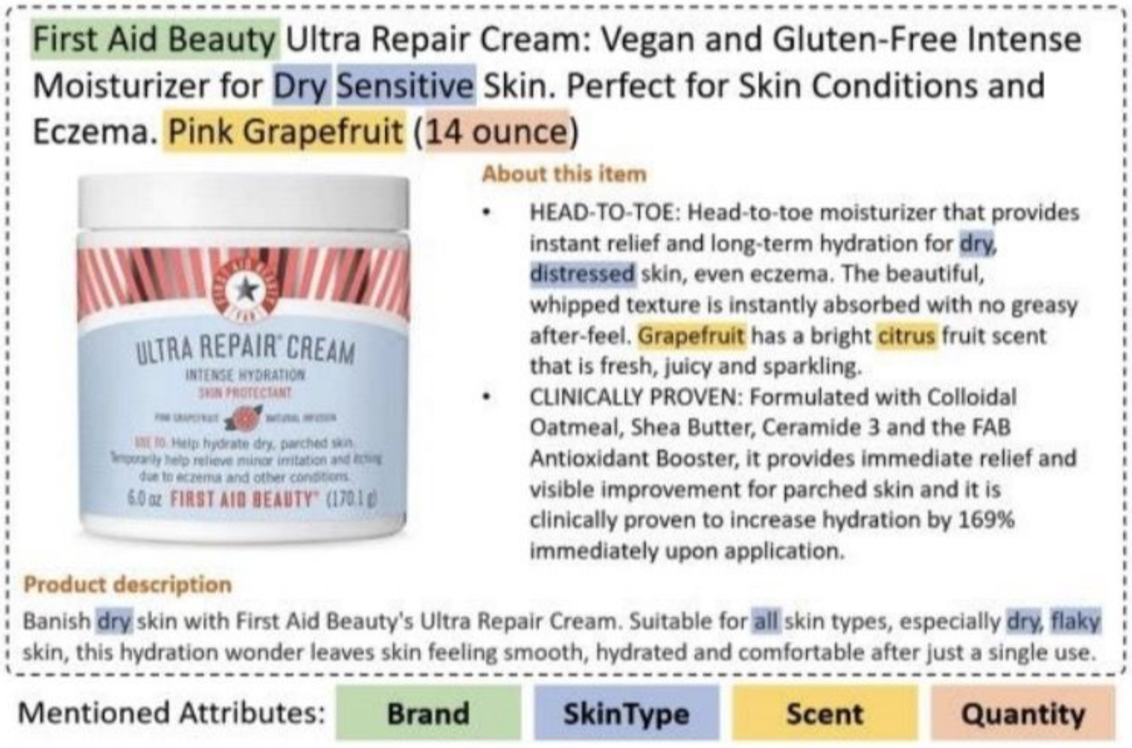
Knowledge Extraction

Questions We Will Answer In This Section

- Task: Attribute value extraction from product profile
  - <u>Attribute value extraction definition</u>
  - Challenges
- Solutions
  - General overview
  - Specific methods

What is Attribute Value Extraction?

Given a product and a list of required attributes, find the attribute values corresponding to each attribute.



Attribute	Attribute Value
Brand	First Aid Beauty
Skin Type	Dry, Sensitive, Distressed, flaky
Scent	Pink Grapefruit, citrus
Quantity	14 ounce

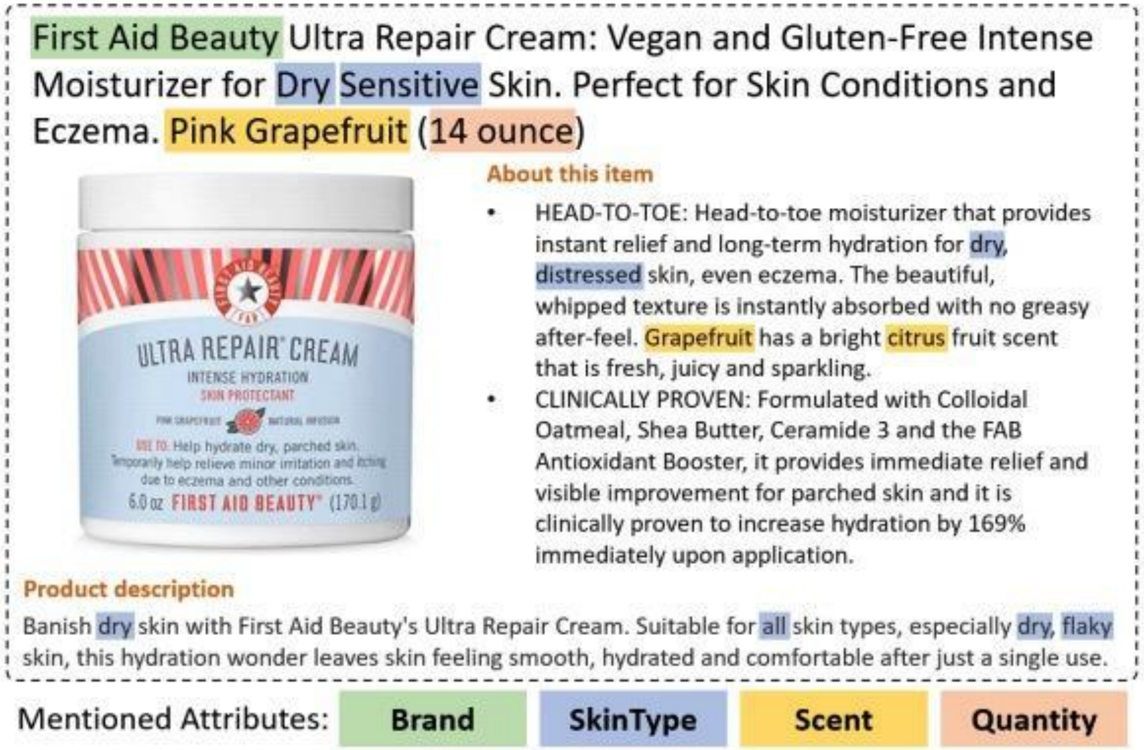
Questions We Will Answer In This Section

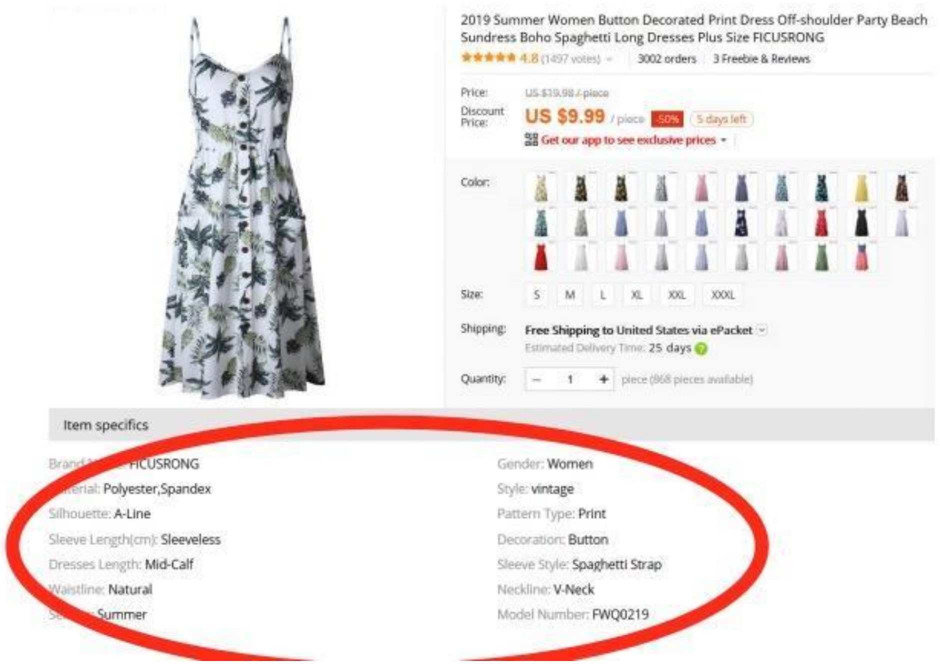
- Task: Attribute value extraction from product profile
  - Attribute value extraction definition
  - <u>Challenges</u>

- Solutions
  - General overview
  - Specific methods

Why Is Attribute Value Extraction Hard?

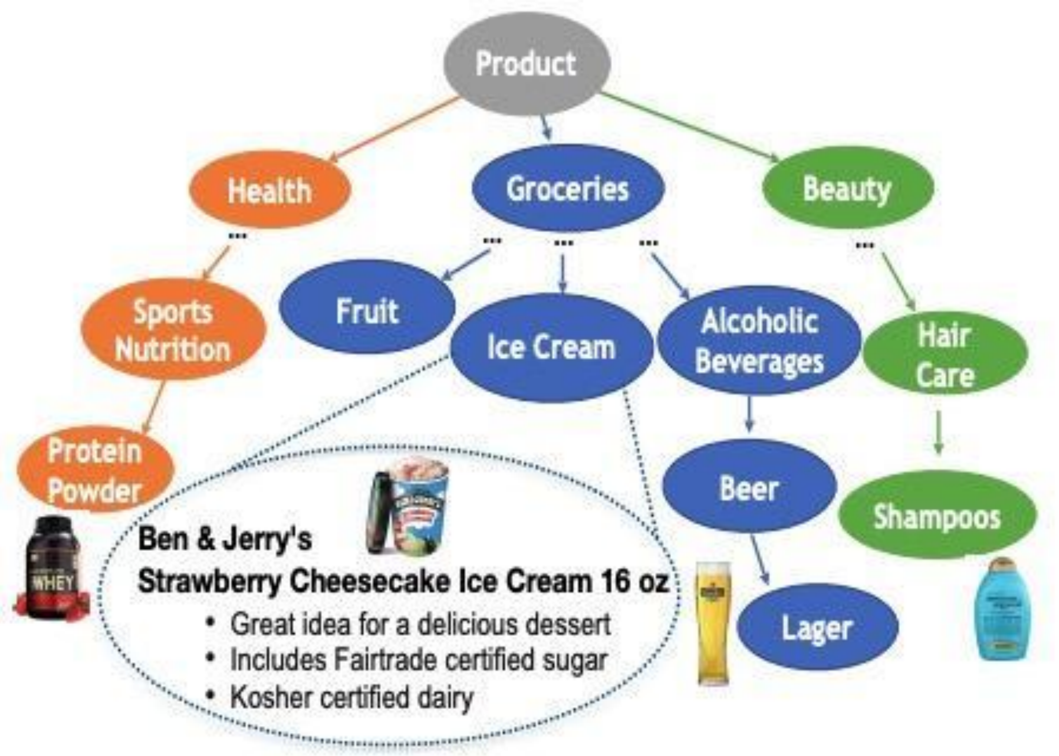
- Values need to be extracted for thousands of attributes and there are evolving new attributes in e-commerce everyday.





Why Is Attribute Value Extraction Hard?

- Attribute values could be different across product types. And there could be thousands of product types.



Eye Shadow product has no scent

Why Is Attribute Value Extraction Hard?

- For a given attribute, there could be multiple attribute values.

Flavor	Assorted
Size	80 Count (Pack of 1)
Brand	Otter Pops
Ingredients	Water, High Fructose Corn Syrup, contains 2% or less of the following: Apple and Pear Juice from Concentrate, Citric Acid, Natural and Artificial Flavors, Sodium Benzoate and Potassium Sorbate (Preservatives), Red... See more

About this item

- FREEZE AT HOME POPS: Pop-Ice Freezer Pops are simple and easy. Just freeze and enjoy!
- FUN FLAVORS: Lemon Lime, Grape, Tropical Punch, Orange, Berry Punch & Strawberry.
- FAT FREE: Pop-Ice freezer popsicles are a zero fat snack or dessert.
- REFRESHING TREAT FOR EVERYONE: Pop-Ice freezer pops are perfect for any age and any occasion.
- 80 FREEZER BARS PER CASE: Each pack has 80 - 1 oz Pop-Ice Freezer Pops.

Why Is Attribute Value Extraction Hard?

- Diversity of textual semantics:
  - “orange” can be a flavor, scent, ingredients, color.
  - “Cleaning ripples” in the category of toilet paper is a special wavy pattern to help with cleaning. “free and clear” in the category of detergent means that it is “scent free”

Why Is Attribute Value Extraction Hard?

- Attribute value can come from multiple resources including: text, text on image and image features itself.

ItemForm for this honey is candy, which can only be inferred from image

Why is Attribute Value Extraction Hard?

- Lack of Training Data
  - Neural network based models require much more annotated data because of huge parameter space.
  - Annotation is an expensive task

Questions we will answer in this section

- Task: Attribute value extraction from product profile
  - Attribute value extraction definition
  - Challenges

- <u>Solutions</u>
  - <u>General overview</u>
  - Specific methods

Generic Information Extraction Method

- Select **features** to represent our raw text
- Select a **model** to take in these features and make a prediction
- **Train** that model

Text features

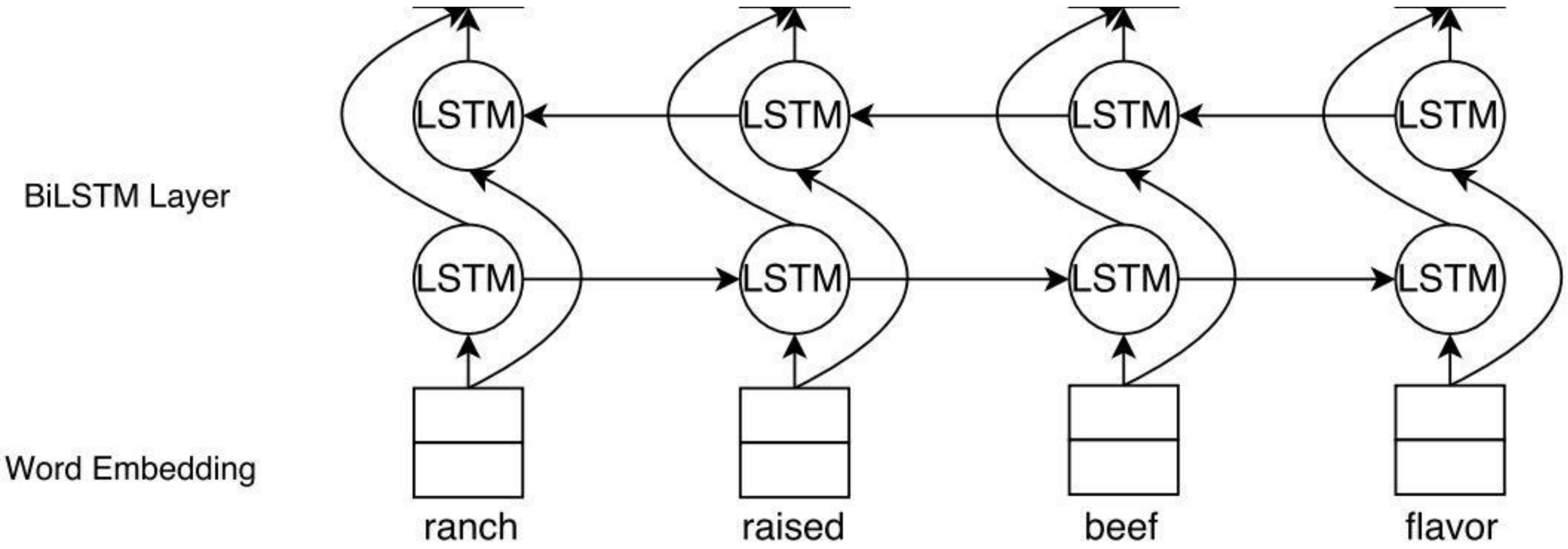
- Understand the meaning of each word
- Understand the meaning of each word in its context
- Understand the meaning of multiple words in a sequence

Futurizing Text

- Bag-of-words, POS tags, syntactic parsing
- Word embeddings: Word2Vec (Mikolov et al, 2013), GloVe (Pennington et al, 2014)
- Pre-trained contextual embedding models

Word Embeddings and LSTMs

- Dense vector representation of a word
  - Bi-LSTMs to encode context



Contextual Word Embeddings

- BERT (Devlin et al, 2019), etc.
  - Builds contextual representation of each token in a sentence
  - Transformer neural net architecture
  - Also builds representation of entire sentence
  - Pre-trained on a large text corpus

Questions we will answer in this section

- Task: Attribute value extraction from product profile
  - Attribute value extraction definition
  - Challenges
- Solutions
  - General overview
  - <u>Specific methods</u>

Method: Sequence Tagging

- ”BIOE Tagging”
  - “Beginning”
  - ”Inside”
  - ”Outside”
  - ”End”

fillet	mignon	and	ranch	raised	lamb	flavor
B	E	O	B	I	E	O

Flavor: fillet mignon, ranch raised lamb

Method: Sequence Tagging

- OpenTag: Open Attribute Value Extraction from Product Profiles (Zheng et al. 2018)
- SUOpenTag: Scaling up Open Tagging from Tens to Thousands: Comprehension Empowered Attribute Value Extraction from Product Title (Xu et al. 2019)
- TXtract: Taxonomy-Aware Knowledge Extraction for Thousands of Product Categories (Karamanolakis et al. 2020)
- AdaTag: Multi-Attribute Value Extraction from Product Profiles with Adaptive Decoding (Yan et al. 2021)

OpenTag

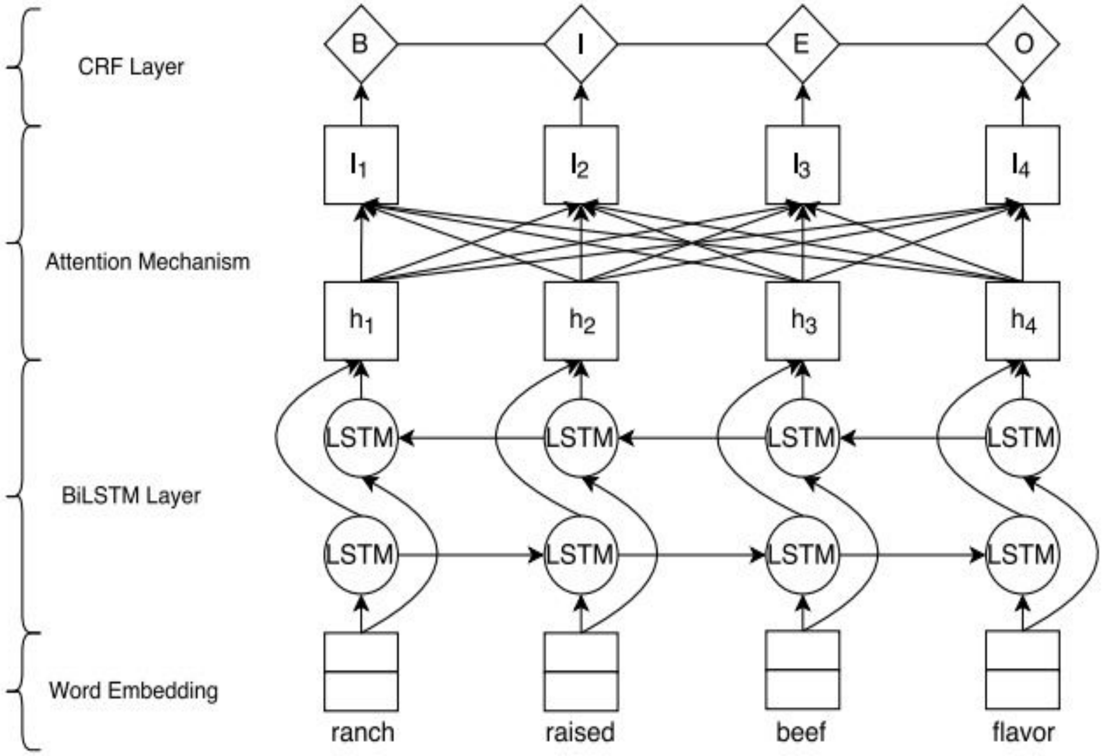
Flavor: Filet Mignon

Flavor: Porterhouse Steak

OpenTag: Open Attribute Value Extraction from Product Profiles (Zheng et al. 2018)

OpenTag

- Word embeddings capture word meaning
- LSTM layer captures word sequence information
- Attention layer allows interaction across sequence
- CRF layer enforces consistency



OpenTag: Open Attribute Value Extraction from Product Profiles (Zheng et al. 2018)

Method: Sequence Tagging

- OpenTag: Open Attribute Value Extraction from Product Profiles (Zheng et al. 2018)
- SUOpenTag: Scaling up Open Tagging from Tens to Thousands: Comprehension Empowered Attribute Value Extraction from Product Title (Xu et al. 2019)
- TXtract: Taxonomy-Aware Knowledge Extraction for Thousands of Product Categories (Karamanolakis et al. 2020)
- AdaTag: Multi-Attribute Value Extraction from Product Profiles with Adaptive Decoding (Yan et al. 2021)

SUOpenTag

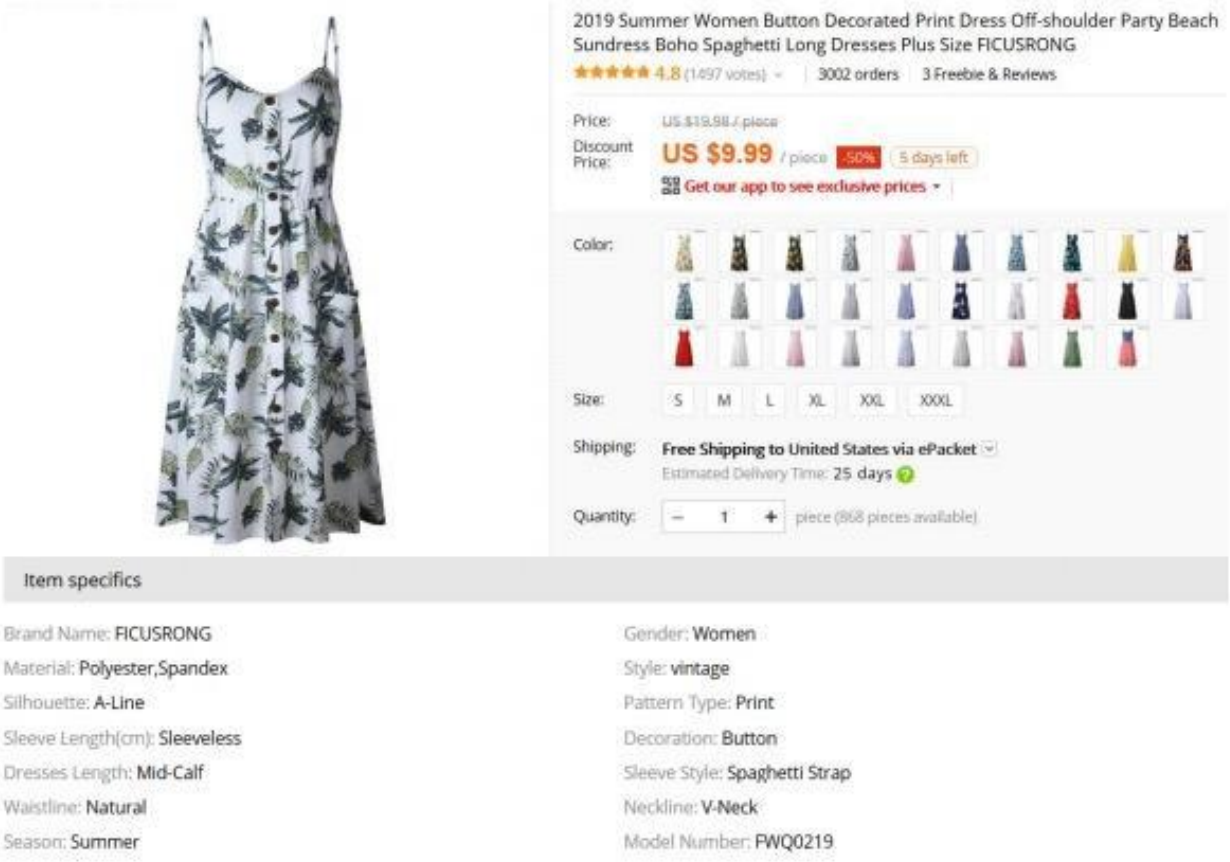
Challenge: Scale up on # of attribute and evolving new attribute

- Scale up to fit the large sized attribute system in the real world
  - The # of attributes typically falls into the range from tens of thousands to millions.
- Extend Open World Assumption to include new attribute
  - Both new attribute and values for newly launched products are emerging everyday.

SUOpenTag: Scaling up Open Tagging from Tens to Thousands: Comprehension Empowered Attribute Value Extraction from Product Title (Xu et al. 2019)

SUOpenTag

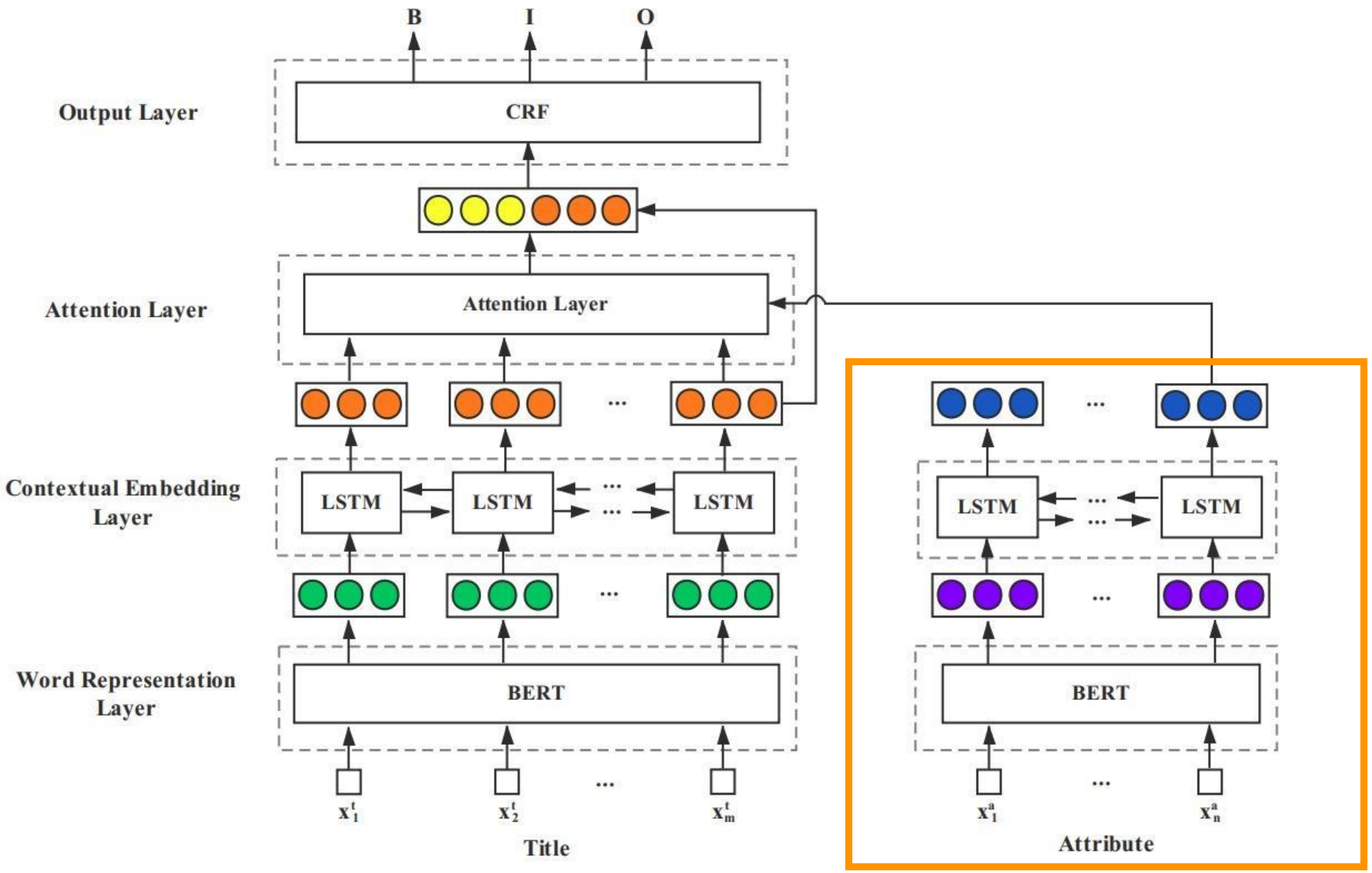
Challenge: Scale up on # of attribute and evolving new attribute



SUOpenTag: Scaling up Open Tagging from Tens to Thousands: Comprehension Empowered Attribute Value Extraction from Product Title (Xu et al. 2019)

SUOpenTag

Challenge: Scale up on # of attribute and evolving new attribute



Using Attribute Embedding Attended to text profile embedding

SUOpenTag: Scaling up Open Tagging from Tens to Thousands: Comprehension Empowered Attribute Value Extraction from Product Title (Xu et al. 2019)

SUOpenTag

Challenge: Scale up on # of attribute and evolving new attribute

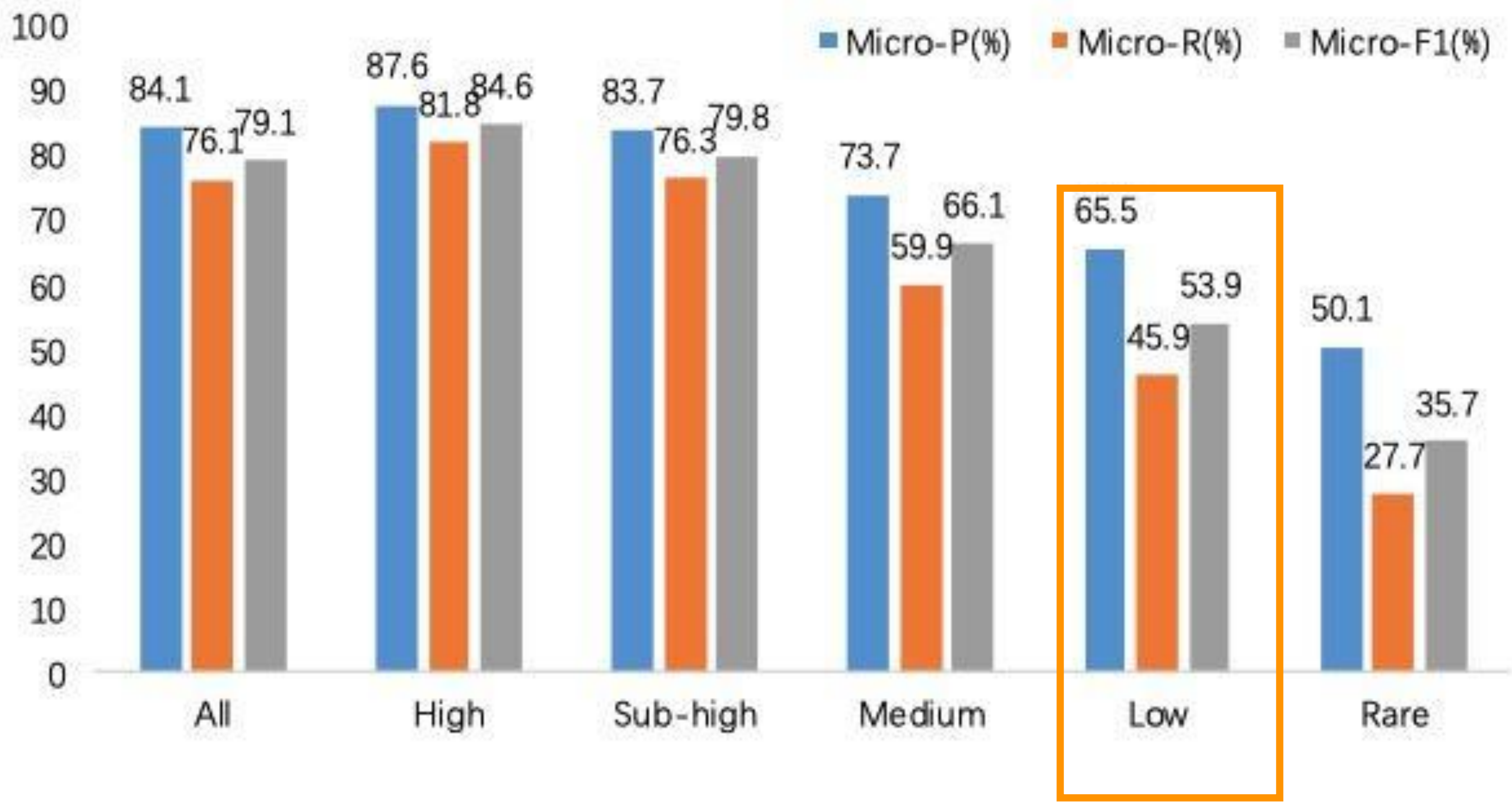
- Dataset:
  - AE-650k, 650K triples which includes 8906 attributes.
  - AE-110K, 110K triples which includes the four frequent attributes. (Brand, Material, Color and Category)

Groups	Occurrence	# of Attributes	Example of attributes
High	$[10{,}000, \infty)$	10	Gender, Brand Name, Model Number, Type, Material
Sub-high	[1000, 10,000)	60	Feature, Color, Category, Fit, Capacity
Medium	[100, 1000)	248	Lenses Color, Pattern, Fuel, Design, Application
Low	[10, 100)	938	Heel, Shaft, Sleeve Style, Speed, Carbon Yarn
Rare	[1, 10)	7,650	Tension, Astronomy, Helmet Light, Flashlight Pouch

SUOpenTag: Scaling up Open Tagging from Tens to Thousands: Comprehension Empowered Attribute Value Extraction from Product Title (Xu et al. 2019)

SUOpenTag

Challenge: Scale up on # of attribute and evolving new attribute



The attributes that have more training data get better results

SUOpenTag: Scaling up Open Tagging from Tens to Thousands: Comprehension Empowered Attribute Value Extraction from Product Title (Xu et al. 2019)

SUOpenTag

Attributes	Models	P (%)	R (%)	F_1 (%)
Brand Name	BiLSTM	95.08	96.81	95.94
	BiLSTM-CRF	95.45	97.17	96.30
	OpenTag	95.18	97.55	96.35
	Our model-110k	**97.21**	96.68	**96.94**
	Our model-650k	**96.94**	**97.14**	**97.04**
Material	BiLSTM	78.26	78.54	78.40
	BiLSTM-CRF	77.15	78.12	77.63
	Opentag	78.69	78.62	78.65
	Our model-110k	**82.76**	**83.57**	**83.16**
	Our model-650k	**83.30**	**82.94**	**83.12**
Color	BiLSTM	68.08	68.00	68.04
	BiLSTM-CRF	68.13	67.46	67.79
	Opentag	71.19	70.50	70.84
	Our model-110k	**75.11**	**72.61**	**73.84**
	Our model-650k	**77.55**	**72.80**	**75.10**
Category	BiLSTM	82.74	78.40	80.51
	BiLSTM-CRF	81.57	79.94	80.75
	Opentag	82.74	80.63	81.67
	Our model-110k	**84.11**	**80.80**	**82.42**
	Our model-650k	**88.11**	**81.79**	**84.83**

Challenge: Scale up on # of attribute and evolving new attribute

The performance on attribute-aware training even gets better compared to non-scalable baseline models

SUOpenTag: Scaling up Open Tagging from Tens to Thousands: Comprehension Empowered Attribute Value Extraction from Product Title (Xu et al. 2019)

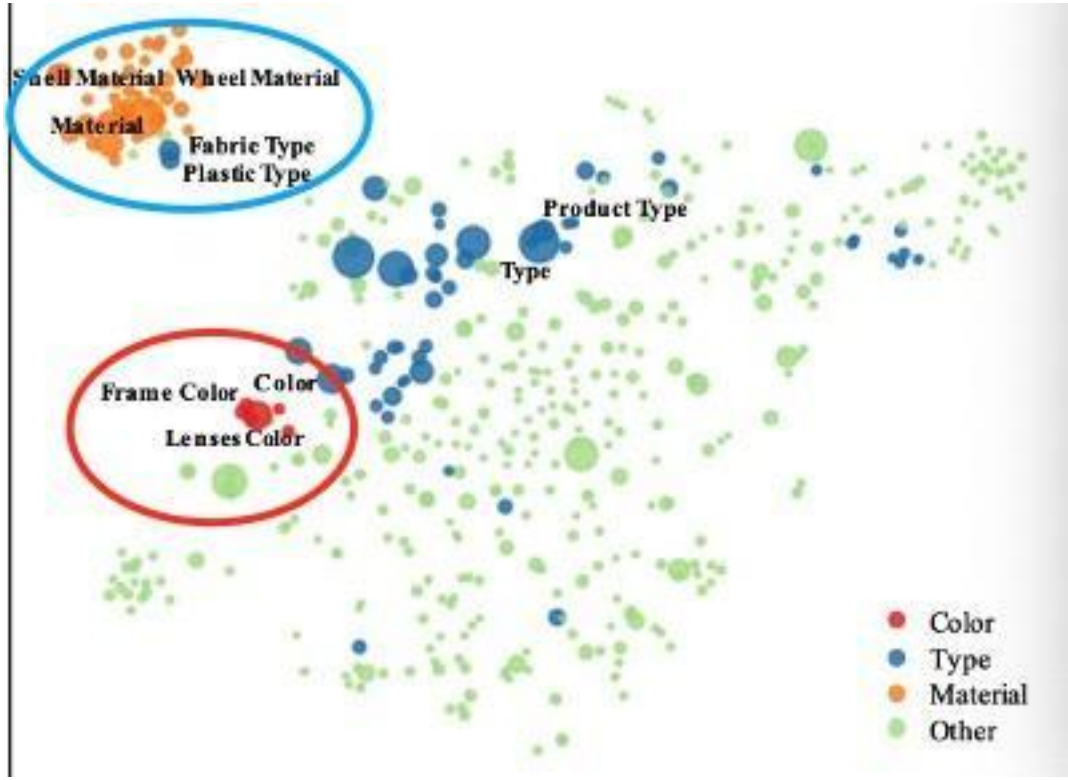
SUOpenTag

Challenge: Scale up on # of attribute and evolving new attribute

- Discovering new attributes

Some attributes in training set are semantically related to unseen attributes and they provide hints to help the extraction

Attributes	P (%)	R (%)	F_1 (%)
Frame Color	63.16	48.00	54.55
Lenses Color	64.29	40.91	50.00
Shell Material	54.05	44.44	48.78
Wheel Material	70.59	37.50	48.98
Product Type	64.86	43.29	51.92



SUOpenTag: Scaling up Open Tagging from Tens to Thousands: Comprehension Empowered Attribute Value Extraction from Product Title (Xu et al. 2019)

Method: Sequence Tagging

- OpenTag: Open Attribute Value Extraction from Product Profiles (Zheng et al. 2018)
- SUOpenTag: Scaling up Open Tagging from Tens to Thousands: Comprehension Empowered Attribute Value Extraction from Product Title (Xu et al. 2019)
- TXtract: Taxonomy-Aware Knowledge Extraction for Thousands of Product Categories (Karamanolakis et al. 2020)
- AdaTag: Multi-Attribute Value Extraction from Product Profiles with Adaptive Decoding (Yan et al. 2021)

TXtract

Challenge: Scale up on # of categories

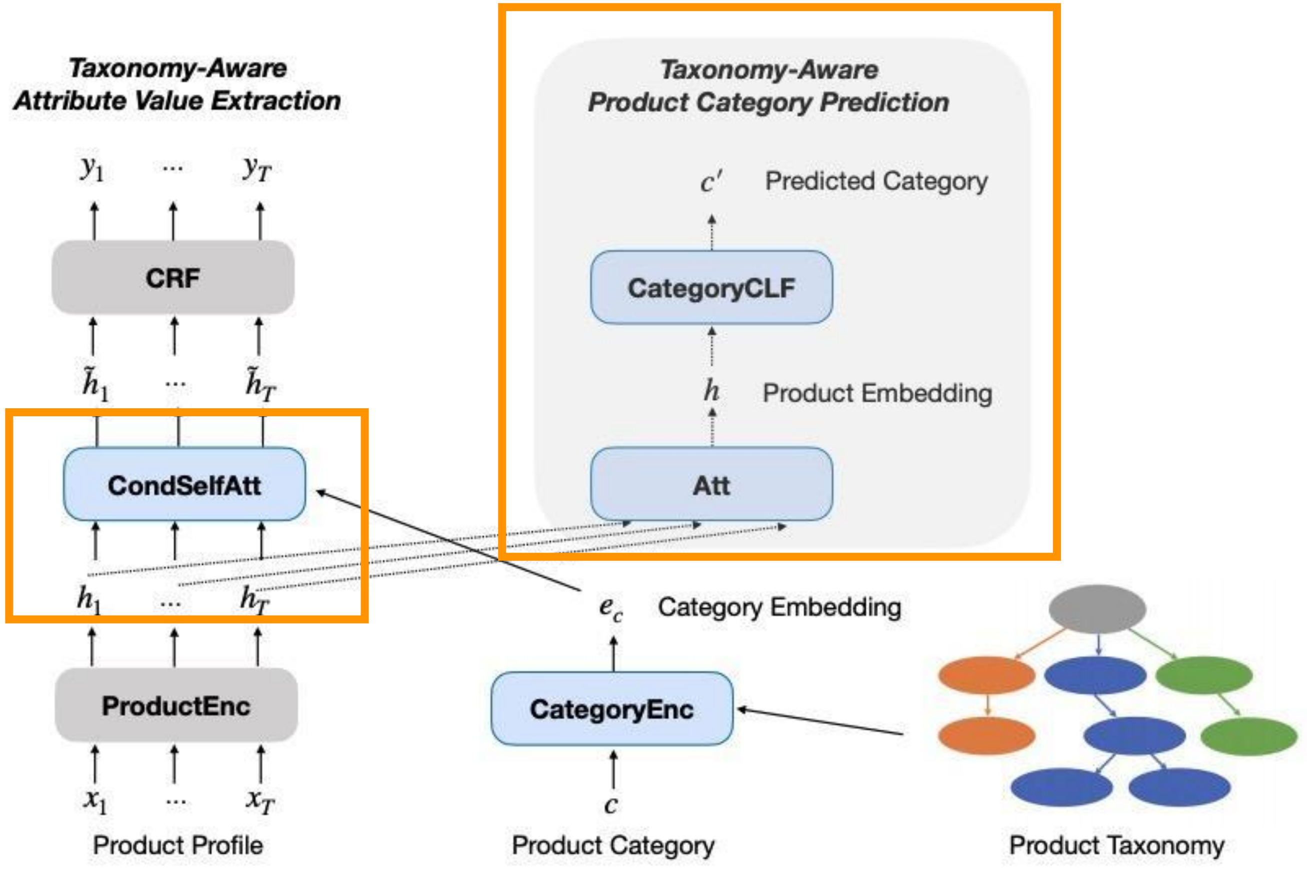
- Capture the hierarchical relations between categories into category embeddings.

- Scaling up the extraction on category by generating category-specific token embeddings.

TXtract: Taxonomy-Aware Knowledge Extraction for Thousands of Product Categories (Karamanolakis et al. 2020)

TXtract

Challenge: Scale up on # of categories



- Tokens are classified to BIOE attribute tags by conditioning to the product's category embedding
- Multi-Task: Attribute Value extraction and Product Type classification

TXtract: Taxonomy-Aware Knowledge Extraction for Thousands of Product Categories (Karamanolakis et al. 2020)

TXtract

Challenge: Scale up on # of categories

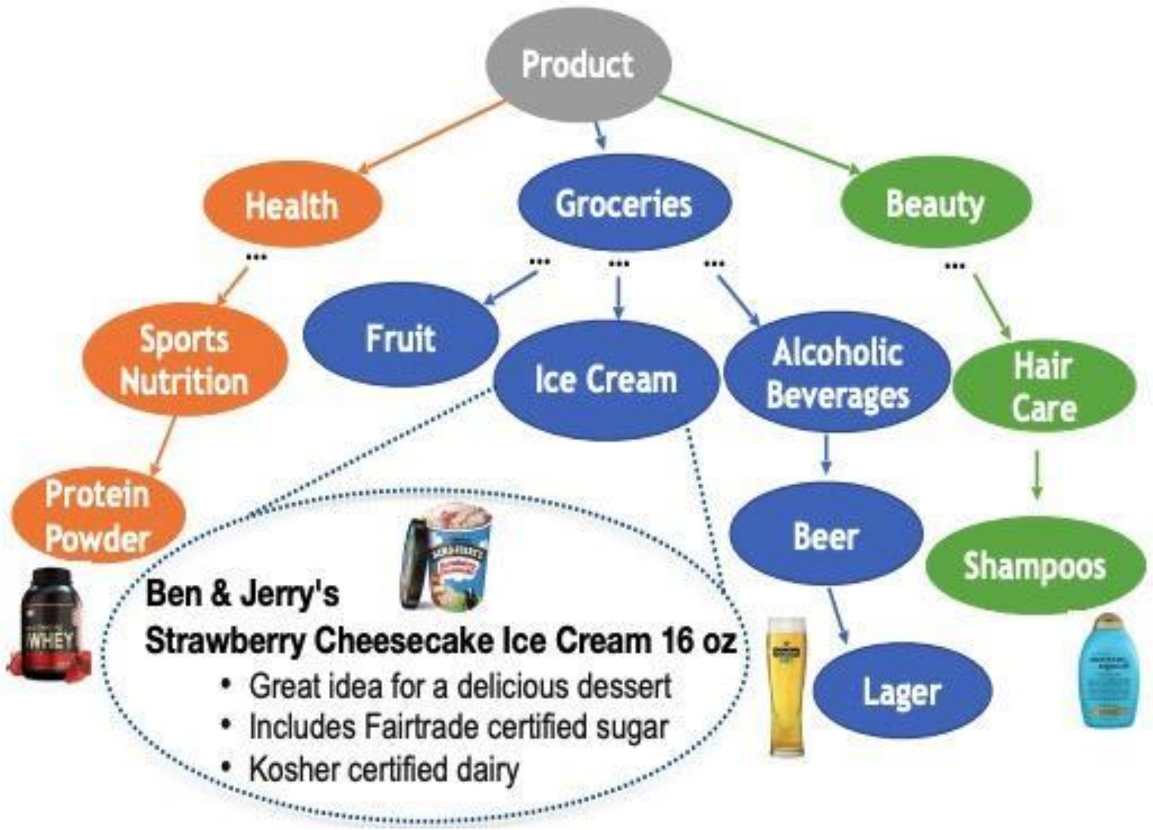
Extraction results for 4 attributes across 4000 categories. Across all attributes, TXtract improves OpenTag by 11.7% in coverage, 6.2% in micro-average F1, and 10.4% in macro-average F1

Attr.	Model	Vocab	Cov	Micro F1	Micro Prec	Micro Rec	Macro F1	Macro Prec	Macro Rec
Flavor	OpenTag	6,756	73.2	57.5	70.3	49.6	54.6	68.0	47.3
	TXtract	**13,093**	**83.9** ↑14.6%	**63.3** ↑10.1%	**70.9** ↑0.9%	**57.8** ↑16.5%	**59.3** ↑8.6%	**68.4** ↑0.6%	**53.8** ↑13.7%
Scent	OpenTag	10,525	75.8	70.6	**87.6**	60.2	59.3	**79.7**	50.8
	TXtract	**13,525**	**83.2** ↑9.8%	**73.7** ↑4.4%	86.1 ↓1.7%	**65.7** ↑9.1%	**59.9** ↑10.1%	78.3 ↓1.8%	**52.1** ↑2.6%
Brand	OpenTag	48,943	73.1	63.4	81.6	51.9	51.7	75.1	41.5
	TXtract	**64,704**	**82.9** ↑13.4%	**67.5** ↑6.5%	**82.7** ↑1.3%	**56.5** ↑8.1%	**55.3** ↑7.0%	**75.2** ↑0.1%	**46.8** ↑12.8%
Ingred.	OpenTag	9,910	70.0	35.7	46.6	29.1	20.9	34.6	16.7
	TXtract	**18,980**	**76.4** ↑9.1%	**37.1** ↑3.9%	**48.3** ↑3.6%	**30.1** ↑3.3%	**24.2** ↑15.8%	**37.4** ↑8.1%	**19.8** ↑18.6%
Average relative increase			↑11.7%	↑6.2%	↑1.0%	↑9.3%	↑10.4%	↑6.8%	↑11.9%

TXtract: Taxonomy-Aware Knowledge Extraction for Thousands of Product Categories (Karamanolakis et al. 2020)

TXtract

Challenge: Scale up on # of categories



How to embed the hierarchical category information into the extraction model

TXtract: Taxonomy-Aware Knowledge Extraction for Thousands of Product Categories (Karamanolakis et al. 2020)

TXtract

Challenge: Scale up on # of categories

Model	TX	MT	Micro F1
OpenTag	-	-	57.5
Title+id	✓	-	55.7 ↓3.1%
Title+name	✓	-	56.9 ↓1.0%
Title+path	✓	-	54.3 ↓5.6%
Concat-wemb-Euclidean	✓	-	60.1 ↑4.5%
Concat-wemb-Poincaré	✓	-	60.6 ↑5.4%
Concat-LSTM-Euclidean	✓	-	60.1 ↑4.5%
Concat-LSTM-Poincaré	✓	-	60.8 ↑5.7%
Gate-Poincaré	✓	-	60.6 ↑5.4%
CondSelfAtt-Poincaré	✓	-	61.9 ↑7.7
MT-flat	-	✓	60.9 ↑5.9%
MT-hier	-	✓	61.5 ↑7.0%
Concat & MT-hier	✓	✓	62.3 ↑8.3%
Gate & MT-hier	✓	✓	61.1 ↑6.3%
CondSelfAtt & MT-hier	✓	✓	**63.3** ↑10.1%

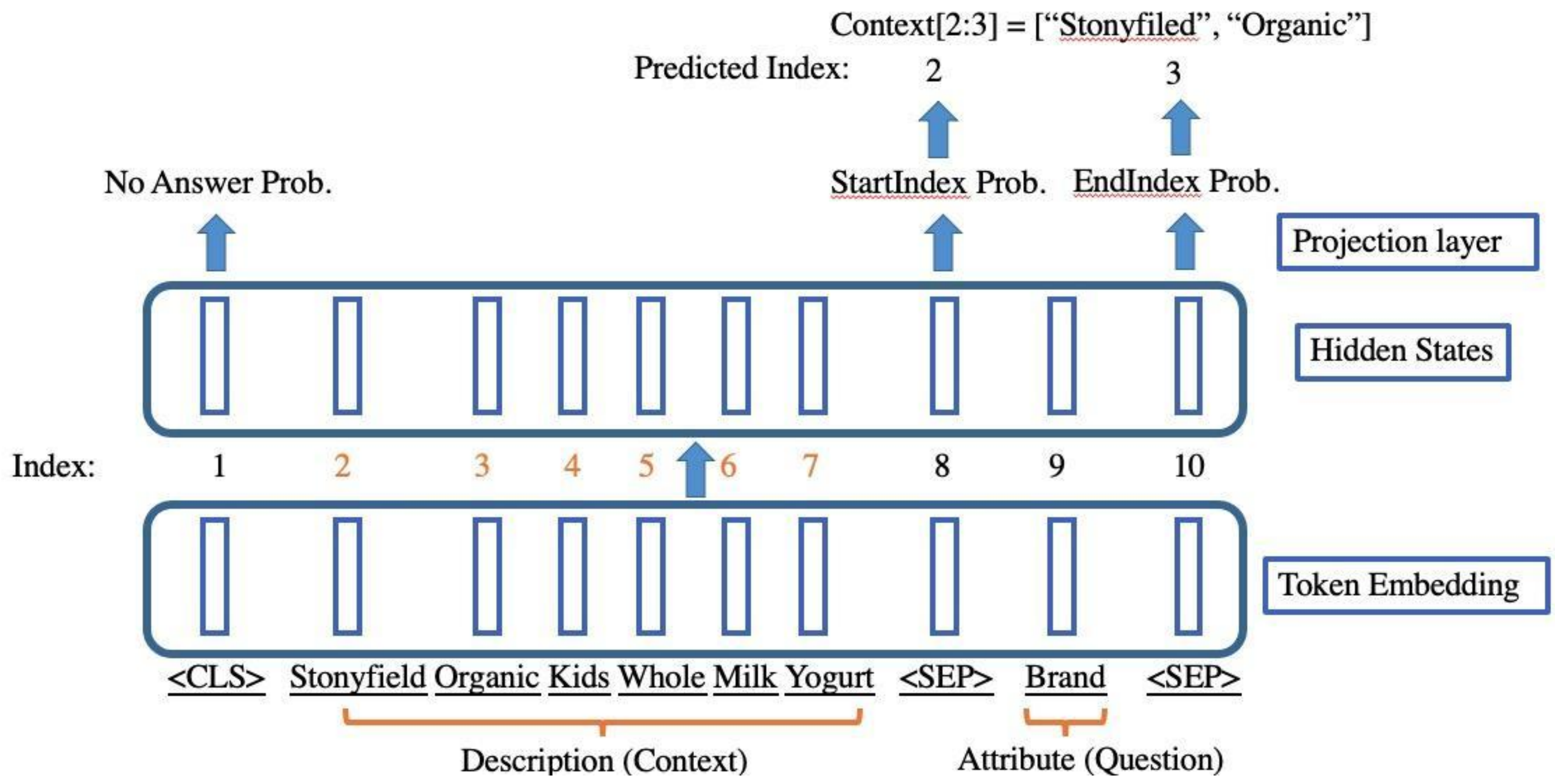
Ablation study on different ways to ingest the category information and effectiveness of multi-task learning

Conditional Self Attention with Poincare embedded product category achieve the best performance

With product category as an auxiliary task, the performance further improved

TXtract: Taxonomy-Aware Knowledge Extraction for Thousands of Product Categories (Karamanolakis et al. 2020)

Method: Question Answering



Method: Question Answering

Challenge : Scale up on # of attribute and evolving new attributes

- AVEQA: Learning to Extract Attribute Value from Product via Question Answering: A Multi-task Approach (Wang et al. 2020)

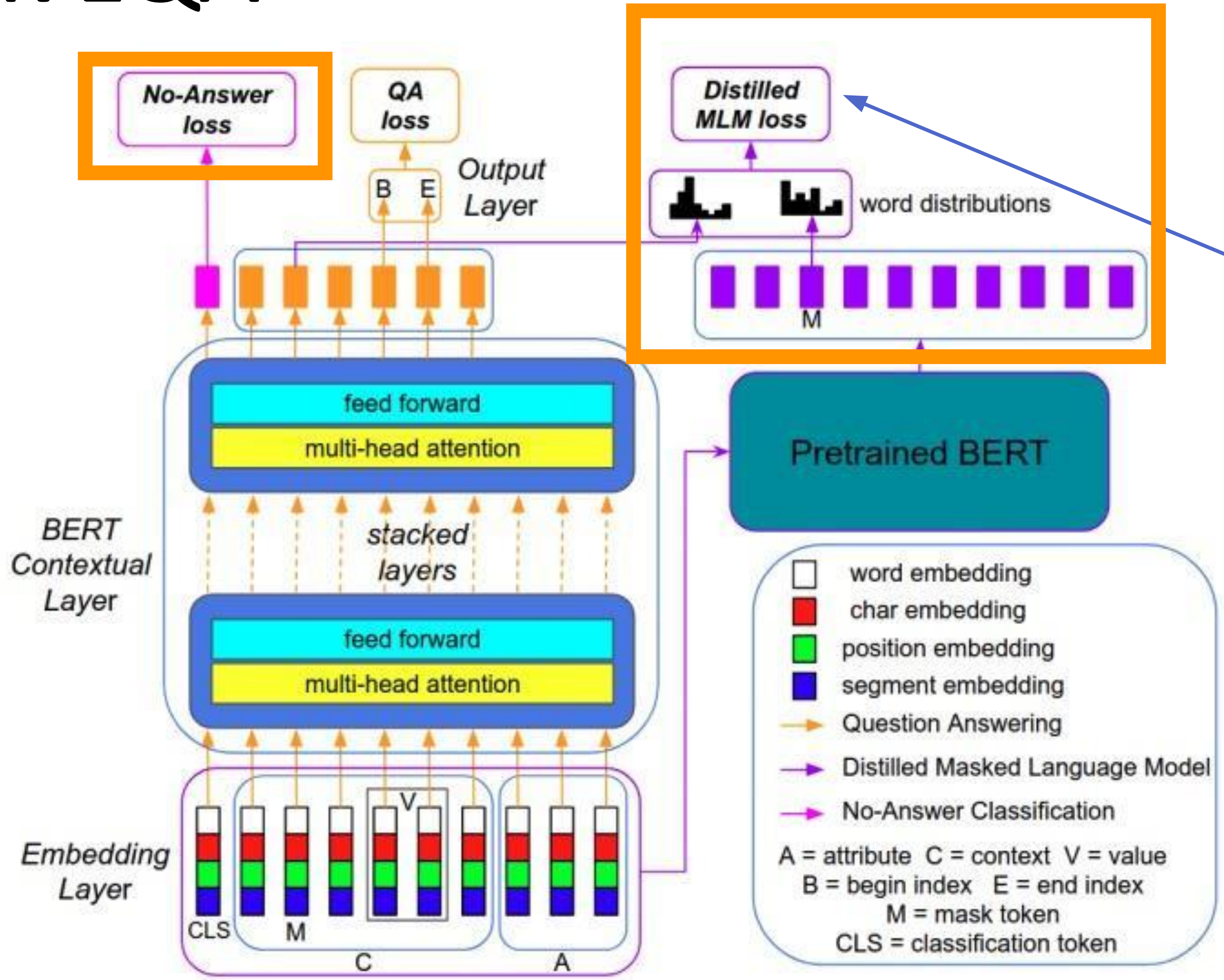
AVEQA

Challenge : Scale up on # of attribute and evolving new attributes

- Formulate the attribute value extraction task as an instance of question answering.
- Distilled mask language model to improve the generalization of the approach on completely unseen attributes.
- Introduce a non-answer classifier to enhance the model ability of predicting no-answers.
- Multi-task approach incorporates all the above tasks.

AVEQA: Learning to Extract Attribute Value from Product via Question Answering: A Multi-task Approach (Wang et al. 2020)

AVEQA



Challenge : Scale up on # of attribute and evolving new attributes

Distilled MLM ensures the encoder to learn effective contextual representations for new attributes, through masking them out and enforcing the predicted distribution to be consistent with the distribution from pretrained BERT.

AVEQA: Learning to Extract Attribute Value from Product via Question Answering: A Multi-task Approach (Wang et al. 2020)

AVEQA

Challenge : Scale up on # of attribute and evolving new attributes

AVEQA further beat the SUOpenTag performance on the frequent seen attributes

	Brand Name			Material			Color			Category		
methods	$P(\%)$	$R(\%)$	$F_1(\%)$	$P(\%)$	$R(\%)$	$F_1(\%)$	$P(\%)$	$R(\%)$	$F_1(\%)$	$P(\%)$	$R(\%)$	$F_1(\%)$
BiLSTM [11]	90.21	90.67	90.44	72.12	62.56	67.00	52.13	48.65	50.33	60.84	50.02	54.89
BiLSTM-CRF [13]	90.45	90.97	90.71	72.40	63.45	67.63	52.68	48.12	50.30	60.48	50.65	55.13
OpenTag [54]	90.32	91.10	90.71	72.56	64.78	68.45	52.83	48.45	50.54	62.17	50.79	55.91
SUOpenTag [50]	91.19	91.57	91.38	74.07	63.86	68.59	57.58	48.72	52.78	62.03	51.58	56.32
AVEQA	**96.41**	**97.00**	**96.70**	**86.34**	**87.20**	**86.76**	**76.47**	**77.68**	**77.06**	**84.43**	**85.70**	**85.05**

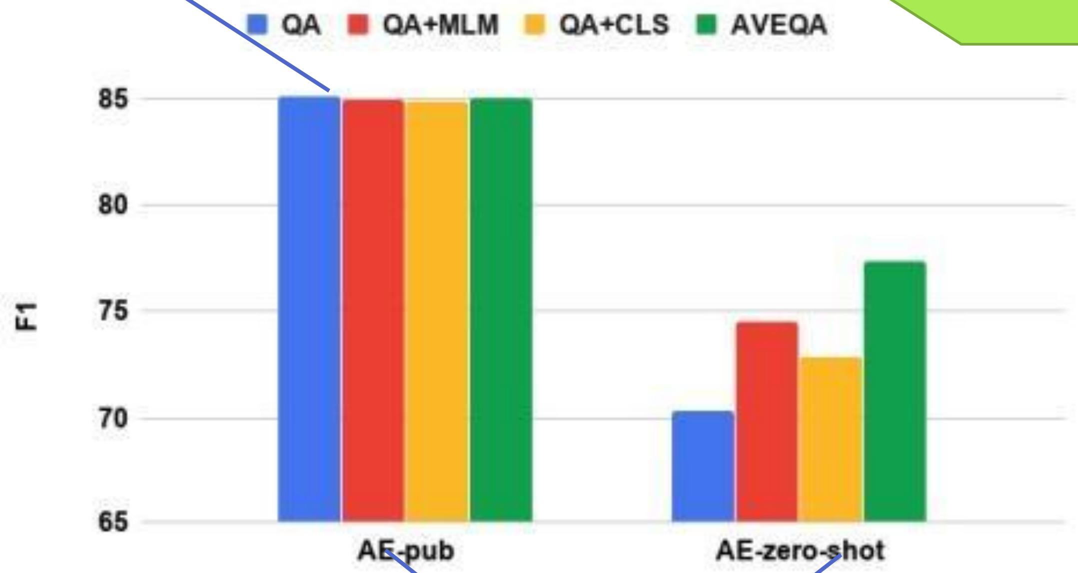
AVEQA: Learning to Extract Attribute Value from Product via Question Answering: A Multi-task Approach (Wang et al. 2020)

AVEQA

QA achieves better performance than SUOpenTag which the micro-f1 is 79.1%

Challenge : Scale up on # of attribute and evolving new attributes

Attributes	Models	$P(\%)$	$R(\%)$	$F_1(\%)$
Frame Color	SUOpenTag	63.16	48.00	54.55
	AVEQA	**86.54**	**48.82**	**62.20**
Lenses Color	SUOpenTag	64.29	40.91	50.00
	AVEQA	**88.42**	**45.91**	**59.94**
Shell Material	SUOpenTag	54.05	44.44	48.78
	AVEQA	**73.96**	**65.76**	**69.52**
Wheel Material	SUOpenTag	70.59	37.50	48.98
	AVEQA	**70.69**	**65.56**	**67.96**
Product Type	SUOpenTag	64.86	43.29	51.92
	AVEQA	**91.79**	**70.69**	**79.82**



QA + MLM does help the most in zero-shot learning setting, aligns with the design of MLM.

AVEQA: Learning to Extract Attribute Value from Product via Question Answering: A Multi-task Approach (Wang et al. 2020)

Method: Multi-Modal

Challenge : Scale up on # of attribute and evolving new attributes

- M-JAVE: Multimodal Joint Attribute Prediction and Value Extraction for E-commerce Product (Zhu et al. 2020)
- PAM: Understanding Product Images in Cross Product Category Attribute Extraction (Lin et al. 2021)

PAM

Challenge 3: The Attribute Value is from multi-source

- Multi-modal learning task that Involves textual, visual and image text features.
- Multi-modal transformer based encoder and decoder
- Multi-task training for attribute value extraction and product category prediction

PAM: Understanding Product Images in Cross Product Category Attribute Extraction (Lin et al. 2021)

PAM

Challenge 3: The Attribute Value is from multi-source

OCR text contains information that textual profile misses

PAM: Understanding Product Images in Cross Product Category Attribute Extraction (Lin et al. 2021)

PAM

Challenge 3: The Attribute Value is from multi-source

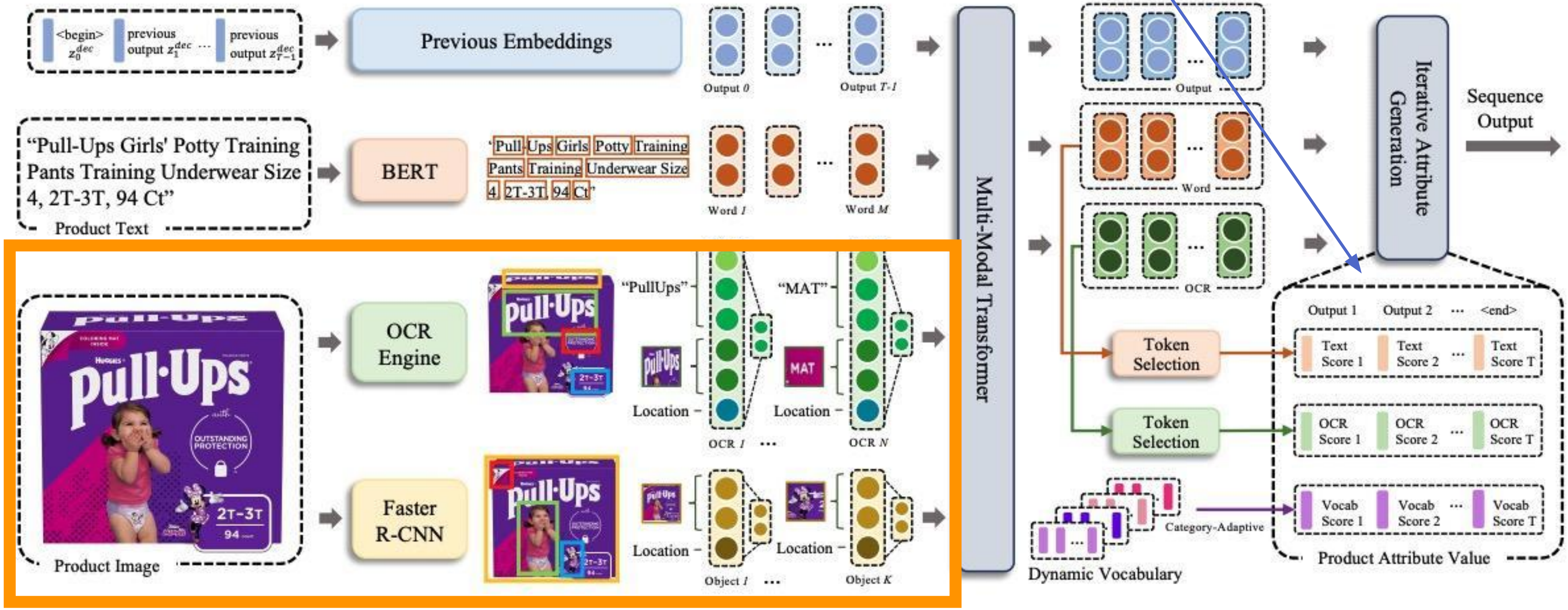
Image features also help identify attribute value

PAM: Understanding Product Images in Cross Product Category Attribute Extraction (Lin et al. 2021)

PAM

Text generation task: text are generated from 3 candidate groups.

Challenge 3: The Attribute Value is from multi-source



PAM: Understanding Product Images in Cross Product Category Attribute Extraction (Lin et al. 2021)

PAM

PAM beats the text baseline model and other multi-modal model since in addition to the multi-modal transform, PAM introduced category type prediction as auxiliary task and also introduced category type based vocabulary.

Challenge 3: The Attribute Value is from multi-source

Attributes	Models	P(%)	R(%)	F1(%)
Item Form	BiLSTM-CRF	90.8	60.2	72.3
	OpenTag	95.5	59.8	73.5
	BUTD	83.3	53.7	65.3
	M4C	89.4	52.6	66.2
	M4C full	90.9	63.4	74.6
	PAM (ours) text-only	94.5	60.1	73.4
	PAM (ours)	91.3	75.3	82.5
Brand	BiLSTM-CRF	81.8	71.0	76.1
	OpenTag	82.3	72.9	77.3
	BUTD	79.7	62.6	70.1
	M4C	72.0	67.8	69.8
	M4C full	83.1	74.5	78.6
	PAM (ours) text-only	81.2	78.4	79.8
	PAM (ours)	86.6	83.5	85.1

Models	P(%)	R(%)	F1(%)
PAM *w/o* text	79.9	63.4	70.7
PAM *w/o* image	88.7	72.1	79.5
PAM *w/o* OCR	82.0	69.4	75.1
PAM	91.3	75.3	82.5

PAM does the best when combining all of the 3 features. Importance ranking the features: Text features > OCR features > image features

PAM: Understanding Product Images in Cross Product Category Attribute Extraction (Lin et al. 2021)

Take Aways

- Modeling attribute value prediction as Sequence Tagging, Question Answering and Text generation task.

- Using the attribute name embedding and product type taxonomy embedding attend to text profile.
  - Improve the performance.
  - Generalizability on few-shot/zero-shot learning.

- Opportunities in combining text, text on image, image feature by utilizing multi-modal transformer to allow interaction between all features.