Neural Models of Response Selection for Bootstrapping Dialogue Systems

Matthew Henderson

PolyAI

Creating Task-based Dialogue Systems

Convincing Application

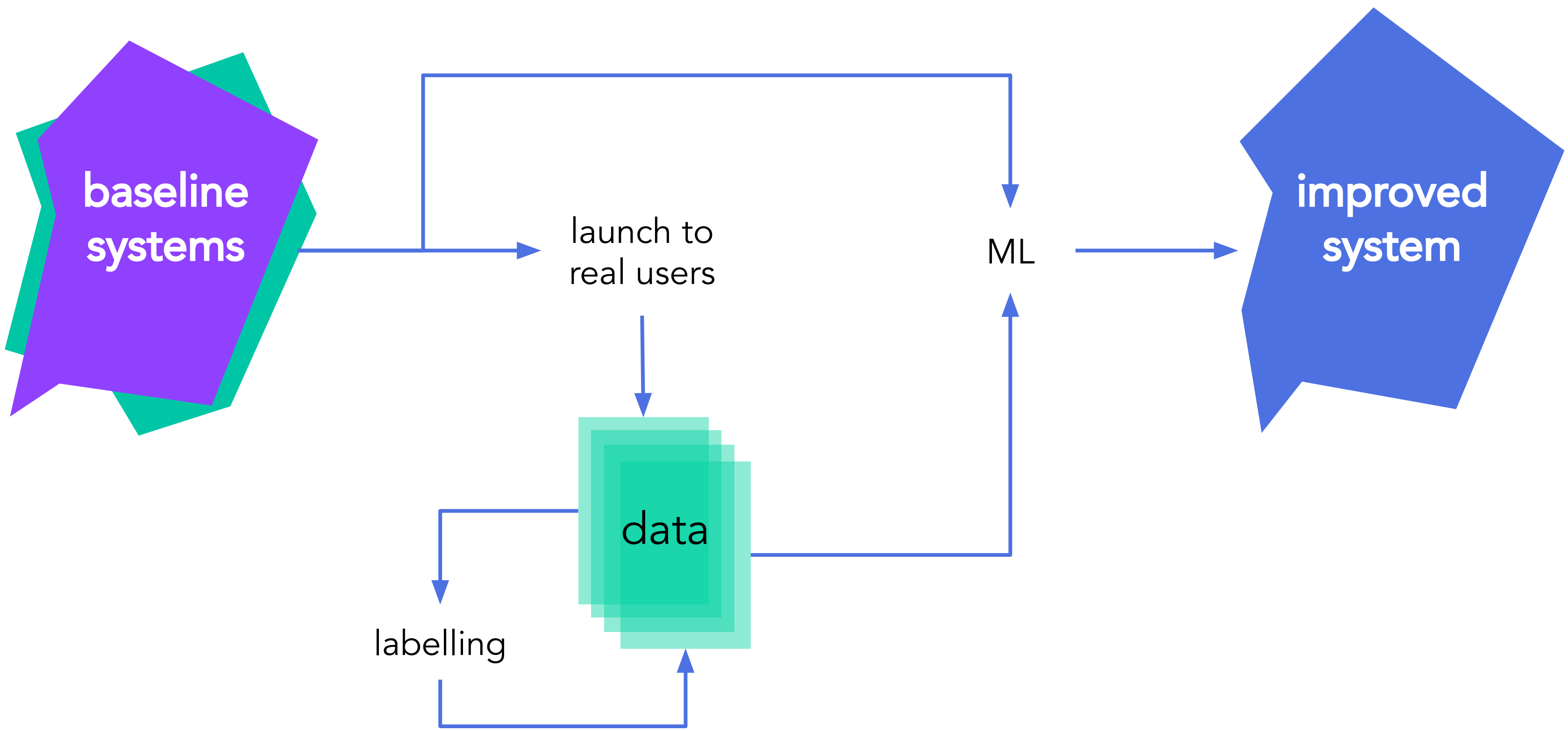
solves a real problem

Meaningful Evaluation

can measure progress

Annotated Data

is machine-learnable


baseline systems
launch to real users
ML
improved system
data
labelling

how do we get a
baseline system?

intent classifiers
slot-value recognisers
response selection/generation

xskills
xdomains
xlanguages

how can we minimise reliance on annotated data?

how can we scale better?
(skills, domains, languages…)

by using large pre-trained models that encapsulate knowledge of conversational response

Pre-training in NLP

- recent trend to pre-train large models of language, then fine-tune

  BERT, ELMo, GPT etc.

- uses unlabelled text + unsupervised objective

  same idea as cbow, skip gram, skip thought etc.

- learns general representations of text, useful for downstream tasks

PolyAI Conversational Datasets

Reddit	OpenSubtitles	AmazonQA
3.7 billion comments from online discussions on many topics	over 400 million lines of subtitles from movies and TV	over 3.6 million product question-answer pairs
↓	↓	↓
727 million examples	316 million examples	3.6 million examples

github.com/PolyAI-LDN/conversational-datasets

Public Conversational Datasets

	~ Turns	Annotations
DSTC 2&3	10^4	response, ASR, SLU
MultiWoz	10^5	response, NLU
DSTC7 Reddit	10^6	response, entities
DSTC7 Ubuntu	10^6	response
PolyAI AmazonQA	10^6	product, response
PolyAI OpenSubtitles	10^8	'response'
PolyAI Reddit	$\mathbf{10^9}$	**response**

Next word prediction

The launch of India's second lunar mission has been

apple
called
halted
celebrate
passport
...

Masked word prediction

The launch of ■ 's second lunar mission has been ???
less than an hour before the scheduled blast- ■ ,
due to a ■ problem.

apple
called
halted
celebrate
passport
...

Response Selection

Any recommendations for short trips from Singapore? →

It doesn't feel like July.
That type of music isn't really my cup of tea.
Bintan is just a quick ferry trip away.
You have to try the vegetarian Haggis!
I'd do a short trip to Paris.
...

Response Selection

- large conversational datasets
- representations encode conversational cues
- encodes full sentences
- directly applicable to retrieval-based dialogue

Language Modelling

- large text datasets
- representations encode word/phrase/sentence cues
- encodes words contextually
- maybe applicable to generation/scoring

a lot of the power of neural techniques is finding good embeddings / encodings

- so learn encoder model on large conversational data

- then use various tricks and small models on the learned vector space for domain specific tasks

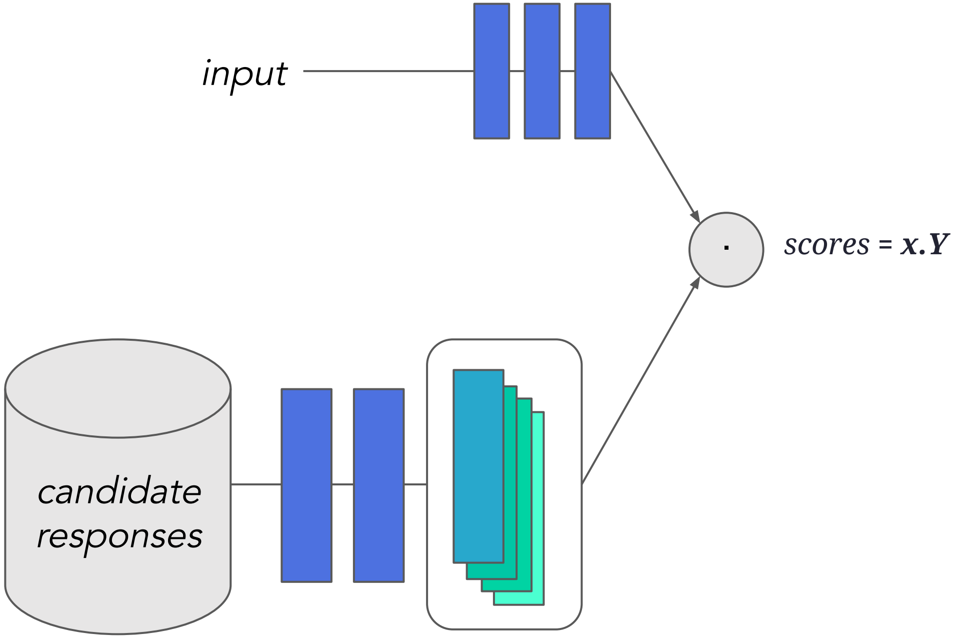
Dual Encoders for Response Selection

dual encoder dot product model

- gmail smart reply
- universal sentence encoder

trained to give a high score for the response found in the data, low score for random responses

final score of an input and response is a dot-product of two vectors



network encodes a batch of inputs to vectors:

$$x_1 \quad x_2 \quad \dots \quad x_N$$

and responses to vectors:

$$y_1 \quad y_2 \quad \dots \quad y_N$$

$x_1 . y_1$	$x_1 . y_2$	$x_1 . y_3$	$x_1 . y_4$	$x_1 . y_5$
$x_2 . y_1$	$x_2 . y_2$	$x_2 . y_3$	$x_2 . y_4$	$x_2 . y_5$
$x_3 . y_1$	$x_3 . y_2$	$x_3 . y_3$	$x_3 . y_4$	$x_3 . y_5$
$x_4 . y_1$	$x_4 . y_2$	$x_4 . y_3$	$x_4 . y_4$	$x_4 . y_5$
$x_5 . y_1$	$x_5 . y_2$	$x_5 . y_3$	$x_5 . y_4$	$x_5 . y_5$

the N x N matrix of all scores is a fast matrix product.

large improvement in 1 of 100 ranking accuracy over binary classification.

$x_1.y_1$	$x_1.y_2$	$x_1.y_3$	$x_1.y_4$	$x_1.y_5$
$x_2.y_1$	$x_2.y_2$	$x_2.y_3$	$x_2.y_4$	$x_2.y_5$
$x_3.y_1$	$x_3.y_2$	$x_3.y_3$	$x_3.y_4$	$x_3.y_5$
$x_4.y_1$	$x_4.y_2$	$x_4.y_3$	$x_4.y_4$	$x_4.y_5$
$x_5.y_1$	$x_5.y_2$	$x_5.y_3$	$x_5.y_4$	$x_5.y_5$

$$\boldsymbol{x}_i = f\,(\text{input } i\ )$$

$$\boldsymbol{y}_j = g\ (\text{response } j\ )$$

$$S_{ij} = \boldsymbol{x}_i \cdot \boldsymbol{y}_j$$

$$P(\text{ response } j \mid \text{ input } i\ ) \propto e^{Sij}$$

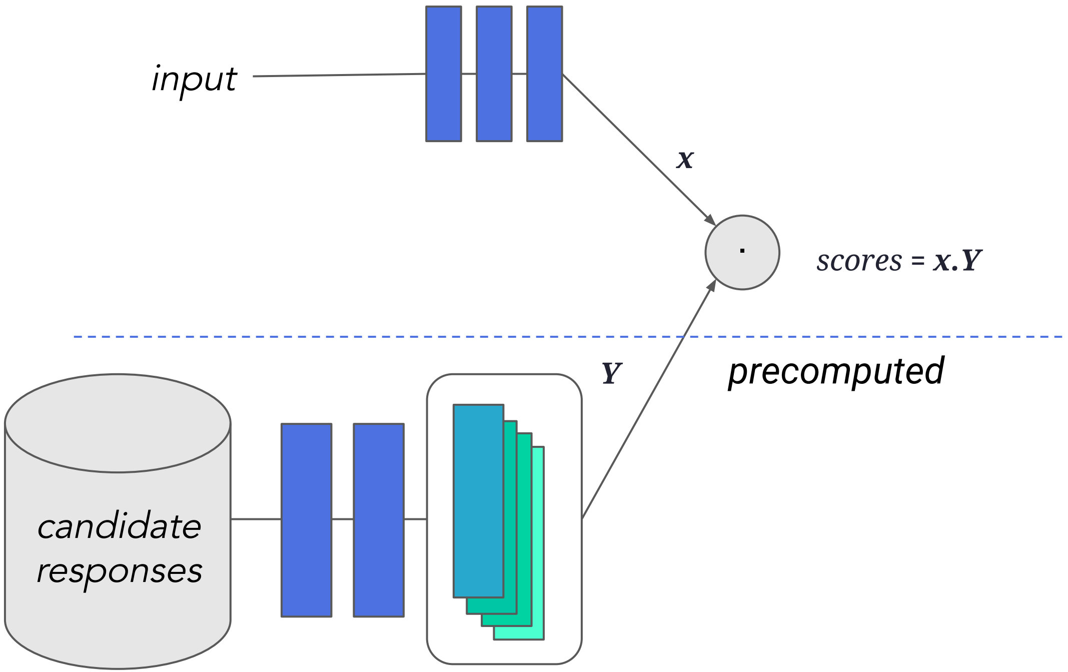
$$-\ log\ P(\text{ example } i\ ) = -\ S_{ii} + log\ \Sigma_j\ e^{Sij}$$

"dot product loss"

Precomputation for dot product model

the representations of the candidates Y can be precomputed

approximate nearest neighbor search can speed up the top N search



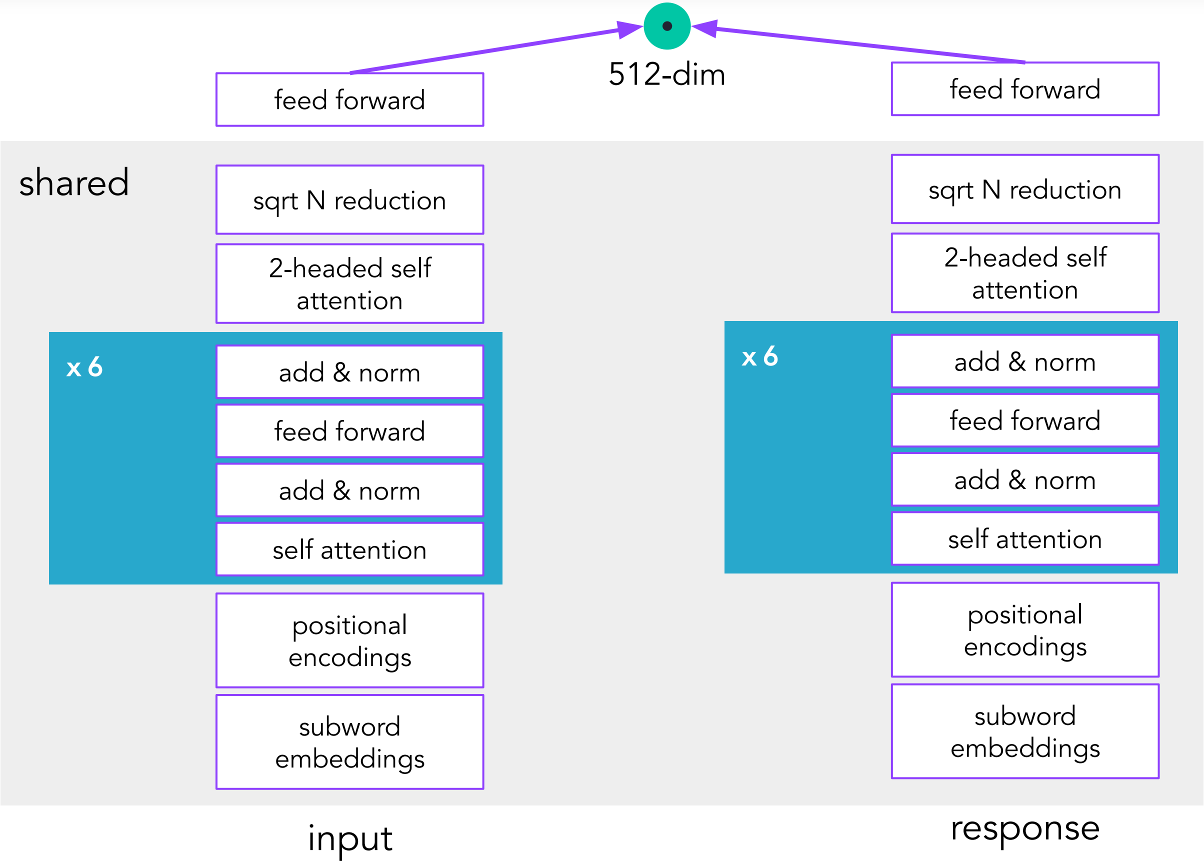
at inference, a user query has N words, there are M responses with N_R words each

- dot product model
  - $O(N)$ to encode input to vector space
  - $O(\log M)$ to find top scoring response with approximate search
- general sequence model (e.g. BERT next sentence scoring)
  - $O(M(N + N_R))$ to score all responses
  - $O(M)$ to find top response

1-of-100 accuracy

how often the correct response is ranked top vs 99 random

PolyAI Encoder



PolyAI Encoder

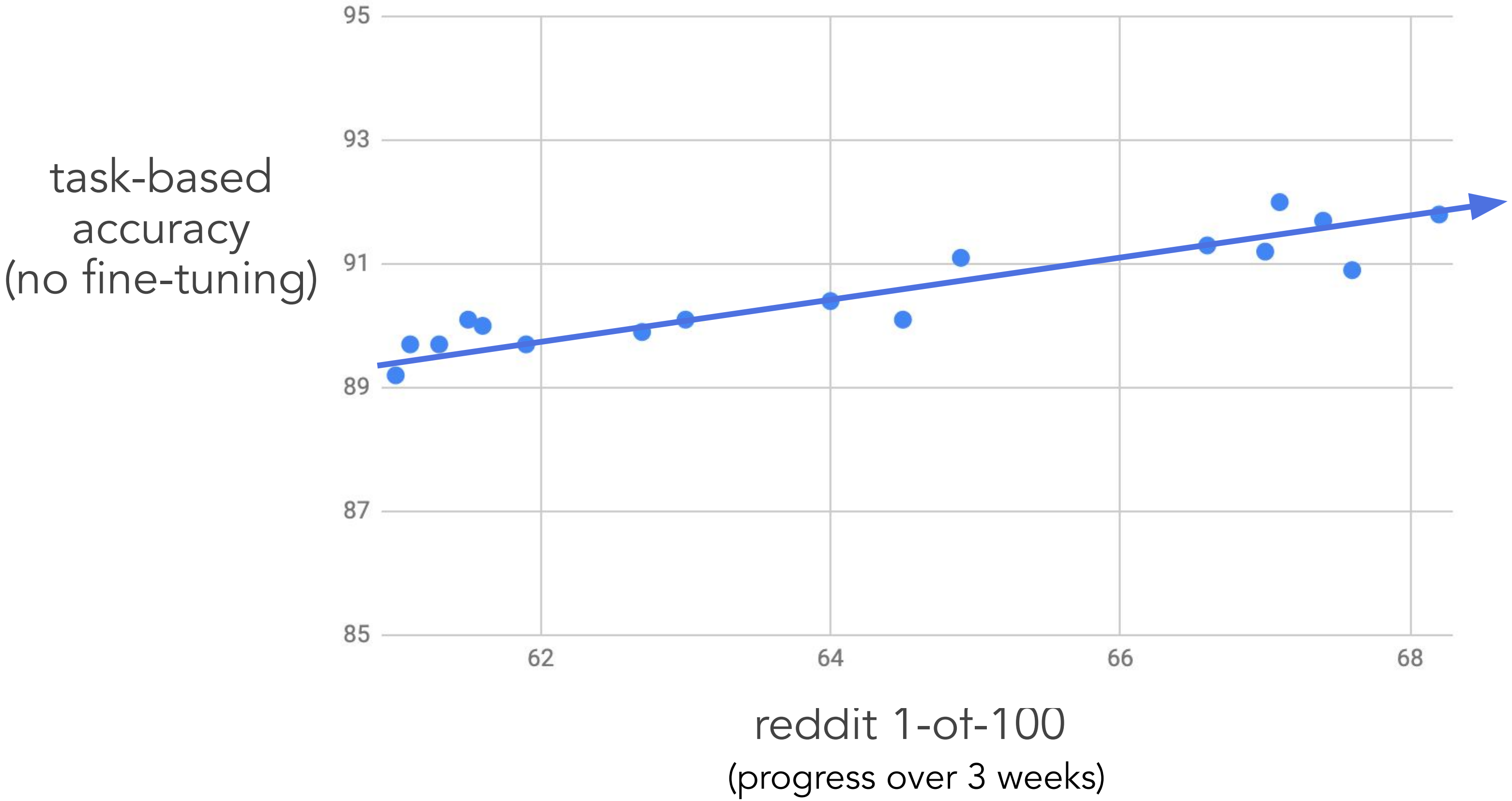
		reddit 1-of-100 accuracy
keyword-based	TF-IDF	26.7%
	BM25	27.6%
MAP dot product models	ELMo	19.3%
	BERT	24.5%
	USE	40.8%
	USE_QA	46.3%
	BERT dot-product model	55.0%
PolyAI Encoders	n-grams	61.3%
	subwords	**68.2%**

PolyAI Encoder

resource-constrained optimization:
pick the best model after training 18 hours on 12 GPUs

- fast ML engineering cycle, rapid progress
- we own the whole training pipeline

- training costs under $100
- model runs fine on CPU
- final model is 40MB


task-based
accuracy
(no fine-tuning)
95
93
91
89
87
85
62
64
66
68
reddit 1-of-100
(progress over 3 weeks)

intent classification

Intent Classification

initiate-booking

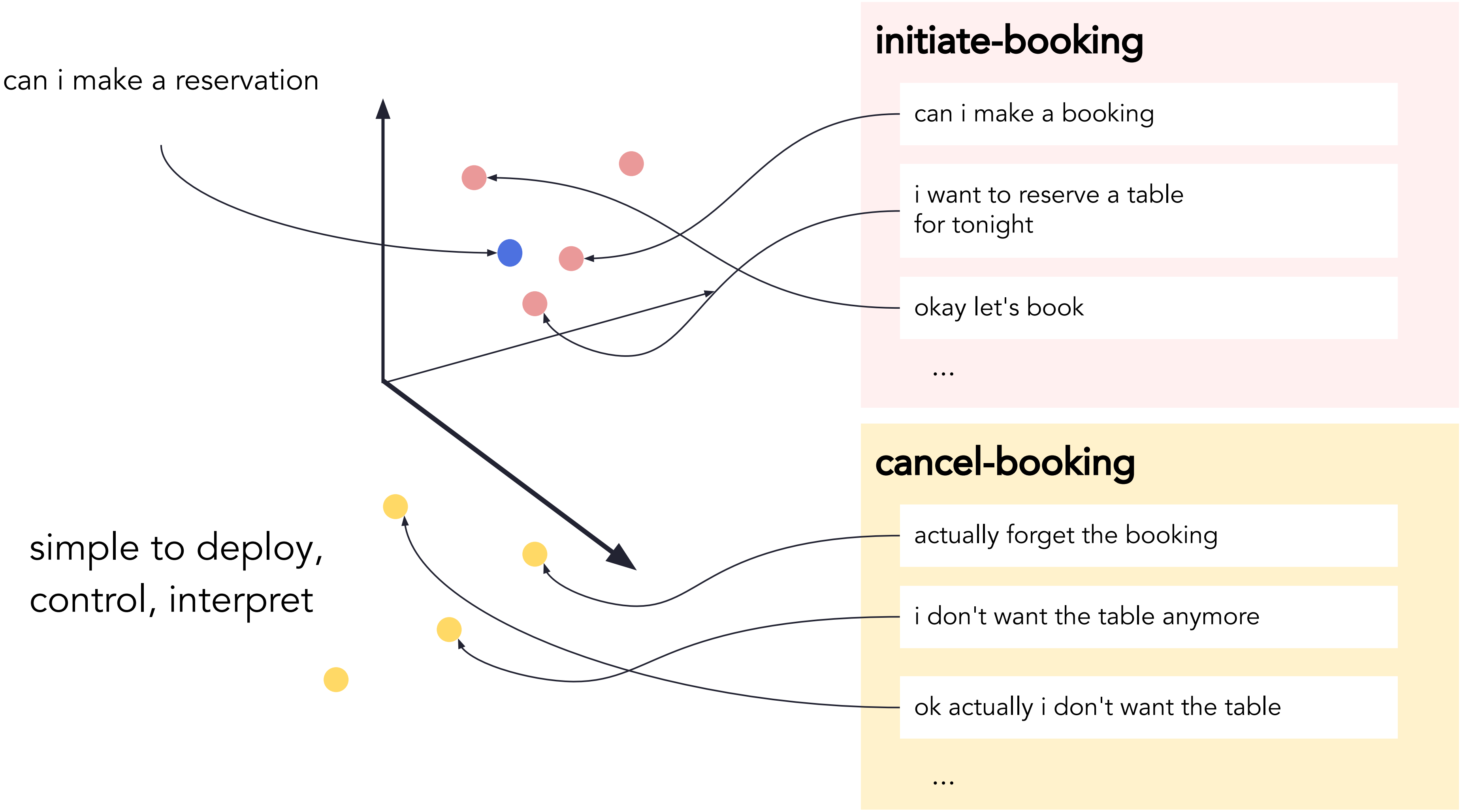
can i make a booking
can i reserve a table
okay i want to book a table for tonight

cancel-booking

cancel it
i don't want the table anymore

restart

let's start over
forget this



\+ simple to deploy, control, interpret

Intent classification

- can train an MLP on top of encoding representation
- can jointly fine-tune the encoding parameters

- can treat similarity in encoding space as as a kernel
    - SVM (more interpretable, encoding-agnostic)

Intent Evaluation

		PolyAI Encoder		
		PolyAI QQ	PolyAI SVM	PolyAI MLP
Banking	small	68.3%	**83.1%**	81.8%
	medium	82.5%	**91.0%**	90.4%
	large	87.7%	**93.1%**	92.7%
	e-commerce	92.0%	94.1%	**94.5%**
	company FAQ	63.8%	64.5%	64.5%

Intent Evaluation

		PolyAI Encoder						
		PolyAI QQ	PolyAI SVM	PolyAI MLP	USE QQ	USE FT	BERT FT	RASA FT
Banking	small	68.3%	**83.1%**	81.8%	67.3%	80.4%	80.4%	63.4%
	medium	82.5%	**91.0%**	90.4%	83.8%	89.8%	90.9%	84.0%
	large	87.7%	**93.1%**	92.7%	87.8%	92.2%	92.9%	89.2%
	e-commerce	92.0%	94.1%	**94.5%**	90.5%	94.0%	94.4%	92.1%
	company FAQ	63.8%	64.5%	64.5%	55.8%	62.4%	**65.0%**	55.4%

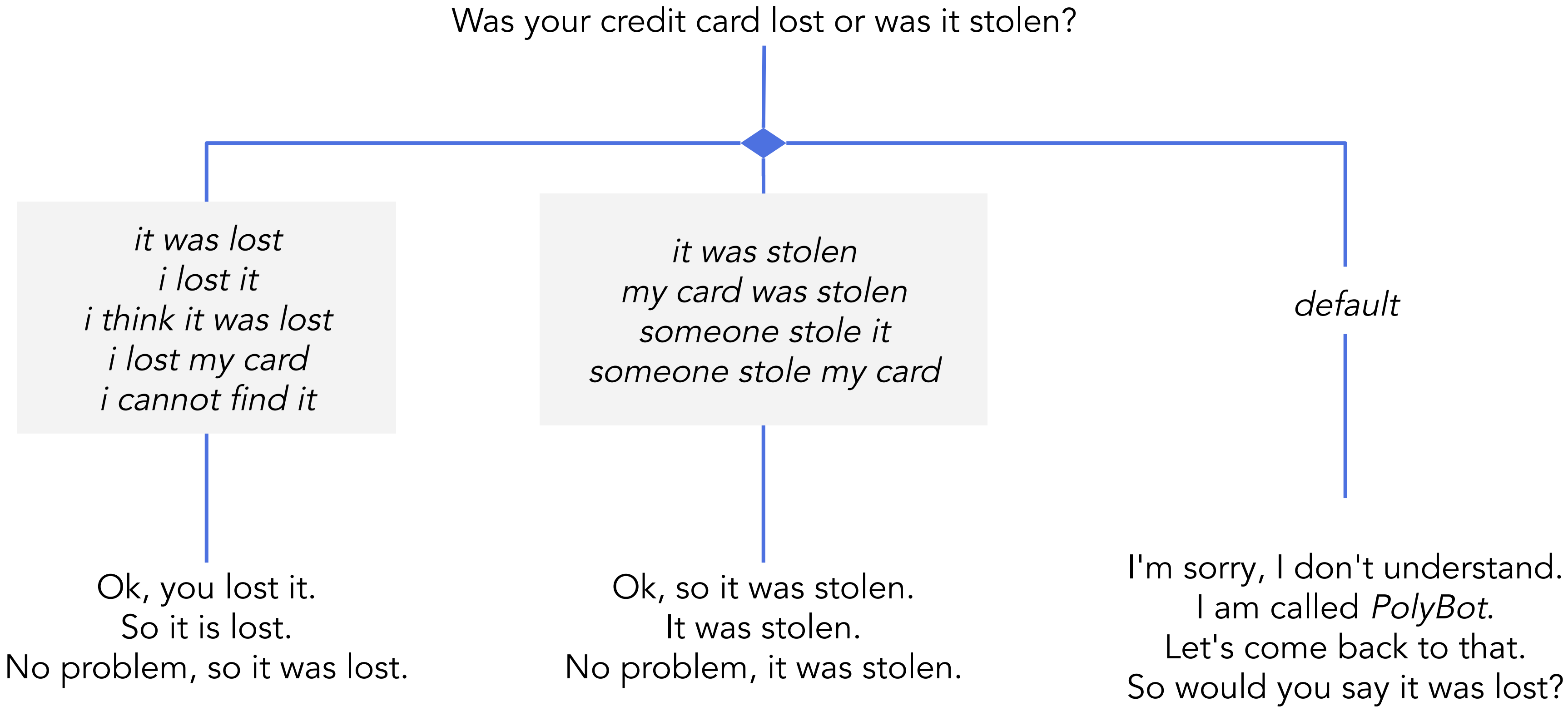
Intent Evaluation

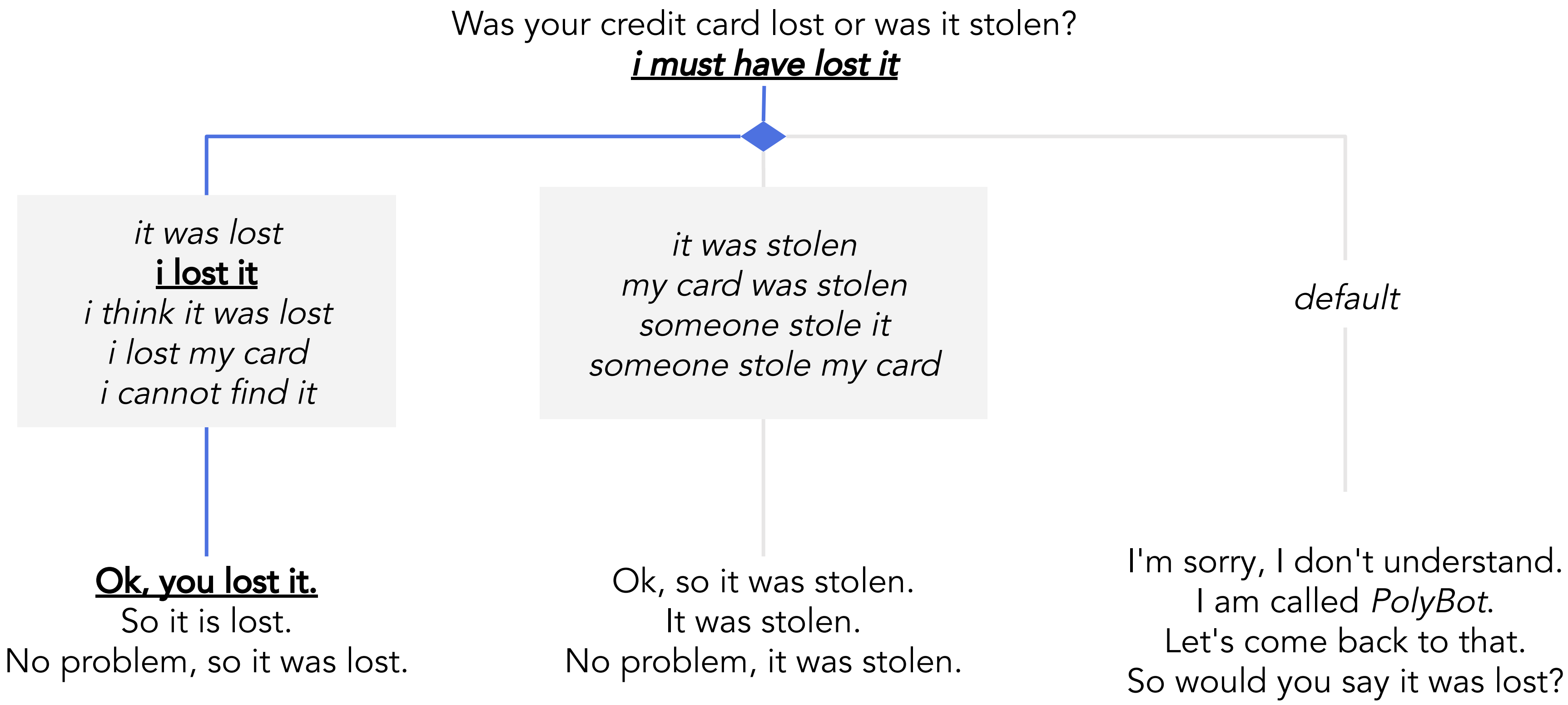
		PolyAI Encoder						
		PolyAI QQ	PolyAI SVM	PolyAI MLP	Twilio	MS Luis	IBM Watson	Dialogflow
Banking	small	68.3%	**83.1%**	81.8%	65.6%	69.0%	73.3%	79.6%
	medium	82.5%	**91.0%**	90.4%	83.7%	80.7%	87.0%	86.4%
	large	87.7%	**93.1%**	92.7%	89.6%	86.9%	90.6%	86.9%
	e-commerce	92.0%	94.1%	**94.5%**	91.3%	92.0%	92.1%	89.8%
	company FAQ	63.8%	64.5%	64.5%	55.7%	55.1%	57.8%	53.2%

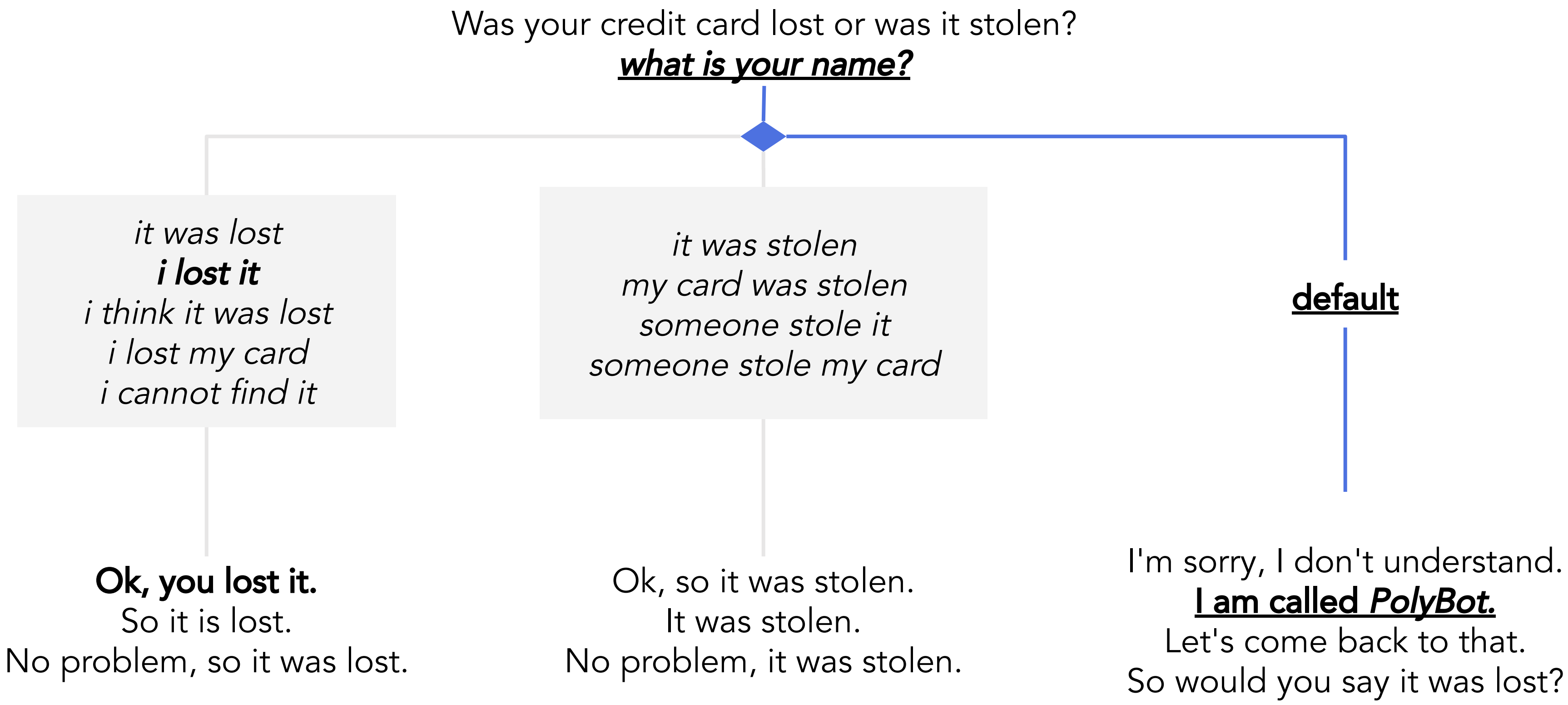
simple bot building

Simple bot building

quick prototyping of dialogues flows using the shared input response embedding space


Was your credit card lost or was it stolen?
it was lost
i lost it
i think it was lost
i lost my card
i cannot find it
it was stolen
my card was stolen
someone stole it
someone stole my card
default
Ok, you lost it.
So it is lost.
No problem, so it was lost.
Ok, so it was stolen.
It was stolen.
No problem, it was stolen.
I'm sorry, I don't understand.
I am called PolyBot.
Let's come back to that.
So would you say it was lost?


Was your credit card lost or was it stolen?
i must have lost it
it was lost
i lost it
i think it was lost
i lost my card
i cannot find it
it was stolen
my card was stolen
someone stole it
someone stole my card
default
Ok, you lost it.
So it is lost.
No problem, so it was lost.
Ok, so it was stolen.
It was stolen.
No problem, it was stolen.
I'm sorry, I don't understand.
I am called *PolyBot*.
Let's come back to that.
So would you say it was lost?


Was your credit card lost or was it stolen?
what is your name?
it was lost
i lost it
i think it was lost
i lost my card
i cannot find it
it was stolen
my card was stolen
someone stole it
someone stole my card
default
Ok, you lost it.
So it is lost.
No problem, so it was lost.
Ok, so it was stolen.
It was stolen.
No problem, it was stolen.
I'm sorry, I don't understand.
I am called *PolyBot.*
Let's come back to that.
So would you say it was lost?

Simple bot building

- fast prototyping / proof of concept

- system design interface is entirely text

  no need to train additional models

- flows can be guaranteed by adding text

restaurant search

DSTC 2 & 3

hello I am looking for a cheap place in the east
> inform(pricerange=cheap, area=east)

sure, what type of food?
> request(food)

i want gastropub food
> inform(food=gastropub)

there are no cheap places serving gastropub in the east.
> inform(name=none, area=east, pricerange=cheap)

how about any pricerange? and i need to know if they have wifi.
> inform(pricerange=dontcare) request(has_wifi)

The King's Arms is a nice place in the east of town serving gastropub food. It has wifi.
> offer(name="The King's Arms", area=east, food=gastropub, has_wifi=true)

DSTC 2 & 3

- explicit semantics forces unnaturally constrained dialogues
    - users need to know the ontology

- requires special annotated data, one specialised model per 'slot'

DSTC 2 & 3

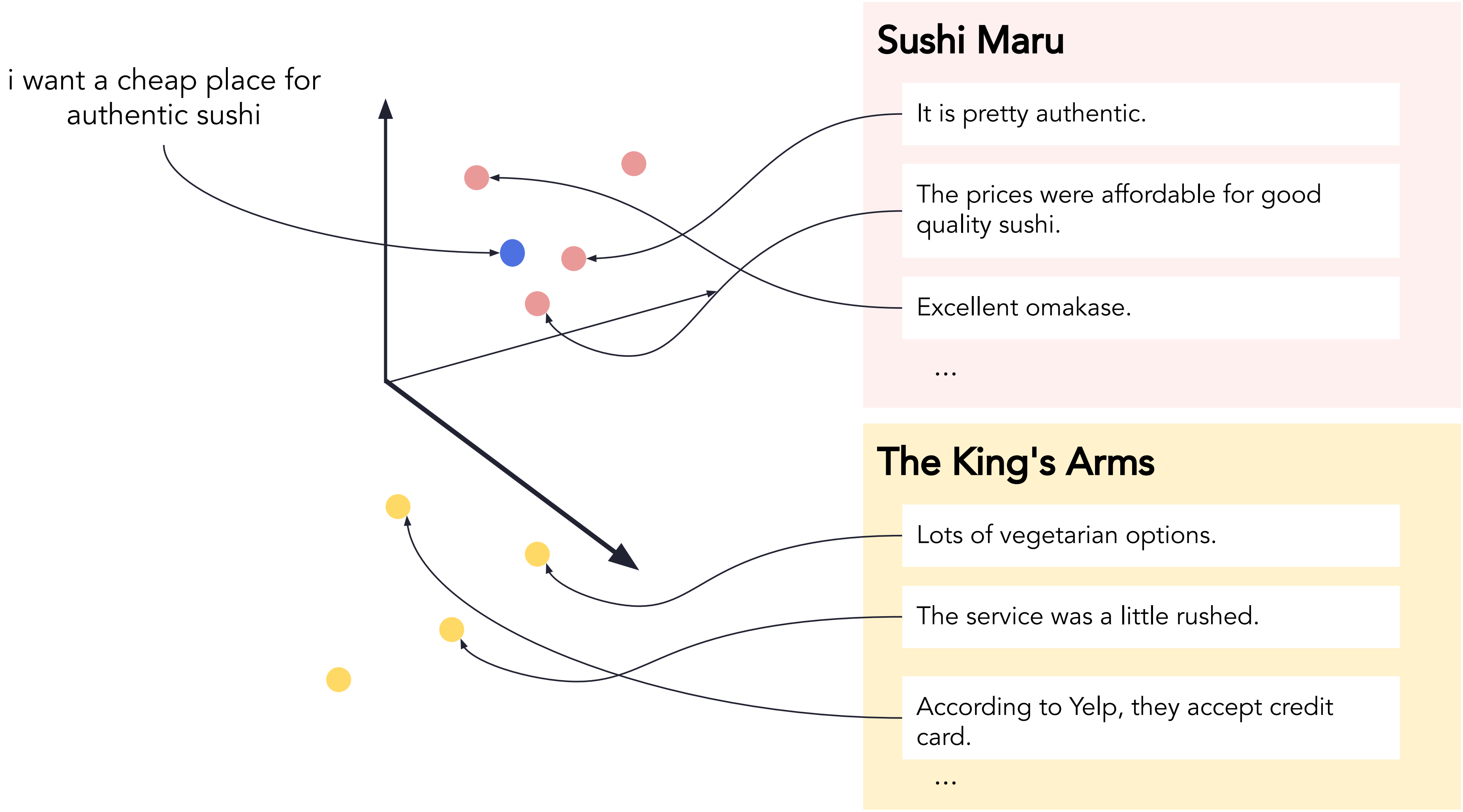
- some slots are necessary

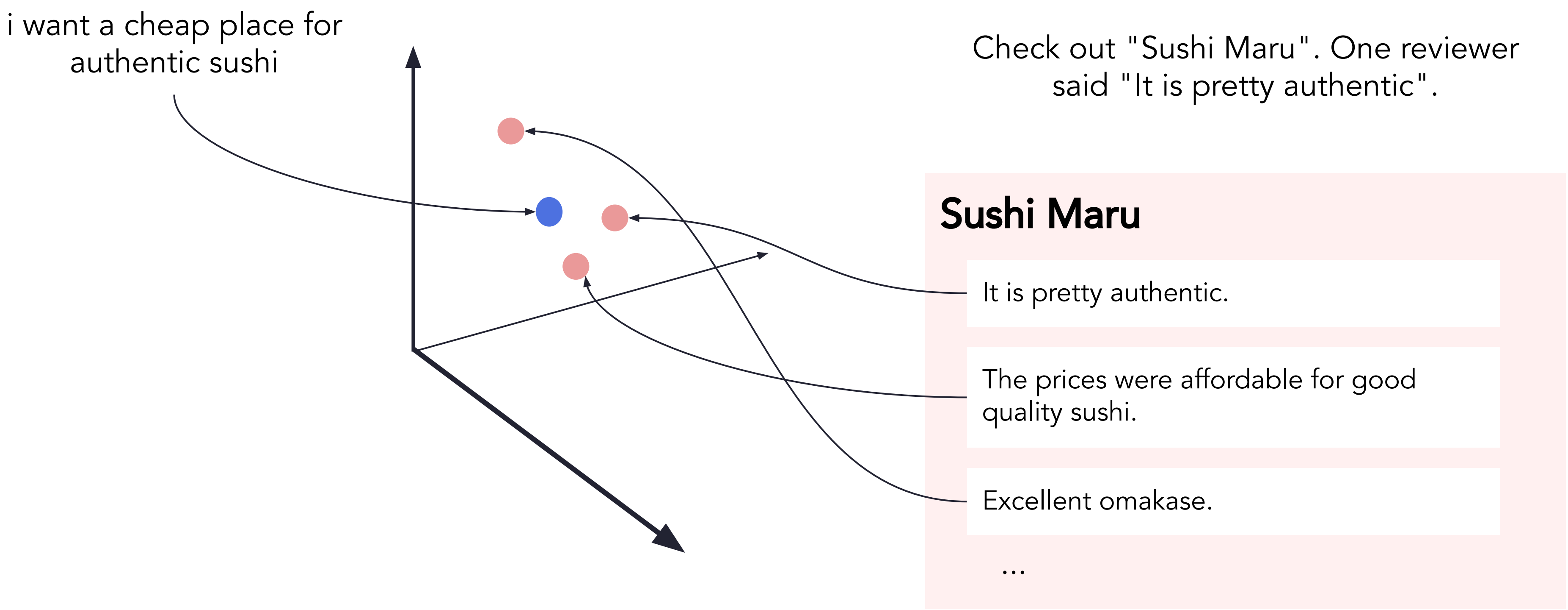
    number of people, booking time, name

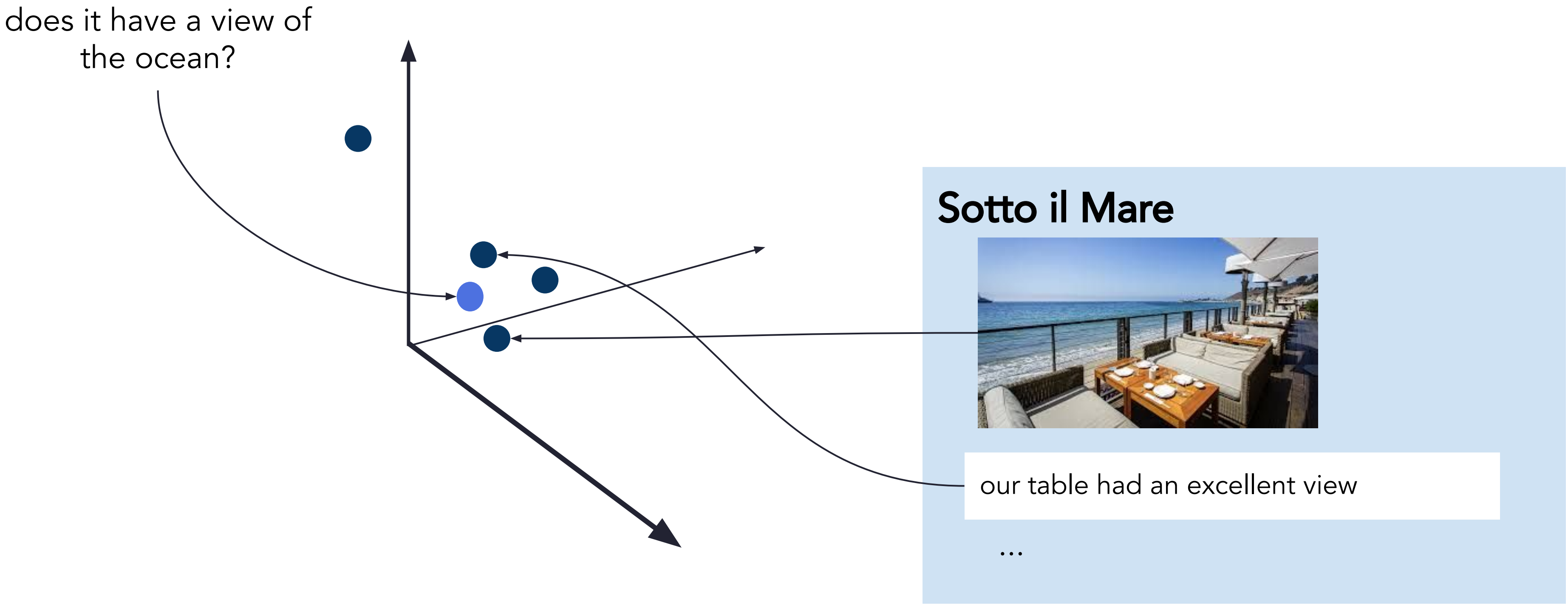
- some might not be

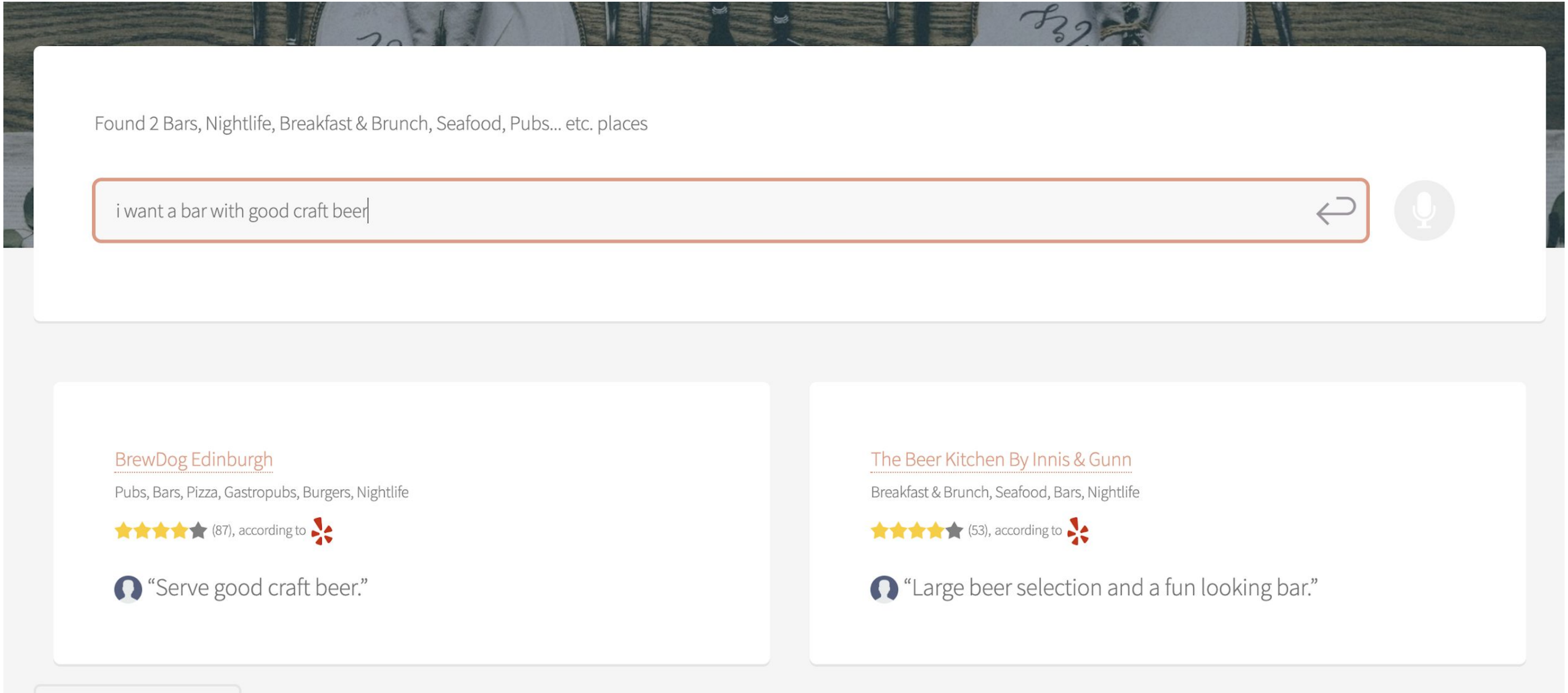
    food, price range, has wifi, has vegetarian, has vegan, serves cocktails….

- use all sentences in all reviews of all restaurants in a city
- treat dialogue as an iterative search
- perform search in implicit vector space learned by encoders


i want a cheap place for authentic sushi
Sushi Maru
It is pretty authentic.
The prices were affordable for good quality sushi.
Excellent omakase.
...
The King's Arms
Lots of vegetarian options.
The service was a little rushed.
According to Yelp, they accept credit card.
...


i want a cheap place for authentic sushi
Check out "Sushi Maru". One reviewer said "It is pretty authentic".
Sushi Maru
It is pretty authentic.
The prices were affordable for good quality sushi.
Excellent omakase.
...


does it have a view of
the ocean?
Sotto il Mare
our table had an excellent view
...


Found 2 Bars, Nightlife, Breakfast & Brunch, Seafood, Pubs... etc. places
i want a bar with good craft beer
BrewDog Edinburgh
Pubs, Bars, Pizza, Gastropubs, Burgers, Nightlife
(87), according to
“Serve good craft beer.”
The Beer Kitchen By Innis & Gunn
Breakfast & Brunch, Seafood, Bars, Nightlife
(53), according to
“Large beer selection and a fun looking bar.”

The Elephant House

Restaurant in Old Town. *Brasseries, Coffee & Tea, Food, British, Sandwiches, Cafes, and Scottish*

where did JK rowling write Harry Potter

“We got to see the spot where it is said JK Rowling wrote Harry Potter.”

“So apparently JK Rowling often wrote Harry Potter here.”

“After all, if JK Rowling was inspired maybe it can rub off?”

The Elephant House

Restaurant in Old Town. *Brasseries, Coffee & Tea, Food, British, Sandwiches, Cafes, and Scottish*

can i book a table for 2

What date and time?

Search procedure

1. encode user text with input encoder

    if restart intent detected, restart

2. find potential answers in restaurant texts

    hundreds of thousands of texts

3. convert scores into probabilities
4. constrain the set of active restaurants to those that make up the head of the distribution

Restaurant search

- entirely powered by a single model, trained on hundreds of millions of examples
- bootstrapped using only raw text representations- restaurants + reviews + facts
- allows more natural search, not bottlenecked by explicit semantics / ontology

Value Extraction

- limit slots to obvious values that the system needs to extract

  booking time & date, your name, number of people

- value extraction can benefit from pre-trained representations

- see our blog post on *Neural language understanding of people's names*

Hello, Thank you for contacting The Haywain, I am a digital agent, how can I help?

do you have any veggie burgers?

Yes, we have a popular Vegan Burger with BBQ Soya for £11.49, and will be happy to help with your requirements.

nice. ok can i book a table for 2

What day would you like to book for?

for tomorrow at 6 in the evening

No problem, we have a table for you for 2 people, tomorrow at 06:00PM. Shall I book that for you?

Response Selection for Bootstrapping Dialogue

efficient task tailored to dialogue

smaller cheaper faster models

robust performance on downstream tasks

competitive intent classification

driven by paraphrase collection

powers conversational search

efficient search

reduced dependency on strict ontology

PolyAI

find your voice