Dialog Policy Learning for Joint Clarification and Active Learning Queries

Aishwarya Padmakumar[2], Raymond J. Mooney[1]

[1]Department of Computer Science, The University of Texas at Austin

[2]Amazon

Dialog Policy Learning for Joint Clarification and Active Learning Queries

Learn a dialog policy for a task oriented dialog system that trades off

- Model Improvement
- Clarification and task completion

Outline

- Introduction
- Task Setup
- User Simulator
- Dialog Policy Model
- Experiments

Outline

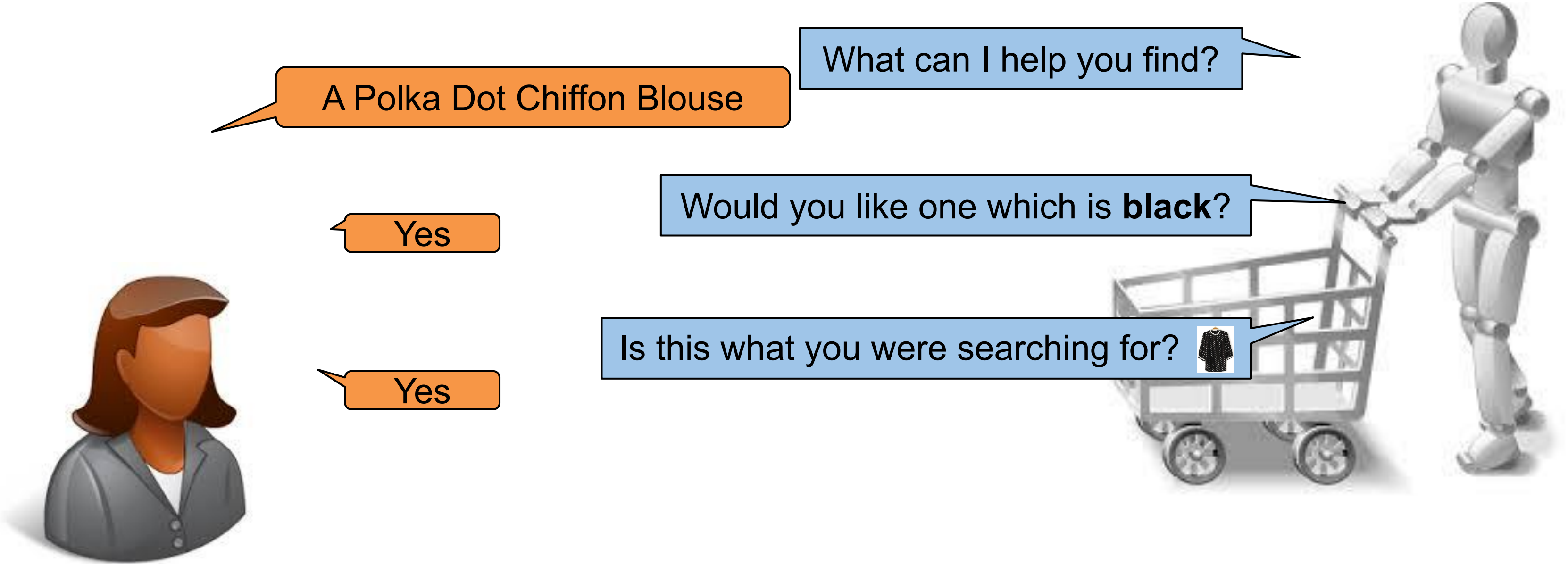
- Introduction
- Task Setup
- User Simulator
- Dialog Policy Model
- Experiments

Dialog Policy Learning for Joint Clarification and Active Learning Queries

Learn a dialog policy for a task oriented dialog system that trades off

- Model Improvement
- Clarification and task completion

Task Oriented Dialog Systems

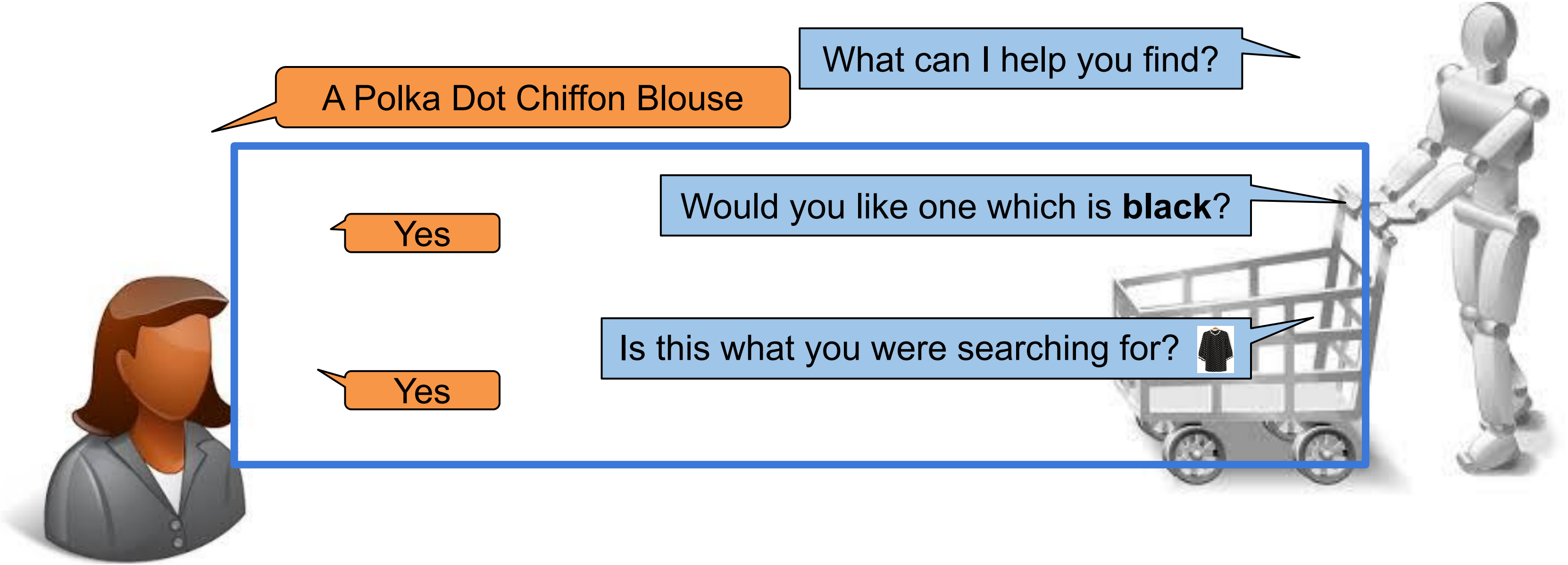


Dialog Policy Learning for Joint Clarification and Active Learning Queries

Learn a dialog policy for a task oriented dialog system that trades off

- Model Improvement
- Clarification and task completion

Task Oriented Dialog Systems

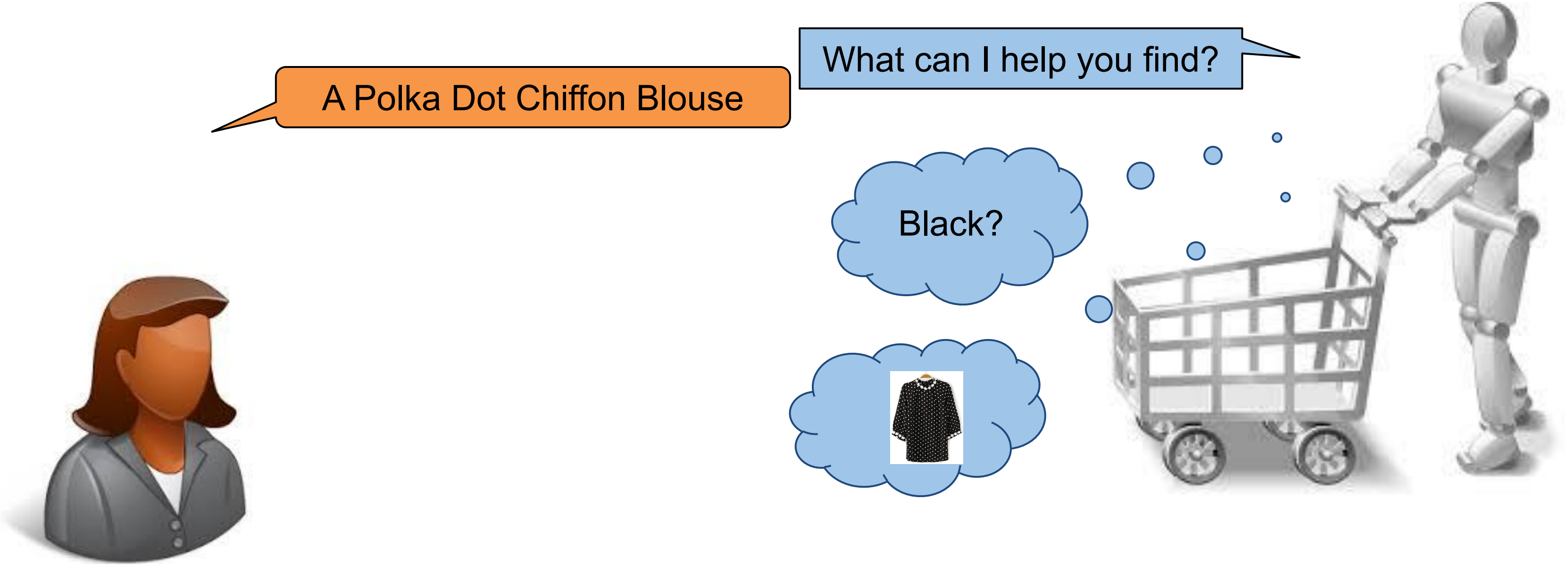


Dialog Policy Learning for Joint Clarification and Active Learning Queries

Learn a dialog policy for a task oriented dialog system that trades off

- Model Improvement
- Clarification and task completion

Task Oriented Dialog Systems

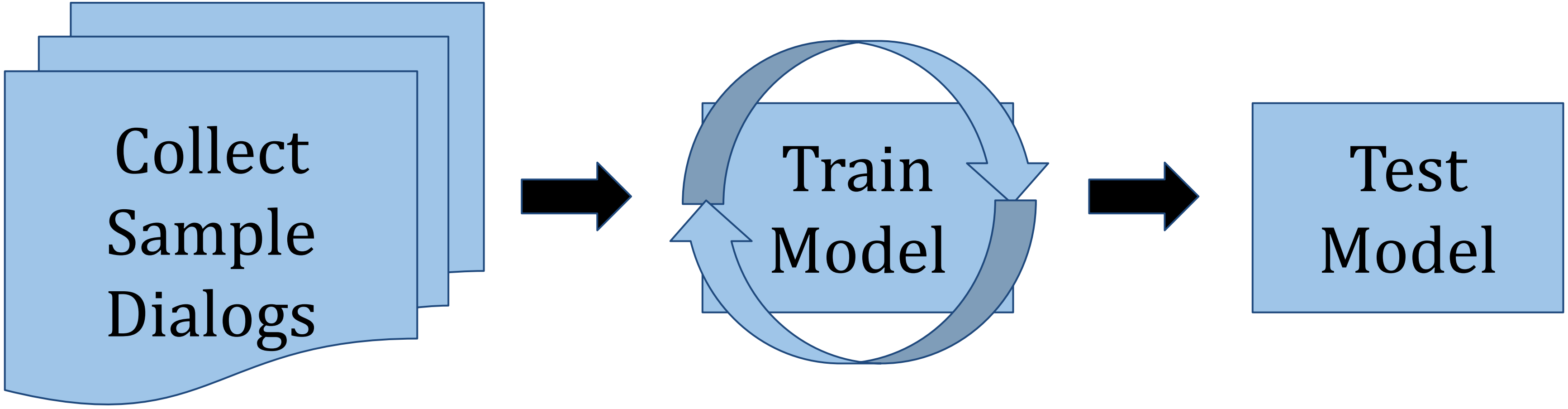


Dialog Policy Learning for Joint Clarification and Active Learning Queries

Learn a dialog policy for a task oriented dialog system that trades off

- Model Improvement
- Clarification and task completion

Standard Supervised Learning Pipeline



Handling a Changing Inventory

Dresses



Tops



Handling a Changing Inventory

Dresses



Tops



Masks



Opportunistic Active Learning

(Thomason et al., CoRL 2017)

Bring the blue mug from Alice's office

Would you use the word "blue" to refer to this object?

Yes

Opportunistic Active Learning

- A framework for incorporating active learning queries into test time interactions.
- Agent asks locally convenient questions during an interactive task to collect labeled examples for supervised learning.
- Questions may not be useful for the current interaction but expected to help future tasks.

Opportunistic Active Learning

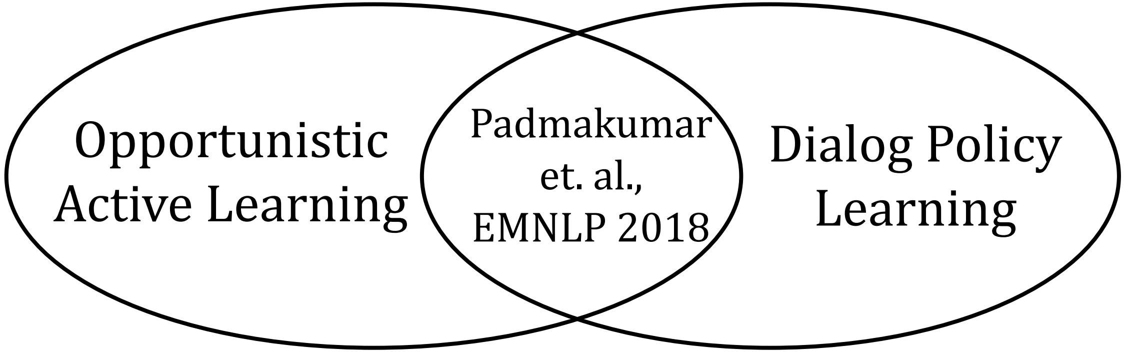
Bring the **blue mug** from Alice’s office

Would you use the word “**tall**” to refer to this object?

Yes

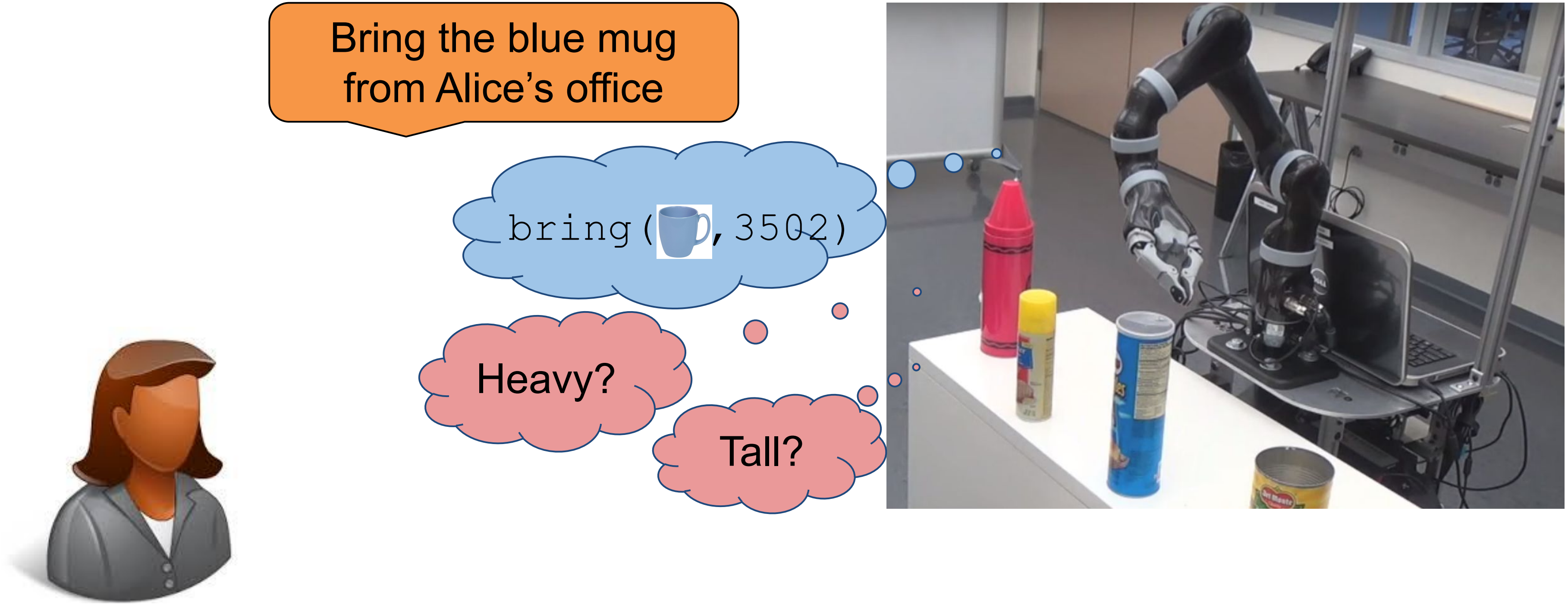
Learning a Policy for Opportunistic Active Learning

Padmakumar et. al., EMNLP 2018

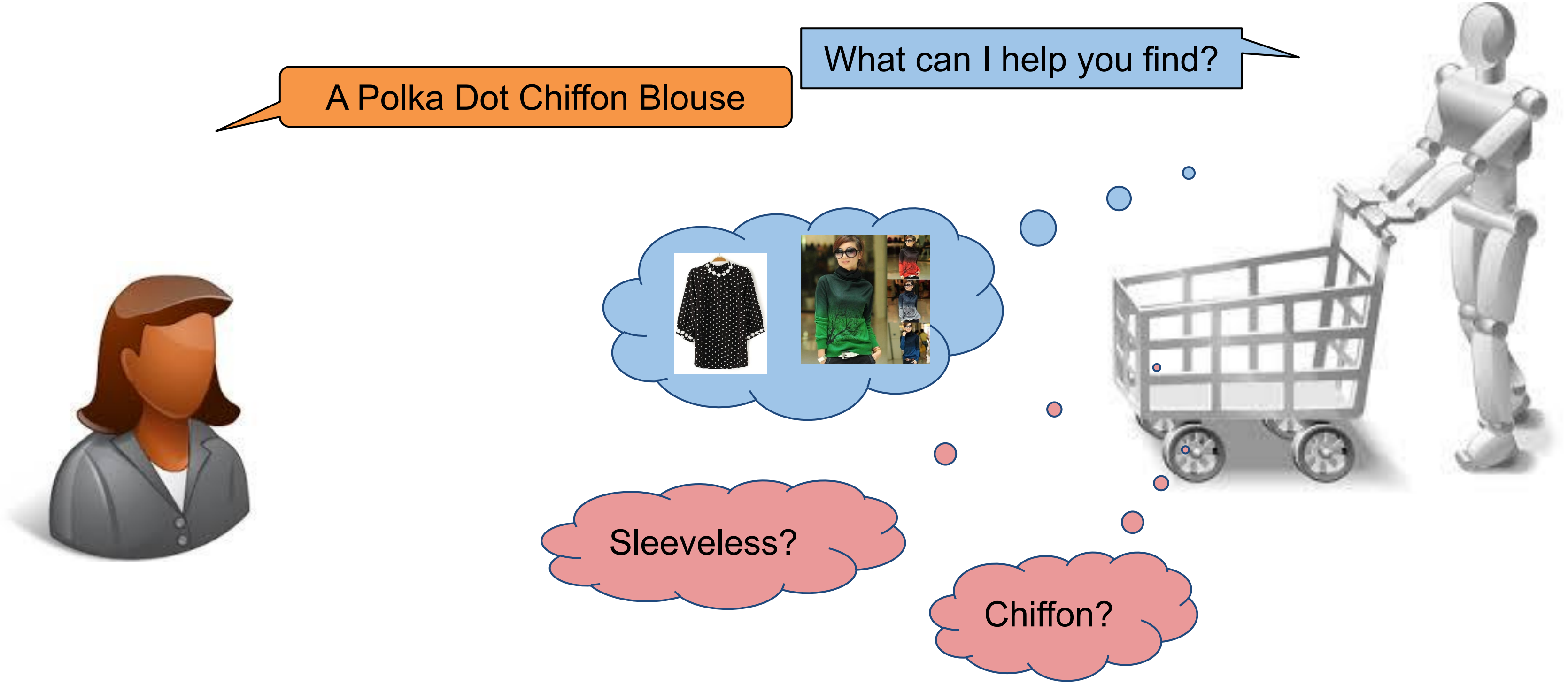


Learns to trade-off between executing an interpreted user command and using opportunistic active learning to improve the underlying models used to understand the command.

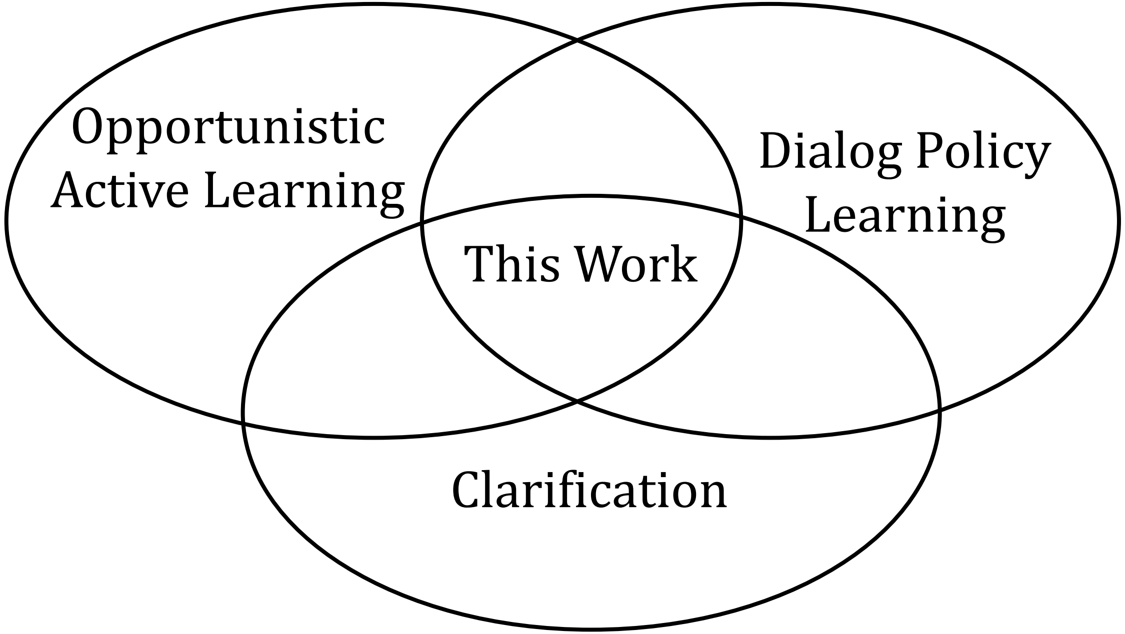
Previous Work



This Work


What can I help you find?
A Polka Dot Chiffon Blouse
Sleeveless?
Chiffon?

Dialog Policy Learning for Joint Clarification and Active Learning Queries



Dialog Policy Learning for Joint Clarification and Active Learning Queries

Learn a dialog policy for a task oriented dialog system that trades off

- **Model Improvement:** Model improvement using opportunistic active learning to better understand future commands
- **Clarification and task completion:** Obtain additional information needed; clarify the completed command; execute a system action

Outline

- Introduction
- Task Setup
- User Simulator
- Dialog Policy Model
- Experiments

Task Setup

- Motivated by an online shopping application
- Use clarifications to help refine search queries
- Use active learning to improve the model that retrieves products based on search queries.

Attribute Based Clarification

- Attribute - any property that can be used to describe a product - categories, colors, shapes, domain specific properties.
- A clarification action corresponds to selecting an attribute.
- Provide ground truth answers to questions for training in simulation.

Outline

- Introduction
- Task Setup
- User Simulator
- Dialog Policy Model
- Experiments

Dataset

- We simulate dialogs using the iMaterialist Fashion Attribute dataset.
- Images have associated product titles and are annotated with binary labels for 228 attributes.
- Attributes: Dress, Shirt, Red, Blue, V-Neck, Pleats, ...

Sample Interaction

Active Training Set

Active Test Set

Sample Interaction

Active Training Set

Active Test Set

Target Image

Sample Interaction

What can I help you find?

A Polka Dot Chiffon Blouse

Simulated User Query: Product title of target image

Sample Interaction

Possible System Actions

- Clarification
- Label Query
- Example Query
- Guess

Sample Interaction

Possible System Actions

- Clarification
- Label Query
- Example Query
- Guess

Sample Interaction

What can I help you find?

A Polka Dot Chiffon Blouse

Would you like one which is **black**?

Yes

Clarification: Does selected attribute apply to target image?

Sample Interaction

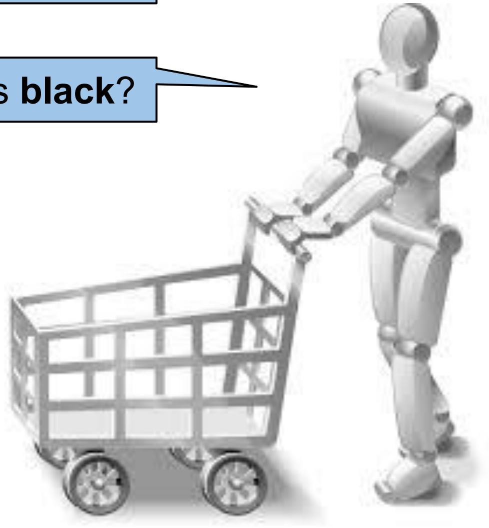
What can I help you find?

A Polka Dot Chiffon Blouse

Would you like one which is **black**?

Yes

Yes/No Response: From attribute labels of target image



Sample Interaction

Possible System Actions

- Clarification
- Label Query
- Example Query
- Guess

Sample Interaction

Active Training Set

Select image for label query

Active Test Set

Sample Interaction

What can I help you find?

A Polka Dot Chiffon Blouse

Would you describe this as **sleeveless**?

Yes

Label Query: *Does selected attribute apply to selected image?*

Sample Interaction

What can I help you find?

A Polka Dot Chiffon Blouse

Would you describe this as **sleeveless**?

Yes

Yes/No Response: From attribute labels of selected image

Sample Interaction

Possible System Actions

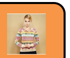
- Clarification
- Label Query
- Example Query
- Guess

Sample Interaction

What can I help you find?

A Polka Dot Chiffon Blouse

Can you show me something you would describe as **chiffon**?



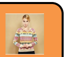
Example Query: Request positive example for selected attribute

Sample Interaction

What can I help you find?

A Polka Dot Chiffon Blouse

Can you show me something you would describe as **chiffon**?



Example Image: From active training set, selected using attribute labels

Sample Interaction

Active Training Set

Active learning queries reference images in the active training set so that the system is forced to learn generaliable classifiers.

Active Test Set

Sample Interaction

Possible System Actions

- Clarification
- Label Query
- Example Query
- Guess

Sample Interaction

What can I help you find?

A Polka Dot Chiffon Blouse

Is this what you were searching for?

Yes

Guess: Selects image from active test set

Sample Interaction

What can I help you find?

A Polka Dot Chiffon Blouse

Is this what you were searching for?

Yes

0/1 Success: *Compare guessed image with target image*

Sample Interaction

What can I help you find?

A Polka Dot Chiffon Blouse

Would you like one which is **black**?

Yes

Would you describe this as **sleeveless**?

Yes

Can you show me something you would describe as **chiffon**?

Is this what you were searching for?

Yes

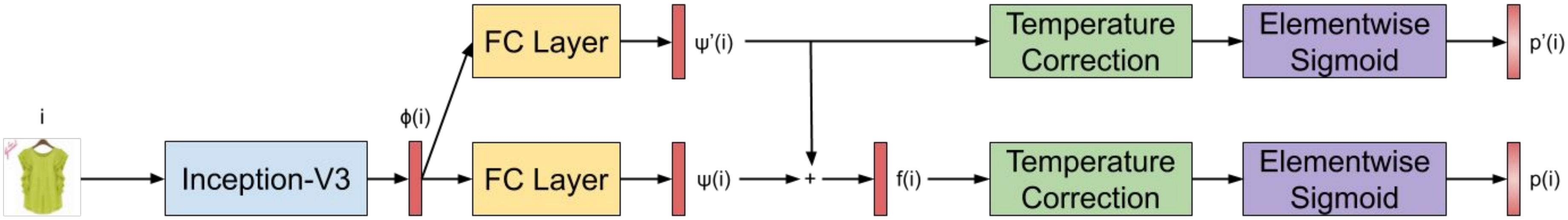
System Goal

- Maximize fraction of successful dialogs.
- Keep dialogs as short as possible.

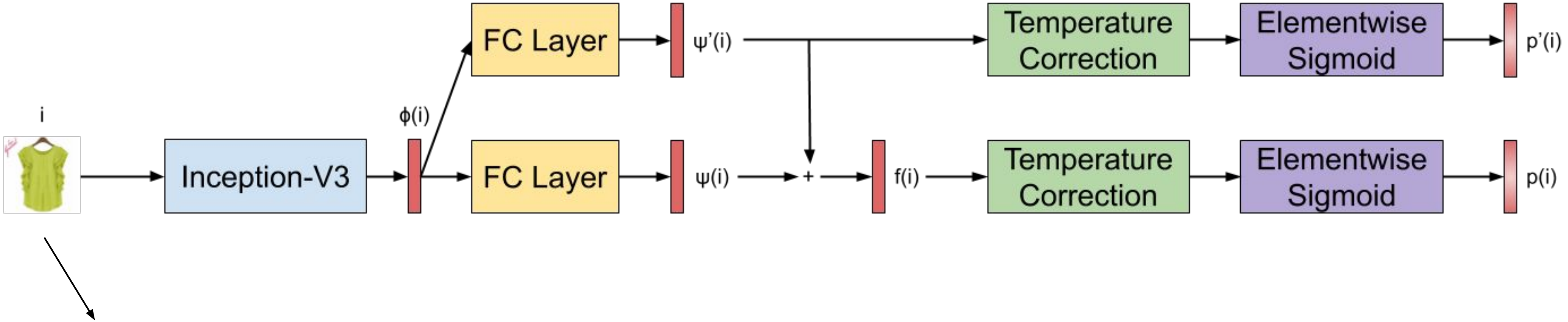
Outline

- Introduction
- Task Setup
- User Simulator
- Dialog Policy Model
- Experiments

Visual Attribute Classifier

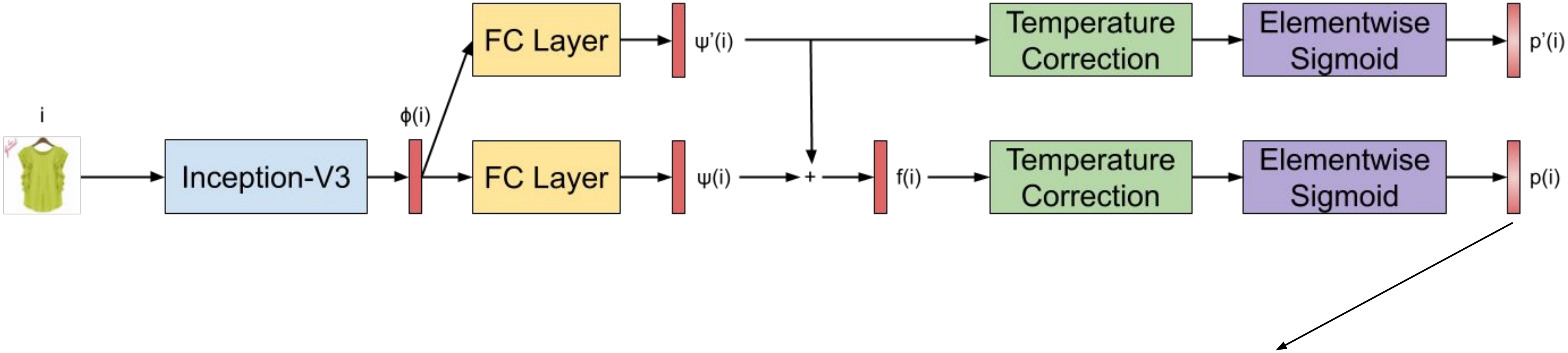


Visual Attribute Classifier



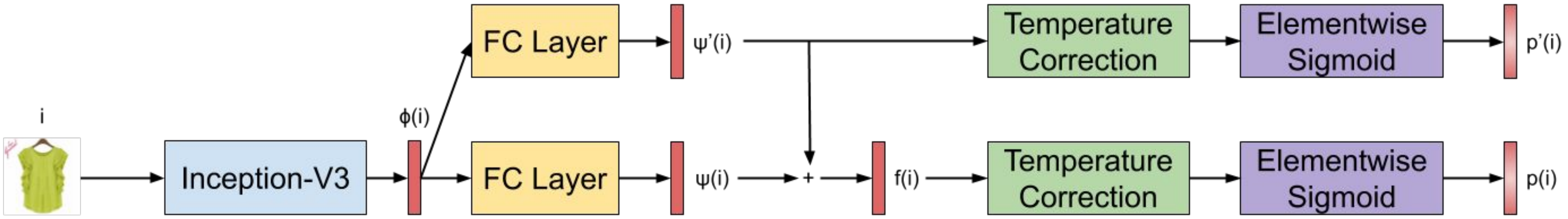
Input image

Visual Attribute Classifier



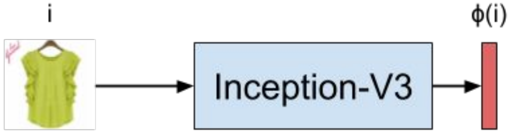
Probability of attributes being positive

Visual Attribute Classifier

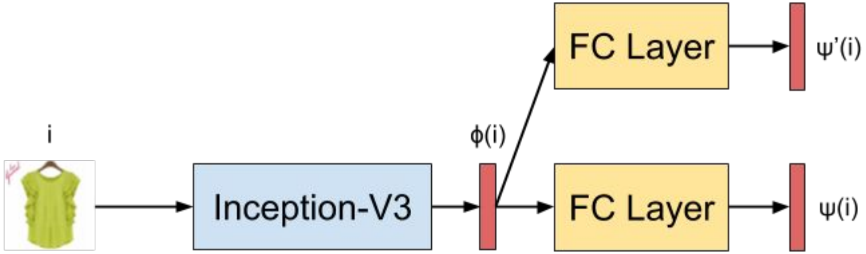


- Imbalanced dataset: Most attributes are negative for most images
- Two branch architecture to up-weight positive examples - more effective than standard up-weighting

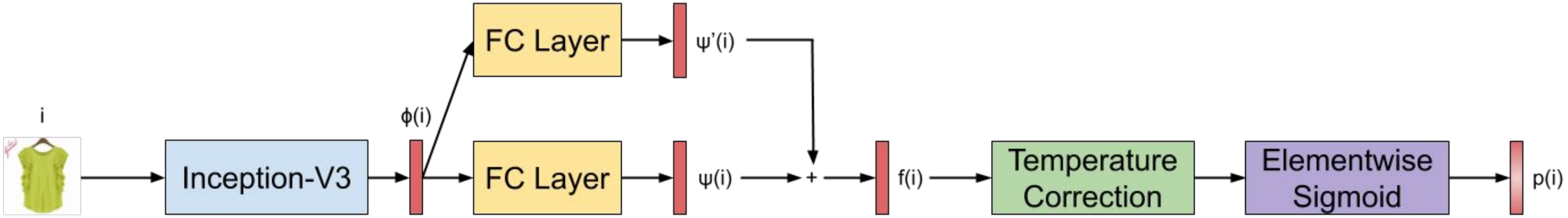
Visual Attribute Classifier



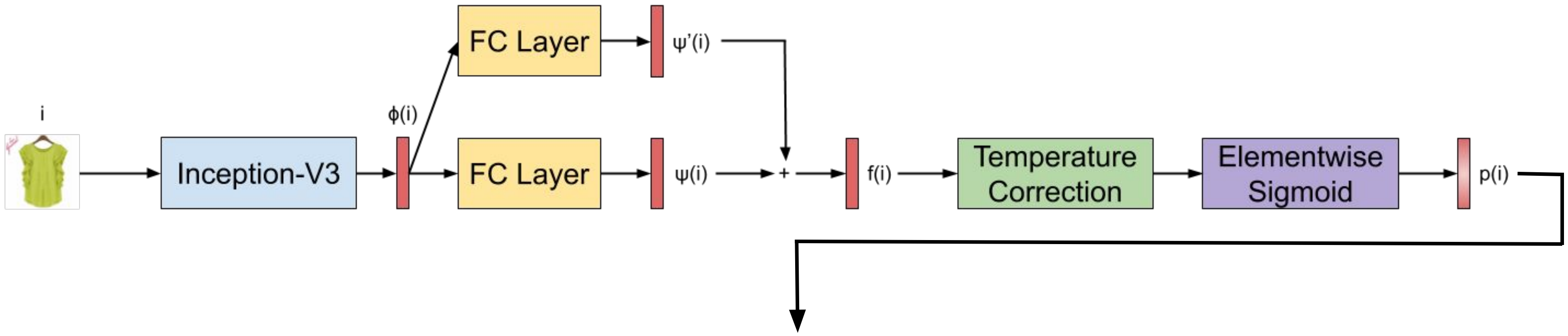
Visual Attribute Classifier



Visual Attribute Classifier



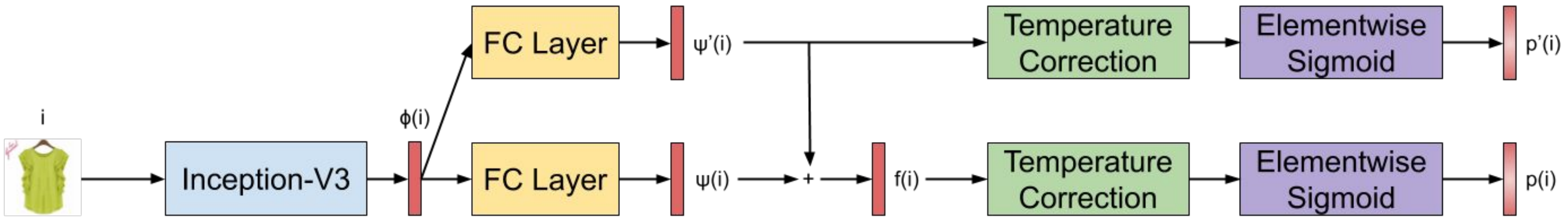
Visual Attribute Classifier



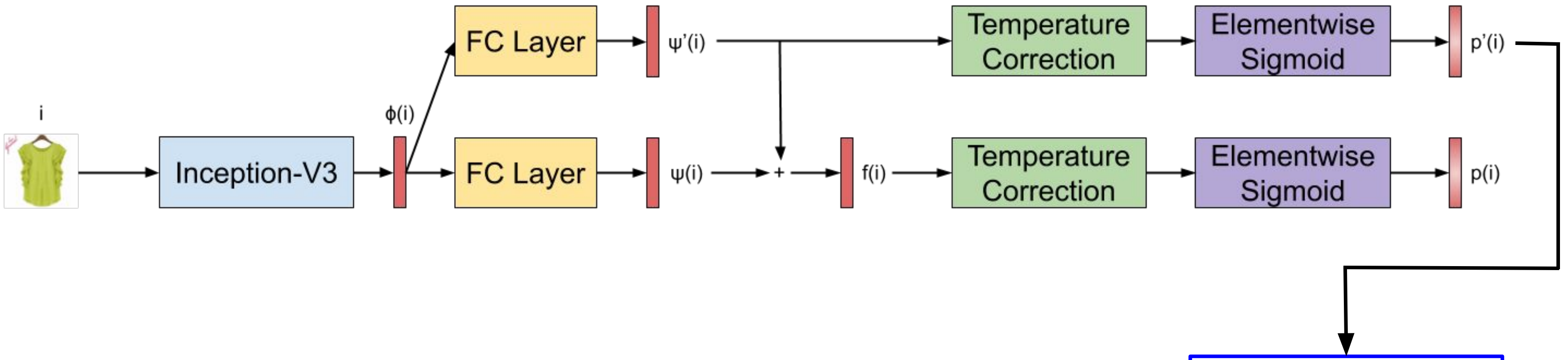
$$L = (1-\lambda)\boxed{\sum_i y_i \log p(i) + (1-y_i)\log(1-p(i))} + \lambda \sum_i y_i \log p'(i)$$

Cross Entropy Loss Over All Examples

Visual Attribute Classifier



Visual Attribute Classifier



$$L = (1-\lambda)\sum_i y_i \log p(i) + (1-y_i)\log(1-p(i)) + \lambda \boxed{\sum_i y_i \log p'(i)}$$

Cross Entropy Loss Over Positive Labels

Grounding Model

A Polka Dot Chiffon Blouse $\longrightarrow$ {Polka Dot, Chiffon, Blouse}

Grounding Model

A Polka Dot Chiffon Blouse $\longrightarrow$ {Polka Dot, Chiffon, Blouse}

Belief: $b(i) = \prod_{\boxed{w \in W_d}} p_w(i)$

Attributes Mentioned in Description

Grounding Model

A Polka Dot Chiffon Blouse $\longrightarrow$ {Polka Dot, Chiffon, Blouse}

Belief: $$b(i) = \prod_{w \in W_d} \boxed{p_w(i)}$$

- Classifier probability that attribute w is positive for image i
- `w`-th value in classifier output for image `i`

Grounding Model

Agent: Would you like one which is black?

User: Yes

<Black, 1>

Belief: $$b(i) = \prod_{w \in W_d} p_w(i) \prod_{\boxed{w \in W_p}} p_w(i)$$

Clarifications that get the answer “Yes”

Grounding Model

Agent: Would you like one which is black?

<Black, 0>

User: No

Belief: $$b(i) = \prod_{w \in W_d} p_w(i) \prod_{w \in W_p} p_w(i) \prod_{\boxed{w \in W_n}} (1 - p_w(i))$$

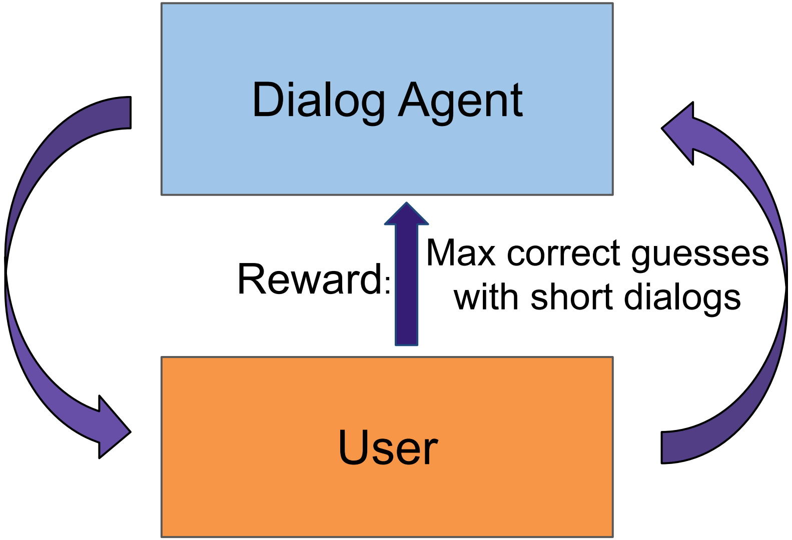
Clarifications that get the answer “No”

Grounding Model

Best guess: Image in active test set with maximum belief b(i)

Dialog as MDP

Action:
- Clarifications
- Label queries
- Example queries
- Guess



State:
- Target description
- Active train and test objects
- Agent's perceptual classifiers

Policy Learning

- Hierarchical Dialog Policy -
  - Clarification policy - chooses best clarification
  - Active learning policy - chooses best active learning query
  - Decision Policy - chooses between guess, best clarification and best active learning query
- Featurize state-action pairs
- Q-Learning and A3C for policy learning

Policy Features

- Clarification Policy Features - Metrics about current beliefs, information gain estimated from classifier probabilities
- Active Learning Policy Features - Margin, Fraction of previous uses and successes
- Decision Policy Features - Metrics about current beliefs, information gain, margin, dialog length

Outline

- Introduction
- Task Setup
- User Simulator
- Dialog Policy Model
- Experiments

Static Baseline

- Clarification: Choose query with maximum estimated information gain
- Active Learning: Uncertainty Sampling
- Decision Policy
  - Fixed dialog length
  - Dialog split equally between clarification and active learning
  - Heuristic checks to ensure usefulness of queries

Experiment Phases

- Classifier Initialization - Train classifier using paired images and labels
- Policy Initialization - Collect experience using the baseline to initialize the policy.
- Policy Training - Improve the policy from on-policy experience.
- Policy Testing - Policy weights are fixed, and we run a new set of interactions, reset classifiers to the state at the end of classifier initialization, over an independent test set containing novel attributes.

Results

Decision Policy Type	Clarification Policy Type	Active Learning Policy Type	Fraction of Successful Dialogs	Average Dialog Length
Q-Learning	A3C	A3C	**0.33**	9.40
Static	Static	Static	0.17	20.00

Results

Decision Policy Type	Clarification Policy Type	Active Learning Policy Type	Fraction of Successful Dialogs	Average Dialog Length
Q-Learning	A3C	A3C	**0.33**	9.40
Static	Static	Static	0.17	20.00

Fully learned policy is significantly more successful than the baseline, while also having significantly shorter dialogs on average

Results

Decision Policy Type	Clarification Policy Type	Active Learning Policy Type	Fraction of Successful Dialogs	Average Dialog Length
Q-Learning	A3C	A3C	**0.33**	9.40
Q-Learning	A3C	Static	0.15	14.16
Q-Learning	Static	A3C	0.09	1.00
Static	Static	Static	0.17	20.00

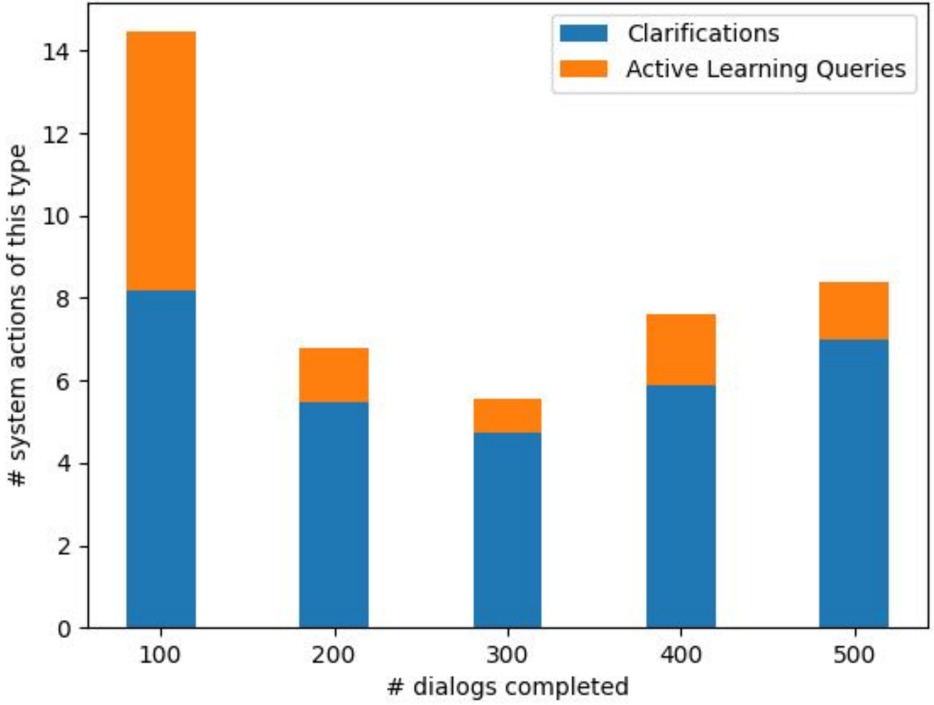
If we replace either the clarification or active learning policies with static policies, we find that the success rate drops considerably.

Results

Decision Policy Type	Clarification Policy Type	Active Learning Policy Type	Fraction of Successful Dialogs	Average Dialog Length
Q-Learning	A3C	A3C	**0.33**	9.40
Static	A3C	A3C	0.27	20.00
Static	Static	Static	0.17	20.00

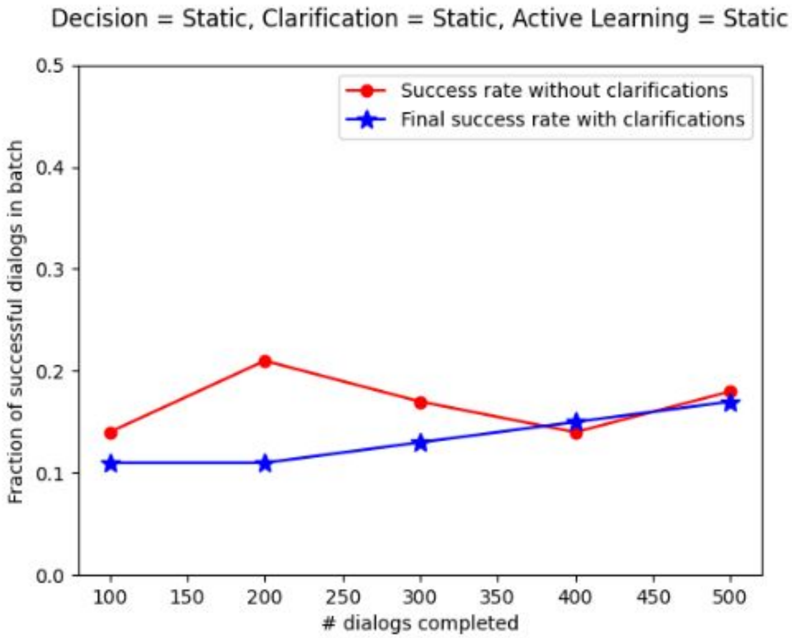
If we replace only the decision policy with a static policy, we find that it remains more successful than the baseline but is unable to shorten dialogs.

Action Types - Learned Policy



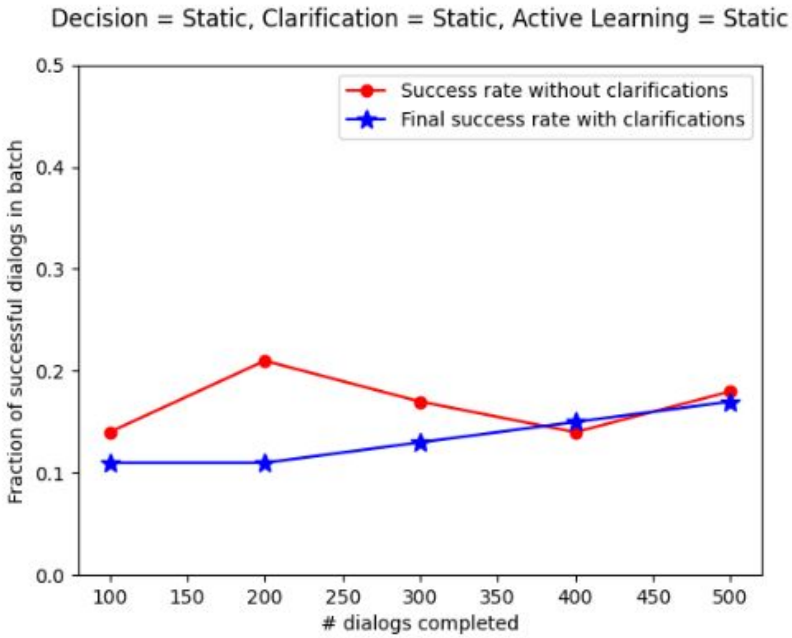
Utility of Clarifications

- Red curve: Success rate if the system just guesses based on the initial user request without any further interaction
- Blue curve: Actual success rate at the end of dialog including clarifications
- Each data point corresponds to one test batch after an additional 100 dialogs.



Utility of Clarifications

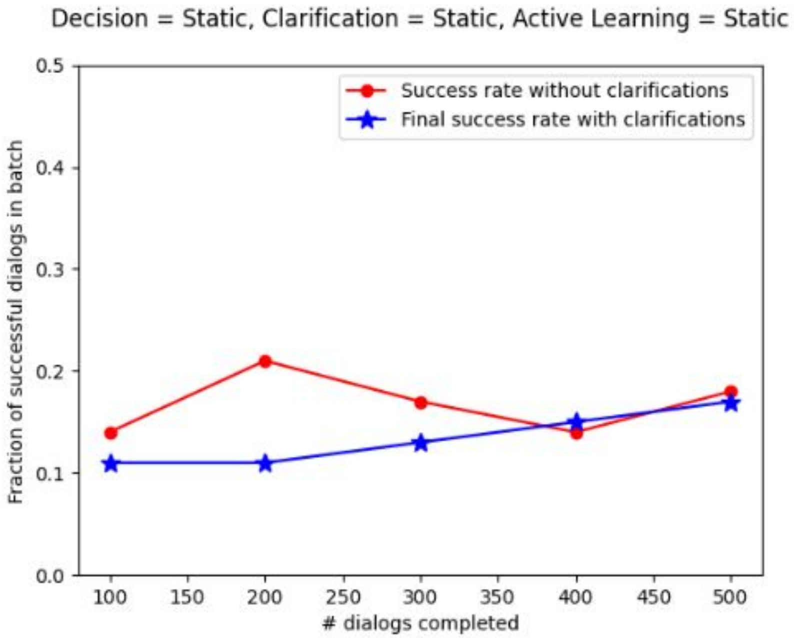
- Classifier updates happen at the end of each batch of dialogs
- First point on each curve corresponds to no active learning.
- Subsequent points correspond to increasing amounts of completed active learning.



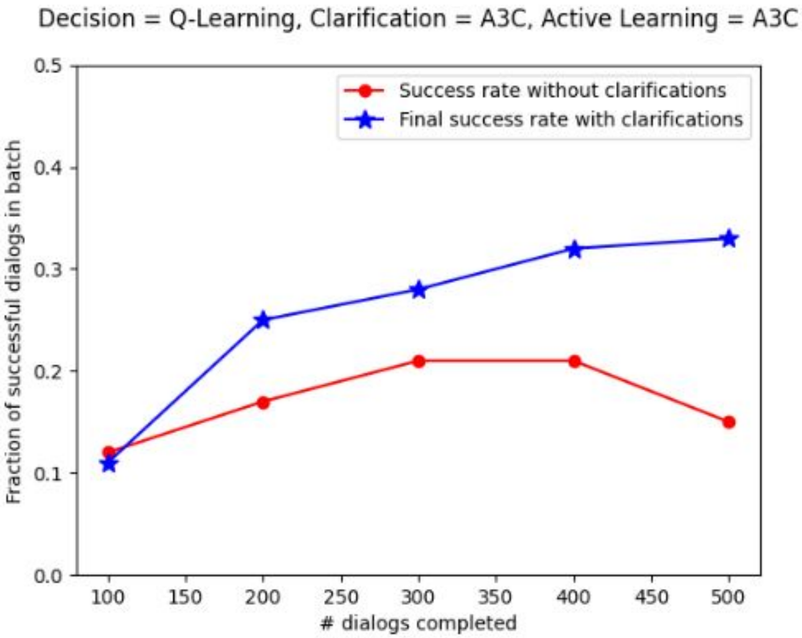
Utility of Clarifications

Fully static policy

- Not much difference between the curves.
    - Clarification does not improve the success rate.
- Not much difference between various points on the curves.
    - Active learning does not affect success rate.



Utility of Clarifications



Fully learned policy

- First test batch - no difference between curves
  - Without active learning, clarification is unhelpful.
- Final test batch - difference between curves
  - Combination of active learning and clarification results in increased success rate.

Human Evaluation - Experiment Changes

- Descriptions from human users contained far fewer attributes than product titles
- Changes in task setup -
  - Provide one attribute from product title as simulated user request
  - Smaller and easier active test set

Results - New Simulated Setup

Policy	Simulation – Fraction of Successful Dialogs	Simulation – Average Dialog Length
Static	0.23	20.0
Learned	**0.65**	20.0

The learned policy is considerably more successful in the new simulated setup but is unable to shorten dialogs compared to the baseline.

Human Evaluation Experiment

- All 4 phases run in new simulated setup
  - Classifier Initialization
  - Policy Initialization
  - Policy Training
  - Policy Testing
- Run a single batch of interactions on Amazon Mechanical Turk with final policy and classifiers

Human Evaluation Interface

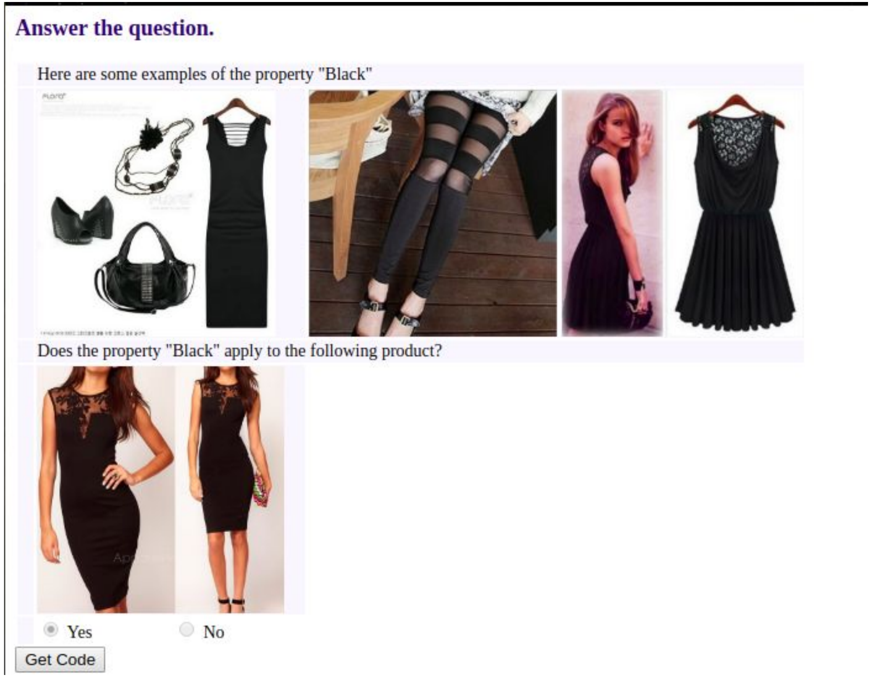
Describe the product in the image.

Describe the product in the image.

red dress

Continue

Human Evaluation Interface



Results - Human Evaluation

Policy	Simulation – Fraction of Successful Dialogs	Simulation – Average Dialog Length	AMT – Fraction of Successful Dialogs	AMT – Average Dialog Length
Static	0.23	20.0	0.06	19.16
Learned	**0.65**	20.0	*0.16*	18.86

- The performance of both policies drops in AMT interactions.
- The learned policy is still somewhat more successful (p <= 0.1)

Summary

- We train a hierarchical dialog policy to trade off opportunistic active learning, attribute based clarification and task completion in a language based image retrieval task in the shopping domain.
- In simulation, our learned policy is more successful than a static baseline while using fewer dialog turns on average.
- In our task setup, both good clarifications and active learning queries are necessary to improve performance over direct retrieval.
- Neither the static nor the learned policies transfer well during human evaluation but the learned policy remains more successful than the static policy.

Dialog Policy Learning for Joint Clarification and Active Learning Queries

Aishwarya Padmakumar, Raymond J. Mooney

padmakua@amazon.com, mooney@cs.utexas.edu