regression model (using e.g. sklearn.linear_model.LinearRegression) to fit a model that predicts the user rating from the objective measure.

2) Which objective measure is most predictive of the user rating, according to your fitted regression model?

3) Assume you record a new dialogue lasting 2 minutes, including 8 clarification requests and 2 repetition requests, for a total of 33 dialogue turns, and that ended up with a successfully completed task, and a user uttering "Thank you" at the end. What would be the predicted user rating according to your model?