Knowledge Extraction

Section Structure

- Problem Definition

  What are unique challenges for PG beyond generic KGs?
- Short answer -- key intuition

  What are key intuitions for attribute value extraction?
- Long answer -- details

  What are practical tips?
- Reflection/short-answer

  Can we apply the techniques to other domains?

What is Attribute Value Extraction?

From the eyes of customers

Amazon Brand - Solimo Oil-free Facial Moisturizer for Sensitive Skin, 4 Fluid Ounce, 1 pack

Visit the Solimo Store

2,745 ratings | 13 answered questions

Price: $5.28 ($1.32 / Fl Oz) prime

Earn 5% back on this purchase (worth $0.26 when redeemed) with your Amazon Prime Store Card.

Size: **4 Fl Oz (Pack of 1)**

Item Form	Liquid
Brand	Solimo
Skin Type	Sensitive
Age Range (Description)	Adult
Scent	Unscented

About this item

- One 4-fluid ounce pump bottle of face moisturizer for sensitive skin
- Moisturizes skin without clogging pores.May be used morning and night
- Alcohol free, paraben free
- If you like Neutrogena Oil-Free Moisture, we invite you to try

Backend data storage

Attribute	Attribute Value
Title	Amazon Brand - Solimo Oil-free Facial Moisturizer for Sensitive Skin, 4 Fluid Ounce, 1 pack
Item Form	Liquid
Skin Type	Sensitive
Brand	SOLIMO
Age Range Description	Adult

What is Attribute Value Extraction?

Problem definition

- Given a product **P**, the product category **PC** optionally, and a list of attributes {**A_1, A_2, … A_n**}.
- For each attribute **A_i**, identify a list of attribute values **{V_ij}** of the product.

What is Attribute Value Extraction?

First Aid Beauty Ultra Repair Cream: Vegan and Gluten-Free Intense Moisturizer for Dry Sensitive Skin. Perfect for Skin Conditions and Eczema. Pink Grapefruit (14 ounce)

About this item

- HEAD-TO-TOE: Head-to-toe moisturizer that provides instant relief and long-term hydration for dry, distressed skin, even eczema. The beautiful, whipped texture is instantly absorbed with no greasy after-feel. Grapefruit has a bright citrus fruit scent that is fresh, juicy and sparkling.
- CLINICALLY PROVEN: Formulated with Colloidal Oatmeal, Shea Butter, Ceramide 3 and the FAB Antioxidant Booster, it provides immediate relief and visible improvement for parched skin and it is clinically proven to increase hydration by 169% immediately upon application.

Product description

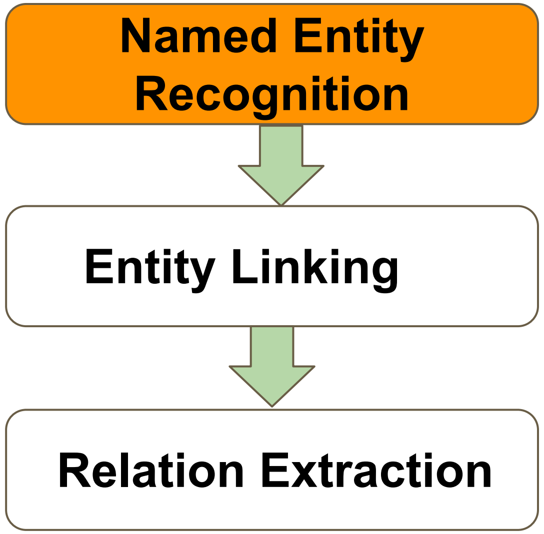
Banish dry skin with First Aid Beauty's Ultra Repair Cream. Suitable for all skin types, especially dry, flaky skin, this hydration wonder leaves skin feeling smooth, hydrated and comfortable after just a single use.

Mentioned Attributes: Brand | SkinType | Scent | Quantity

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

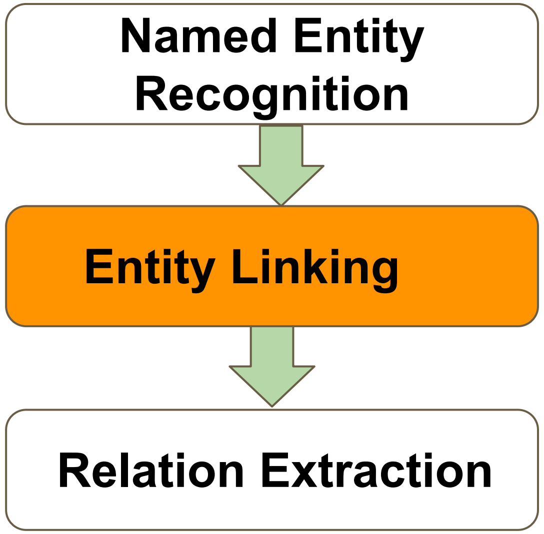
Generic Solution

Extraction output: (subject, predicate, object) triple



Generic Solution

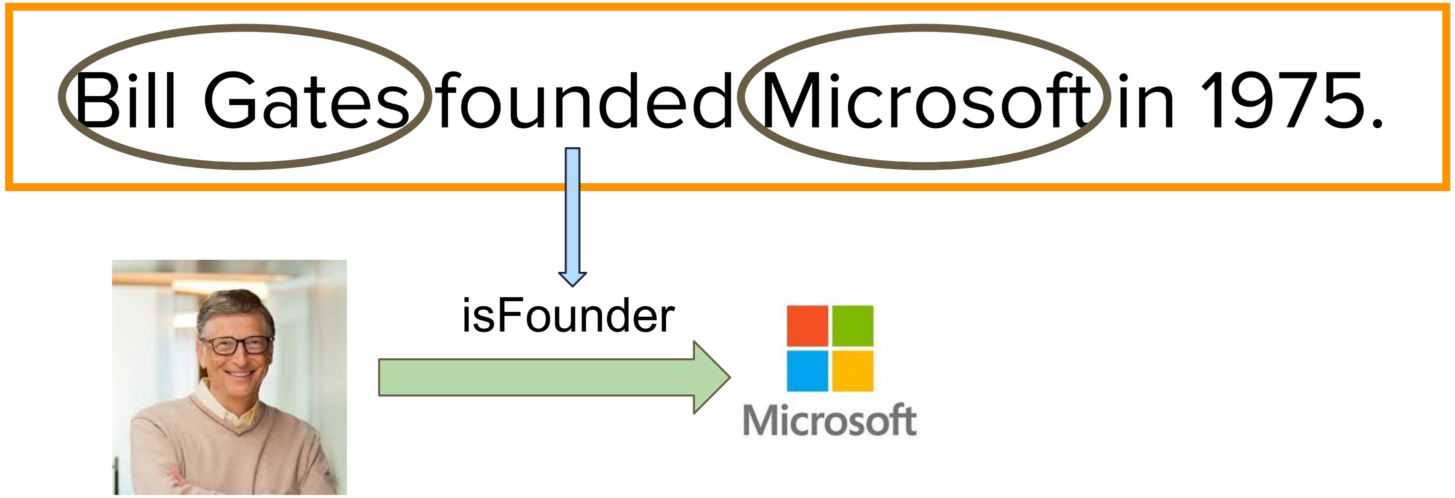
extraction output: (subject, predicate, object) triple



Bill Gates founded Microsoft in 1975.

Generic Solution

Extraction output: (subject, predicate, object) triple



Named Entity Recognition

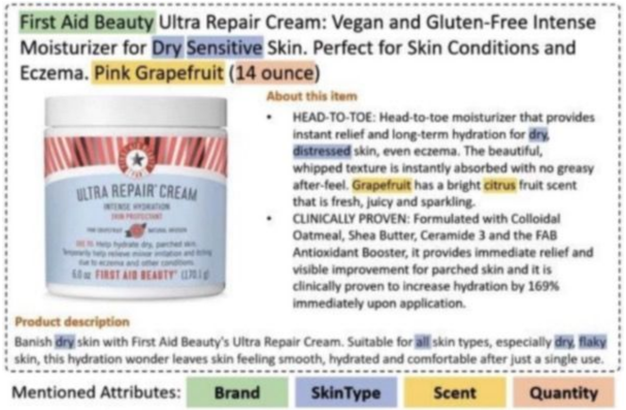
Entity Linking

Relation Extraction

Generic Solution v.s., PG Specific

Bill Gates founded Microsoft in 1975.

First Aid Beauty Ultra Repair Cream: Vegan and Gluten-Free Intense Moisturizer for Dry Sensitive Skin. Perfect for Skin Conditions and Eczema. Pink Grapefruit (14 ounce)



The differences:

- The subject is given.
- The objects are often not entities.

Why is Attribute Value Extraction Hard?

- Diversity of textual semantics:
  - “Orange” can be a flavor, scent, ingredients, color.
  - “Free and clear” in the category of detergent means that it is “scent free”.

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

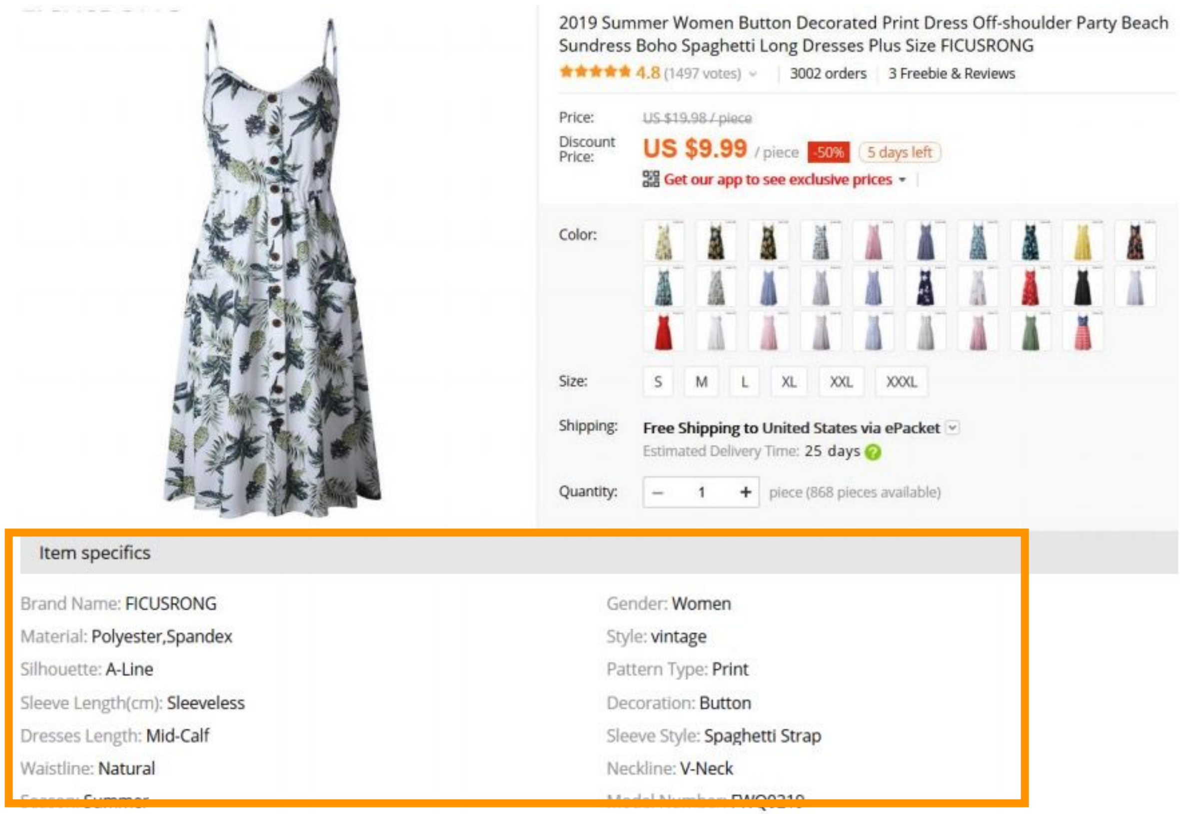
About this item

- 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?

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

From the eyes of customers

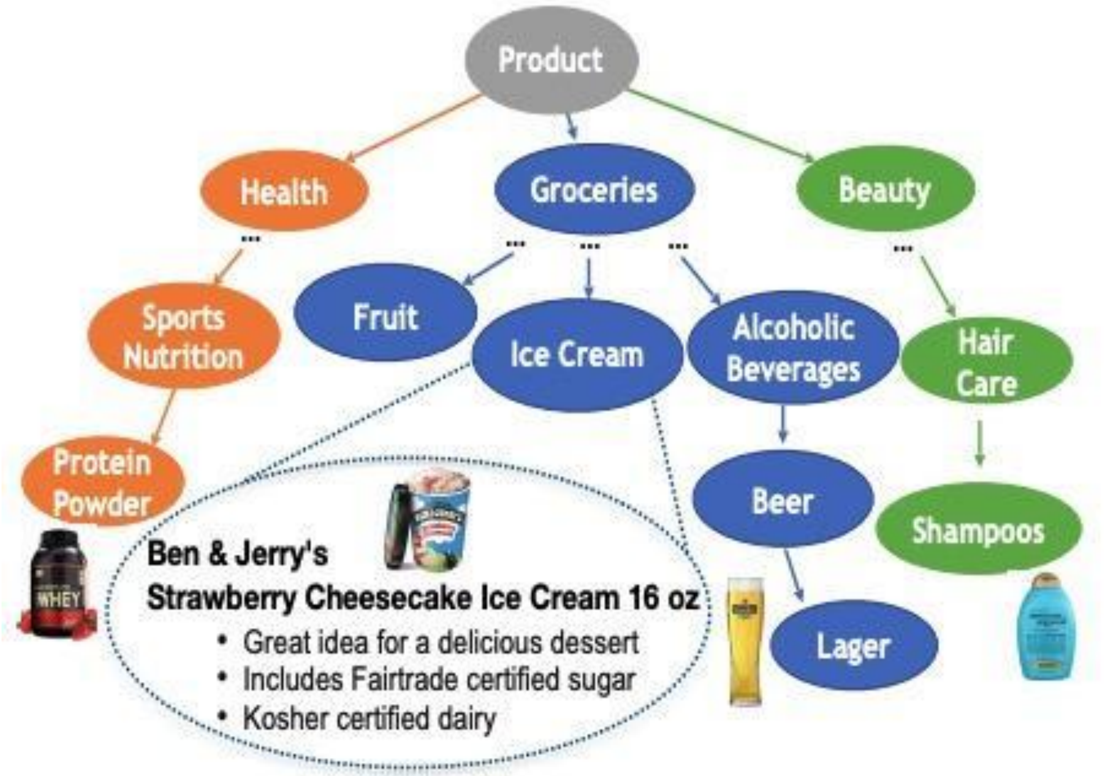


backend data storage

Attribute	Attribute Value
Gender	Women
Neckline	V-Neck
Style	Vintage
…	…

Why is Attribute Value Extraction Hard?

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



Category	Flavor Vocab
Ice Cream	Vanilla, Matcha, Chocolate, Coconut, Strawberry, Banana, Mango, Oreo ...
Beer	Crisp & Clean, Hop - Hoppy & Bitter, Malt & Sweet , Dark & Roasty, Smoke...
Shampoos	Flavor is not applicable

Why is Attribute Value Extraction Hard?

- Lack of training data
    - Neural network based models require much more annotated data because of the large parameter space.
    - Manual annotation is an expensive task.

Short Answer/Solution

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

Short answer/solution

- **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

Short answer/solution

- **Featurizing 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

Short Answer/Solution: Tagging

- "BIOE Tagging"
  - "Beginning"
  - "Inside"
  - "Outside"
  - "End"

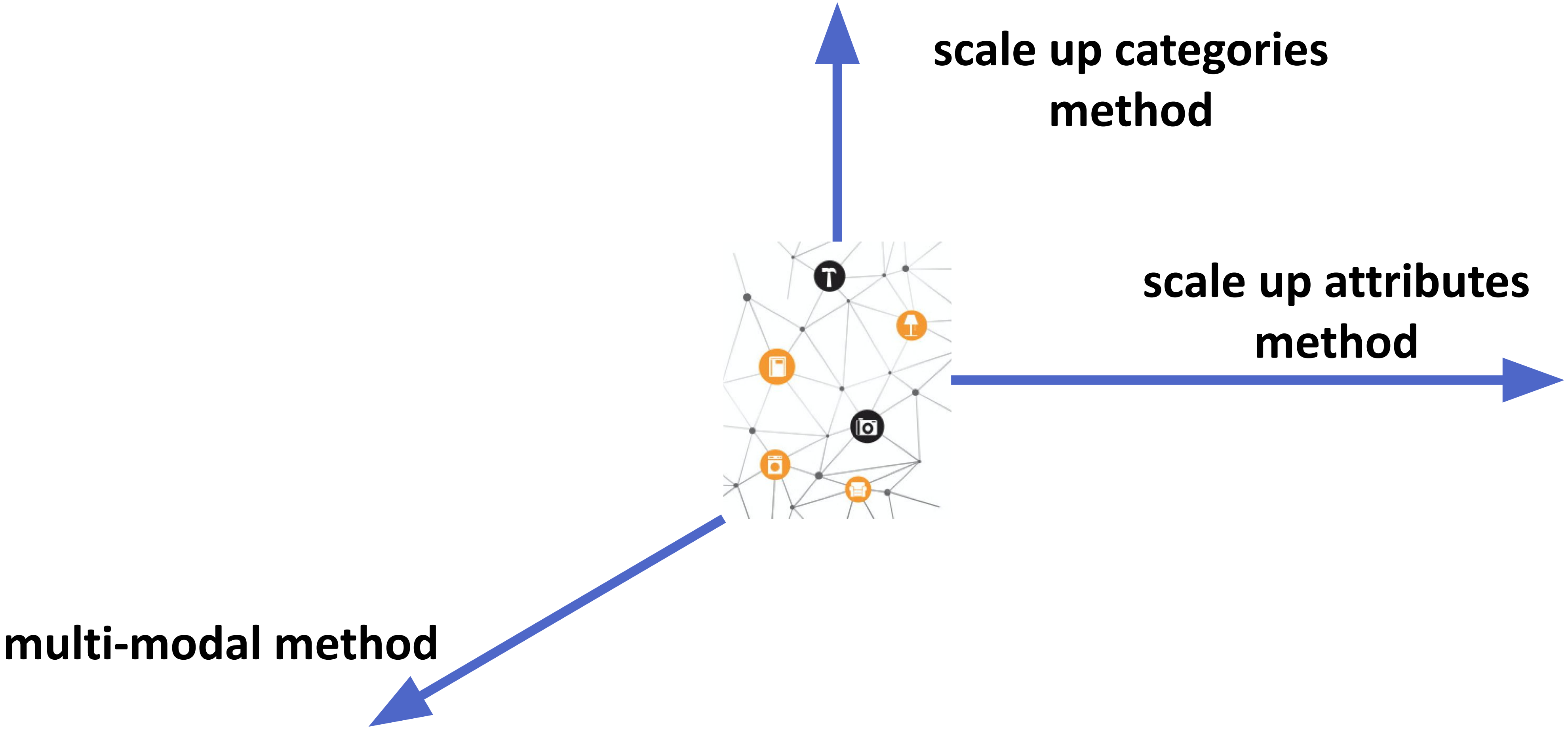
Variety	Pack	Filet	Mignon	and	Ranch	Raised	Lamb	Dog	Food	12	count
O	O	B	E	O	B	I	E	O	O	O	O

Flavor: Filet Mignon

Flavor: Ranch Raised Lamb

Zheng et al., OpenTag: Open Attribute Value Extraction from product profiles, SigKDD, 2018

Short Answer/Solution

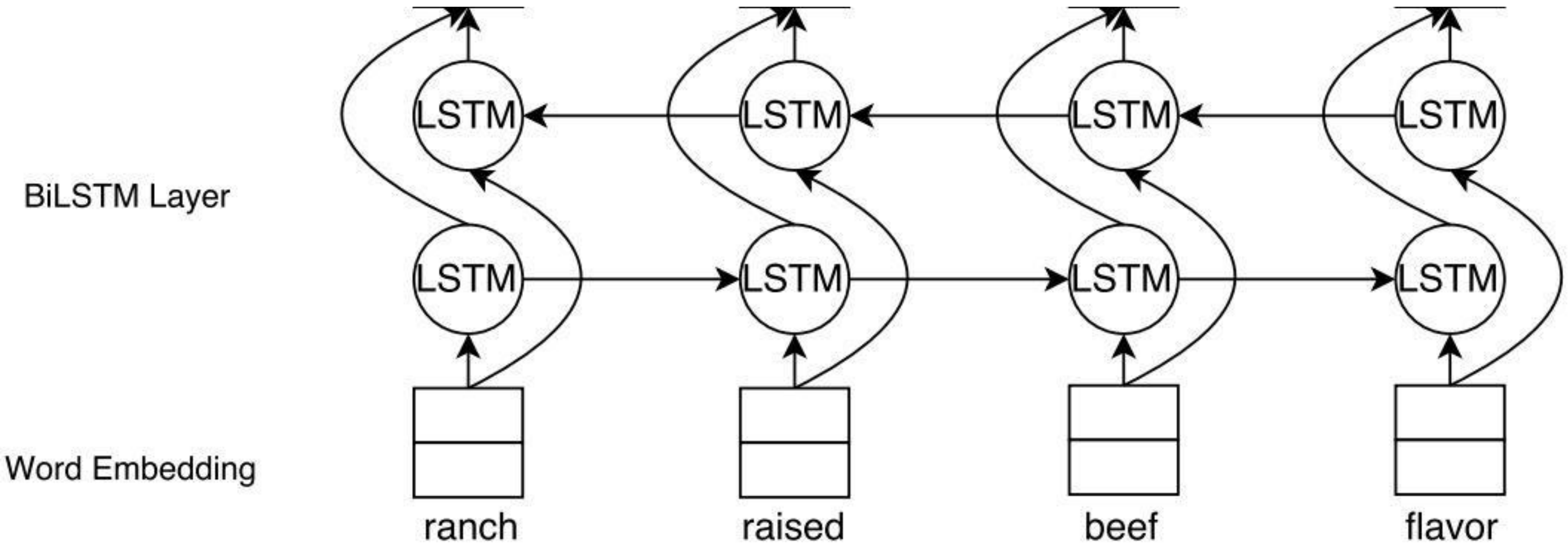


Short Answer/Solution

- Taking attribute name and category as first-class citizen
- Multi-modal extraction
- Semi-supervised learning for training data generation

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-based neural network architecture.
  - Also builds representation of entire sentence.
  - Pre-trained on a large textual corpus.

Long Answer: OpenTag

CRF

Captures correlations between BIOE tags

Attention

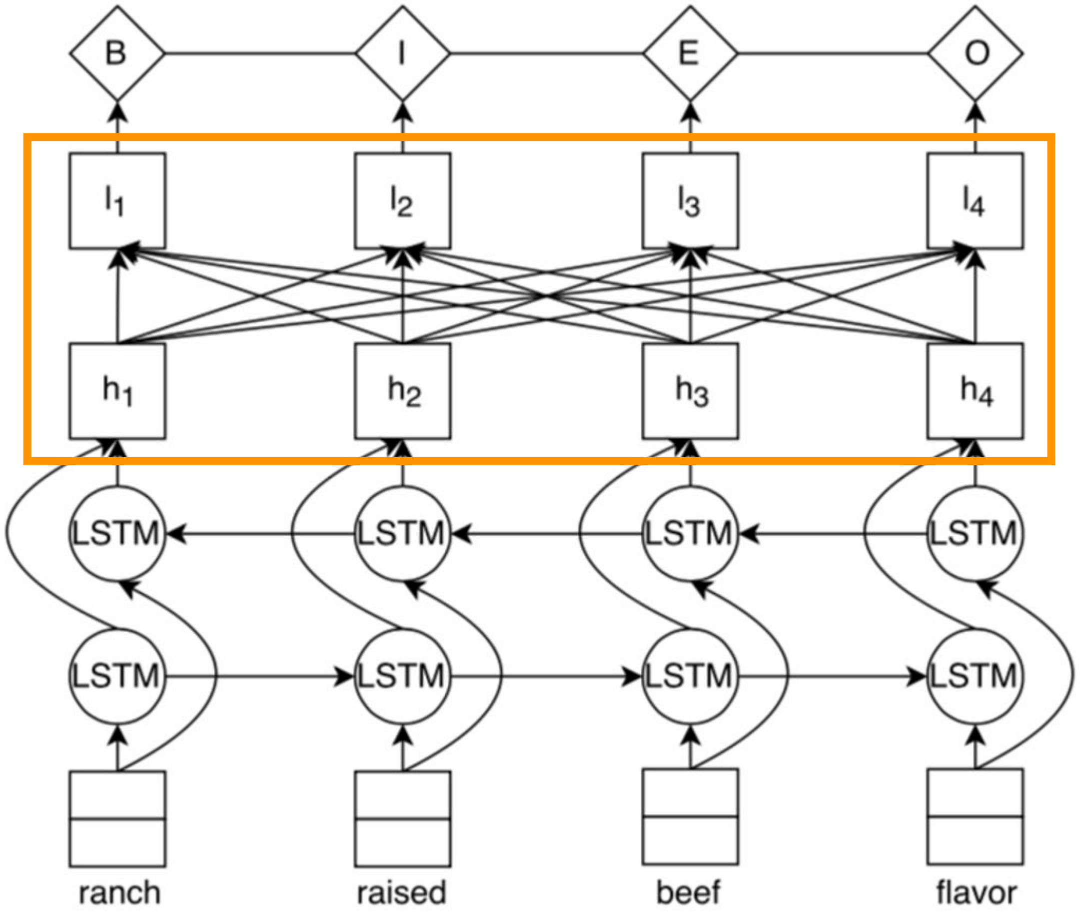
Identifies important terms leading to attribute values

Bi-LSTM

Captures sequence info

Word Embedding

Captures semantics of each token



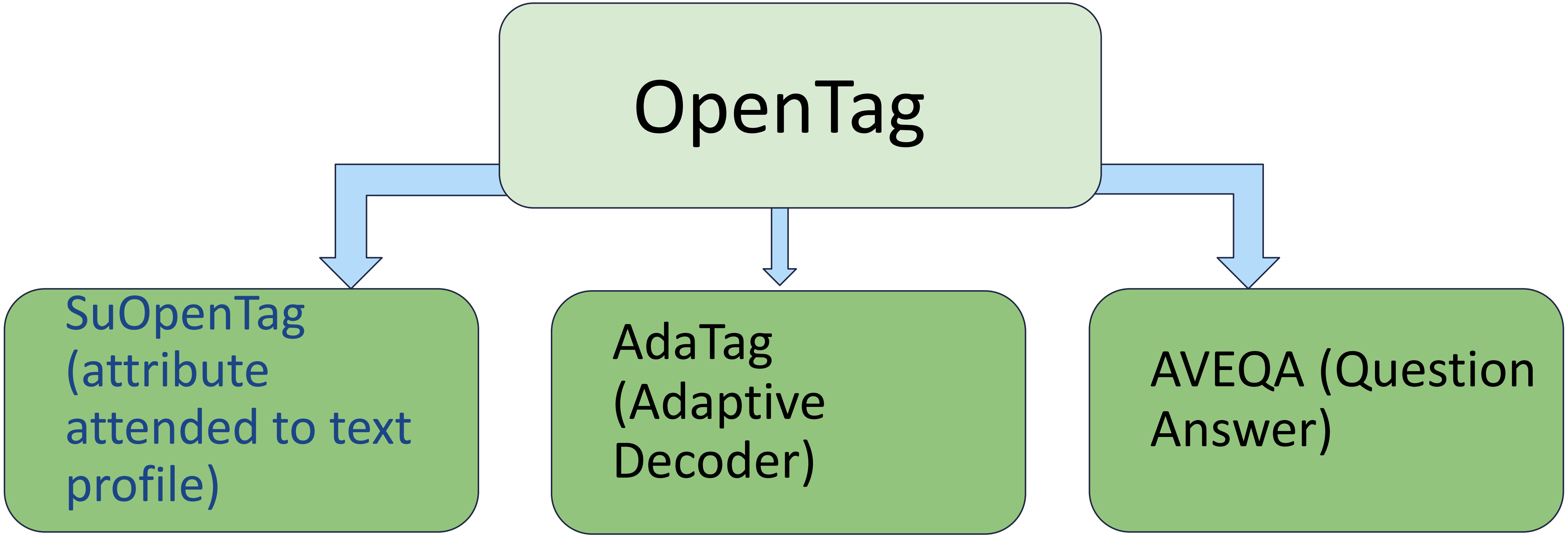
Zheng et al., OpenTag: Open Attribute Value Extraction from product profiles, SigKDD, 2018

Long Answer: OpenTag

Datasets/Attribute	Models	Precision	Recall	Fscore
Dog Food: Title Attribute: Flavor	BiLSTM	83.5	85.4	84.5
	BiLSTM-CRF	83.8	85.0	84.4
	OpenTag	**86.6**	**85.9**	**86.3**
Camera: Title Attribute: Brand	BiLSTM	94.7	88.8	91.8
	BiLSTM-CRF	91.9	**93.8**	92.9
	OpenTag	**94.9**	93.4	**94.1**
Detergent: Title Attribute: Scent	BiLSTM	81.3	82.2	81.7
	BiLSTM-CRF	**85.1**	82.6	83.8
	OpenTag	84.5	**88.2**	**86.4**
Dog Food: Description Attribute: Flavor	BiLSTM	57.3	58.6	58
	BiLSTM-CRF	62.4	51.5	56.9
	OpenTag	**64.2**	**60.2**	**62.2**
Dog Food: Bullet Attribute: Flavor	BiLSTM	93.2	94.2	93.7
	BiLSTM-CRF	94.3	94.6	94.5
	OpenTag	**95.7**	**95.7**	**95.7**

OpenTag improves F1 for all attributes

Long Answer: Scaling Up Attribute Extraction



SUOpenTag

Challenge: Multi-Attributes

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

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

Long Answer: SUOpenTag

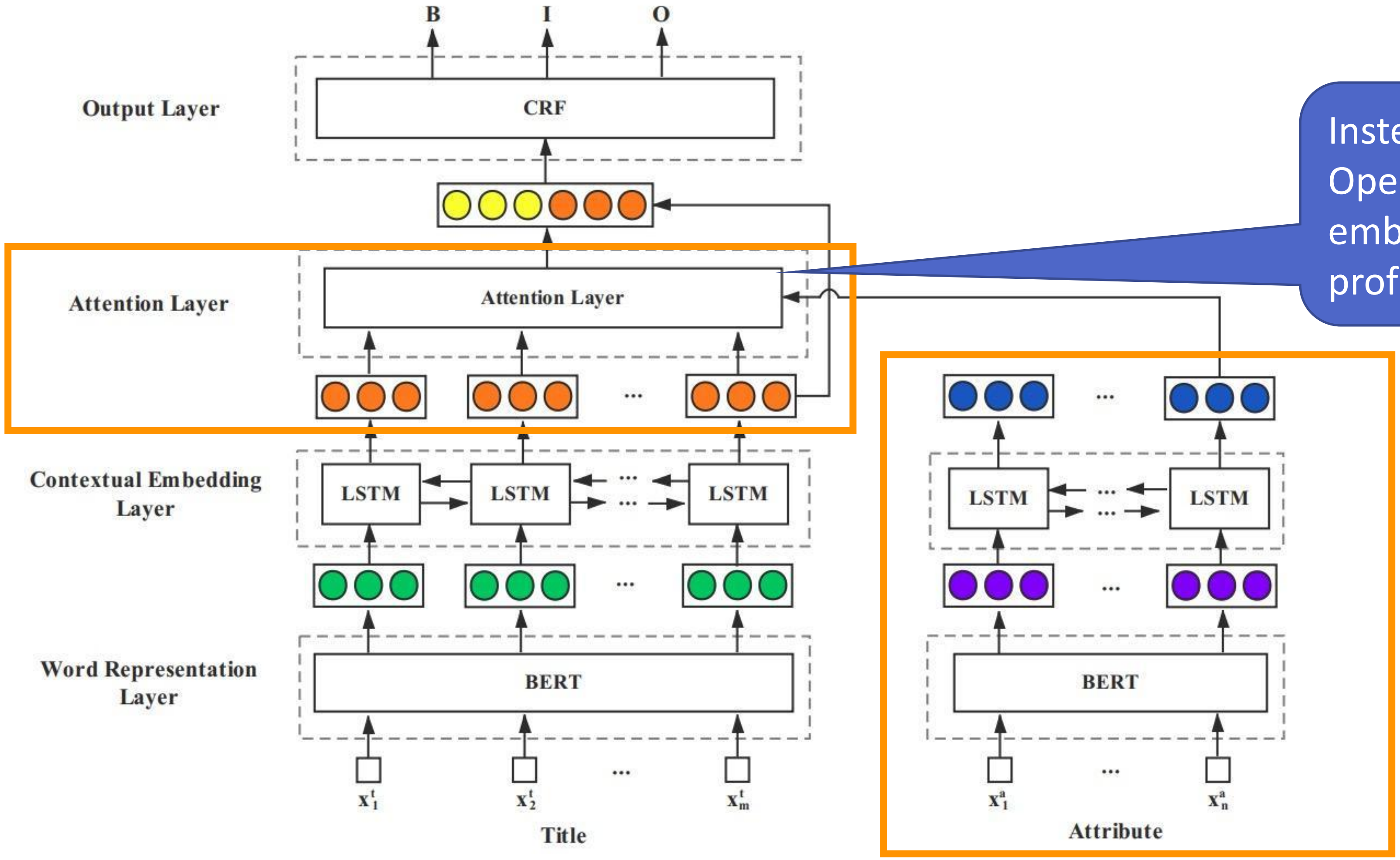
Challenge:
Multi-Attributes

- One model for all attributes (OpenTag used one model for each attribute).
- Attribute name attends to product profile (OpenTag used product profile self-attention).

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

Long Answer: SUOpenTag

Challenge: Multi-Attributes

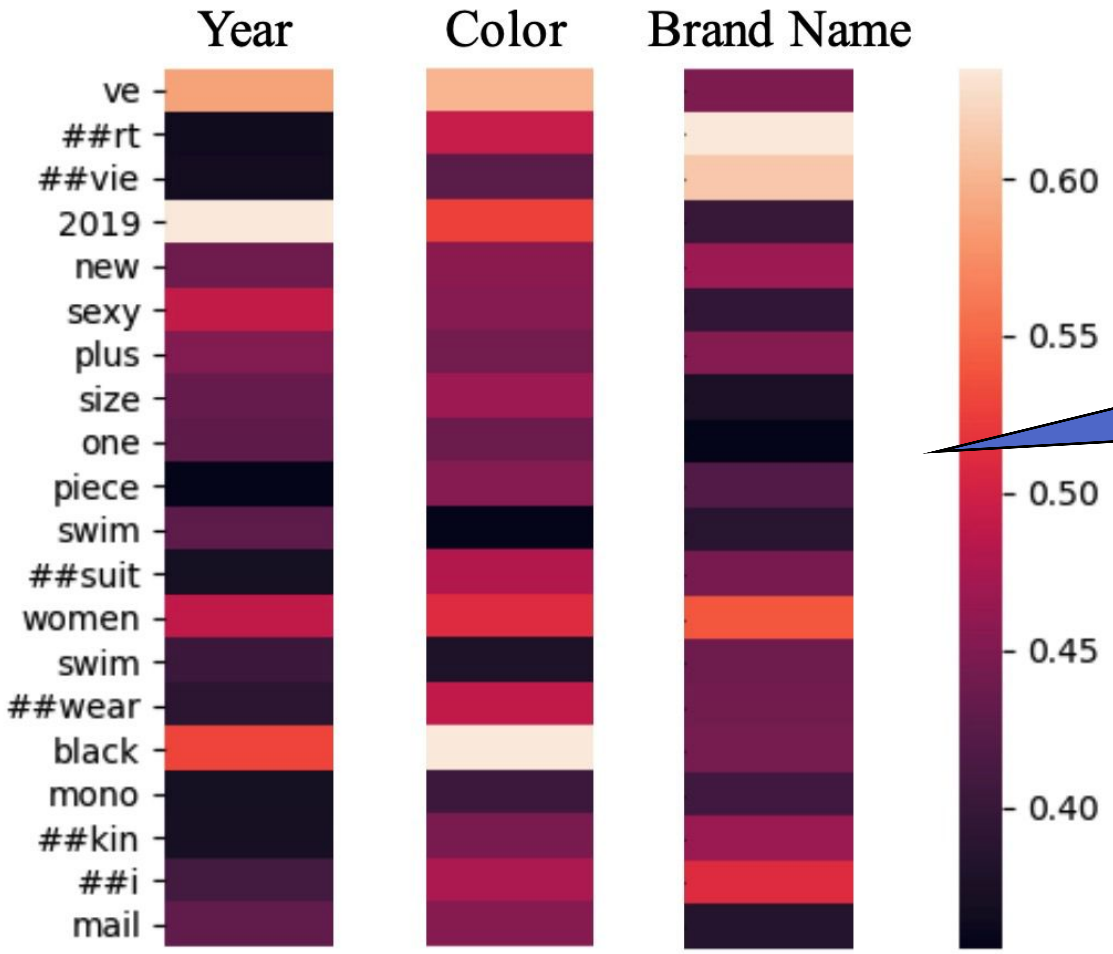


Instead of using self-attention like OpenTag, it uses attribute embedding attended to text profile embedding.

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

Long Answer: SUOpenTag

Challenge: Multi-Attributes



Attribute name has higher attention on the corresponding attribute value

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

Long Answer: SUOpenTag

Challenge:
Multi-Attributes

- Dataset:
  - AE-650k, 650K triples which includes 8906 attributes.
    - Positive: Negative = 4:1
    - Train: Dev: Test = 7:2:1
  - AE-110K, 110K triples which includes the four frequent attributes. (Brand, Material, Color and Category).
    - Designd for fair comparison between SuOpenTag and not #attributes scalable model.

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

Long Answer: SUOpenTag

Challenge:
Multi-Attributes

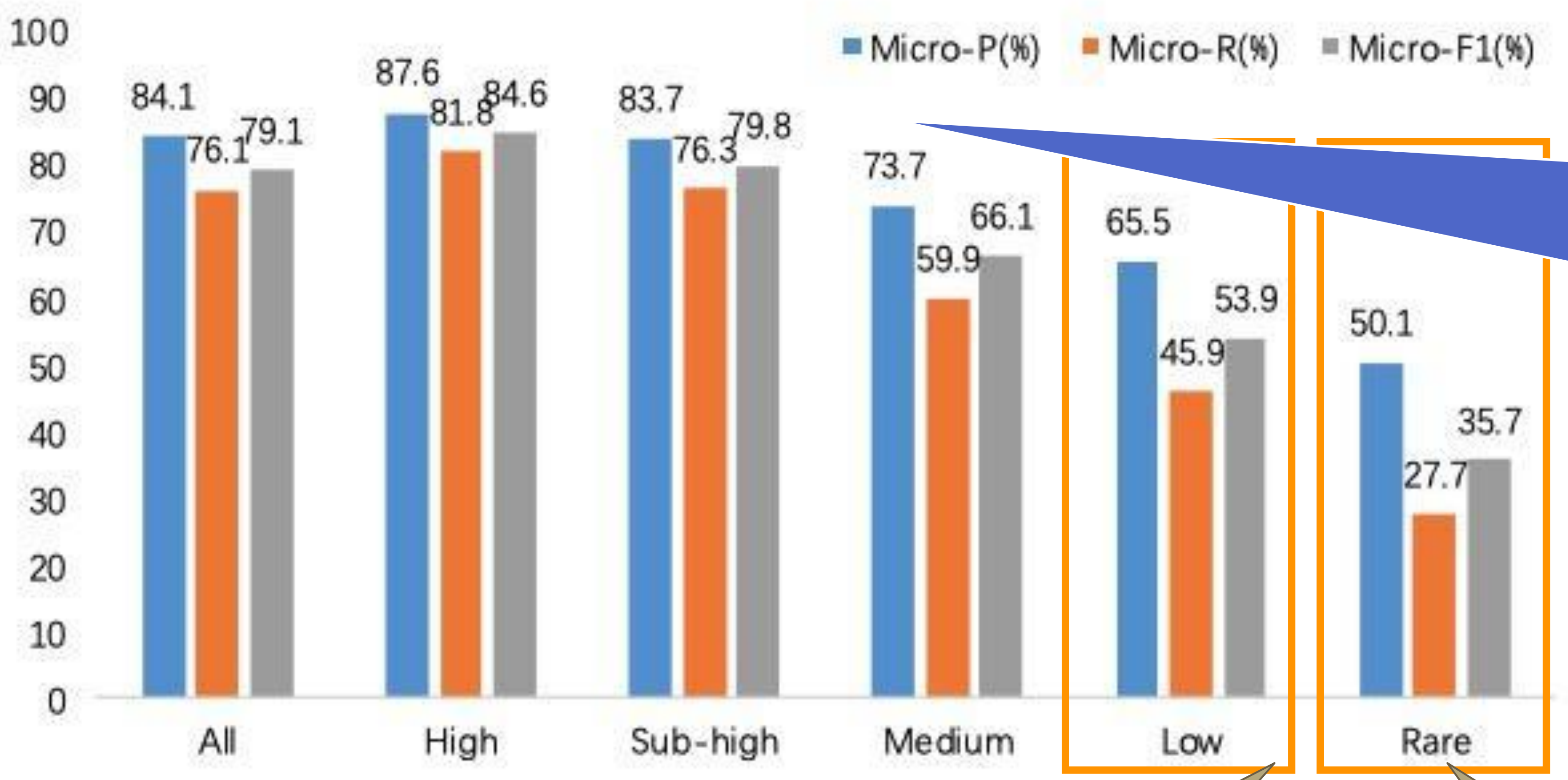
Group the attributes by their occurrences in AE-650k.

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

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

Long Answer: SUOpenTag

Challenge:
Multi-Attributes



The attributes that have more training data get better results

labels (10,100)

labels (1,10)

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

Long Answer: SUOpenTag

Challenge: Multi-Attribute

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**

110K has 4 attributes, 650k has 8906 attributes

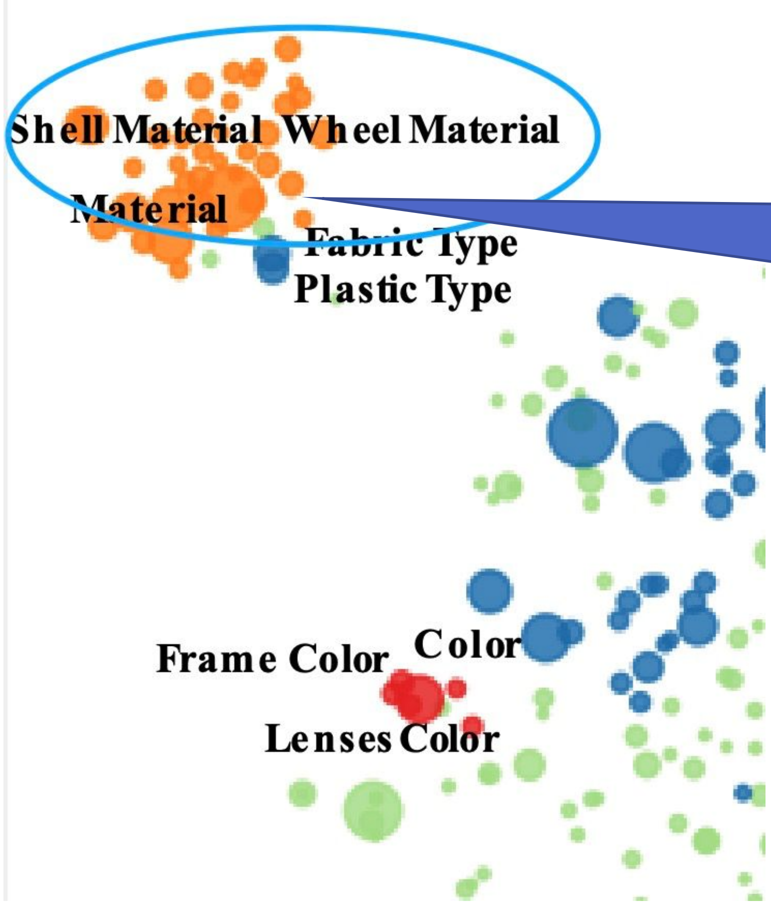
- Both precision and recall goes up.
- More attributes in training data gives better P/R.

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

SUOpenTag: Discover New Attributes

Challenge:
Multi-Attribute

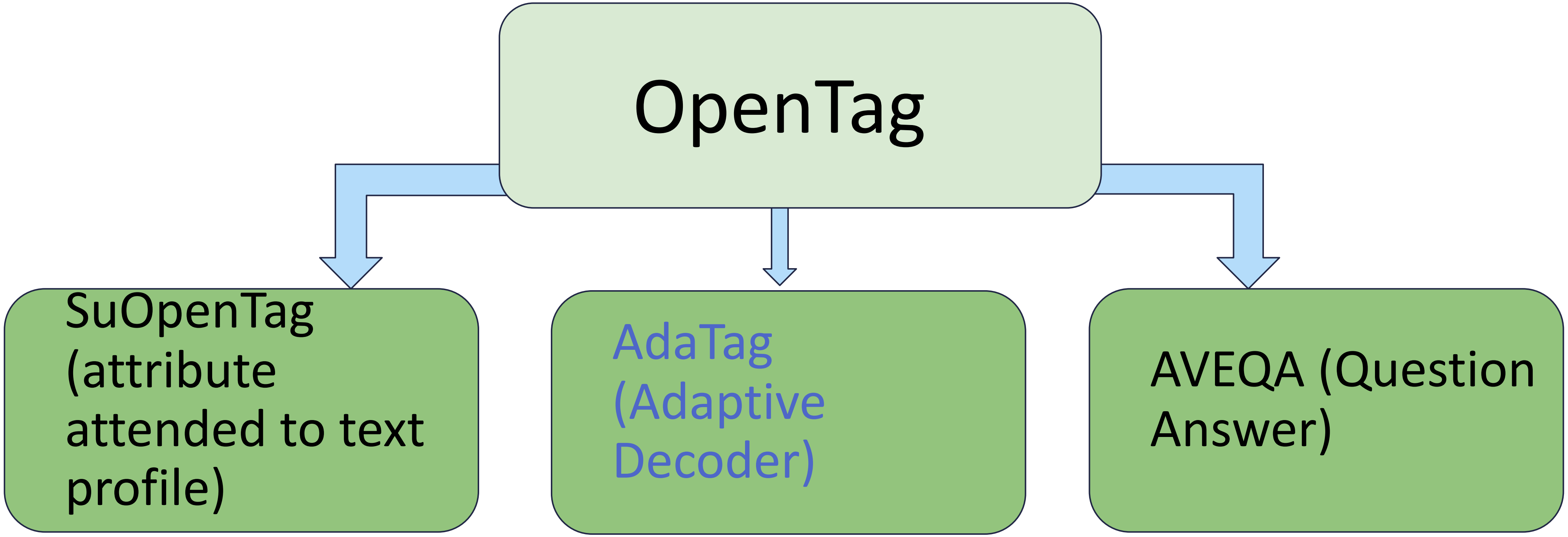
projecting attribute name embedding by t-sne. Material attribute are semantically related to unseen attributes and they provide hints to help the extraction



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

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

Long Answer: Attribute Scaling Up



Long Answer: AdaTag

Challenge: Multi-Attribute

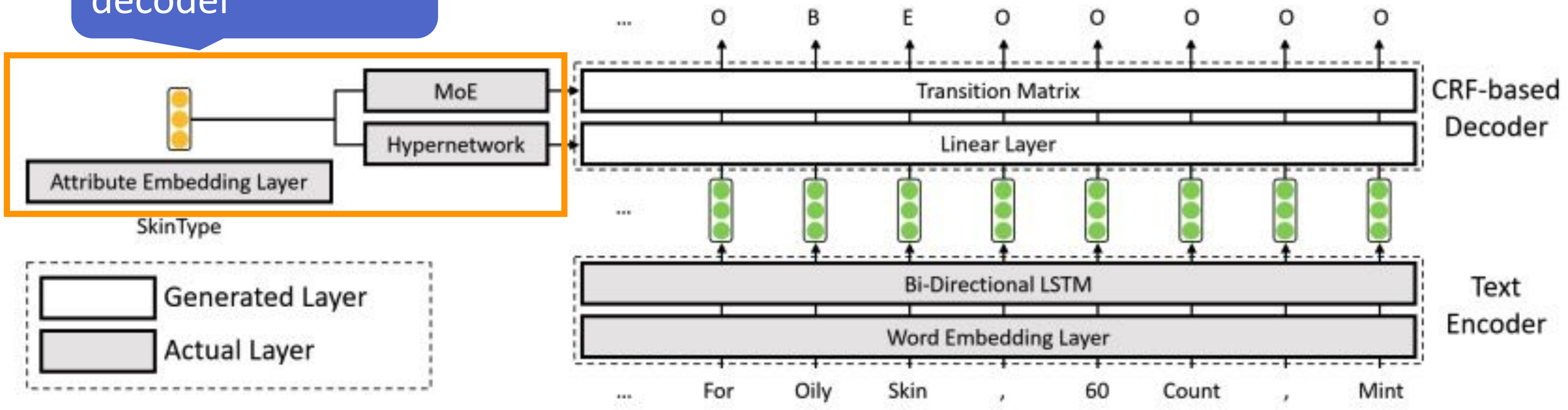
- One model handles all attributes.
- The encoder is shared across all attributes.
  - The representation can be enhanced through learning with different subtasks.
- The decoder is parameterized with pretrained attribute embeddings.
  - Separate, but semantically correlated, decoders to be generated on the fly.

Yan et al. AdaTag: Multi-Attribute Value Extraction from Product Profiles with Adaptive Decoding. ACL, 2021

Long Answer: AdaTag

Challenge: Multi-Attribute

Attribute-Aware decoder



Yan et al. AdaTag: Multi-Attribute Value Extraction from Product Profiles with Adaptive Decoding. ACL, 2021

Long Answer: AdaTag

Challenge:
Multi-Attribute

>=1000 training instances

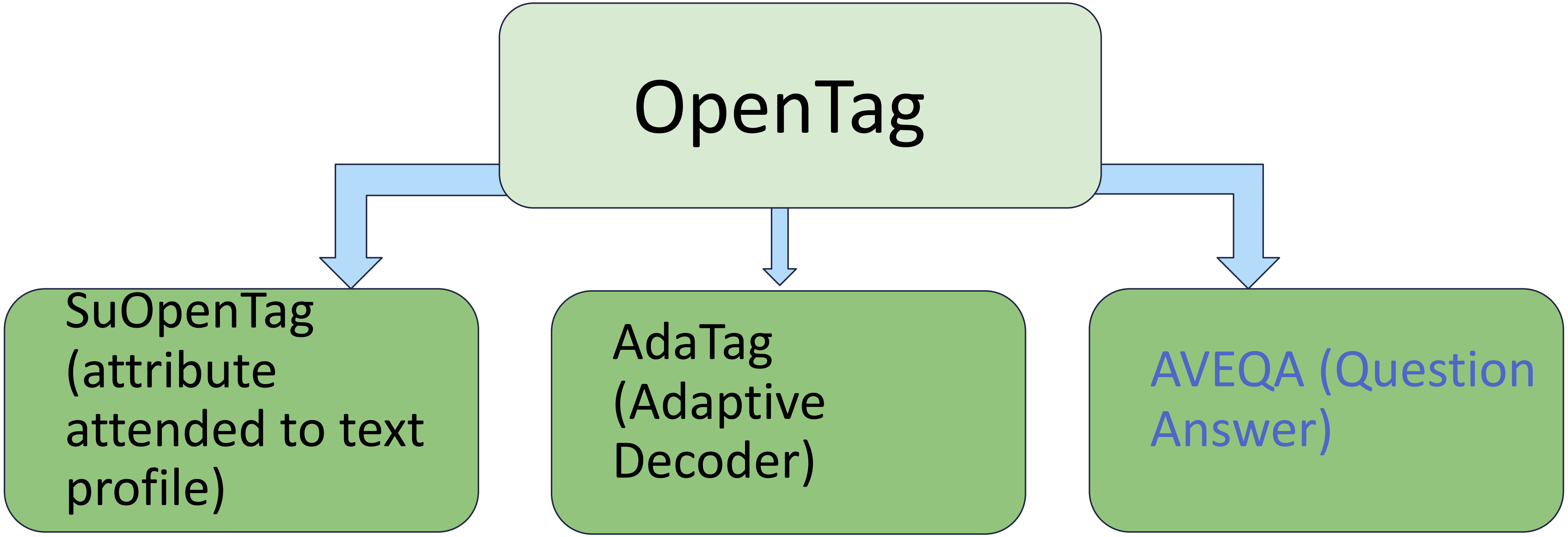
<1000 training instances

Methods	High-Resource Att.			Low-Resource Att.		
	P(%)	R(%)	F_1(%)	P(%)	R(%)	F_1(%)
BiLSTM-CRF (N models)	54.04	75.66	61.57	83.72	65.08	71.19
BiLSTM-MultiCRF	54.38	74.42	60.23	**84.70**	67.29	73.97
SUOpenTag	55.34	72.94	60.49	80.16	69.13	73.31
AdaTag (Our Model)	**56.05**	**76.07**	**62.00**	82.90	**75.48**	**78.45**

High-Resource/Low-Resource AdaTag beats SUOpenTag

Yan et al. AdaTag: Multi-Attribute Value Extraction from Product Profiles with Adaptive Decoding. ACL, 2021

Long Answer: Attribute Scaling Up



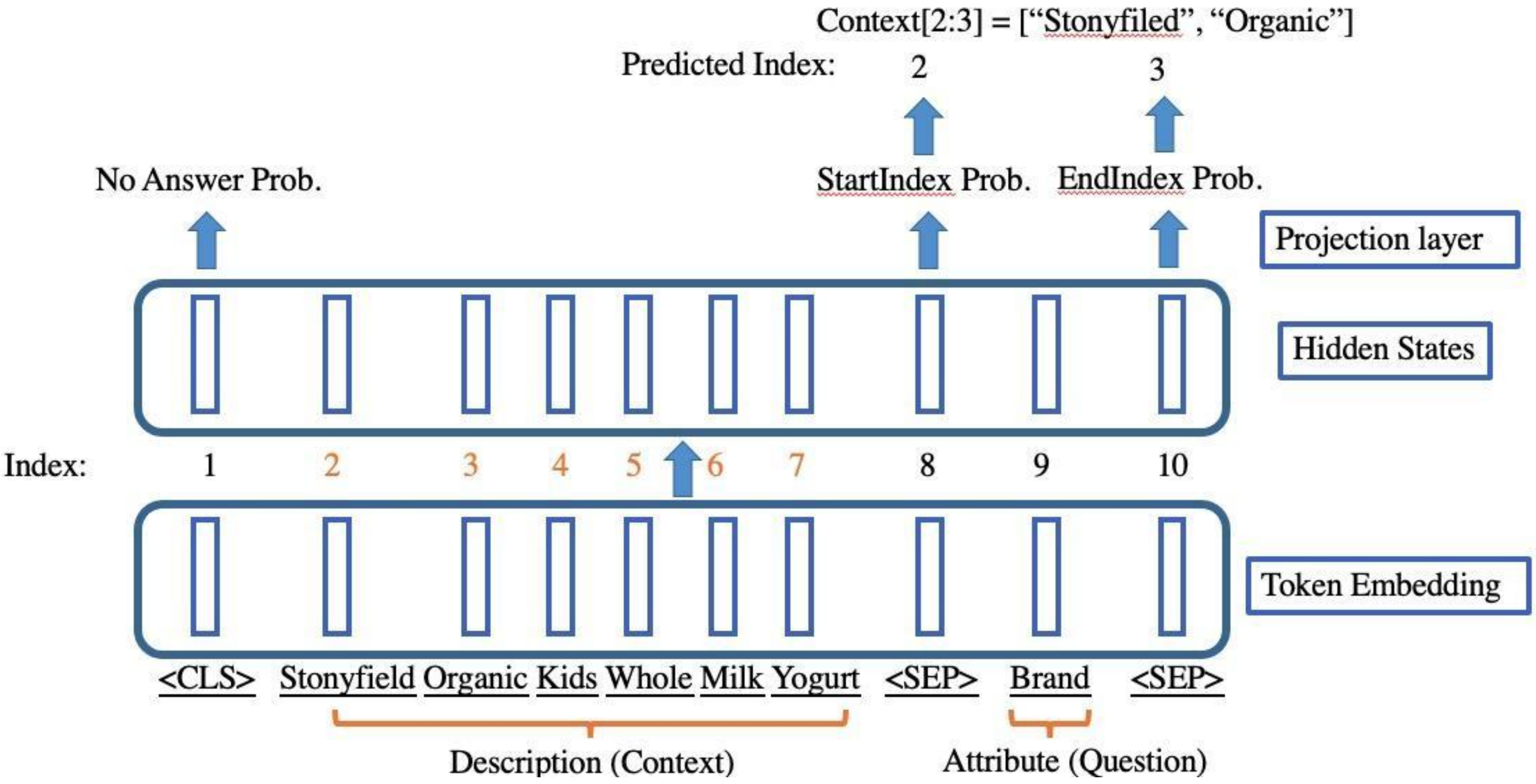
Long Answer: AVEQA

Challenge : Multi-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.

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

Method: Question Answering



Long Answer: AVEQA

Challenge : Multi-Attributes

<CLS> Stonyfield Organic Kids Whole Milk Yogurt <SEP> Brand <SEP>

context: product profile

question: attribute name

<CLS>	Stonyfield	Organic	Kids	Whole	Milk	Yogurt	<SEP>	Brand	<SEP>
1	2	3	4	5	6	7	8	9	10

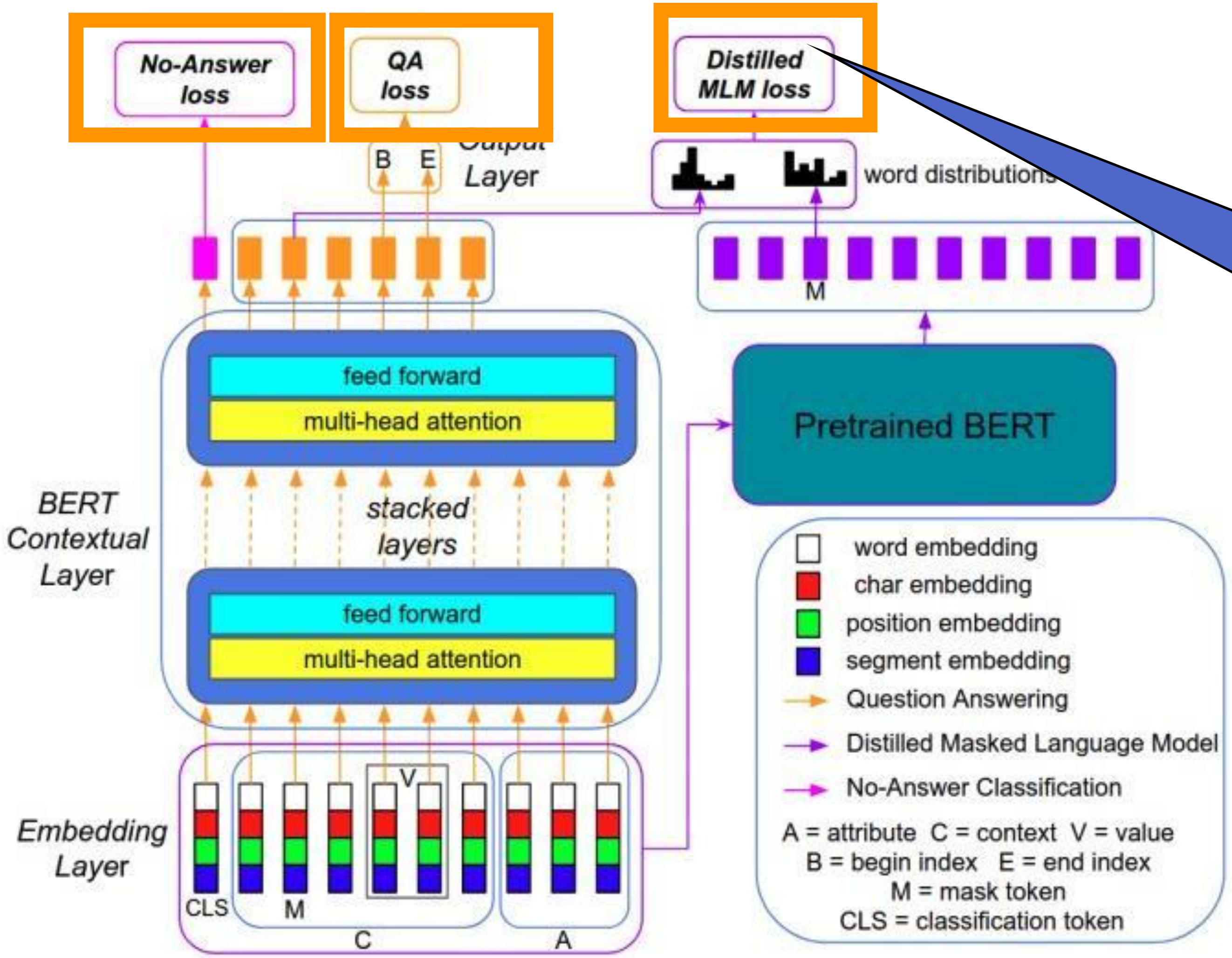
begin index: 2
end index: 2

Brand: Stonyfield

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

Long Answer: AVEQA

Challenge : Multi-Attributes

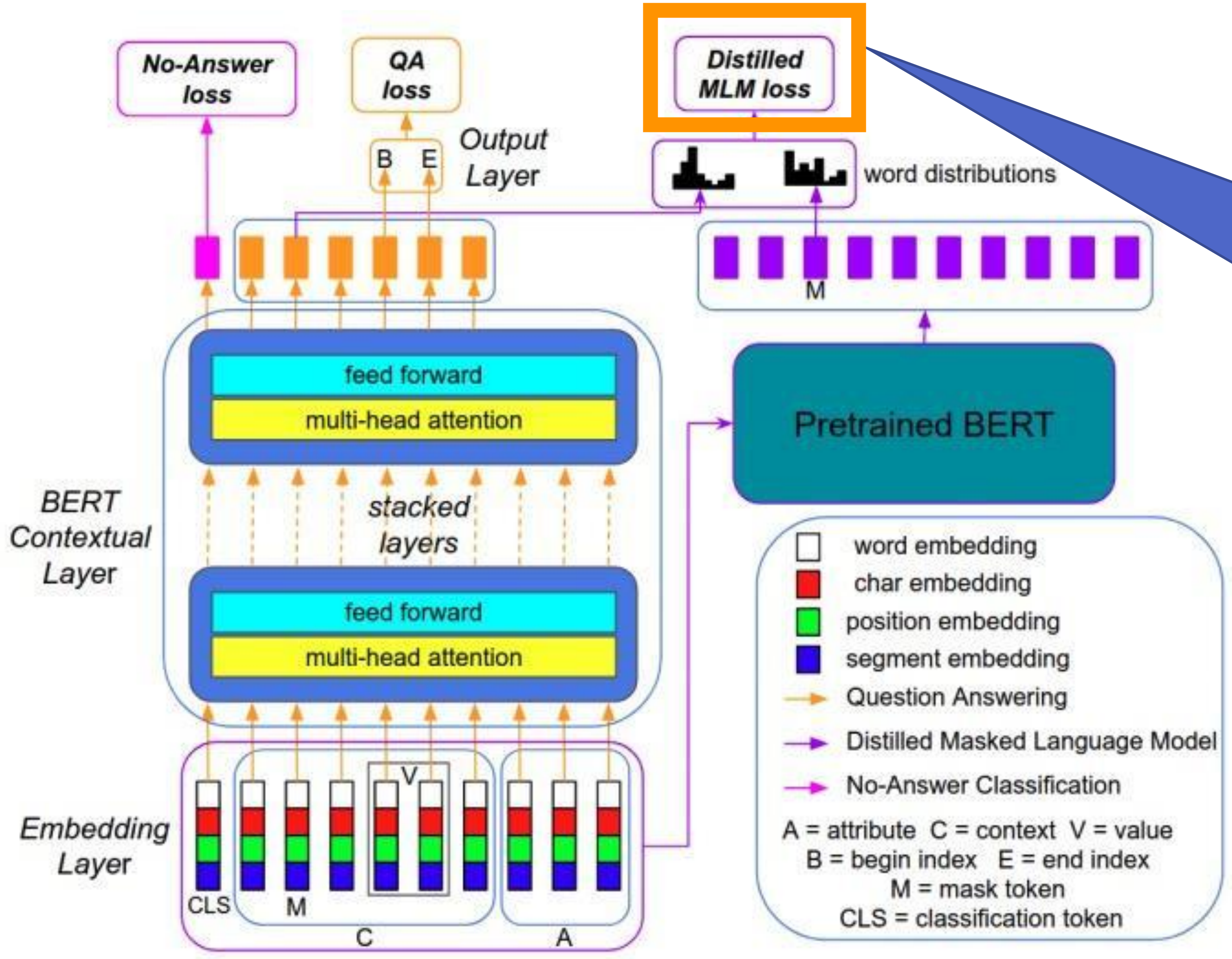


Multi-Task: combines QA loss, No-Answer loss, Distilled Masked Language Model loss

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

Long Answer: AVEQA

Challenge : Multi-Attributes



Distilled Masked Language Model ensures that the encoder learns effective contextual representations for new attributes

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

Long Answer: AVEQA

Challenge : Multi-Attributes

AVEQA further beats the SUOpenTag on precision and recall for the **frequently 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**

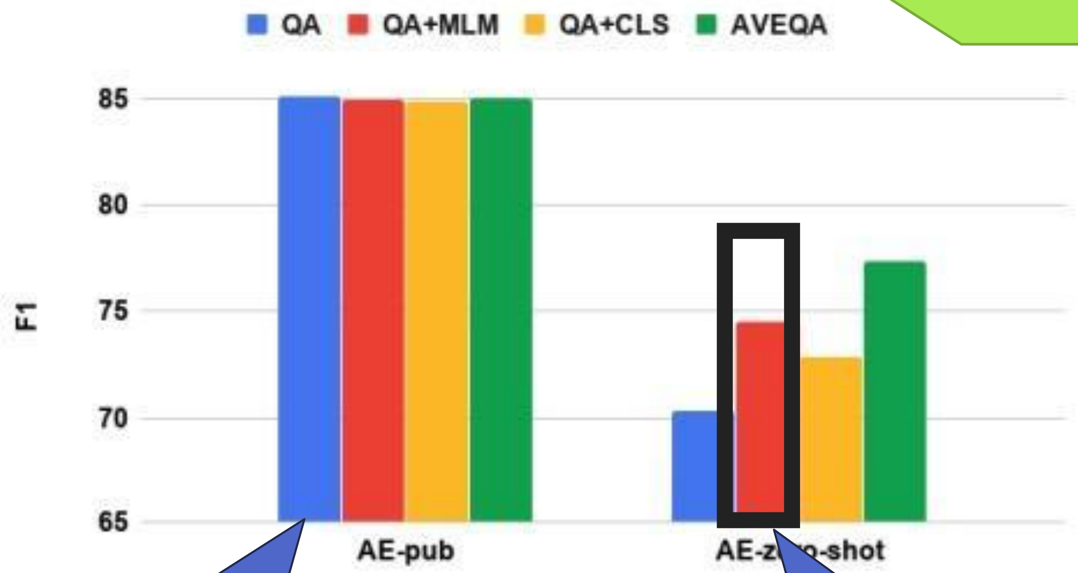
Wang et al., AVEQA: Learning to Extract Attribute Value from Product via Question Answering: A Multi-task Approach, SIGKDD, 2020

Long Answer: AVEQA

Challenge : Multi-Attributes

AVEQA beats the SUOpenTag on precision and recall for the **zero-shot** 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**

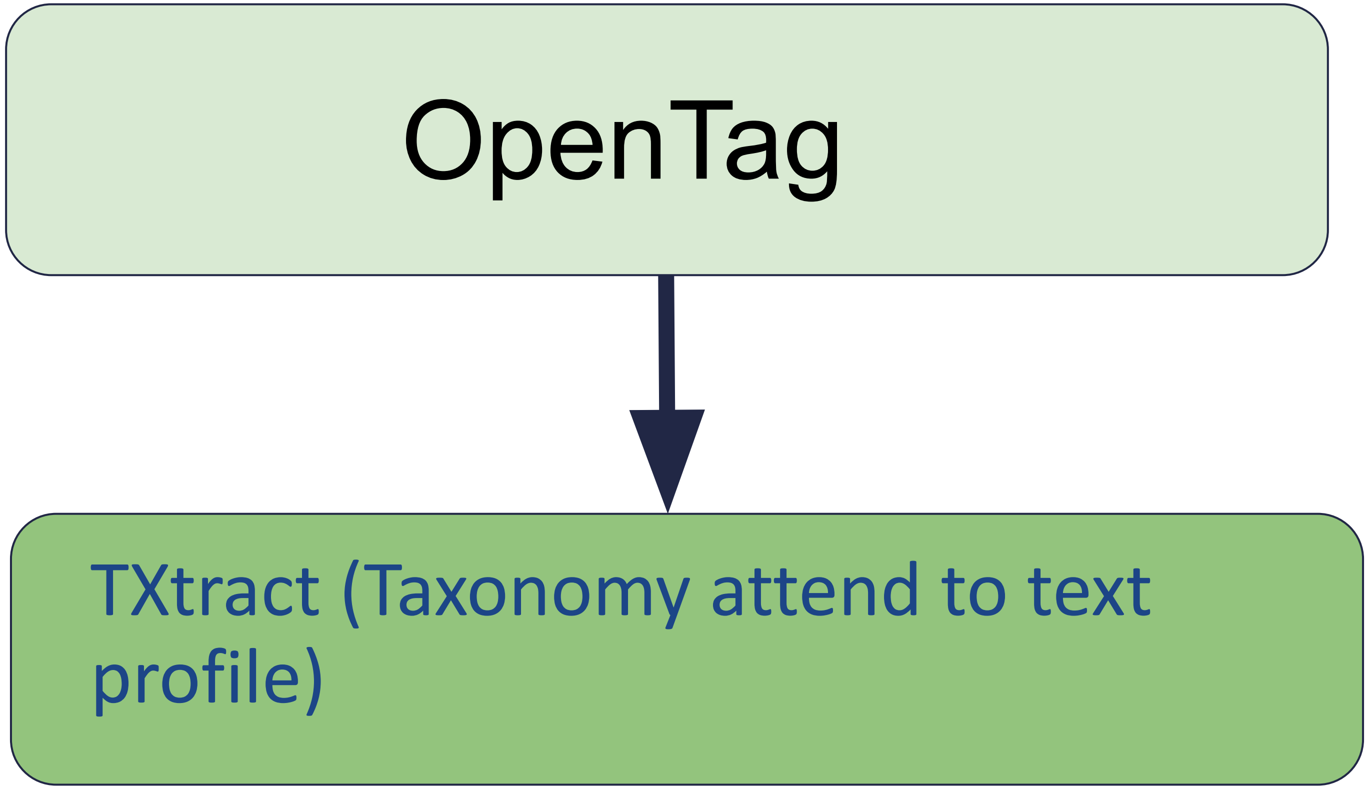


Beats SUOpenTag whose f1: 79.1%

Masked Language Modeling helps the most in zero-shot setting

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

Long Answer: Categories Scaling Up



Long Answer: TXtract

Challenge: Multi-Categories

- Capturing the hierarchical relations between categories into category embeddings.
- Scaling up the extraction on category by generating category embedding attended to product profile embeddings.
- Multi-task learning: attribute value extraction + product type prediction.

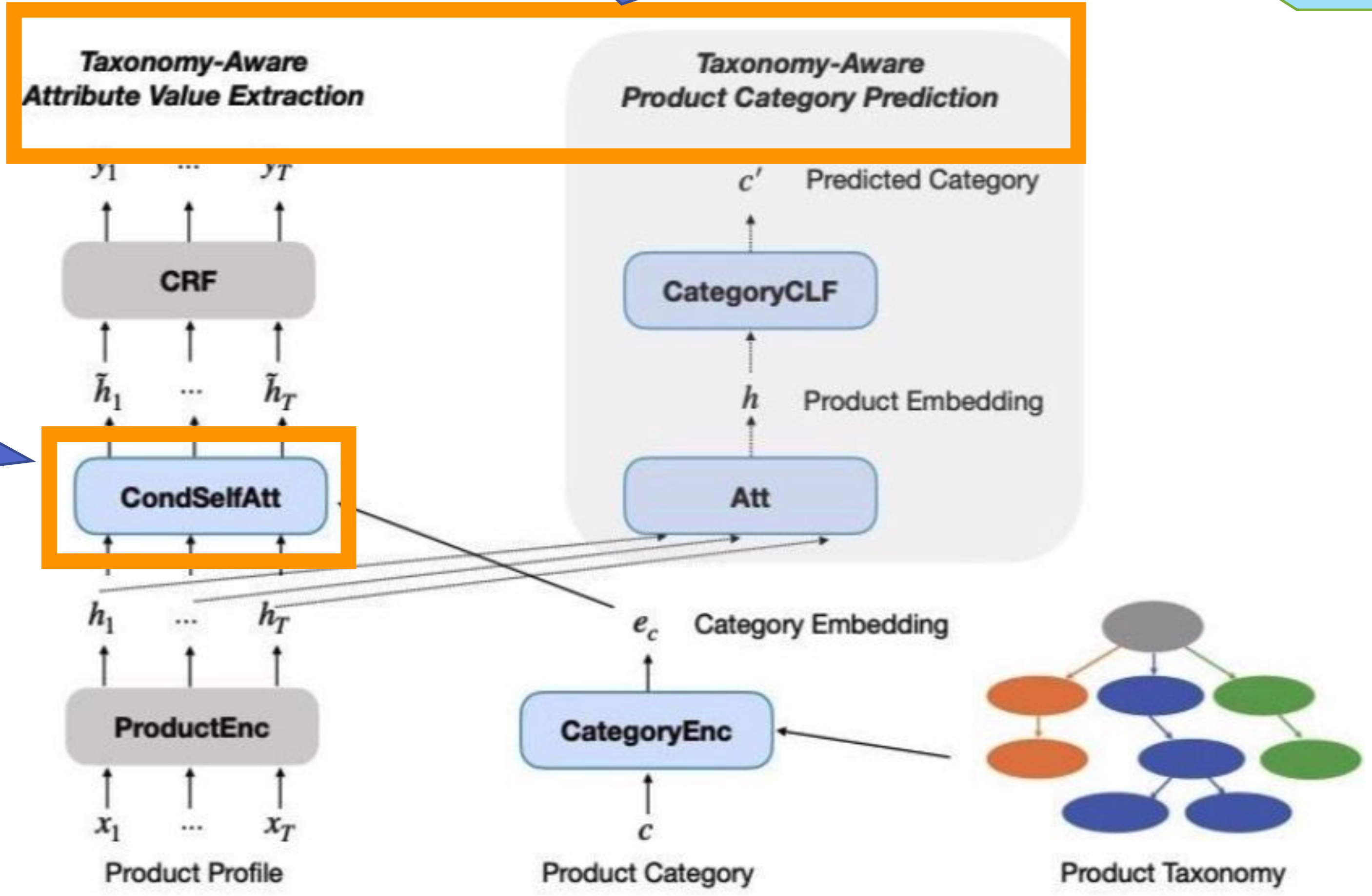
Karamanolakis et al.,Taxonomy-Aware Knowledge Extraction for Thousands of Product Categories, ACL, 2020

Long Answer: TXtract

Challenge: Multi-Categories

Multi-task learning

Category embeddings attend to the product profile embedding



Karamanolakis et al.,Taxonomy-Aware Knowledge Extraction for Thousands of Product Categories, ACL, 2020

Long Answer: TXtract

Challenge: Multi-Categories

Across 4 attributes and 4000 categories, TXtract improves F1 by 6.2%

Attr.	Model	Vocab	Cov	Micro F1	Micro Prec	Micro Rec
Flavor	OpenTag	6,756	73.2	57.5	70.3	49.6
	TXtract	**13,093**	**83.9** ↑14.6%	**63.3** ↑10.1%	**70.9** ↑0.9%	**57.8** ↑16.5%
Scent	OpenTag	10,525	75.8	70.6	**87.6**	60.2
	TXtract	**13,525**	**83.2** ↑9.8%	**73.7** ↑4.4%	86.1 ↓1.7%	**65.7** ↑9.1%
Brand	OpenTag	48,943	73.1	63.4	81.6	51.9
	TXtract	**64,704**	**82.9** ↑13.4%	**67.5** ↑6.5%	**82.7** ↑1.3%	**56.5** ↑8.1%
Ingred.	OpenTag	9,910	70.0	35.7	46.6	29.1
	TXtract	**18,980**	**76.4** ↑9.1%	**37.1** ↑3.9%	**48.3** ↑3.6%	**30.1** ↑3.3%
Average relative increase			↑11.7%	↑6.2%	↑1.0%	↑9.3%

Karamanolakis et al.,Taxonomy-Aware Knowledge Extraction for Thousands of Product Categories, ACL, 2020

Long Answer: TXtract

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 multi-task learning

Challenge: Multi-Categories

Poincare embedded category attended to product profile achieves the best performance

Product category prediction as an auxiliary task further improves the performance

Karamanolakis et al.,Taxonomy-Aware Knowledge Extraction for Thousands of Product Categories, ACL, 2020

Opportunity: Images

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

Long Answer: Multi-Modality

OpenTag

Text features

M-JAVE: (Multimodal Joint Attribute Prediction and Value Extraction for E-commerce Product)

Text features + image features

PAM

Text features + image features + image OCR text features

Long Answer: PAM

Multi-Modality

- Multi-modal learning that involves textual, visual and image text features.
- Multi-modal transformer-based encoder and decoder.
- Formulate attribute value extraction task as a text generation task.

Lin et al.,PAM: Understanding Product Images in Cross Product Category Attribute Extraction, SIGKDD, 2021

Long Answer: PAM

Multi-Modality

OCR text contains information that textual profiles may miss

Lin et al.,PAM: Understanding Product Imagess in Cross Product Category Attribute Extraction, SIGKDD, 2021

Long Answer: PAM

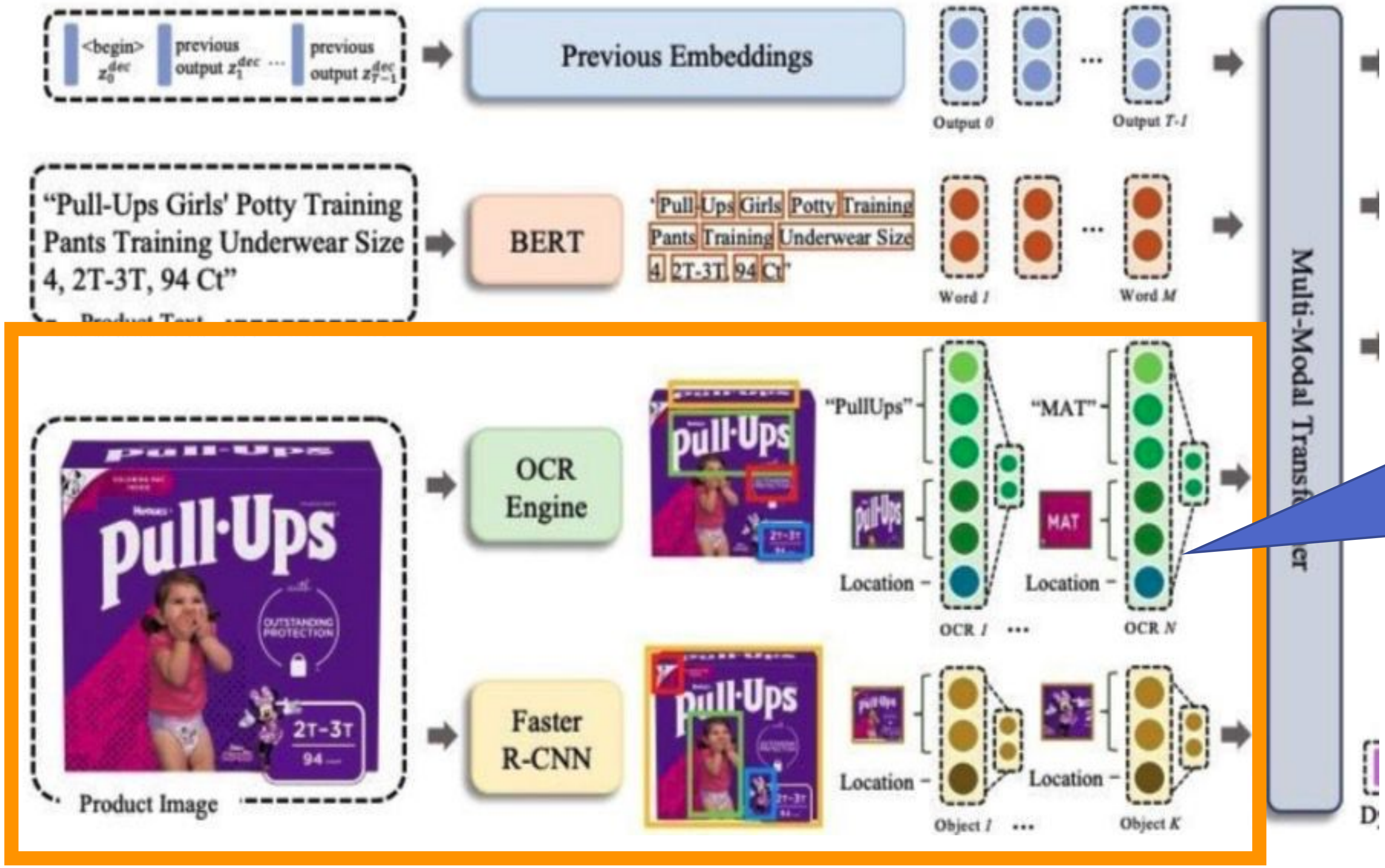
Multi-Modality

Powder or **Stick**? Image features also help identify attribute value

Lin et al.,PAM: Understanding Product Imagess in Cross Product Category Attribute Extraction, SIGKDD, 2021

Long Answer: PAM

Multi-Modality

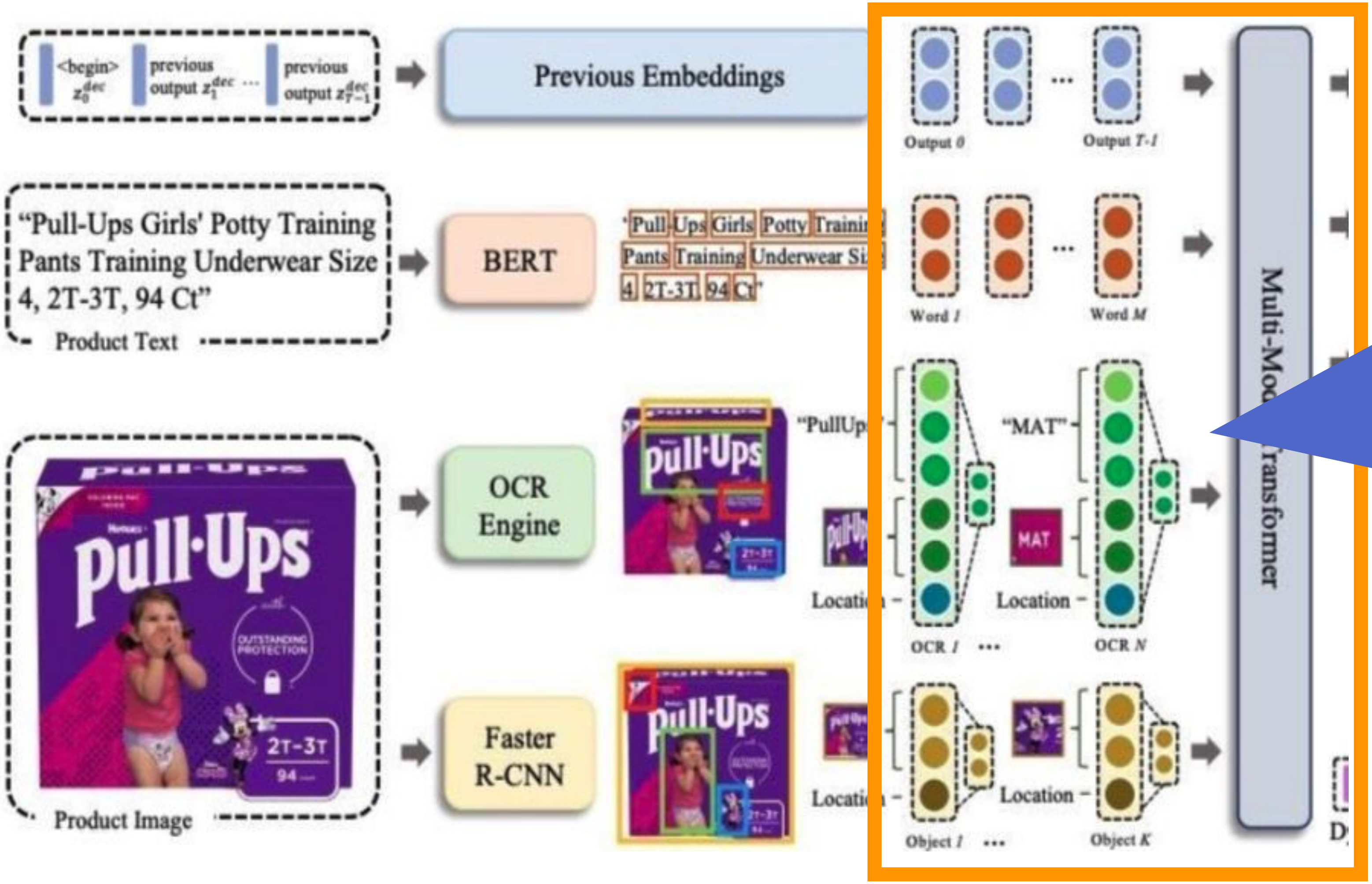


Two kinds of image information: OCR texts from image, image features

Lin et al.,PAM: Understanding Product Imagess in Cross Product Category Attribute Extraction, SIGKDD, 2021

Long Answer: PAM

Multi-Modality



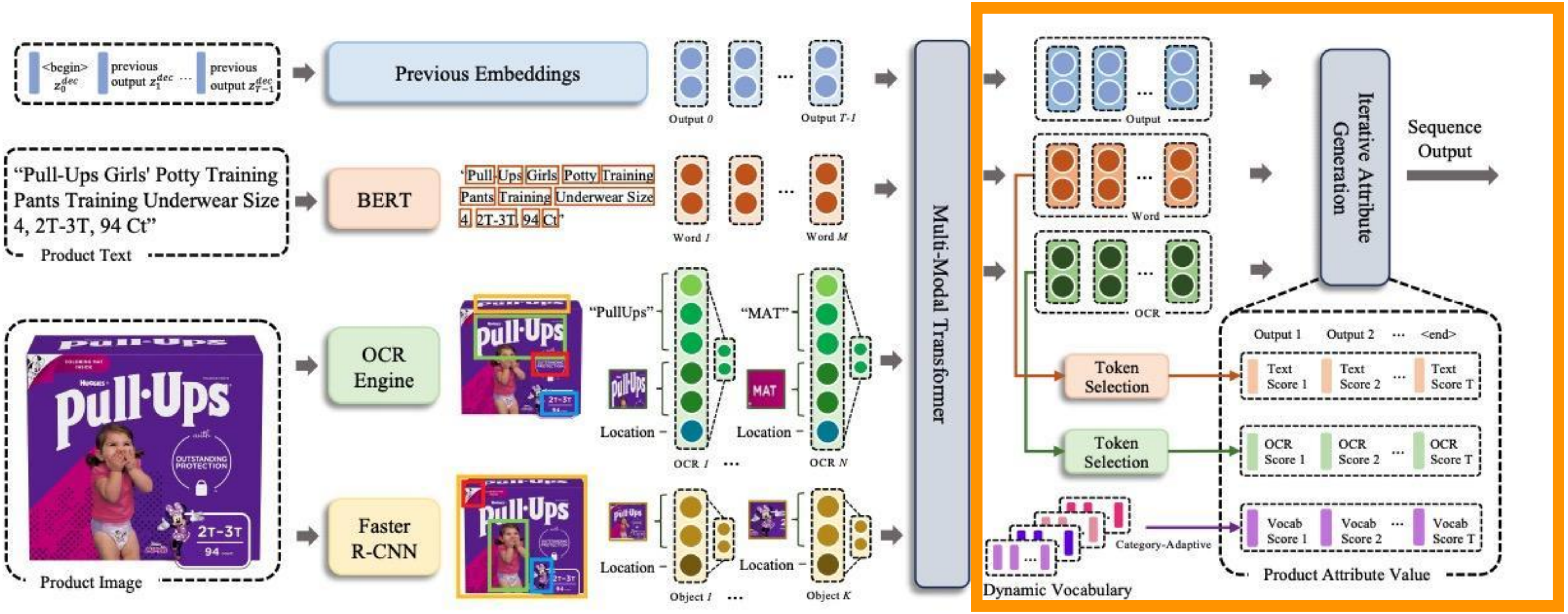
All modalities are concatenated as a single sequence of embeddings.
Each modality is free to attend to each other

Lin et al.,PAM: Understanding Product Imagess in Cross Product Category Attribute Extraction, SIGKDD, 2021

Long Answer: PAM

Multi-Modality

Text generation task has 3 token candidate sources. The decoder generates product type first



Lin et al.,PAM: Understanding Product Imagess in Cross Product Category Attribute Extraction, SIGKDD, 2021

Long Answer: PAM

Multi-Modality

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

Since PAM introduced category type prediction as auxiliary task and also introduced category type based vocabulary

Lin et al.,PAM: Understanding Product Imagess in Cross Product Category Attribute Extraction, SIGKDD, 2021

Long Answer: PAM

Multi-Modality

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 improves both precision and recall compared to text-only model

The ranking of feature importance: text features > OCR features > image features for P/R

Lin et al.,PAM: Understanding Product Imagess in Cross Product Category Attribute Extraction, SIGKDD, 2021

Short Answer/Solution

- Taking attribute name and category as first-class citizen.
- Multi-modal extraction.
- Semi-supervised learning for training data generation.

Long Answer: Data Programming

Challenge: Lack of Training Data

- Often may have multiple sources of weak supervision.
  - Distant supervision from a Knowledge Base.
  - Heuristics / regular expressions.
  - Noisy crowd-labeled data.
  - Manually defined constraints.
  - Extractions from an existing (and imperfect) IE system.
- How can we most effectively learn from noisy data from different sources?

Ratner at.el, Data Programming: Creating Large Training Sets, Quickly. NeurIPS, 2016

Challenge: Lack of Training Data

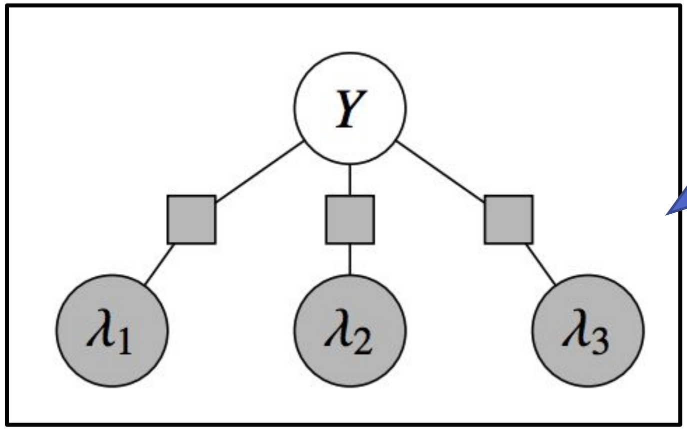
Long Answer: Data Programming

- Generative model to “de-noise” training data.
- Learns which labeling functions are best for which data points.

```
def lambda_1(x):
    return 1 if (x.gene,x.pheno) in KNOWN_RELATIONS_1 else 0

def lambda_2(x):
    return -1 if re.match(r'.*not␣cause.*', x.text_between) else 0

def lambda_3(x):
    return 1 if re.match(r'.*associated.*', x.text_between)
              and (x.gene,x.pheno) in KNOWN_RELATIONS_2 else 0
```



Denoise by re-weight the label functions for each data point

Ratner, et.al, Snorkel: Rapid Training Data Creation with Weak Supervision. VLDB, 2017

Long Answer: Rapid Training Data Creation with Weak Supervision (Snorkel)

Challenge: Lack of Training Data

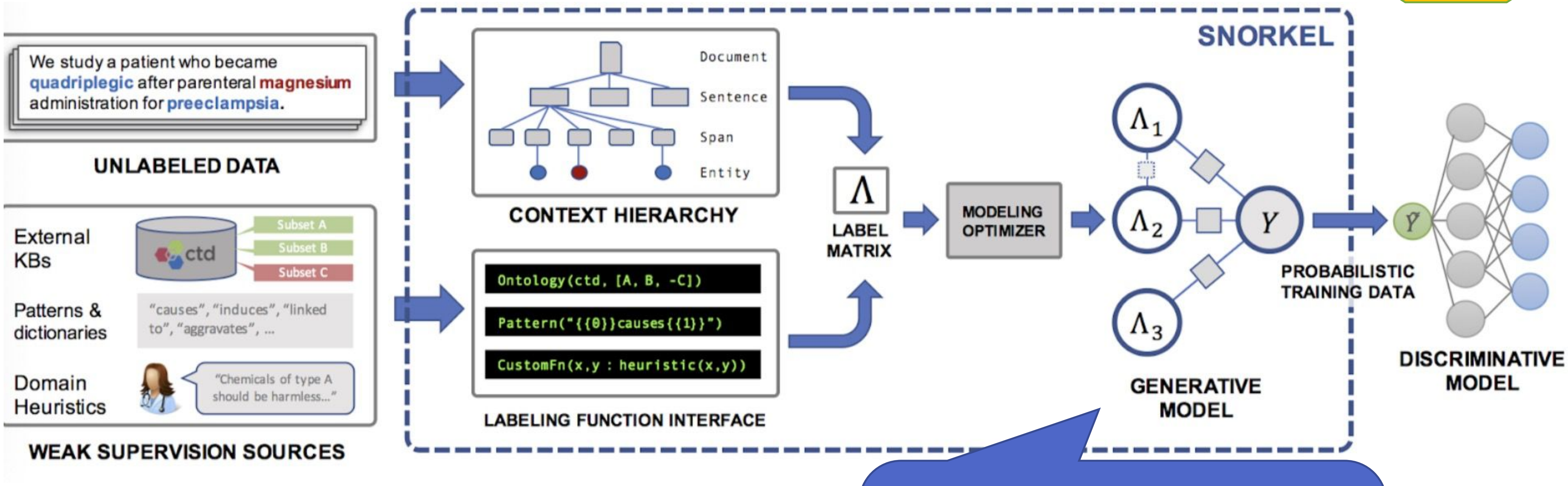
- Open source system implementing the Data Programming paradigm.

- Interface allows users to easily create labeling functions.

Ratner, et.al, Snorkel: Rapid Training Data Creation with Weak Supervision. VLDB, 2017

Long Answer: Snorkel

Challenge: Lack of Training Data



Denoise by re-weight the label functions for each data point

Ratner, et.al, Snorkel: Rapid Training Data Creation with Weak Supervision. VLDB, 2017

Long Answer: Snorkel

Challenge: Lack of Training Data

Task	# LFs	% Pos.	# Docs	# Candidates
Chem	16	4.1	1,753	65,398
EHR	24	36.8	47,827	225,607
CDR	33	24.6	900	8,272
Spouses	11	8.3	2,073	22,195
Radiology	18	36.0	3,851	3,851
Crowd	102	-	505	505

Relatively small # of labeling functions

Ratner, et.al, Snorkel: Rapid Training Data Creation with Weak Supervision. VLDB, 2017

Long Answer: Snorkel

Challenge: Lack of Training Data

	Distant Supervision			Snorkel (Gen.)				Snorkel (Disc.)				Hand Supervision		
Task	P	R	**F1**	P	R	**F1**	**Lift**	P	R	**F1**	**Lift**	P	R	**F1**
Chem	11.2	41.2	17.6	78.6	21.6	33.8	+16.2	87.0	39.2	54.1	+36.5	-	-	-
EHR	81.4	64.8	72.2	77.1	72.9	74.9	+2.7	80.2	82.6	81.4	+9.2	-	-	-
CDR	25.5	34.8	29.4	52.3	30.4	38.5	+9.1	38.8	54.3	45.3	+15.9	39.9	58.1	47.3
Spouses	9.9	34.8	15.4	53.5	62.1	57.4	+42.0	48.4	61.6	54.2	+38.8	47.8	62.5	54.2

Up to 39% F1 improvement over distant supervision

Competitive with human labels

Ratner, et.al, Snorkel: Rapid Training Data Creation with Weak Supervision. VLDB, 2017

Practical Tips

- Tuning probability score.
- Cleaner version of training data.
- One major model covering majority of categories/attributes, additional models for hard categories/attributes.

Practical Tips

- Model categorical attributes using classifiers and open vocab attributes using text extraction models.
- Rule-based post-processing step to further improve precision.
- Two-step evaluations:
  - benchmark dataset evaluation.
  - pre-publishing evaluation.

Reflection/short answer

- Attribute Value extraction task can be modeled 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 modalities.
- The techniques used here can also be applied to other domains like finance, biomedical etc, when the “subject” is known.

Reflections/short-answers

- Definition: Given a product, its category as optional, a list of attributes. For each attribute, find a list of attribute values for the product.
- Key intuition:
    - Make model attributes and categories aware to share/transfer knowledge between attributes and categories in order to scale up.
    - Opportunities also exist for combining text, text on image, image features with multi-modal model to allow interactions across all features.
- PG related techniques apply to:
    - Domains like finance, biomedical etc, when the “subject” is known.

Reflections/Short-Answers

- **Definition**: Find values for a given product and a set of attributes.
- **Recipe**: Sequence tagging.
- **Key to Success**: Scale up in different dimensions (#attributes, #categories).
- **Applicability to other domains**: Domains like finance, biomedical etc, where the “subject” is known.

Future Directions

- Scaling up jointly on attribute and category dimensions.
- Scaling up to multi-lingual extraction.
- Ensemble methods to handle categorical and open-vocab attributes.
- Improving the training data quality using data programming methods.