Procedural Reasoning Networks for Understanding Multimodal Procedures

Mustafa Sercan Amac, Semih Yagcioglu, Aykut Erdem, Erkut Erdem

Hacettepe University

CoNLL 2019

Procedural Understanding and Commonsense Reasoning

Pan-fried Salmon Cooking Recipe

Preheat a non-stick pan, spray it with canola oil, and add the spices of your choice.

Rinse your salmon and pat it dry. Drop it skin-side down in the pan and cook over medium-high heat.

Spray the top surface of the fish with the canola oil. Add the same spices that you originally added to the pan.

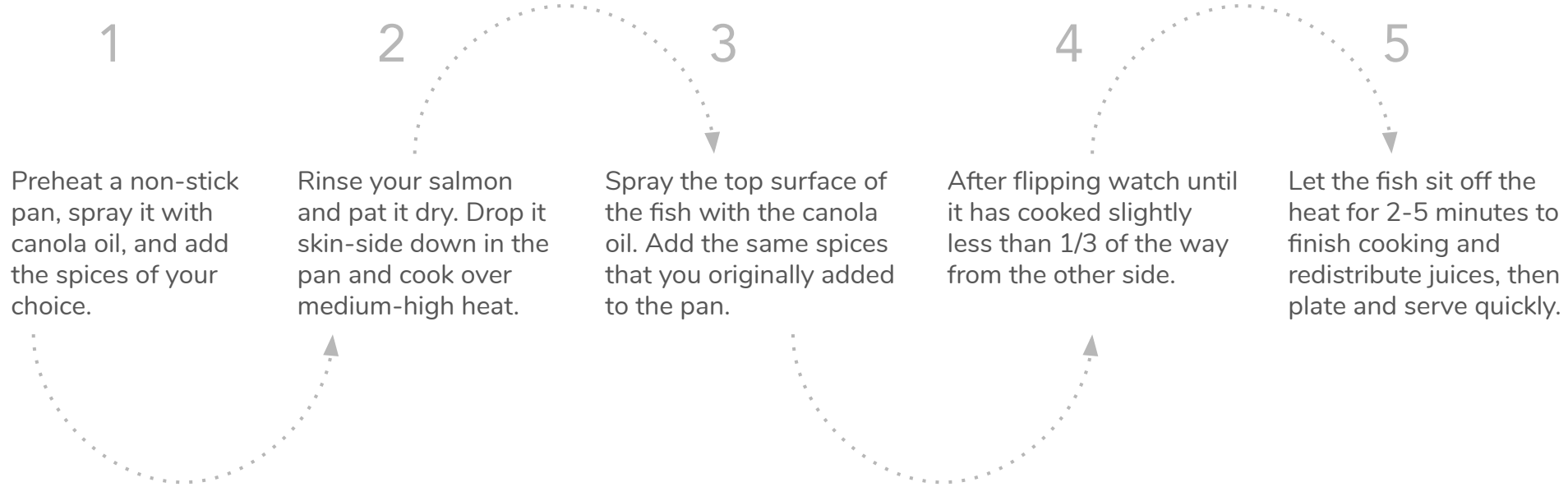
After flipping watch until it has cooked slightly less than 1/3 of the way from the other side.

Let the fish sit off the heat for 2-5 minutes to finish cooking and redistribute juices, then plate and serve quickly.

Understanding and reasoning about procedural texts *e.g.* cooking recipes, how-to guides, scientific processes is a very challenging problem

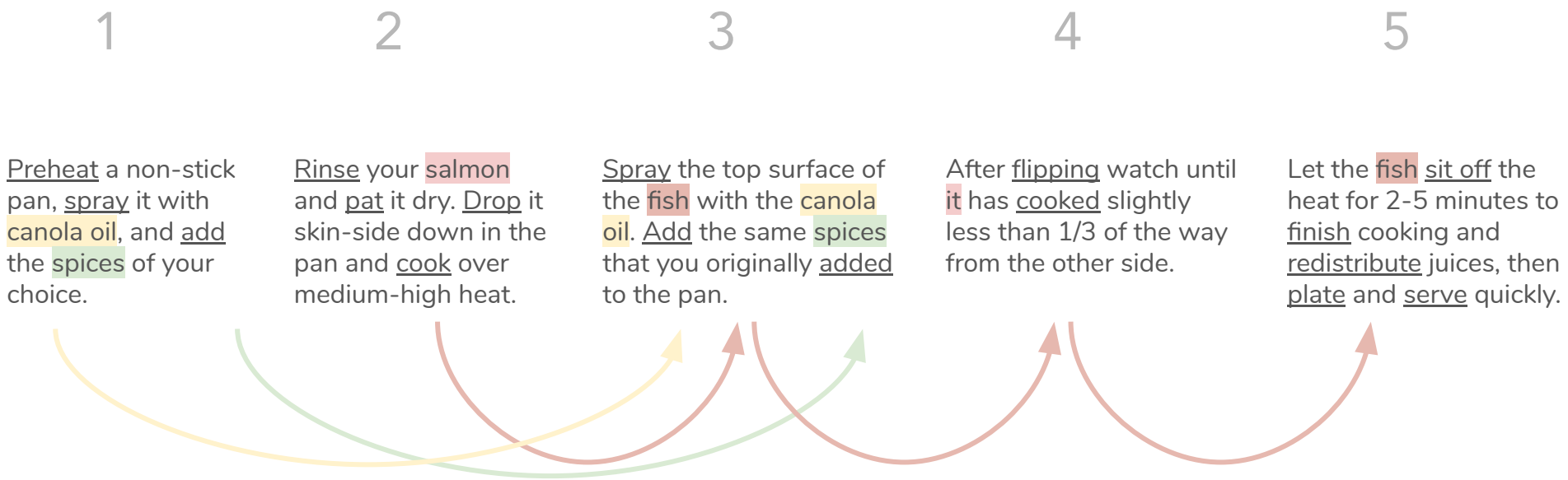
Credit: canida from instructibles.com

Procedural Understanding and Commonsense Reasoning



requires understanding causal relations where one process or state, contributes to the production of another process or state, along with understanding temporal relations

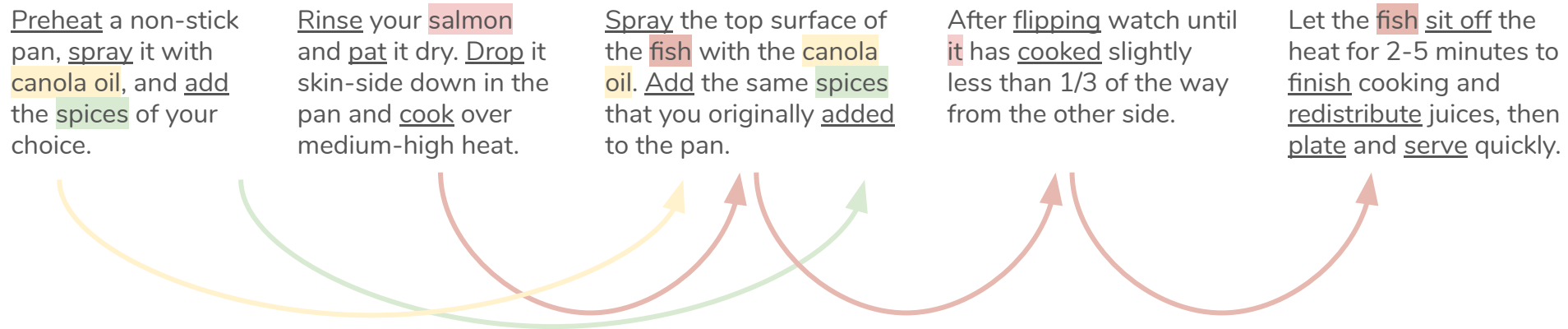
Procedural Understanding and Commonsense Reasoning



it demands modeling the intrinsic dynamics of the procedures *e.g.* identifying key entities and actions, tracking state changes, inferring relations between them, understanding common sense knowledge

Procedural Understanding and Commonsense Reasoning

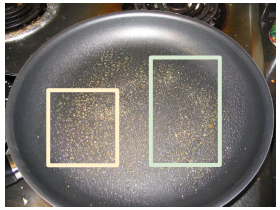
Can we model these intrinsic dynamics and better comprehend procedures?



it demands modeling the intrinsic dynamics of the procedures *e.g.* identifying key entities and actions, tracking state changes, inferring relations between them, understanding common sense knowledge

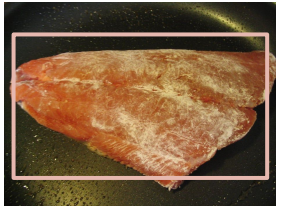
Procedural Understanding and Commonsense Reasoning

1



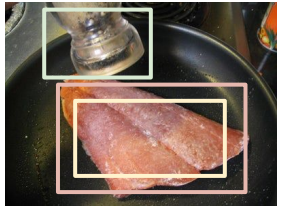
Preheat a non-stick pan, spray it with canola oil, and add the spices of your choice.

2



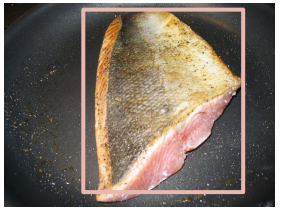
Rinse your salmon and pat it dry. Drop it skin-side down in the pan and cook over medium-high heat.

3



Spray the top surface of the fish with the canola oil. Add the same spices that you originally added to the pan.

4



After flipping watch until it has cooked slightly less than 1/3 of the way from the other side.

5

Let the fish sit off the heat for 2-5 minutes to finish cooking and redistribute juices, then plate and serve quickly.

we explore the question of how multimodality can be exploited to provide a complementary semantic signal to address the problem of comprehending procedural commonsense knowledge

Credit: Basic Steps of How to Make Cheese - The Milk Maid (Instructibles)

RecipeQA

- around **20K unique recipes**
- with arbitrary number of steps
- **3 visual reasoning tasks**
- over **27K recipe-question-answer** triplets

RecipeQA: A Challenge Dataset for Multimodal Comprehension of Cooking Recipes,
S. Yagcioglu, *A. Erdem*, *E. Erdem*,
N. Ikizler-Cinbis, EMNLP 2018

https://hucvl.github.io/recipeqa/

Visual Reasoning Tasks in RecipeQA

Step 1: Select and Prepare Your Bread Slices
Cut your bread sticks into thickish slices diagonally and arrange on a tray (I cover the tray with foil for easy clean up afterwards).liberally sprinkle olive oil on the slices and [...]

Step 2: Prepare the Garlic Butter
Right - while the bread is toasting, its time to prepare the garlic butter.Choose a microwave safe cup or ramakin, put some butter or marg in it and zap it in the microwave for about 30-40 seconds. [...]

Recipe (R)

Step 3: Butter Up Your Slices
Your bread should now be nicely toasted, remove the tray and flip your slices. Add a good teaspoon of the butter/garlic mix to each slice, stir the mix well as the garlic tends to sink [...]

Step 4: Cheese It Up
The final step is to add your favourite cheese topping and melt it again under the grill. I like to add a light sprinkling of herbs on top of the cheese for appearance. Once the cheese is all melted and bubbling - its time to dish them out and collect the thanks of those you share [...]

Question: Choose the best image for the missing blank to **correctly complete the recipe.**

Choices:

A) B) C) D)

Visual Cloze Task

Easy Garlic Bread and Cheese

Step 1: Select and Prepare Your Bread Slices
Cut your bread sticks into thickish slices diagonally and arrange on a tray (I cover the tray with foil for easy clean up afterwards).liberally sprinkle olive oil on the slices and [...]

Step 2: Prepare the Garlic Butter
Right - while the bread is toasting, its time to prepare the garlic butter.Choose a microwave safe cup or ramakin, put some butter or marg in it and zap it in the microwave for about 30-40 seconds. [...]

Step 3: Butter Up Your Slices
Your bread should now be nicely toasted, remove the tray and flip your slices. Add a good teaspoon of the butter/garlic mix to each slice, stir the mix well as the garlic tends to sink [...]

Step 4: Cheese It Up
The final step is to add your favourite cheese topping and melt it again under the grill. I like to add a light sprinkling of herbs on top of the cheese for appearance. Once the cheese is all melted and bubbling - its time to dish them out and collect the thanks of those you share [...]

Question: Choose the best image for the missing blank to **correctly complete the recipe**.

Choices:

A) B) C) D)

Visual Cloze Task

Easy Garlic Bread and Cheese

Step 1: Select and Prepare Your Bread Slices
Cut your bread sticks into thickish slices diagonally and arrange on a tray (I cover the tray with foil for easy clean up afterwards).liberally sprinkle olive oil on the slices and [...]

Step 2: Prepare the Garlic Butter
Right - while the bread is toasting, its time to prepare the garlic butter.Choose a microwave safe cup or ramakin, put some butter or marg in it and zap it in the microwave for about 30-40 seconds. [...]

Step 3: Butter Up Your Slices
Your bread should now be nicely toasted, remove the tray and flip your slices. Add a good teaspoon of the butter/garlic mix to each slice, stir the mix well as the garlic tends to sink [...]

Step 4: Cheese It Up
The final step is to add your favourite cheese topping and melt it again under the grill. I like to add a light sprinkling of herbs on top of the cheese for appearance. Once the cheese is all melted and bubbling - its time to dish them out and collect the thanks of those you share [...]

Question: Choose the best image for the missing blank to **correctly complete the recipe**.

Choices:

A) B) C) D)

Visual Coherence Task

Chicken Jelly Cake

Step 1: Chicken Bullion
Put a chicken thigh in a saucepan and pour a hot water over it (just enough to cover the chicken), add some onions and black pepper. Simmer for about an hour or a bit more [...]

Step 2: Cutting and Separating Everything
Drain the broth in a separate small container and put it aside. Cut the carrots in a shape you like and arrange them beautifully at the bottom of the dish (so when you flip it - the shape will be on top) [...]

Step 3: Final Bit
Lay the chicken on top of the carrots very carefully (so the shape will not get distorted/messed up). Put all the chicken in and press the spoon gently on top. [...]

Step 4: Flip
Take a knife and slide it through the edges of the chicken jelly, so it will unstick. Tip it over on a plate and decorate it. Wa lah - it's done! :-)

Question: Select the **incoherent image** in the following sequence of images.

Choices:

A) B) C) D)

Visual Coherence Task

Chicken Jelly Cake

Step 1: Chicken Bullion
Put a chicken thigh in a saucepan and pour a hot water over it (just enough to cover the chicken), add some onions and black pepper. Simmer for about an hour or a bit more [...]

Step 2: Cutting and Separating Everything
Drain the broth in a separate small container and put it aside. Cut the carrots in a shape you like and arrange them beautifully at the bottom of the dish (so when you flip it - the shape will be on top) [...]

Step 3: Final Bit
Lay the chicken on top of the carrots very carefully (so the shape will not get distorted/messed up). Put all the chicken in and press the spoon gently on top. [...]

Step 4: Flip
Take a knife and slide it through the edges of the chicken jelly, so it will unstick. Tip it over on a plate and decorate it. Wa lah - it's done! :-)

Question: Select the **incoherent image** in the following sequence of images.

Choices:

A) B) C) D)

Visual Ordering Task

Pepperoni Pizza Dip

Step 1: Taking Your Positions
You are going to need: 1 package (8 oz.) cream cheese 1/2 cup sour cream 1/8 teaspoon dried oregano 1/8 teaspoon garlic powder [...]

Step 2: 1...2...3...
Mix the cream cheese, the sour cream, the oregano, garlic powder and cayenne pepper in a mixing bowl. Stir until it has the [...]

Step 3: ...HUT!
Spread into an ungreased pan and spread the pizza sauce on top.

⋮

Step 5: Run for the Touchdown!
Take out of the oven, sprinkle on the cheese and put it in the oven for another 5 or so minutes.

Question: What is the **correct order** of the images?

A)

B)

C)

D)

Visual Ordering Task

Pepperoni Pizza Dip

Step 1: Taking Your Positions
You are going to need: 1 package (8 oz.) cream cheese 1/2 cup sour cream 1/8 teaspoon dried oregano 1/8 teaspoon garlic powder [...]

Step 2: 1...2...3...
Mix the cream cheese, the sour cream, the oregano, garlic powder and cayenne pepper in a mixing bowl. Stir until it has the [...]

Step 3: ...HUT!
Spread into an ungreased pan and spread the pizza sauce on top.

⋮

Step 5: Run for the Touchdown!
Take out of the oven, sprinkle on the cheese and put it in the oven for another 5 or so minutes.

Question: What is the **correct order** of the images?

A)

B)

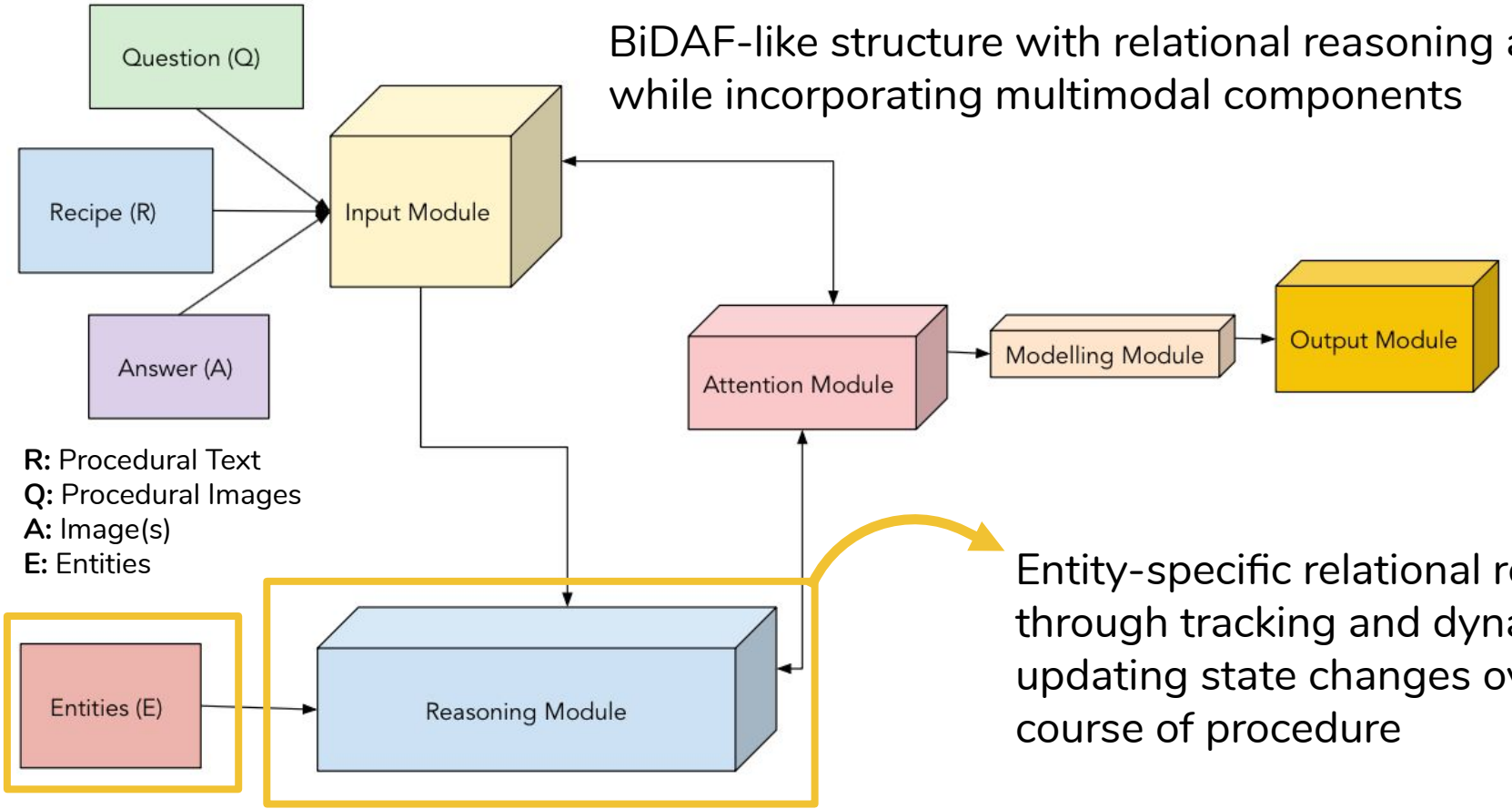
C)

D)

Introducing

Procedural Reasoning Networks (PRN)

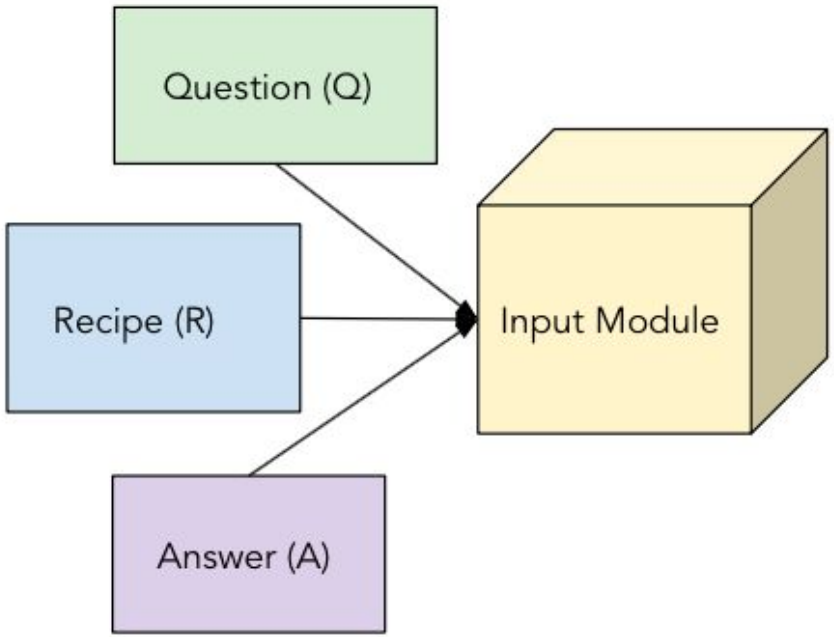
PRN Architecture



BiDAF-like structure with relational reasoning abilities while incorporating multimodal components

Entity-specific relational reasoning through tracking and dynamically updating state changes over the course of procedure

Input Module



extracts vector representations of inputs at different levels of granularity by using several different encoders

for encoding Q and A we used a Resnet-50 model pretrained on ImageNet

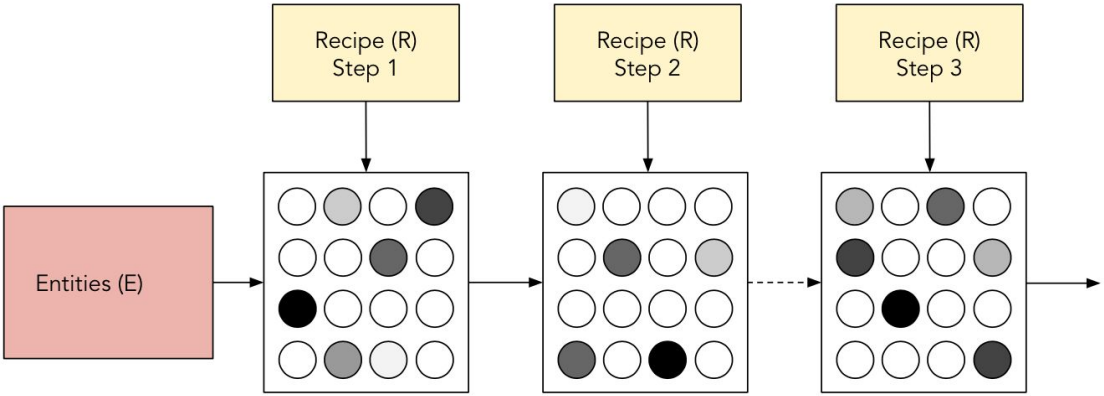
encode recipe R at character, word, and step levels

$$S = (s_1, s_2, \ldots, s_T)$$

we can think of S as **step embedding**...

...from the final state of a BiLSTM encoding of word and char embeddings

Reasoning Module

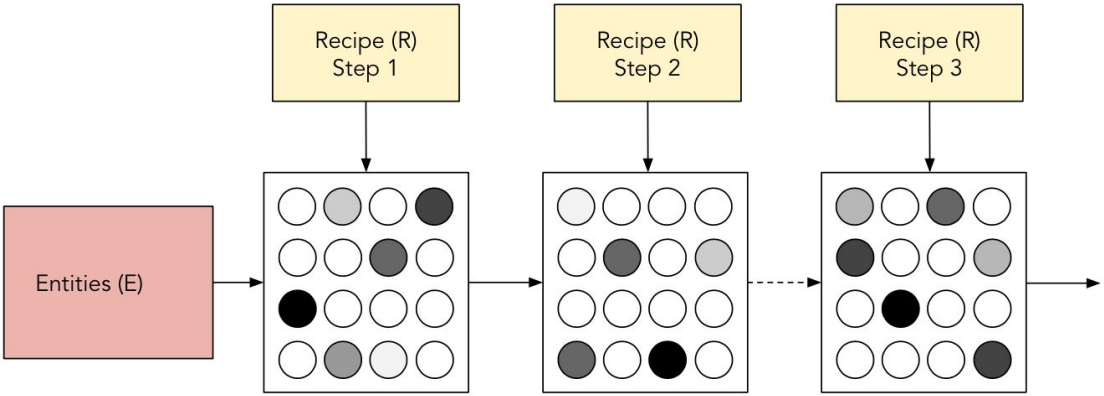


Relational Memory Core (Santoro et al., NeurIPS 2018)

scans the procedural text,
tracks the states of the entities, their relations through a Relational-RNN (Santoro et al., NeurIPS 2018)

memory cells initialized via GloVe vectors of entities

Reasoning Module



Relational Memory Core (Santoro et al., NeurIPS 2018)

R-RNN is 2-d LSTM model whose matrix of cell states represent our memory matrix E

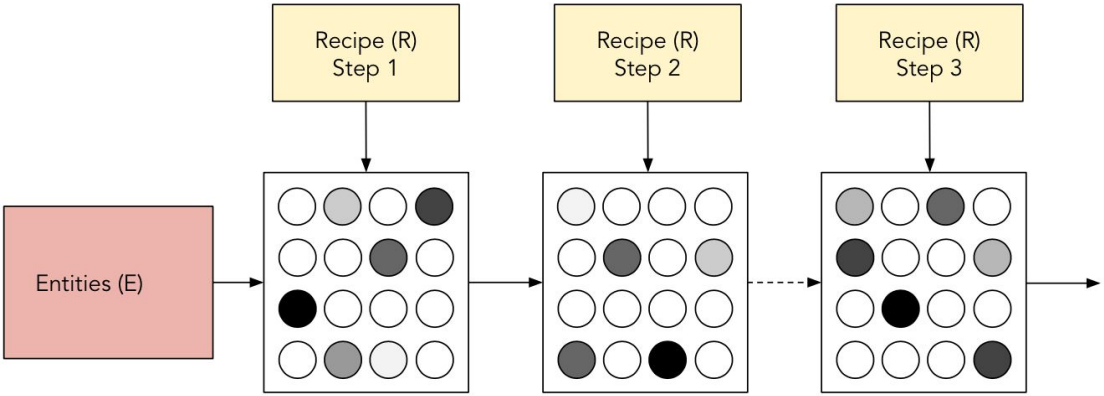
each row refers to a specific entity

$$\phi_{i,t} = \text{R-RNN}(\phi_{i,t-1}, \mathbf{s}_t)$$

we can think of φ as **cell states**...

...of relational RNN conditioned on previous cell states and new step

Reasoning Module



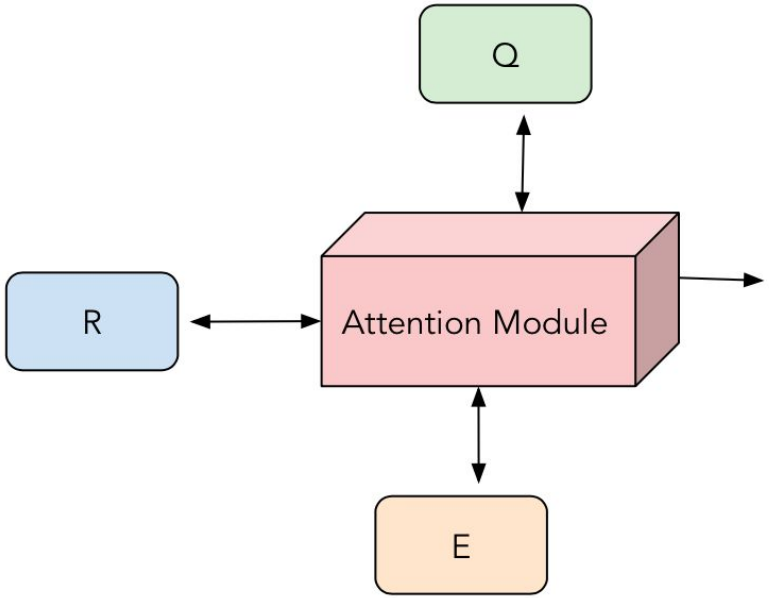
Relational Memory Core (Santoro et al., NeurIPS 2018)

we update the memory cells after each step, reflecting the state changes happened on the entities

multi-head dot attention product (Vaswani et al., NeurIPS 2017) allows memories to interact

dynamically updating entity states in relation to each other and allowing memories to interact enables relational reasoning between states and their interactions in time

Attention Module



Bidirectional Attention Flow (Seo et al., ICLR 2016)

takes Q and R representations from the input module, and E from the reasoning module

constructs the question-aware recipe representation G and the question-aware entity representation Y

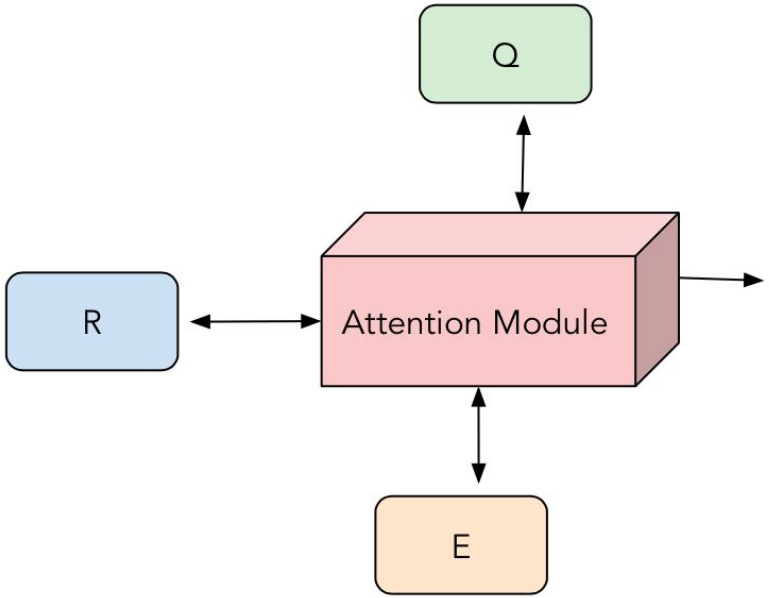
$$\mathbf{S}^R_{i,j} = \mathbf{w}_R^\top [\mathbf{R}'_i; \mathbf{Q}'_j; \mathbf{R}'_i \circ \mathbf{Q}'_j]$$

$$\mathbf{S}^E_{i,j} = \mathbf{w}_E^\top [\mathbf{E}_i; \mathbf{Q}'_j; \mathbf{E}_i \circ \mathbf{Q}'_j]$$

we can think of S as **a shared affinity matrix**...

...where w indicates trainable weights o, is dot product and ; is concatenation

Attention Module



Bidirectional Attention Flow (Seo et al., ICLR 2016)

$$\tilde{\mathbf{Q}}_i = \sum_j \mathbf{a}_{ij} \mathbf{Q}'_j$$

for question-aware representation of recipe R
we calculate a **recipe-to-question attention matrix** $\tilde{\mathbf{Q}}$

$$\mathbf{a}_i = \mathrm{softmax}(\mathbf{S}_i^R)$$

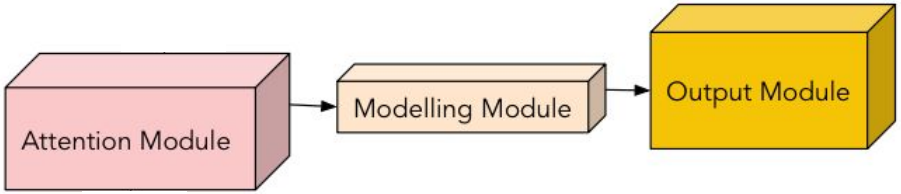
where row-wise attention is computed via affinity matrix S

$$\mathbf{G}_i = [\mathbf{R}'_i; \tilde{\mathbf{Q}}_i; \mathbf{R}'_i \circ \tilde{\mathbf{Q}}_i; \mathbf{R}'_i \circ \tilde{\mathbf{R}}_i]$$

G is a question-aware representation of recipe R

attentions from Q to E, and from E to Q are computed in a way similar to the ones described here, resulting with Y

Modelling Module



$$\mathbf{o} = \varphi_o([\mathbf{c}; \mathbf{f}])$$

we can think of o as **outputs of previous layers**...

...where φo is an MLP with tanh activation function

takes the question-aware representations of the recipe G and the entities Y

employs two multi-layered BiLSTMs to encode previous layers outputs

project them to a fixed size vector

where c and f indicating latent vectors obtained from G and Y

Output Module

scores a candidate answer from the given multiple-choice list

we employ a hinge ranking loss (Collobert et al., 2011)

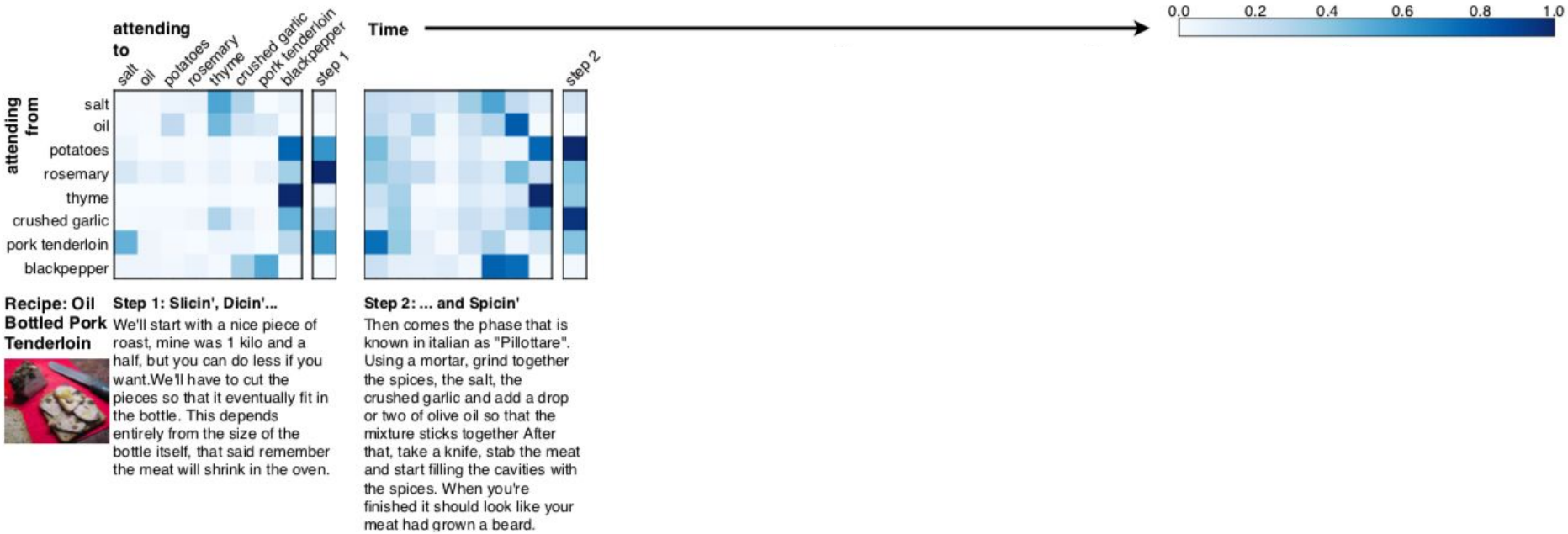
$$L = \max\{0, \gamma - \cos(\mathbf{o}, \mathbf{a}_+) + \cos(\mathbf{o}, \mathbf{a}_-)\}$$

where γ is the margin parameter, a_+ and a_- are the correct and the incorrect answers, respectively.

Relational State Changes in the Memory

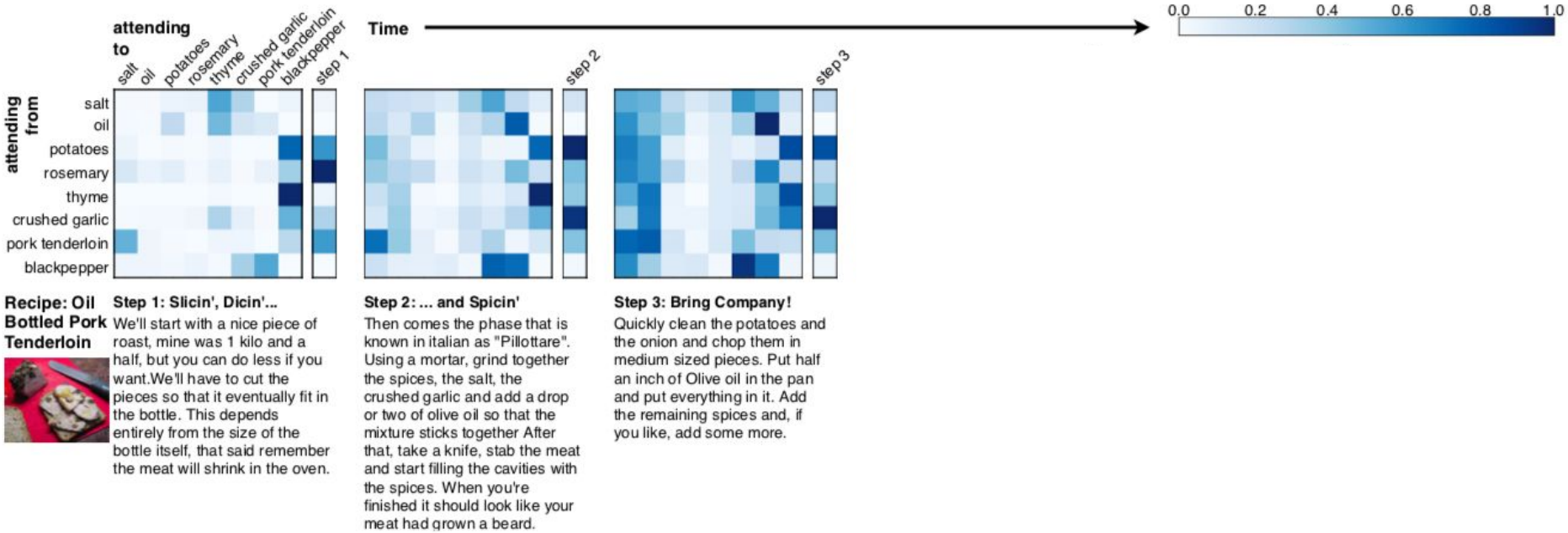
attention maps from one entity to another and their interactions while reading recipe

Relational State Changes in the Memory



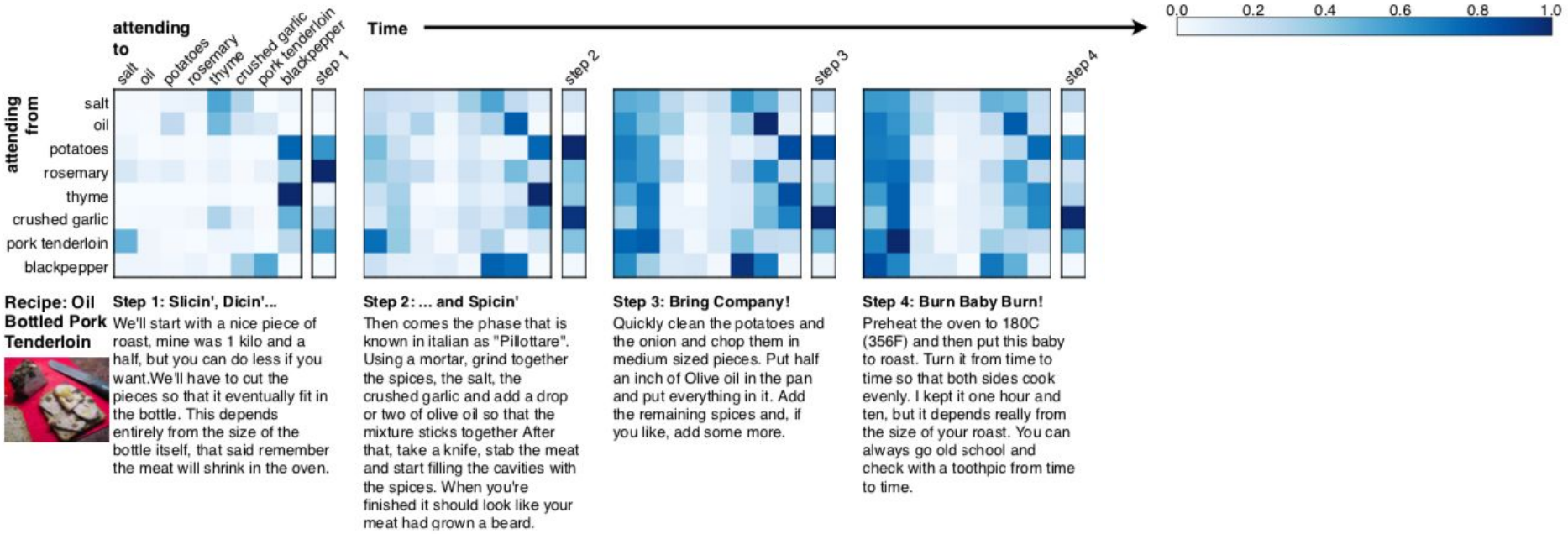
attention maps from one entity to another and their interactions while reading recipe

Relational State Changes in the Memory



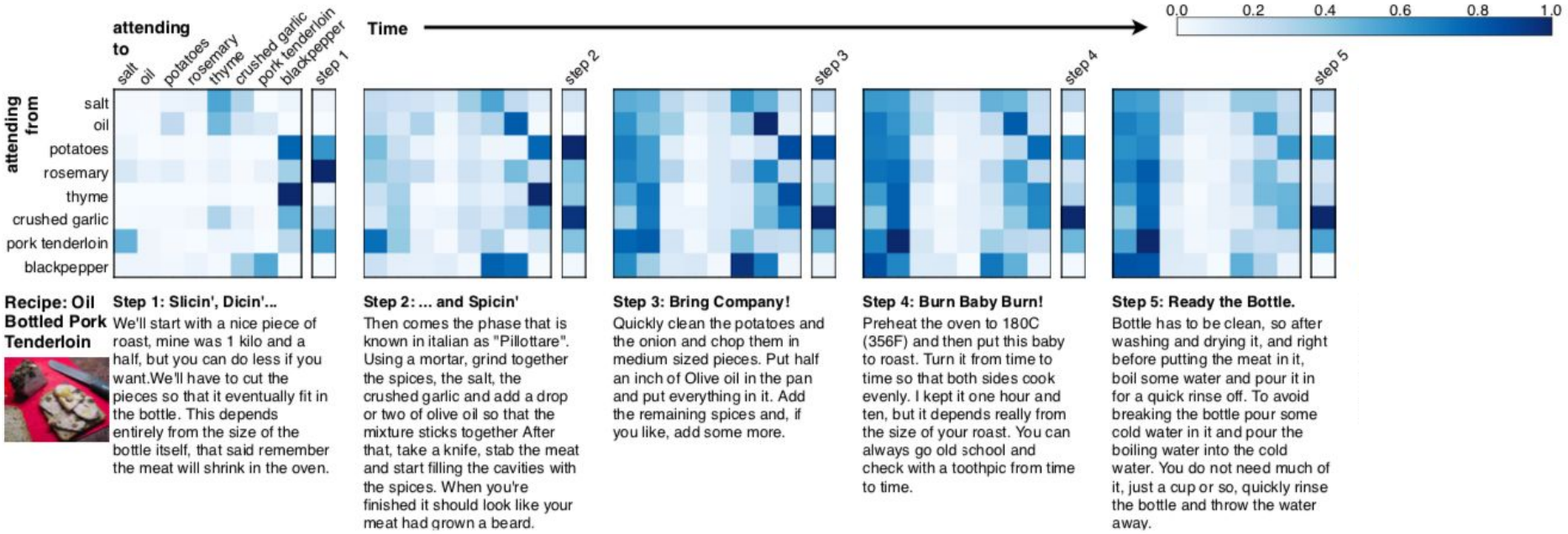
attention maps from one entity to another and their interactions while reading recipe

Relational State Changes in the Memory



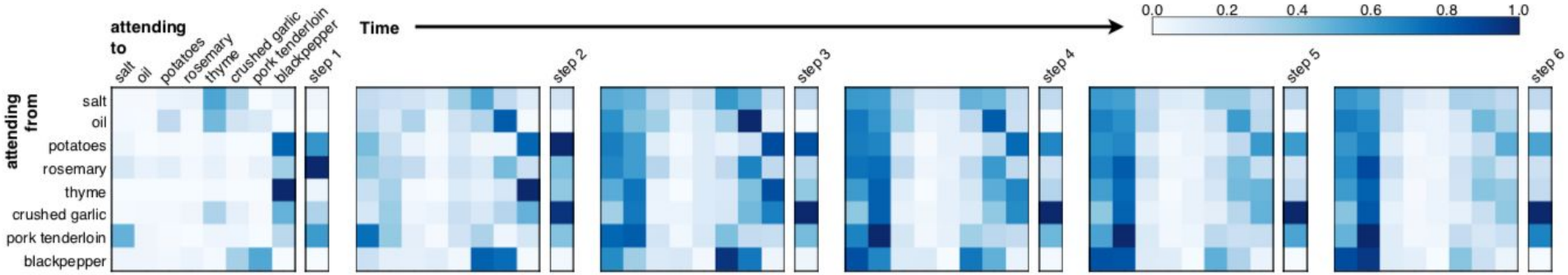
attention maps from one entity to another and their interactions while reading recipe

Relational State Changes in the Memory



attention maps from one entity to another and their interactions while reading recipe

Relational State Changes in the Memory



Recipe: Oil Bottled Pork Tenderloin

Step 1: Slicin', Dicin'...
We'll start with a nice piece of roast, mine was 1 kilo and a half, but you can do less if you want.We'll have to cut the pieces so that it eventually fit in the bottle. This depends entirely from the size of the bottle itself, that said remember the meat will shrink in the oven.

Step 2: ... and Spicin'
Then comes the phase that is known in italian as "Pillottare". Using a mortar, grind together the spices, the salt, the crushed garlic and add a drop or two of olive oil so that the mixture sticks together After that, take a knife, stab the meat and start filling the cavities with the spices. When you're finished it should look like your meat had grown a beard.

Step 3: Bring Company!
Quickly clean the potatoes and the onion and chop them in medium sized pieces. Put half an inch of Olive oil in the pan and put everything in it. Add the remaining spices and, if you like, add some more.

Step 4: Burn Baby Burn!
Preheat the oven to 180C (356F) and then put this baby to roast. Turn it from time to time so that both sides cook evenly. I kept it one hour and ten, but it depends really from the size of your roast. You can always go old school and check with a toothpic from time to time.

Step 5: Ready the Bottle.
Bottle has to be clean, so after washing and drying it, and right before putting the meat in it, boil some water and pour it in for a quick rinse off. To avoid breaking the bottle pour some cold water in it and pour the boiling water into the cold water. You do not need much of it, just a cup or so, quickly rinse the bottle and throw the water away.

Step 6: Put the Piggies to Sleep.
Wait till the meat is cold, then put it into the freshly sterilized bottle and cover in olive oil. The meat has to rest for at least two days, then you can start eating it.

attention maps from one entity to another and their interactions while reading recipe

Results

	Single-task Training				Multi-task Training			
Model	Cloze	Coherence	Ordering	Average	Cloze	Coherence	Ordering	All
Human*	77.60	81.60	64.00	74.40	–	–	–	–
Hasty Student	27.35	**65.80**	40.88	44.68	–	–	–	–
Impatient Reader	27.36	28.08	26.74	27.39	–	–	–	–
BIDAF	53.95	48.82	62.42	55.06	44.62	36.00	**63.93**	48.67
BIDAF w/ static memory	51.82	45.88	60.90	52.87	**47.81**	40.23	62.94	**50.59**
PRN	**56.31**	53.64	**62.77**	**57.57**	46.45	**40.58**	62.67	50.17

*Taken from the RecipeQA project website, based on 100 questions sampled randomly from the validation set.

Entity Arithmetics

onions (Flowerpot Chicken)

Step 1:
This is a cheap and easy method of an ancient cooking technique known as clay pot cooking using a common terra cotta flowerpot and saucer. You can spend over $100 on a clay cooker at a gourmet kitchen gadget store, or about $20 at a garden supply. You choose. Some of you may already have the pot lying in your yard, garage or shed. Once you try this you will probably be cooking all kinds of things in it!

to illustrate our model's capacity to capture the procedural semantics we extracted entity embeddings from memory cells and applied vector arithmetics
e.g. an onion entity from a recipe's first step

Entity Arithmetics

onions (Flowerpot Chicken)

Step 1:
This is a cheap and easy method of an ancient cooking technique known as clay pot cooking using a common terra cotta flowerpot and saucer. You can spend over $100 on a clay cooker at a gourmet kitchen gadget store, or about $20 at a garden supply. You choose. Some of you may already have the pot lying in your yard, garage or shed. Once you try this you will probably be cooking all kinds of things in it!

onions (Flowerpot Chicken)

Step 3: Prepare Vegetables.
Chop your vegetables while the pot is soaking. You can use whatever you like for this, root vegetables mixed with onions are always a nice base. This time I used leeks, bell peppers, garlic and red onions.

if we subtract the same entity's embedding vector from step 3, in which a chopping action is applied on the entity

Entity Arithmetics

onions (Flowerpot Chicken)

Step 1:
This is a cheap and easy method of an ancient cooking technique known as clay pot cooking using a common terra cotta flowerpot and saucer. You can spend over $100 on a clay cooker at a gourmet kitchen gadget store, or about $20 at a garden supply. You choose. Some of you may already have the pot lying in your yard, garage or shed. Once you try this you will probably be cooking all kinds of things in it!

onions (Flowerpot Chicken)

Step 3: Prepare Vegetables.
Chop your vegetables while the pot is soaking. You can use whatever you like for this, root vegetables mixed with onions are always a nice base. This time I used leeks, bell peppers, garlic and red onions.

tomatoes (Flowerpot Chicken)

Step 1:
This is a cheap and easy method of an ancient cooking technique known as clay pot cooking using a common terra cotta flowerpot and saucer. You can spend over $100 on a clay cooker at a gourmet kitchen gadget store, or about $20 at a garden supply. You choose. Some of you may already have the pot lying in your yard, garage or shed. Once you try this you will probably be cooking all kinds of things in it!

and add tomatoes entity from the first step of the same recipe

Entity Arithmetics

onions (Flowerpot Chicken)

Step 1:
This is a cheap and easy method of an ancient cooking technique known as clay pot cooking using a common terra cotta flowerpot and saucer. You can spend over $100 on a clay cooker at a gourmet kitchen gadget store, or about $20 at a garden supply. You choose. Some of you may already have the pot lying in your yard, garage or shed. Once you try this you will probably be cooking all kinds of things in it!

−

onions (Flowerpot Chicken)

Step 3: Prepare Vegetables.
Chop your vegetables while the pot is soaking. You can use whatever you like for this, root vegetables mixed with onions are always a nice base. This time I used leeks, bell peppers, garlic and red onions.

+

tomatoes (Flowerpot Chicken)

Step 1:
This is a cheap and easy method of an ancient cooking technique known as clay pot cooking using a common terra cotta flowerpot and saucer. You can spend over $100 on a clay cooker at a gourmet kitchen gadget store, or about $20 at a garden supply. You choose. Some of you may already have the pot lying in your yard, garage or shed. Once you try this you will probably be cooking all kinds of things in it!

=

?

tomatoes (Chilli Con Carne)

Step 1: Prepping the Vegetables.
The first step is to have all the Vegetables prepped and ready to go in the pan, so finely dice the Garlic, onions and Peppers. Don't worry about mixing them up in the bowl, all of these items are going to be sauteed in a small amount of oil at the next stage. Picture 1. Finely dice up the Garlic, you want it to be almost puree consistency. Picture 2. Finely dice up the Onions, this doesn't need to be as fine as the garlic but you should ensure that they are all roughly the same size. Picture 3. Lastly dice up the bell pepper, I show you how i cut this in the video, but i will go over it quickly. Firstly i take off the four walls of the pepper, flatten them then cut them in to strips, then simply cut the other way so i have them diced.

tomatoes (Seven Layer Seven Grain Bread)

Step 1: Ingredients
...
pepperoni (I used what was left in a package which was enough for one layer) 1/2 onion 2 roma tomatoes dried rosemary shredded mozarella and parmesan fresh savory, basil, tarragon, and thyme 2 or 3 cloves of garlic salt (sea or kosher salt are best) and pepper

Slice the tomatoes and onion as thin as is reasonable, slice the garlic as thin as possible. Thoroughly wash the fresh herbs and pull the leaves from the stems. Discard the stems.

tomatoes (How to Make Chicken Cacciatore)

Step 1: Gather Your Ingredients...
...
1 teaspoon dried oregano, 1/8 teaspoon red pepper flakes (see step five for a bit of humor on this note), 3/4 to 1 cup wine - Honestly, folks, don't be too particular about the wine. Red or white is fine. (you may substitute chicken broth, or even add broth in addition to the wine. Be creative!)(you may substitute chicken broth, or even add broth in addition to the wine. Be creative!) 1 - 28 ounce can diced tomatoes (save the juice!)
1/2 teaspoon dried Porcino mushrooms (Optional, see step #2)

the retrieved nearest neighbours are tomato entities from **various recipes** which are in states such as sliced or diced

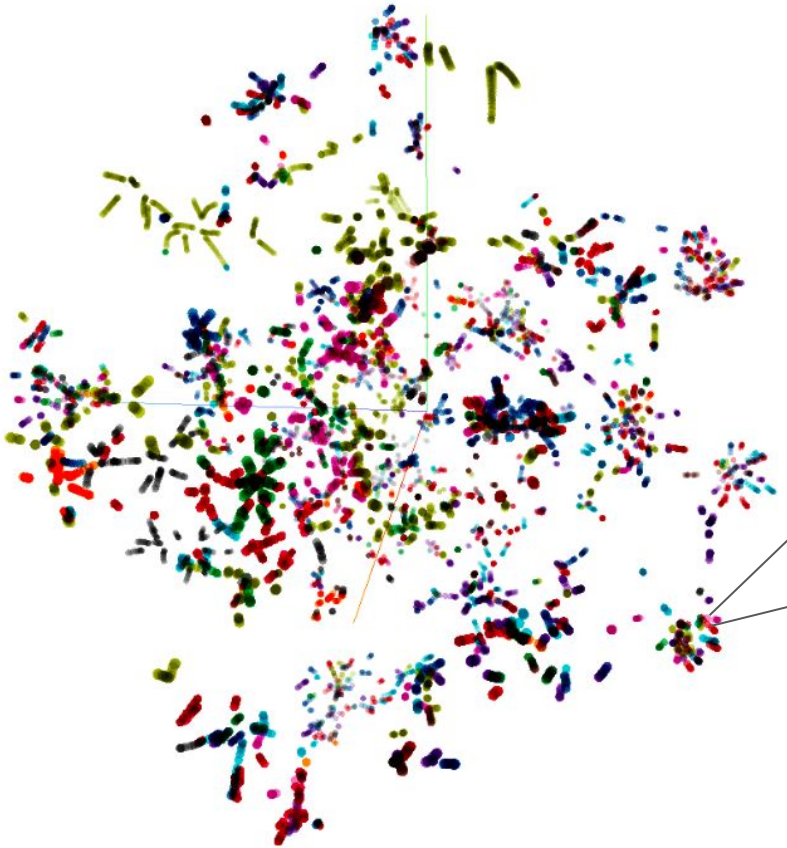
t-SNE Mappings

explored entity state changes using t-SNE mapping the learned entity embeddings neighbouring entities are in similar states

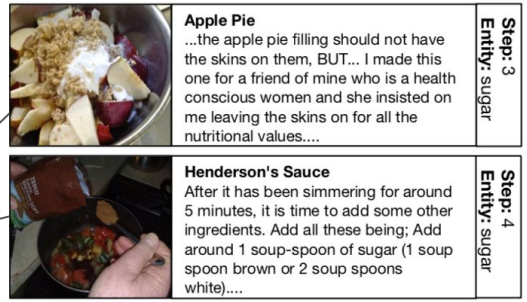
Food Categories

- bread
- vegetarian-and-vegan
- snacks-and-appetizers
- salad
- cocktails-and-mocktails
- sandwiches
- pizza
- soups-and-stews
- pie
- coffee
- canning-and-preserves
- bbq-and-grilling
- dessert
- bacon
- homebrew
- cupcakes
- cake
- breakfast
- pasta
- main-course
- beverages
- cookies
- recipes
- candy

t-SNE Mappings



such that same entities with similar transformations becomes neighbors



Summary

- propose PRNs for **multi-modal understanding** of step-by-step instructions
- explicit memory unit for **relational reasoning**
- implicit mechanism to keep track of state changes of entities
- significantly improves previous model results
- better understanding for **procedural text with accompanying images**
- meaningful dynamic representations of entities
- **no entity-level supervision**
- code and dataset is available!

Q/A

Thank you!

Code: http://hucvl.github.io/prn

M. Sercan Amac, Semih Yagcioglu, Aykut Erdem, Erkut Erdem