An Improved Approach of Intention Discovery with Machine Learning for POMDP-based Dialogue Management

MASTER'S THESIS PROPOSAL PRESENTATION

FRIDAY, 18TH JANUARY, 2019

BY

RUTURAJ R. RAVAL

SCHOOL OF COMPUTER SCIENCE, UNIVERSITY OF WINDSOR

Committee members

Dr. Xiaobu Yuan	Dr. Luis Rueda	Dr. Gokul Bhandari
Supervisor	***Internal Reader***	***External Reader***

Outline

- Introduction
- Background – Literature
- Related work
- Proposed method
- Architecture
- Algorithm
- Experiment & Design
- Timeframe of further work
- Conclusion
- Future work
- References

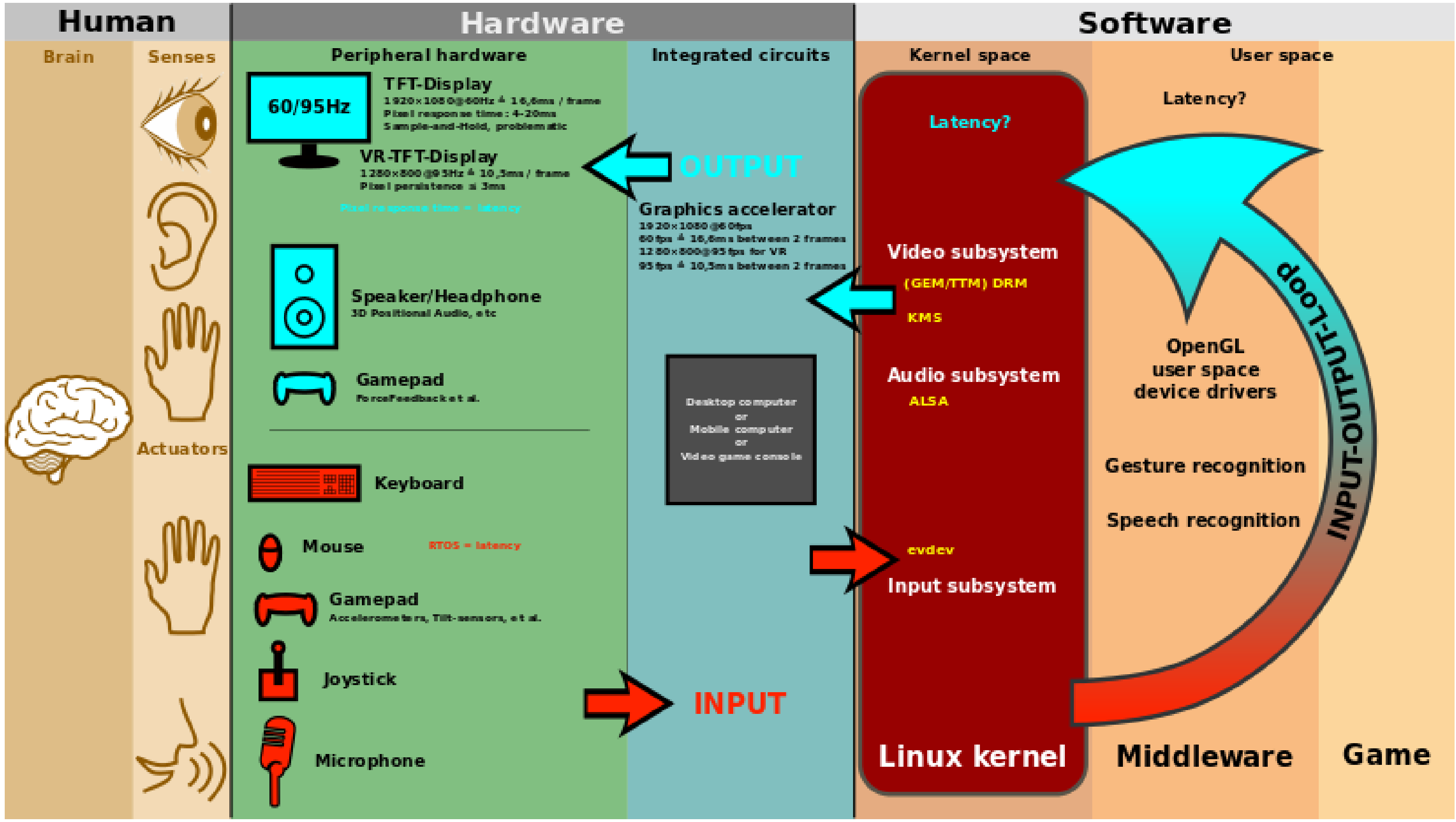
Introduction

OVERVIEW- HCI & ECA

DIALOGUE MANAGEMENT

DECISION MAKING PROCESS & MDP

Human Computer Interaction



https://commons.wikimedia.org/w/index.php?title=File:Linux_kernel_INPUT_OUPUT_evdev_gem_USB_framebuffer.svg&oldid=180540123

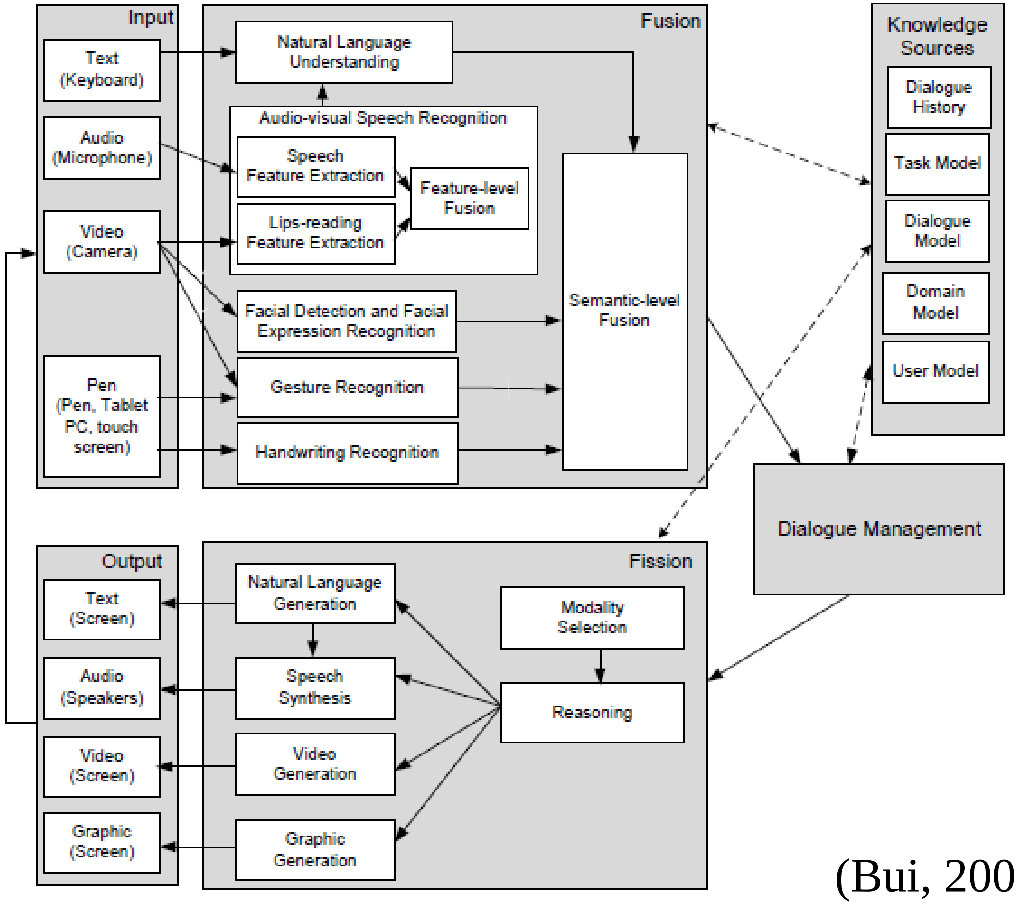
- Focus on interface between people (users) and computers
- HCI situated at the interaction of,
  - Computer Science
  - Behavioural Science
  - Design
- HCI confronts difficult challenges like, (Hollan et. Al., 2000)
  - Supporting complex tasks
  - Mediating network interaction
  - Managing and exploiting the ever increasing availability of digital information
- User interacts directly with the hardware for the human input and output such as displays. (Fehrenbacher, 2017)

Embodied Conversational Agent

- ECAs are the artificial agents, known as interface agents (Serenko et. Al., 2007)

- ECA have embodied agents with a graphical front-end as opposed to the robotic body, capable enough to engage in a conversation with the humans employing verbal and nonverbal medium such as gesture, facial expression, etc. (W. Contributors, 2018)

- Avatar (ECA) is the graphical representation of the user or the user's alter ego or character. Some of the ECAs use sophisticated natural language processing systems, but many simpler systems scan for keywords within the input then pull a reply with the most matching keywords or the most similar wording patterns from the database. (Russell et. Al., 2003)



(Bui, 2008)

Dialogue Management (DM)

- A system consisting of a dialogue manager
- To track the dialogue states and maintain a dialogue policy which decides how the system reacts on given dialogue state
- Statistical approaches to dialogue modelling (Mnih et. Al., 2013)
- The dialogue: MDP a process- for each state, DM has to select an action; and possible rewards from each action
- Dialogue author needs to define reward function, for example:
  - ***Tutorial dialogues***: the reward is the increase in the student grade
  - ***Information seeking dialogues***: the reward is positive if human receives the information, but also a negative reward can be assigned
- Different models in DM domain: Finite State Machine (FSM) | Frame-based | POMDP (Seron et. Al., 2016)

Finite State Machine (FSM)

- Advantages
  - Clear structure
  - Easy to develop to effectively control the dialogue process
- Disadvantages
  - Not suitable for a complex dialogue task

Frame-based

- Advantages
  - Can handle more complex inputs
- Disadvantages
  - Dialogues turns out to be unnatural

POMDP

- Advantages
  - Very popular in theoretical studies
  - Proven to be a good model to deal with uncertain problems in speech recognition and language understanding
  - Can be applied to more fields by factoring its state space
- Disadvantages
  - Cannot handle multitopic tasks and the tagging and training corpus are expensive and time consuming

Markov Decision Process (MDP)

- MDP is an output of the continuous cast of dialogue management, composed of finite set of actions, continuous multivariate belief state space and a reward function. (Zhao et. Al., 2016)

- MDP is a 5-tuple process (S, A, P_a, R_a, γ)
  - ***S***: finite set of states
  - ***A***: finite set of actions
  - P_a: probability of action ***a***, in the state ***s***, at time ***t***, leading to state ***s'*** at time ***t+1***
  - R_a: immediate reward received after transition from state ***s*** to ***s'***, due to action ***a***
  - *γ*: discount factor, represents difference between future and present rewards

- Limitations
  - MDP cannot optimize in the stochastic environment when combined with the policy, to solve we need POMDP

Decision-making process		
Observable	***Deterministic***	***Stochastic***
Fully	BFS, DFS, A*	MDP
Partial	-	POMDP

<table>
<tr><th colspan="4"><i><b>Decision-making process</b></i></th></tr>
<tr><td colspan="2" rowspan="2"><i><b>Markov Models</b></i></td><td colspan="2"><i><b>Have control over the state transitions?</b></i></td></tr>
<tr><td>NO</td><td>YES</td></tr>
<tr><td rowspan="2"><i><b>Are states completely Observable?</b></i></td><td>YES</td><td>Markov Chain</td><td>MDP (Markov Decision Process)</td></tr>
<tr><td>NO</td><td>HMM (Hidden Markov Model)</td><td>POMDP (Partially Observable Markov Decision Process)</td></tr>
</table>

Background - Literature

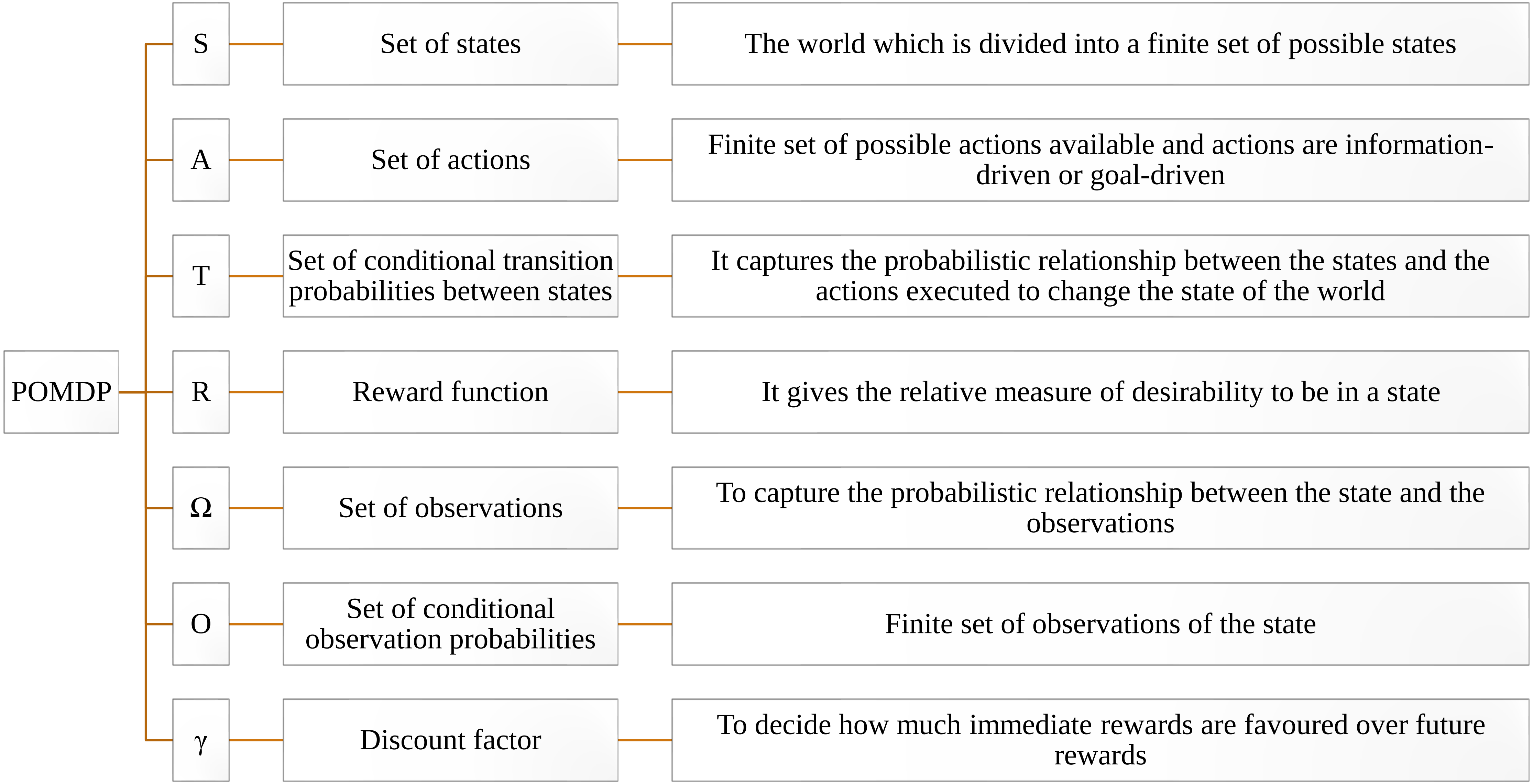
POMDP

FUZZY LOGIC

BELIEF STATE HISTORY

Partially Observable Markov Decision Process (POMDP)

- In the MDP context, assume state ***s*** is known when action is to be taken, otherwise policy ***π(s)*** cannot be calculated
- If the assumption is not true, then it will be partially observable
- POMDP: generalization of MDP (Ultes et. Al., 2017)
  - Assumption made that the agent models decision process using MDP dynamics of determination
  - Agent can't observe the underlying state
  - Instead, must maintain a probability distribution over the set of possible states, based on a set of observations and observations probabilities and underlying MDP
- POMDP framework is general enough to model a variety of real-world sequential decision processes (Ultes et. Al., 2017)
- POMDP builds discrete-time relationship between the agent and the user (environment)
- POMDP is a 7-tuple ***(S, A, T, R, Ω, O, γ)***
- Limitation
  - Because of Markovian property, POMDP model refrains to capture the history of actions taken and observations made which is valuable data

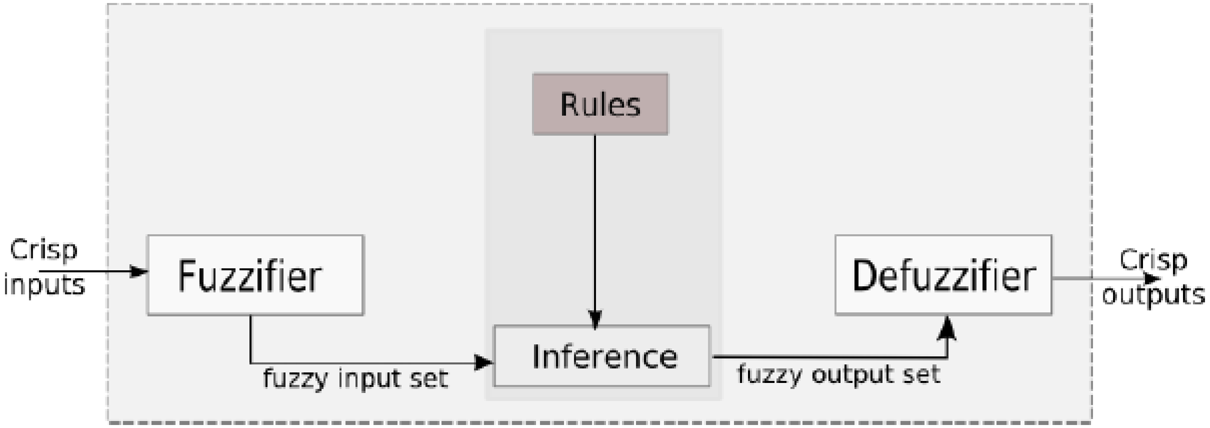


POMDP

S	Set of states	The world which is divided into a finite set of possible states
A	Set of actions	Finite set of possible actions available and actions are information-driven or goal-driven
T	Set of conditional transition probabilities between states	It captures the probabilistic relationship between the states and the actions executed to change the state of the world
R	Reward function	It gives the relative measure of desirability to be in a state
Ω	Set of observations	To capture the probabilistic relationship between the state and the observations
O	Set of conditional observation probabilities	Finite set of observations of the state
γ	Discount factor	To decide how much immediate rewards are favoured over future rewards

Fuzzy logic

- Fuzzy logic is a form of many-valued logic in which the truth values of variables; between real numbers 0 and 1 (Liu et. Al., 2008)
- Employed to handle the concept of partial truth
  - Truth value may range between completely true and completely false
- A fuzzy logic system can be defined as non-linear mapping of an input dataset to a scalar output data consisting 4 components such as: fuzzifier | rules | inference engine | defuzzifier (Mendel, 1995)

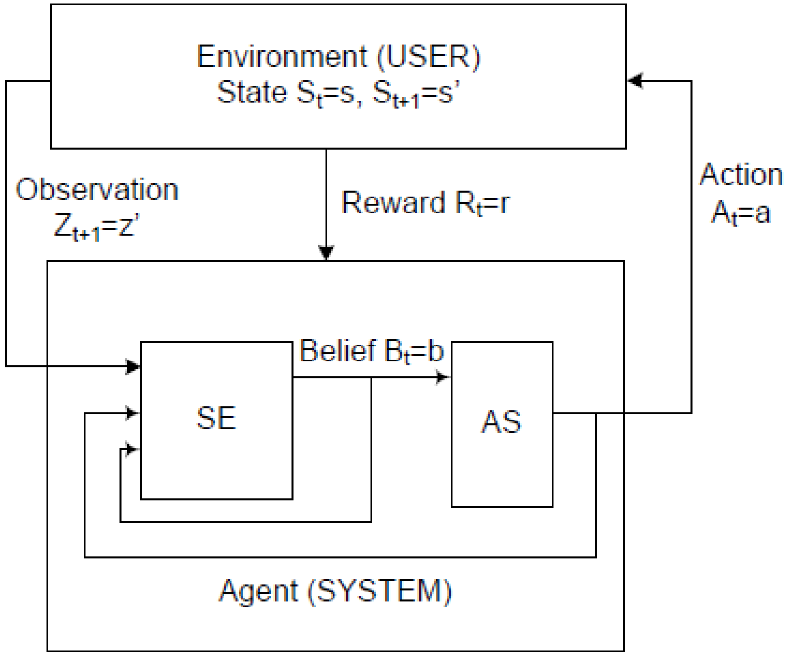


http://cs.bilkent.edu.tr/~zeynep/files/short_fuzzy_logic_tutorial.pdf

Belief state history

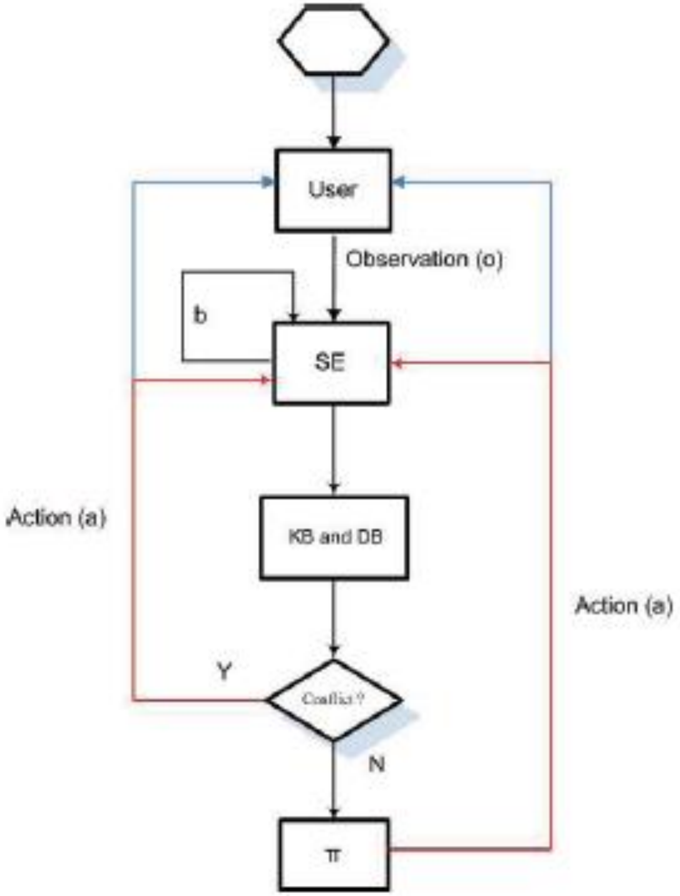
- ➢Belief state: it is a probability distribution over all possible states which gives as much information as the entire action-observation history (Alexandersson et. Al., 2014)
- ➢Using the dynamics of belief state and its history and outperforming shortcomings of traditional POMDP, where the compressed form of every new action taken on the state comes with the loss of important information, and retaining its advantages of current POMDP-based approach (Cuayahuitil et. Al., 2015) (Yuan et. Al., 2010)
- ➢Although the history information of observations and actions are not maintained explicitly, the negative effects of Markov assumptions and diminished and POMDP-based DM is allowed to plan for actions (Yuan et. Al., 2010)
  - ➢Not only for current belief state but also the updated history before reaching the current state

POMDP-based architecture



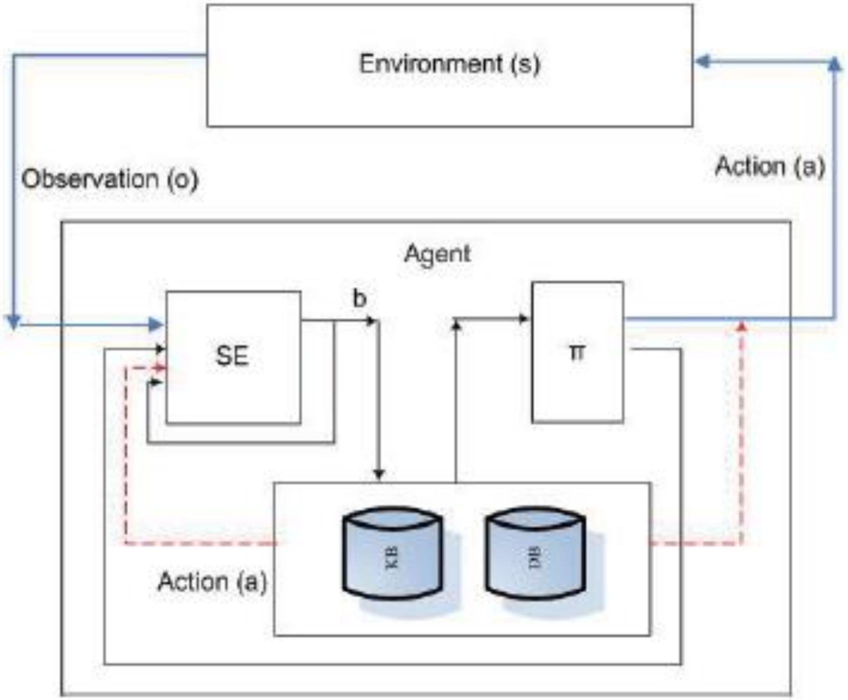
(Bui, 2008)

Flow-chart of POMDP using belief state history



(Yuan et. Al., 2010)

POMDP-based architecture using belief state history



(Yuan et. Al., 2010)

Related work

TREND ANALYSIS

SURVEY

PROBLEM STATEMENT

Trend analysis

- A widespread practice of collecting information and attempting to spot a pattern; often used to predict future events; could be used to estimate uncertain events in the past, already stored using POMDP model
- Trend analysis often refers to techniques for extracting an underlying pattern behaviour in a time series which would otherwise be partly or nearly completely hidden by noise (Bui, 2008)
- Different approaches in trend analysis (Mulpuri, 2016)
    - ***Sampling***: historical data is split into training and testing datasets (e.g. random sampling)
    - ***Histogram***: historical data constructed by histogram (e.g. V-optimal)
    - ***Sketches***: frequency distribution of historical data is summarized using hash function (e.g. count sketches)
    - ***Wavelets***: mathematical transformations applied to transform data into a set of coefficients to analyse trend (e.g. discrete wavelet transform)

Survey			
Index	***Reference***	***Contribution***	***Key points***
1	R. Harel, Z. Yumak and F. Dignum, "**Towards a generic framework for multi-party dialogue with virtual humans**," *CASA: 31st International Conference on Computer Application and Social Agents*, Beijing, China, 2018.	• A genetic framework to aid in development of multi-modal, multi-party dialogue • It contains mechanisms inspired by social practice theory for both action selection and timing – including handling of interruption • Expectations are utilized to paint a rough sketch of what a socially-acceptable interaction looks like • The agent's aim is to follow the set of candidate actions under exceptional circumstances as a result	• Multi-modal; multi-party dialogue management • Action selection and timing based mechanism • Real-time acceptable interaction • Aim is to follow candidate actions in exceptional circumstances
2	M. Igl, L. Zintgraf, T. A. Le, F. Wood and S. Whiteson, "**Deep variational reinforcement learning for POMDPs,**" in *Proceedings of the 35th International Conference on Machine Learning*, Stockholm, Sweden, 2018.	• The model, DVRL (Deep Variational Reinforcement Learning) introduces an inductive bias that allows an agent to learn a generative model of the environment and perform inference in the model to effectively aggregate the available information • This method solves POMDPs for given only a stream of observations, without knowledge of the latent state space or the transition and observation functions operating in that space	• DVRL allows agent to learn generative model • Solves POMDP for given stream of observation without any knowledge of latent state space

Index	***Reference***	***Contribution***	***Key points***
3	S. Omidshafiei, J. Pazis, C. Amato, J. P. How and J. Vian, "**Deep decentralized multi-task multi-agent reinforcement learning under partial observability**," in *Proceedings of the 34th International Conference on Machine Learning*, Sydney, Australia, 2017.	• Addresses the problem of multi-task multi-agent reinforcement learning under partial observability using decentralized single-task learning approach that is robust to concurrent interactions of teammates • Also presents an approach for distilling single-task policies into a unified policy that performs well across multiple related tasks, without explicit provision of task identity	• Multi-task; multi-agent RL • Under partial observability • Decentralized single-task learning • Without task identity, single-task policies are unified
4	Y.-N. Chen, A. Celikyilmaz and D. Hakkani-Tur, "**Deep learning for dialogue systems**," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics-Tutorial Abstracts*, Vancouver, Canada, 2017.	• Focuses on the motivation of the work on conversation-based intelligent agents in which the core underlying system is task-oriented dialogue systems • Different aspects are considered as a part of the survey like, language understanding; dialogue management; natural language generation; end-to-end learning for dialogue system; dialogue breadth; dialogue depth, etc.	• Conversation-based intelligent agent • Core is task-oriented dialogue systems • Consists of, language understanding | dialogue management | natural language generation, etc.

Index	***Reference***	***Contribution***	***Key points***
5	J. N. Foerster, Y. M. Assael, N. d. Freitas and S. Whiteson, "**Learning to communicate with deep multi-agent reinforcement learning**," in *30th Conference on Neural Information Processing Systems (NIPS)*, Barcelona, Spain, 2016.	• Proposed two approaches: Reinforced Inter-Agent Learning (RIAL) and Differentiable Inter-Agent Learning (DIAL) • RIAL uses deep Q-learning; DIAL exploits the fact that, during learning agents can backpropagate error derivatives through communication channels • One of the first attempt at learning communication and language with deep learning approaches • Offers novel environments and successful techniques for learning communication protocols	• RIAL, DIAL models • RIAL: deep Q-learning • One of the first attempt in communication learning and deep learning approach • Novel environments to learn communication
6	P. Shah, D. Hakkani-Tur and L. Heck, "**Interactive reinforcement learning for task-oriented dialogue management**," in *Workshop on deep learning for action and interaction*, Barcelona, Spain, 2016.	• Investigates policy gradient based methods for interactive reinforcement learning where the agent receives action-specific feedback from the user and incorporates the feedback into the policy • Using feedback, to shape the policy directly, enables dialogue manager to learn new interactions faster compared to interpreting the feedback as a reward value	• Interactive RL • Agent receives action-specific feedback from user • Feedback helps in improving policy • Feedback helps in learning new interactions

Index	*Reference*	*Contribution*	*Key points*
7	M. Gasic, N. Mrksic, L. M. Rojas-Barahona, P.-H. Su, S. Ultes, D. Vandyke, T.-H. Wen and S. Young, "**Dialogue manager domain adaption using gaussian process reinforcement learning**," *Computer speech and language*, pp. 552-569, 2016.	• Data-driven machine learning methods have been applied to dialogue modelling and the results achieved for limited-domain applications are comparable to outperform the traditional approaches • Method based on Gaussian processes are particularly effective as they enable good models to be estimated from limited training data • They provide an explicit estimate of the uncertainty particularly useful for reinforcement learning • Gaussian process RL is an elegant framework that naturally supports a range of methods including prior knowledge	• Data-driven machine learning approach • Results based on limited-domain application • Traditional approaches are outperformed • Estimates uncertainty, useful for RL • Proposed an elegant framework of GPRL
8	M. Gasic, D. Kim, P. Tsiakoulis and S. Young, "**Distributed dialogue policies for multi-domain statictical dialogue management**," *IEEE: ICASSP*, pp. 5371-5375, 2015.	• Hierarchical distributed dialogue architecture in which policies are organized in a class hierarchy aligned to an underlying knowledge graph • Gaussian process-based RL is used to represent within the framework, generic policies can be constructed which provides acceptable user performance	• Hierarchical distributed dialogue architecture • GPRL based RL is used in constructing generic policies

Problem Statement

- Shortcomings of the POMDP model are outperformed, at the same time retaining the advantages of POMDP model by iterating belief state history to find the trend using DWT (Discrete Wavelet Transformation) can improve the intention discovery
- How to make the agent more intelligent in terms of generating natural dialogues?
  - Goal-driven dialogue conversation
  - Sentiment learning
  - Policy improvement, etc.
- There is a need to implement constant learning for the agent to reduce the dialogue length using machine learning approach- Reinforcement Learning (RL), which encourages a belief or a pattern of behaviour in terms of the rewards
- Using sentiment analysis and RL techniques can improve the intention discovery of user's aim of reaching goal using DM and context-driven communication

Proposed method

SENTIMENT ANALYSIS

REINFORCEMENT LEARNING

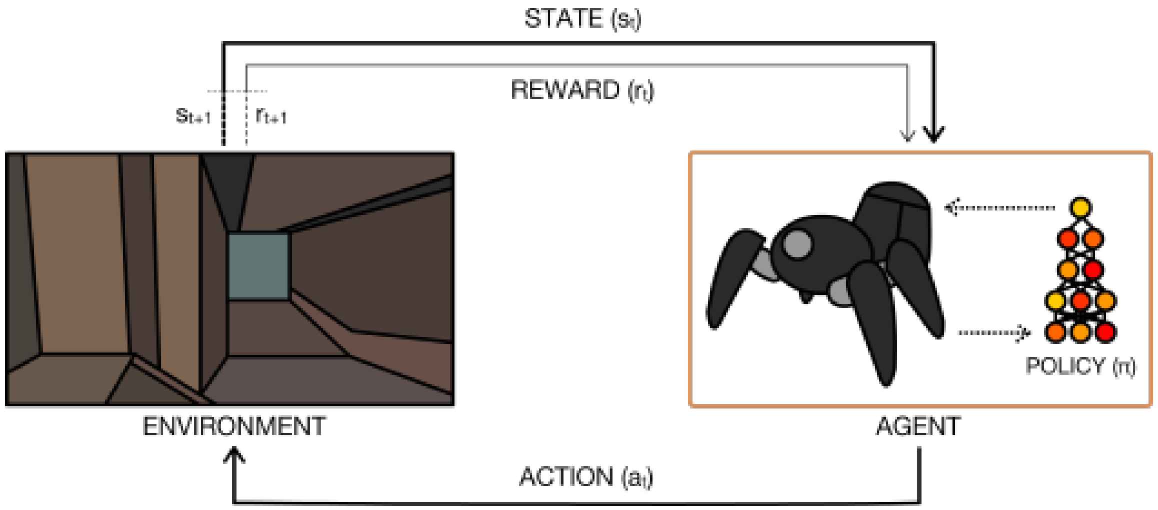
RL USING Q-LEARNING

Sentiment analysis

- The process of computationally identifying and categorizing emotion (sentiment) expressed from a piece of text
- To determine the intention using emotion discovery
- Helps in achieving end-to-end task-completion (Li et. Al., 2018)
- Sentiment analysis has a wide appeal as providing information about the subjective dimension of texts. It can be regarded as a classification technique, either binary (polarity classification into positive/negative) or multi-class categorization (e.g. positive/neutral/negative) (Klein, 2015) (Bird et. Al., 2009)
- Most approaches use a sentiment lexicon as a component (sometimes the only component). Lexicons can either be general purpose, or extracted from a suitable corpus, such as movie reviews with explicit ranking information (Klein, 2015) (Bird et. Al., 2009)

Reinforcement Learning

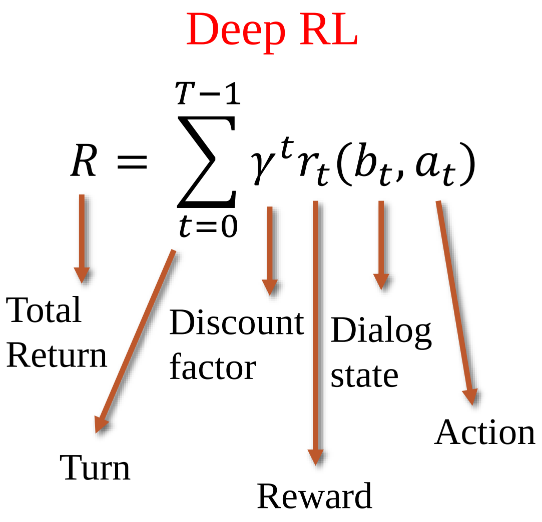
- Primary goal: AI is to produce fully autonomous agents that interact with their users (environments) to learn optimal behaviours, improving over time through trial-and-error (Arulkumaran et. Al., 2017)
- A principle mathematical framework for experience-driven autonomous learning is Reinforcement Learning (RL) (Sutton et. Al., 1998)
- Limitations of previous traditional RL approaches (Strehl et. Al., 2006)
  - Memory complexity
  - Computational complexity
  - Scalability
- Limitations outperformed (Arulkumaran et. Al., 2017)
  - Rise of deep learning using powerful *function approximation* and *representation learning* properties of DNN (Deep Neural Network)
- Essence of RL is learning through the *interaction* (Arulkumaran et. Al., 2017)
- RL agent interacts with its environment and upon observation, consequences of actions, can learn to alter its own behaviour in response to rewards received



The perceptron-action-learning loop, at time $\boldsymbol{t}$, the agent receives state $\boldsymbol{s_t}$ from the environment, the agent uses its policy to choose an action $\boldsymbol{a_t}$. After action execution, environment transitions a step, providing the next step $\boldsymbol{s_{t+1}}$ as well as the feedback in the form of a reward $\boldsymbol{r_{t+1}}$. Agent uses knowledge of state transitions, of the form $\boldsymbol{(s_t, a_t, s_{t+1}, r_{t+1})}$, in order to learn and improve its policy.

https://arxiv.org/pdf/1708.05866.pdf

- The dynamics of trial-and-error-learning has its roots in behaviourist psychology, being one of the main foundation of RL (Sutton et. Al., 1998)
- The best sequence of actions are determined by the rewards provided by the environment, every time the environment transitions to a new state, it also provides scalar reward $\boldsymbol{r_{t+1}}$ to the agent as a feedback. The goal of the agent is to learn policy π, that maximizes the expected return as reward
- For given state, a policy returns an action to perform an optimal policy (any policy) that maximizes the expected return in the environment, RL aims to solve the problem of optimal control, where the agent needs to learn about the consequences of actions by trial-and-error (Arulkumaran et. Al., 2017)



Deep RL

$$R = \sum_{t=0}^{T-1} \gamma^t r_t(b_t, a_t)$$

RL using Q-Learning

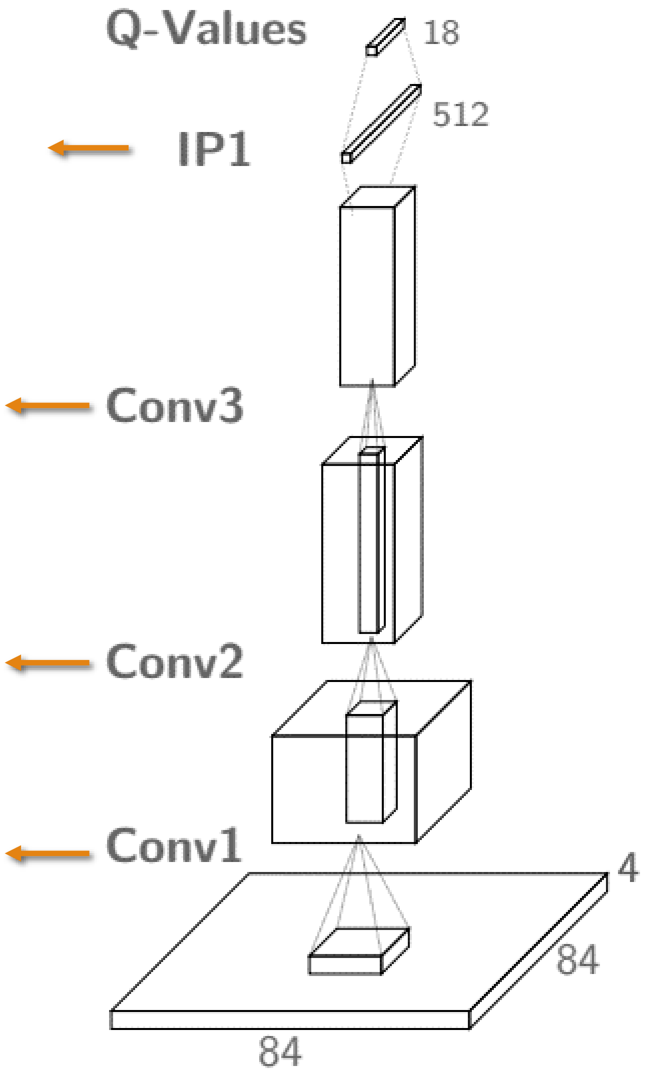
- Q-learning is a RL technique used in Machine Learning
- "Q" names the function that returns the reward used to provide the reinforcement and can be said to stand for the "quality" of an action taken in a given state. (Matiisen, 2015)
- Q-learning can identify an optimal action-selection policy for any, given infinite exploration time and a partly-random policy. (Melo, 2007)
- Q-learning finds a policy that is optimal in the sense that it maximizes the expected value of the total reward over all successive steps, starting from the current state. (Melo, 2007)
- Deep Q-learning (Deep RL)
  - RL method using deep neural network as Q-Value function approximator (Mnih et. Al., 2015)
  - A neural network is used to approximate the Q-values in a decision process (Egorov, 2015)
  - Q-values are parameterized by the belief and the action; belief state history. These modified Q-values can be learned by a neural network

Inner Product Layer

64 filters
3 X 3
Stride 1

64 filters
4 X 4
Stride 2

32 filters
8 X 8
Stride 4



$$\overset{\text{New Q}}{Q(s,a)} = \overset{\text{Old value}}{Q(s,a)} + \underset{\text{Learning rate}}{\alpha}\overbrace{\left(\underset{\text{Reward}}{r} + \underset{\text{Discount factor}}{\gamma}\underbrace{\max_{a'} Q(s',a') - Q(s,a)}_{\text{Estimation of Optimal future value}}\right)}^{\text{Learned value}}$$

Hyperparameter
Value- 0.001

Value- 1
To maximize
the future sum
of rewards

https://pdfs.semanticscholar.org/f5f3/23e62acb75f785e00b4c90ace16f1690076f.pdf

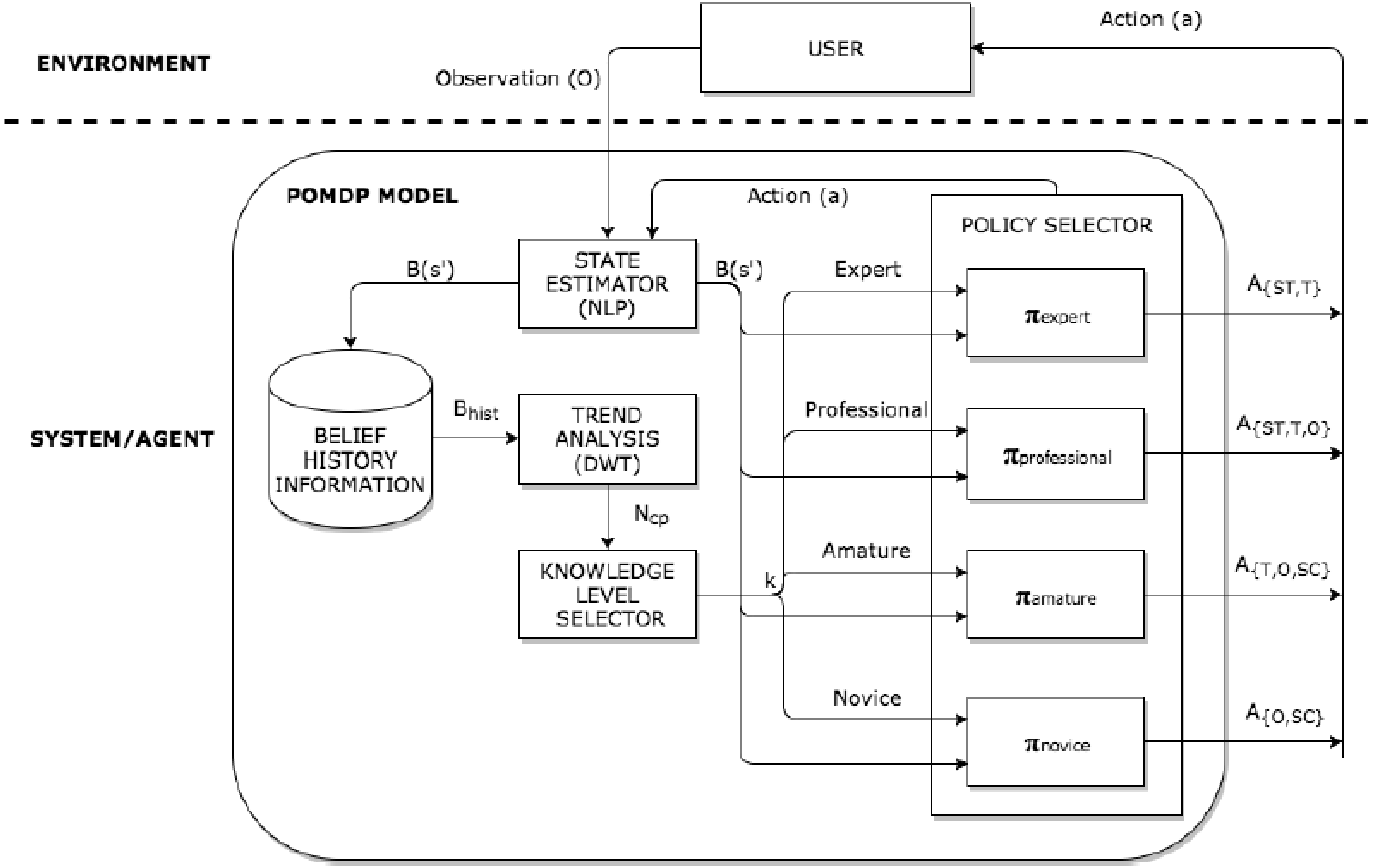
Proposed Architecture

DISCUSSION- COMPARISON & CONTRIBUTION

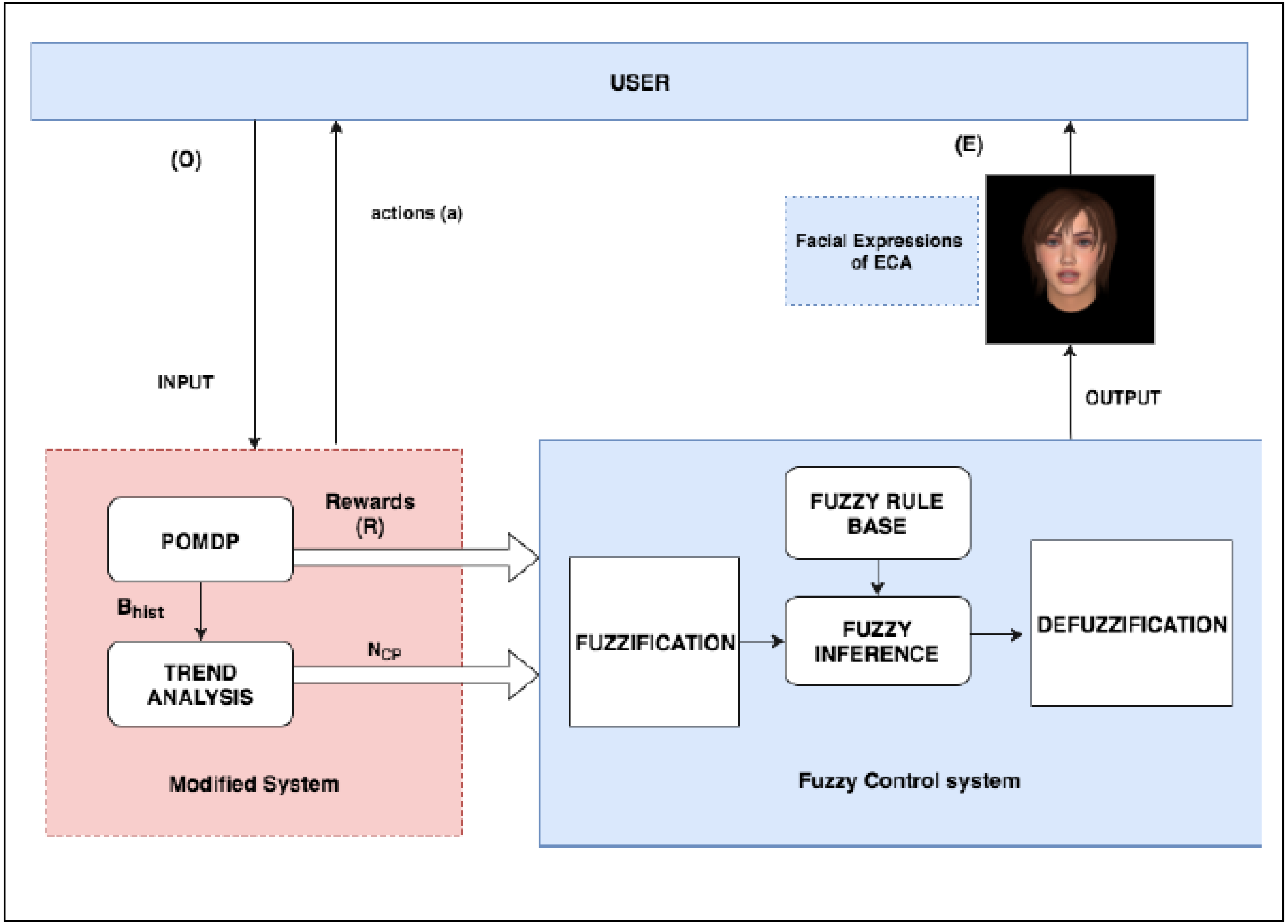
ARCHITECTURE DIAGRAM

ANALYSIS

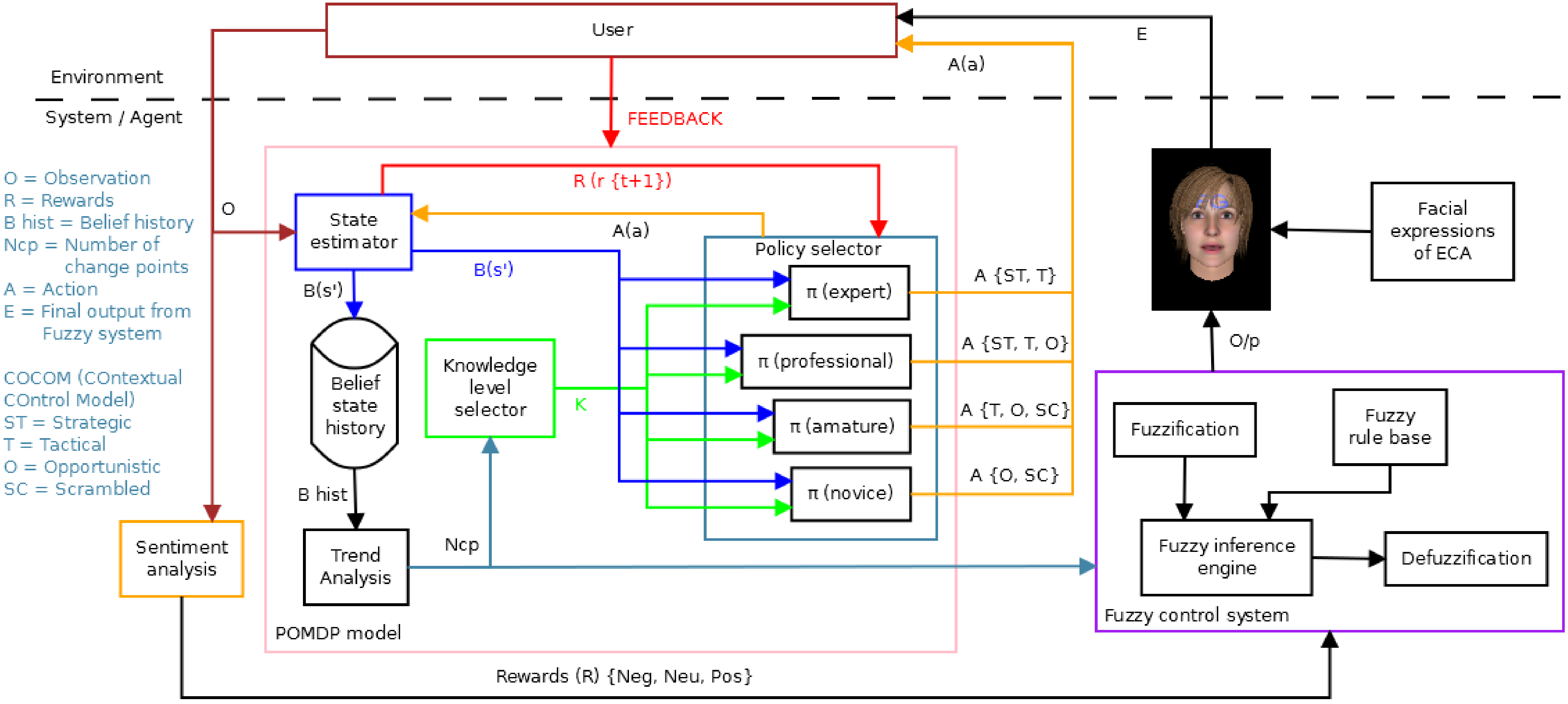
Discussion- Comparison and Contribution		
Bui, 2008	Worked on traditional POMDP model represents belief updating finding; finding the optimal policy Value-iteration based POMDP is used to compute optimal or near-optimal policy Affective dialogue system is used to issue two inputs of observations of action and state Rapid Dialogue Prototyping Methodology is presented using factored POMDP	Accuracy for 1000 runs (goal achievement) was measured – 67.9%
Mulpuri, 2016 (UoW)	Automation of requirement elicitation in software product line Decision-making algorithm for automation Used POMDP model; trend analysis on belief-state history to anticipate user knowledge level Knowledge level used in addressing policy selection to perform appropriate action	Accuracy for 1000 runs (goal achievement) was measured – 79.3%
Kaur, 2016 (UoW)	ECA with emotions offers better understanding of user import Fuzzy logic system used to generate facial expressions Usability study was conducted to improve the user interface Based on user's opinions, user satisfaction has been improved	Usability results were driven based on effectiveness, efficiency, satisfaction, learnability, easiness, necessity, etc.
Ruturaj R. Raval, 2019 (UoW)	Extended work of *Mulpuri* and *Kaur* Proposed architecture to improve user import using ***Sentiment Analysis*** to improve intention discovery; decreasing dialog length ***Reinforcement Learning*** is implemented using Q-learning technique to acquire optimal policy to reach user's goal Sentiment Analysis feeds Negative, Neutral, Positive values to fuzzy logic system to improve ECA emotion and Reinforcement Learning operates on feedback provision from user to agent based on trial-and-error action learning to update the knowledge	Anticipated planning to improve the accuracy of the policy and ECA embedded with emotions response will be improved.



(Mulpuri, 2016)


USER
(O)
INPUT
actions (a)
(E)
Facial Expressions of ECA
OUTPUT
POMDP
Rewards (R)
B_{hist}
TREND ANALYSIS
N_{CP}
Modified System
FUZZY RULE BASE
FUZZIFICATION
FUZZY INFERENCE
DEFUZZIFICATION
Fuzzy Control system

(Kaur, 2016)



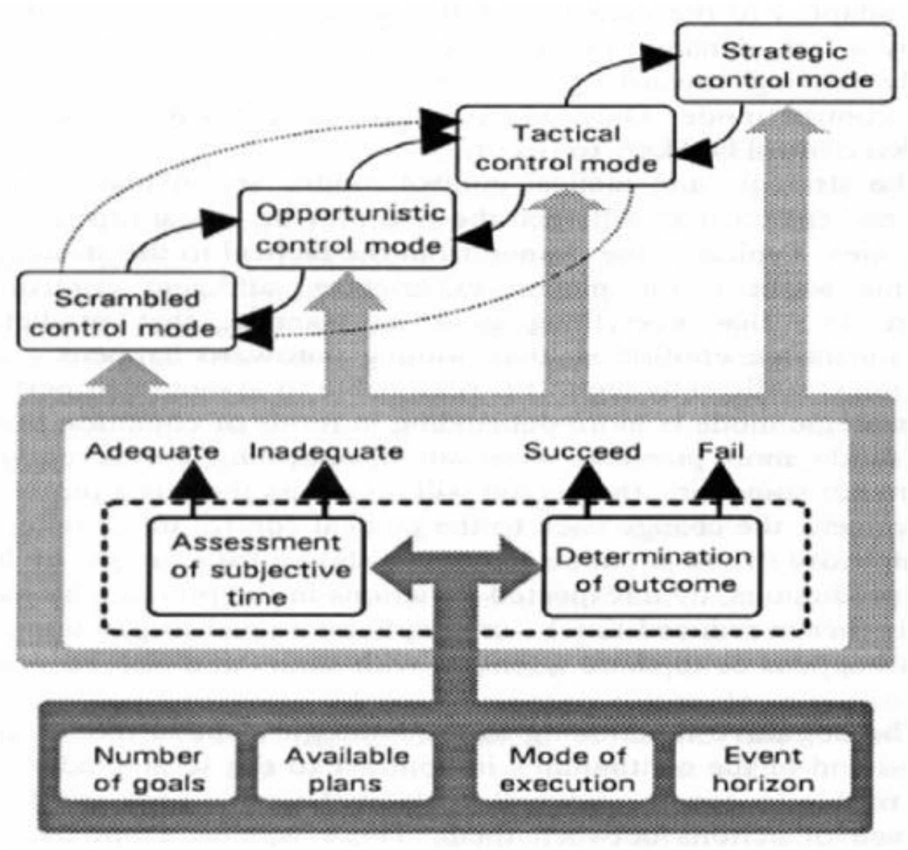
O = Observation
R = Rewards
B hist = Belief history
Ncp = Number of change points
A = Action
E = Final output from Fuzzy system

COCOM (COntextual COntrol Model)
ST = Strategic
T = Tactical
O = Opportunistic
SC = Scrambled

Feedback = reward r_{t+1} and knowledge of state transitions, of the form $(s_t, a_t, s_{t+1}, r_{t+1})$, in order to learn and improve the policy

(Diagram- Proposed Architecture, 2019)

Analysis



FAMM (Fuzzy Associative Memory Matrix)	***Fuzzy RULE base selection based on Trend Analysis performed on Belief-state history using COCOM***			
Sentiment Analysis Reward	Strategic	Tactical	Opportunistic	Scrambled
Negative	Disgust	Anger		Fear
Neutral	Fear	Sad	Surprise	Sad
Positive	Happy		Surprise	

COCOM – *CO*ntextual *CO*ntrol *M*odel

Algorithm

ALGORITHM

DATASET & TOOLS

Algorithm		

Initialization	1
	2
	3
	4
	5
	6
	7
	8
	9
		isGoalState ← false
		belief ← 1
		CREATE empty LIST bhist
		ADD belief to LIST bhist
		WHILE isGoalState NOT EQUAL true
		input ← READ(observation)
		IF input EQUAL ‘exit’ THEN
		isGoalState ← true
		ELSE

StateEstimator	10
	11
	12
	13
	14
	15
		StateEstimator(input, belief) ← USER_FEEDBACK
		b(s’) ← StateEstimator(input, belief) T.C.- $O(n^2)$
		tokens ← NLP(input)
		s’ ← MATCH_SERVICES(tokens)
		b(s’) ← Pr(s’ | belief, action, input)
		return b(s’)

Feedback	16 17 18	Rewards (R(r{t+1})) ← StateEstimator(input, belief) π{m} ← Rewards (R(r{t+1})) ADD b(s’) to LIST bhist

TrendAnalysis	19
	20
	21
		Ncp ← TrendAnalysis(bhist) T.C.- $O(n^2)$
		Ncp ← DWT(bhist)
		return Ncp

T.C. = Time Complexity
n = input string length

KnowledgeLevelSelector	22	k ← KnowledgeLevelSelector(Ncp) T.C.- O(1)
	23	expertThreshold ← READ_FROM_TRAINED_MODEL
	24	professionalThreshold ← READ_FROM_TRAINED_MODEL
	25	amateurThreshold ← READ_FROM_TRAINED_MODEL
	26	noviceThreshold ← READ_FROM_TRAINED_MODEL
	27	IF Ncp < expertThreshold THEN
	28	k ← ‘expert’
	29	ELSE IF Ncp >= expertThreshold AND Ncp < professionalThreshold
	30	k ← ‘professional’
	31	ELSE IF Ncp >= professionalThreshold AND Ncp < amateurThreshold
	32	k ← ‘amateur’
	33	ELSE
	34	k ← ‘novice’
	35	ENDIF
	36	return k
SentimentAnalysis	37	R ← SentimentAnalysis{Neg, Neu, Pos} T.C.- O(1)
FuzzyLogicSystem	38	Ncp ← {ST, T, O, SC}
	39	F_Ncp ← FUZZIFICATION (Ncp)
	40	F_R ← FUZZIFICATION (R)
	41	FAMM ← LOAD_RULE_BASE ()
	42	W ← INFERENCE (F_Ncp, F_R, FAMM)
	43	output ← DEFUZZIFICATION (W, FAMM)
	44	GENERATE_FACIAL_EXPRESSION (output)

T.C. = Time Complexity
n = input string length

PolicySelector	45 46 47 48 49 50 51	π{m} ← PolicySelector(k) T.C.- O(1) CASE k OF expert: return πexpert professional: return πprofessional amateur: return πamateur novice: return πnovice ENDCASE
MakeAction	52 53 54 55 56	action ← MakeAction(π{m}, b(s’)) T.C.- O(1) action ← GET(Action from Transition of state s to s’) return action belief ← b(s’) // updating the belief value PRINT action

End	57
	58
		ENDIF
		END WHILE

T.C. = Time Complexity
n = input string length

Datasets and Tools	
Item	***Details***
Project name	Avatar Interaction
OS	Android & Windows
Languages	Java, XML & Python
IDEs	Android Studio 3.1, Notepad++ & IDLE & R studio
Database	SquliteDB, CouchDB
3D models	*Library*: jpct_ae.jar & *Software*: Facegen
Simulator	Android studio ADB, Genymotion & Terminal
Sentiment Analysis	NLTK (Natural Language Tool Kit) Python library
Word(s) embedding	Word2vec (pretrained model- GloVe)
Training dataset	*Emotions*: 7500+ sentences; 1500+ words to train with emotions & *Dialogues*: ConvAI2; VisDial (354:17:10M)
Deep Q-Learning	Tensorflow; PyTorch
Wizard-of-Oz	WoZ: Relational Strategies in Customer Service Dataset
R studio	fuzzyR (Fuzzy), waveslim (Trend), pomdp (POMDP)

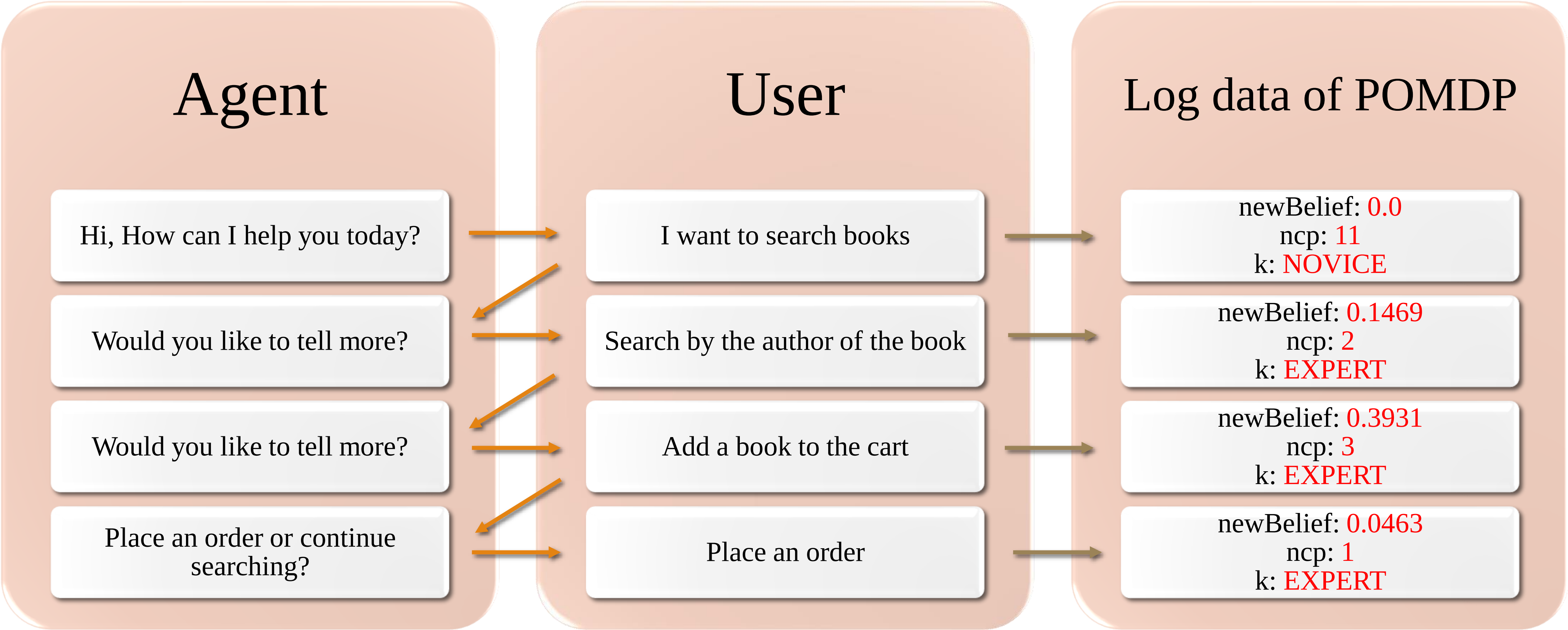
Experiment & Design

EXAMPLE

PROTOTYPE ENVIRONMENT

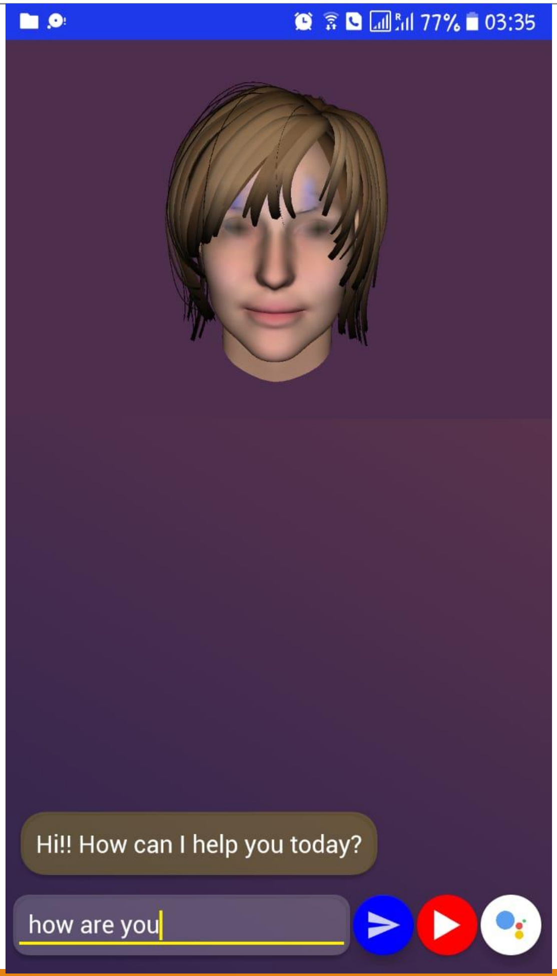
CHALLENGES

Example



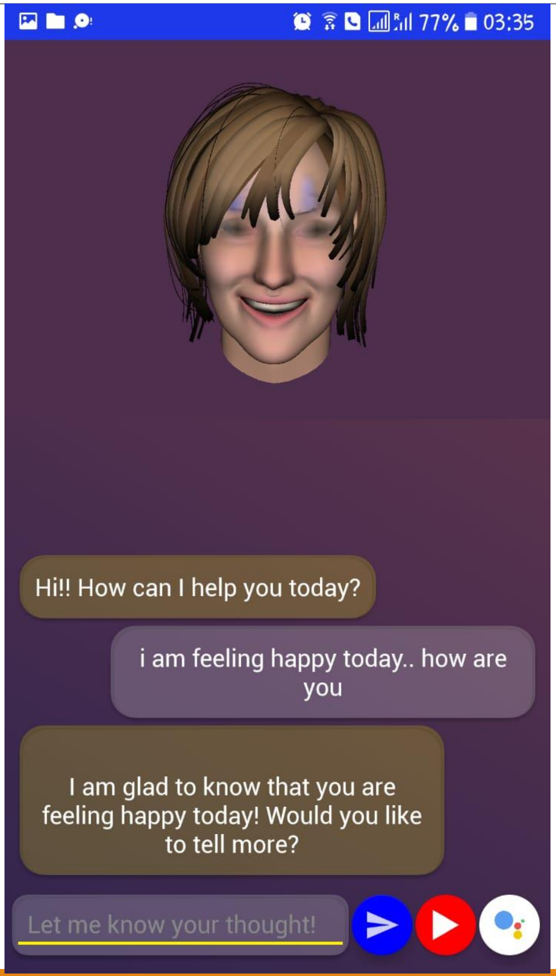
Agent	User	Log data of POMDP
Hi, How can I help you today?	I want to search books	newBelief: 0.0 ncp: 11 k: NOVICE
Would you like to tell more?	Search by the author of the book	newBelief: 0.1469 ncp: 2 k: EXPERT
Would you like to tell more?	Add a book to the cart	newBelief: 0.3931 ncp: 3 k: EXPERT
Place an order or continue searching?	Place an order	newBelief: 0.0463 ncp: 1 k: EXPERT

Prototype environment



Hi!! How can I help you today?

how are you

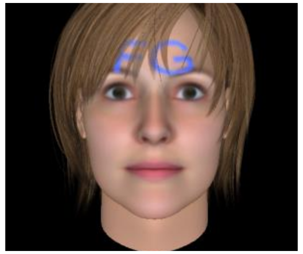


Hi!! How can I help you today?

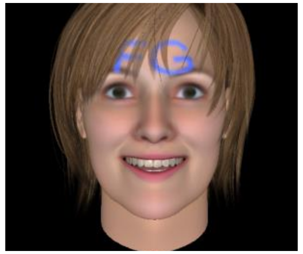
i am feeling happy today.. how are you

I am glad to know that you are feeling happy today! Would you like to tell more?

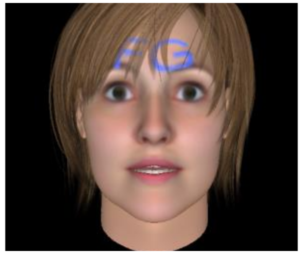
Let me know your thought!



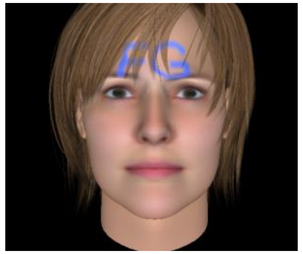
Neutral



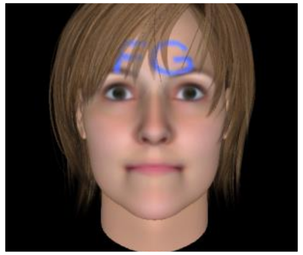
Joy



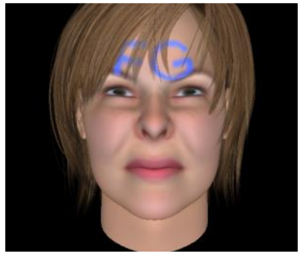
Surprise



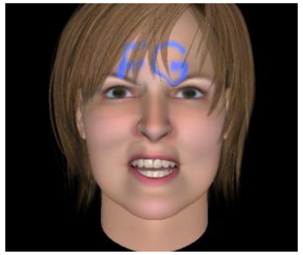
Sad



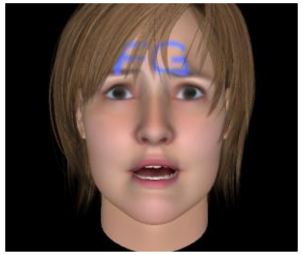
Shame



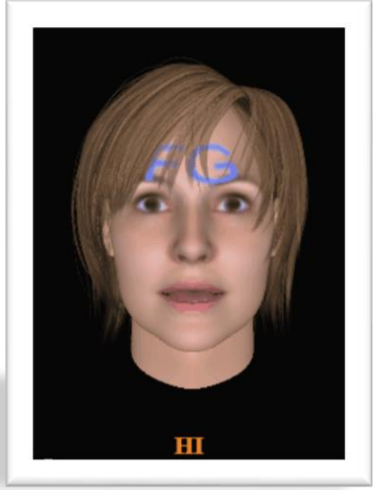
Angry

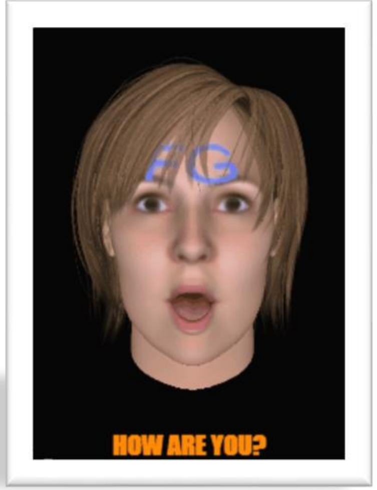


Disgust

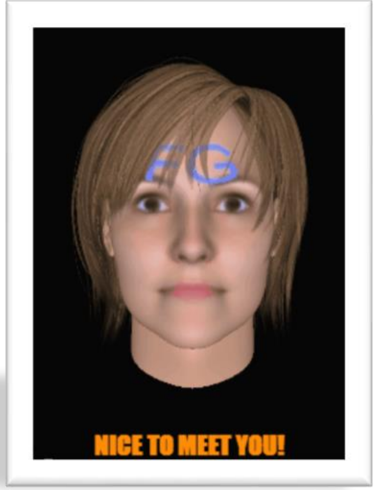


Fear





I AM FINE.
THANK YOU!



```
Hi, How are you?
neg: 0.0, neu: 1.0, pos: 0.0,
I am fine, not bad! you say!
neg: 0.0, neu: 0.433, pos: 0.567,
You know what you are good for nothing, I have no good reviews..
neg: 0.0, neu: 0.539, pos: 0.461,
I am glad that you brought this up!!
neg: 0.0, neu: 0.626, pos: 0.374,
Well, this is the most untidy thing I have seen urghhh..!!
neg: 0.0, neu: 0.77, pos: 0.23,
I don't care, I don't have trust...
neg: 0.396, neu: 0.604, pos: 0.0,
I don't feel good after this, this is so horrific..
neg: 0.231, neu: 0.769, pos: 0.0,
```

```
Most automated sentiment analysis tools are shit.
neg: 0.375, neu: 0.625, pos: 0.0,
VADER sentiment analysis is the shit.
neg: 0.0, neu: 0.556, pos: 0.444,
Sentiment analysis has never been good.
neg: 0.325, neu: 0.675, pos: 0.0,
Sentiment analysis with VADER has never been this good.
neg: 0.0, neu: 0.703, pos: 0.297,
```

FAMM	***Fuzzy RULE base selection based on Trend Analysis performed on Belief-state history using COCOM***			
Sentiment Analysis Reward	Strategic	Tactical	Opportunistic	Scrambled
Negative	Disgust	Anger		Fear
Neutral	Fear	Sad	Surprise	Sad
Positive	Happy		Surprise	

FAMM	Strategic	Tactical	Opportunistic	Scrambled
Negative	W1	W4	W7	W10
Neutral	W2	W5	W8	W11
Positive	W3	W6	W9	W12

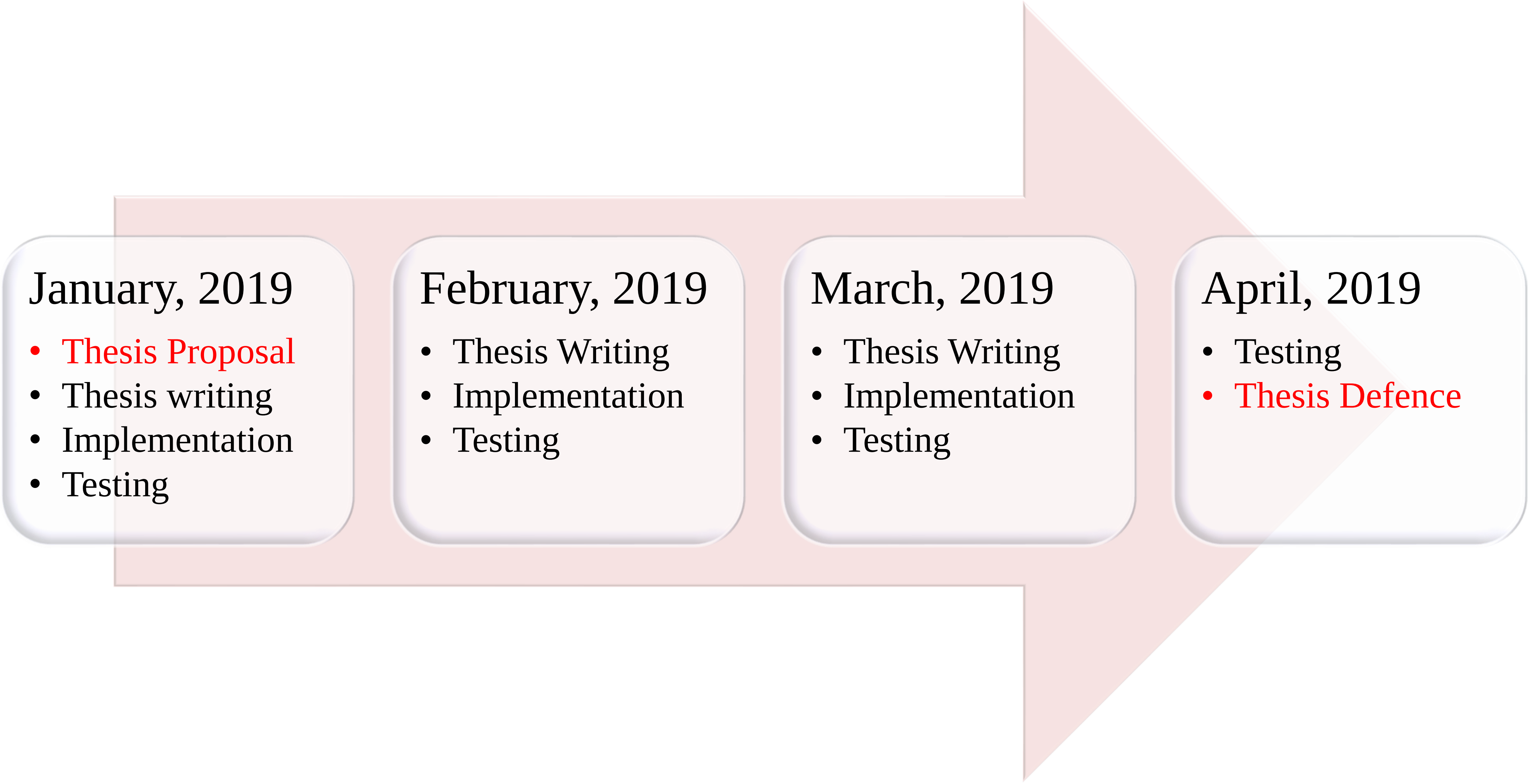
Most automated sentiment analysis tools are shit.				
FAMM	Strategic	Tactical	Opportunistic	Scrambled
Negative 0.375	W1*Disgust	W4*Anger	W7*Anger	W10*Fear
Neutral 0.625	***W2*Fear***	***W5*Sad***	W8*Surprise	W11*Sad
Positive 0.0	W3*Happy	W6*Happy	W9*Surprise	W12*Surprise

Challenges

- User import in slang will not offer good accuracy to reach optimal policy, as sentiment can’t be detected and misunderstanding might occur
- RL operates on trial-and-error while solving the optimal control aim, which generates uncertain consequences as actions in the environment are still under training
- The optimal policy must be inferred by trial-and-error interaction with the environment, the only learning signal the agent receives is the reward
- The observations of the agent depend on its actions and can contain strong temporal correlations
- Agents must deal with long-range time dependencies
  - The consequences of an action only materialize after many transitions of the environment known as *credit assignment problem*

Timeframe of Further work

TIMELINE OF FUTURE WORK



January, 2019

- Thesis Proposal
- Thesis writing
- Implementation
- Testing

February, 2019

- Thesis Writing
- Implementation
- Testing

March, 2019

- Thesis Writing
- Implementation
- Testing

April, 2019

- Testing
- Thesis Defence

Conclusion

CONCLUSION UNTIL THESIS PROPOSAL

Conclusion

- Preliminary implementation is achieved
- POMDP-based dialogue management helps in improving intention discovery for ECA (agent) eventually helps in improving policy
- Sentiment analysis helps in understanding emotion detection from user input to decide goal-driven aim achievement
- Reinforcement learning helps in learning optimal policy, which improves intention discovery which helps in reducing dialogue length, making dialogue management conversation smooth than the former

Future work

FUTURE WORK FOR THESIS DEFENCE

Future work

- Future work will be carried out before thesis defence will focus on implementing proposed approach to perform experiments with listed datasets using tools and relevant libraries, etc.
- Additionally, final results will be obtained to showcase the improved intention to reach user’s goal to compare with other existing models

References

Reference 1

- J. Hollan, E. Hutchins and D. Kirsh, “Distributed cognition: toward a new foundation for human-computer interaction research,” ACM Transactions on Computer-Human Interaction (TOCHI), vol. 7, no. 2, pp. 174-196, 2000.
- W. C. Contributors, “File:Linux kernel INPUT OUPUT evdev gem USB framebuffer.svg,” Wikimedia Commons, the free media repository., 29 November 2015. [Online]. Available: https://commons.wikimedia.org/w/index.php?title=File:Linux_kernel_INPUT_OUPUT_evdev_gem_USB_framebuffer.svg&oldid=180540123. [Accessed 20 December 2018]
- D. D. Fehrenbacher, “Affect Infusion and Detection through Faces in Computer-mediated Knowledge-sharing Decisions,” Journal of the Association for Information Systems, vol. 18, no. 10, 2017.
- A. Serenko, N. Bontis and B. Detlor, “End-user adoption of animated interface agents in everyday work applications,” Behaviour and Information Technology, vol. 26, no. 2, pp. 119-132, 2007.
- W. Contributors, “Embodied agent,” Wikipedia, The Free Encyclopedia., 01 August 2018. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Embodied_agent&oldid=853012404. [Accessed 22 December 2018].
- S. J. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, New Jersey: Prentice Hall: (2nd ed.), Upper Saddle River, Chapt. 2, 2003.
- T. Zhao and M. Eskenazi, “Towards end-to-end learning for dialogue state tracking and management using deep reinforcement learning,” in SIGDIAL, 2016.

Reference 2

- V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra and M. Riedmiller, “Playing Atari with Deep Reinforcement Learning,” in NIPS Deep Learning Workshop, 2013.
- F. J. Seron and C. Bobed, “VOX system: a semantic embodied conversational agent exploiting linked data,” Multimedia tools appl, vol. 75, pp. 381-404, 2016.
- S. Ultes, L. Rojas-Barahona, P.-H. Su, D. Vandyke, D. Kim, I. Casanueva, P. Budzianowski, N. Mrksic, T.-H. Wen, M. Gasic and S. Young, “PyDial: A multi-domain statistical dialogue system toolkit,” Association for computational linguistics-system demonstrations, pp. 73-78, 2017.
- Z. Liu, S. He and W. Xiong, "A fuzzy logic based emotion model for virtual human," in Cyberworlds: International conference on IEEE, 2008.
- J. M. Mendel, "Fuzzy logic systems for engineering: a tutorial," in Proceedings of the IEEE, 1995.
- Z. C. Yildiz, "A Short Fuzzy Logic Tutorial," 08 April 2010. [Online]. Available: http://cs.bilkent.edu.tr/~zeynep/files/short_fuzzy_logic_tutorial.pdf. [Accessed 29 December 2018].
- J. Alexandersson, A. Girenko, V. Petukhova, D. Klakow, N. Taatgen, N. Campbell, A. Stricker, D. Spiliotopoulos, D. Koryzis, M. Specht, M. Aretoulaki and M. Gardner, “Metalogue: a multiperspective multimodal dialogue system with metacognitive abilities for highly adaptive and flexible dialogue management,” in IEEE: Intelligent Environments, 2014.

Reference 3

- H. Cuayahuitil, S. Keizer and O. Lemon, “Strategic dialogue management via deep reinforcement learning,” NIPS'15 Workshop on Deep Reinforcement Learning, vol. 1511, no. 08099v1, 2015.
- X. Yuan and L. Bian, “A modified approach of POMDP-based dialogue management,” IEEE, 2010.
- T. H. Bui, Toward affective dialogue management using partially observable Markov decision processes, Enschede, Netherlands: University of Twente, 2008.
- V. K. Mulpuri, “Trend Analysis of Belief-State History with Discrete Wavelet Transform for Improved Intention Discovery,” Electronic Theses and Dissertations, Windsor, 2016.
- K. Kaur, “An Approach of Facial Expression Modeling with Changing Trend in the History of Belief States,” Electronic Theses and Dissertations, Windsor, 2016.
- X. Li, Y.-N. Chen, L. Li, J. Gao and A. Celikyilmaz, “End-to-end task-completion neural dialogue systems,” Computation and Language, 2018.
- E. Klein, "Sentiment Analysis," Github: nltk, 20 October 2015. [Online]. Available: https://github.com/nltk/nltk/wiki/Sentiment-Analysis. [Accessed 31 December 2018].
- Bird, Steven, E. Loper and E. Klein, Natural Language Processing with Python, O'Reilly Media Inc., 2009.

Reference 4

- K. Arulkumaran, M. P. Deisenroth, M. Brundage and A. A. Bharath, "A brief survey of deep reinforcement learning," IEEE signal processing magazine, special issue on deep learning for image understanding, vol. 1708, no. 05866v2, 2017.
- R. S. Sutton and A. G. Barto, "Reinforcement learning: an introduction," MIT press, 1998.
- A. L. Strehl, L. Li, E. Wiewiora, J. Langford and M. L. Littman, "PAC model-free reinforcement learning," ICML, 2006.
- T. Matiisen, "DEMYSTIFYING DEEP REINFORCEMENT LEARNING," COMPUTATIONAL NEUROSCIENCE LAB, 19 December 2015. [Online]. Available: https://neuro.cs.ut.ee/demystifying-deep-reinforcement-learning/. [Accessed 05 January 2019].
- F. S. Melo, "Convergence of Q-learning: a simple proof," Institute for Systems and Robotics, Instituto Superior Técnico, Lisboa, PORTUGAL, 2007.
- V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg and D. Hassabis, "Human-level control through deep reinforcement learning," Nature: International Journal of Science, pp. 529-533, 2015.
- M. Egorov, "Deep Reinforcement Learning with POMDPs," 11 December 2015. [Online]. Available: http://cs229.stanford.edu/proj2015/363_report.pdf. [Accessed 05 January 2019].

THANK YOU!

ANY QUESTIONS / SUGGESTIONS?