PAVE: Lazy-MDP based Ensemble to Improve Recall of Product Attribute Extraction Models

Kushal Kumar, Anoop Saladi,
International Machine Learning (IML),
amazon

Outline

- Attribute Extraction (AE) task
- Challenges with AE models
- PAVE
- Results

Attribute Extraction (AE) Task

Title	Garnier Skin Naturals, Charcoal, Face Serum Sheet Mask (Black), 28g & Garnier Skin Naturals
Bullet	Garnier Skin Naturals, Charcoal, Face Serum Sheet Mask (Black), 28gGarnier introduces a new generation of face masks for women that infuses skin with 1 week of serum with 1 mask. Garnier black serum mask is a breakthrough black tissue mask technology that offers double purifying and hydrating efficacy. Use Garnier black serum mask if you have dull skin with clogged and enlarged pores.
Scent	Blank

Similar item to consider

Amazon Brand - Solimo Solimo Banana Sheet Mask for Face
★★★★☆ (667)
₹ 49.00 (₹49.00/count) prime

Brand	Garnier
Skin Type	Combination
Item Form	Sheet
Product Benefits	Hydrating, Tightening
Scent	Charcoal
Material Type Free	Petroleum Free
Number of Items	1
Included Components	Mask
Material Feature	Purifying and Hydrating
Active Ingredients	Black Algae, LHA & Hyaluronic Acid

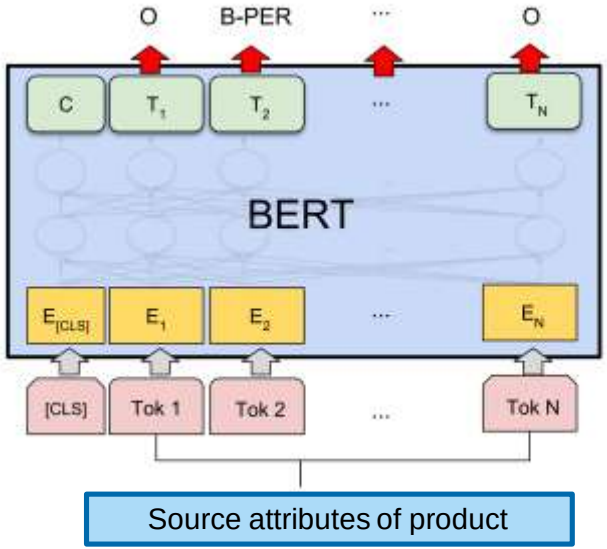
See less

About this item

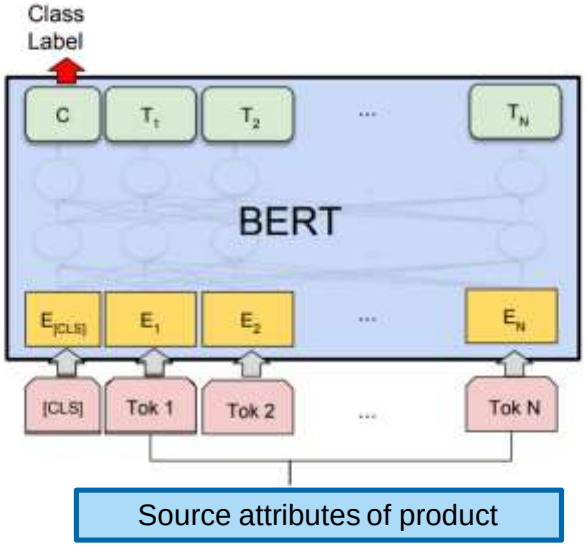
- Pure Charcoal Sheet Mask that helps with oily skin and enlarged pores
- Hydrates and detoxifies the skin, leaving it feeling fresh and with tighter pores

Attribute Extraction (AE) Solutions

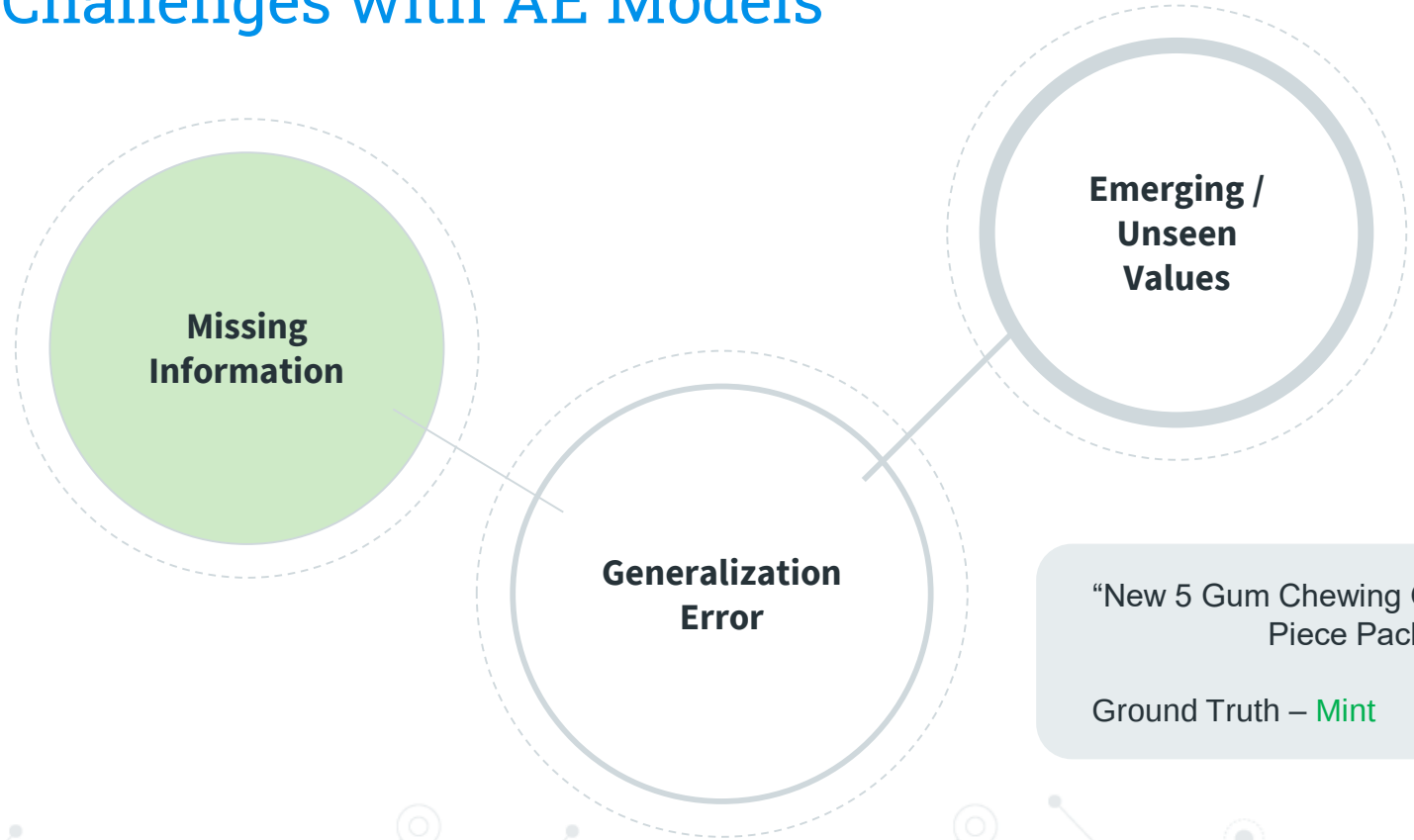
OPEN ATTRIBUTES



CLOSED ATTRIBUTES



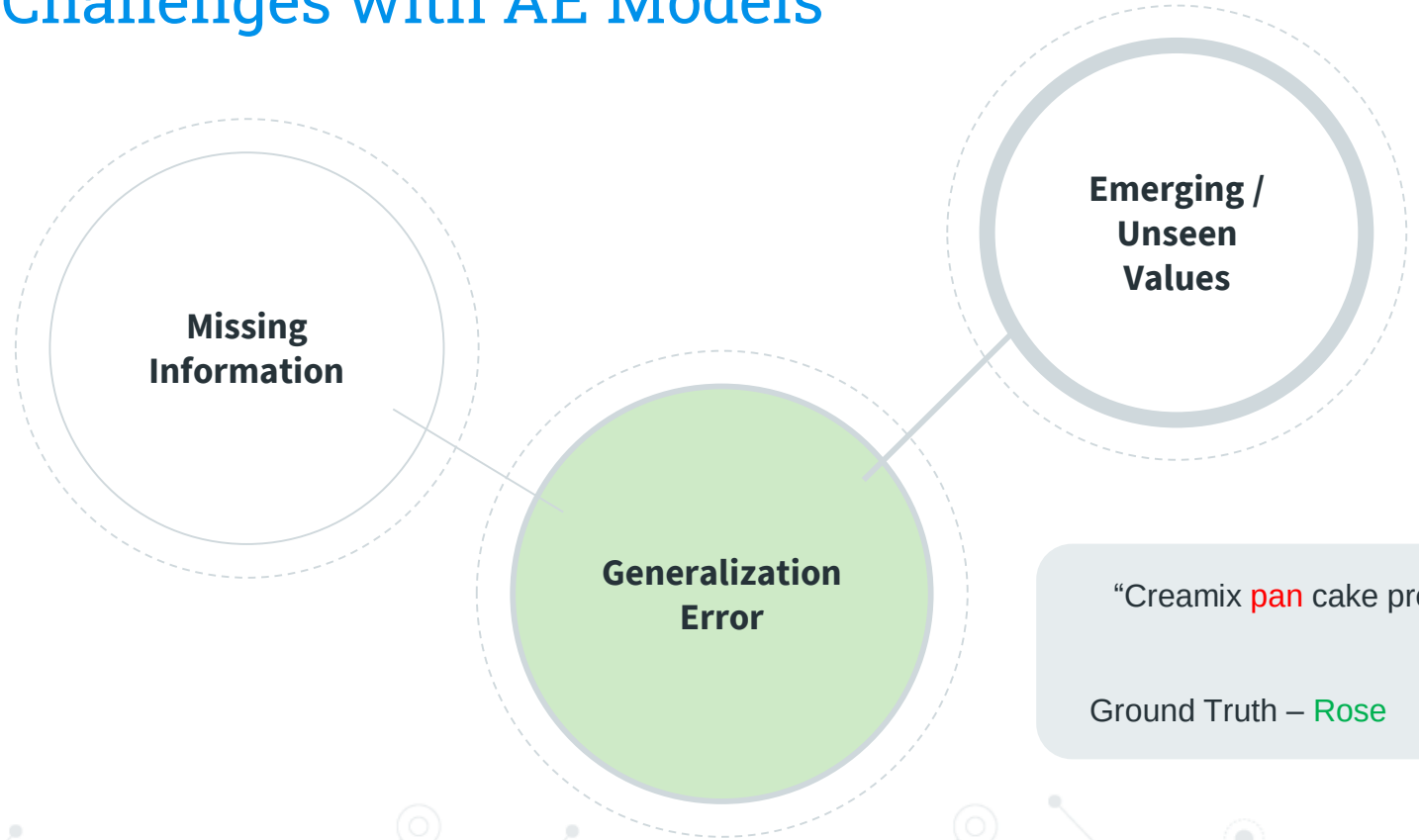
Challenges with AE Models



"New 5 Gum Chewing Gum Pack of 10 15 Piece Packages"

Ground Truth – Mint Attribute - Scent

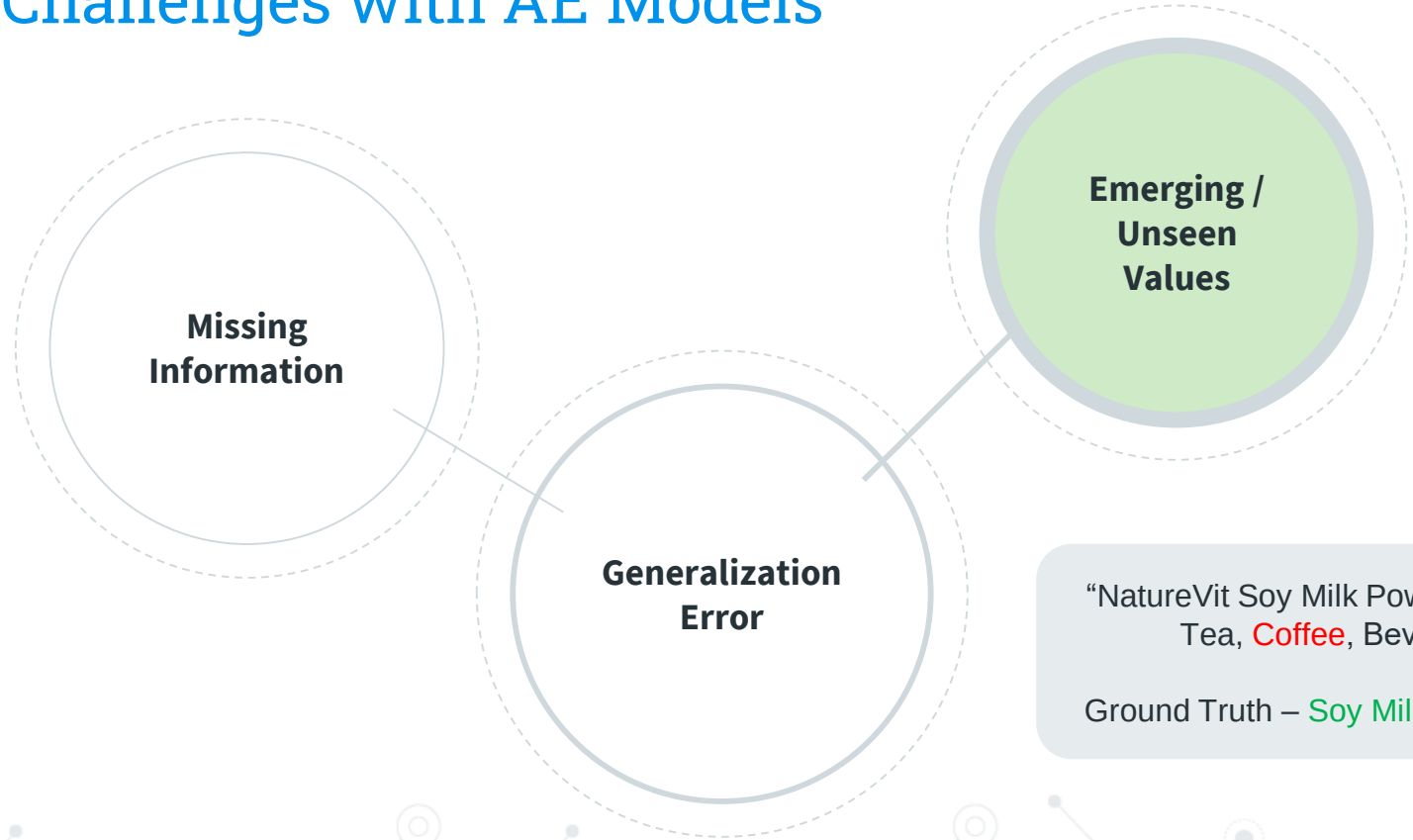
Challenges with AE Models



"Creamix pan cake premix rose 4000g"

Ground Truth – Rose Attribute - Flavor

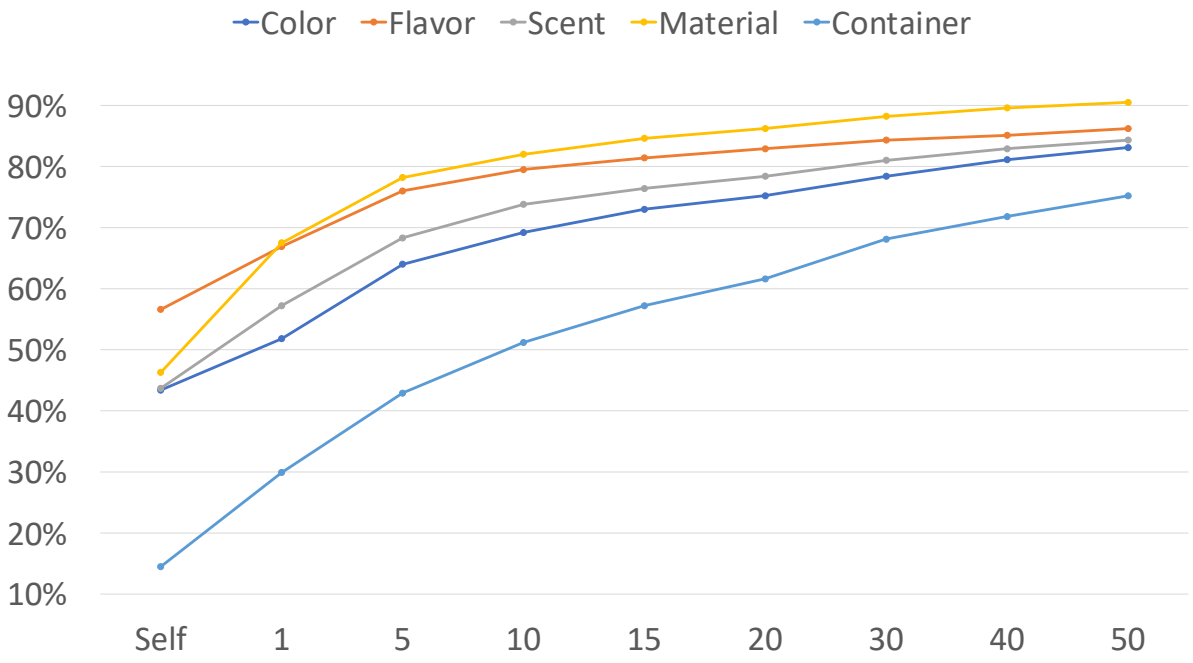
Challenges with AE Models



"NatureVit Soy Milk Powder can be used in Tea, Coffee, Beverages, etc."

Ground Truth – Soy Milk Attribute - Flavor

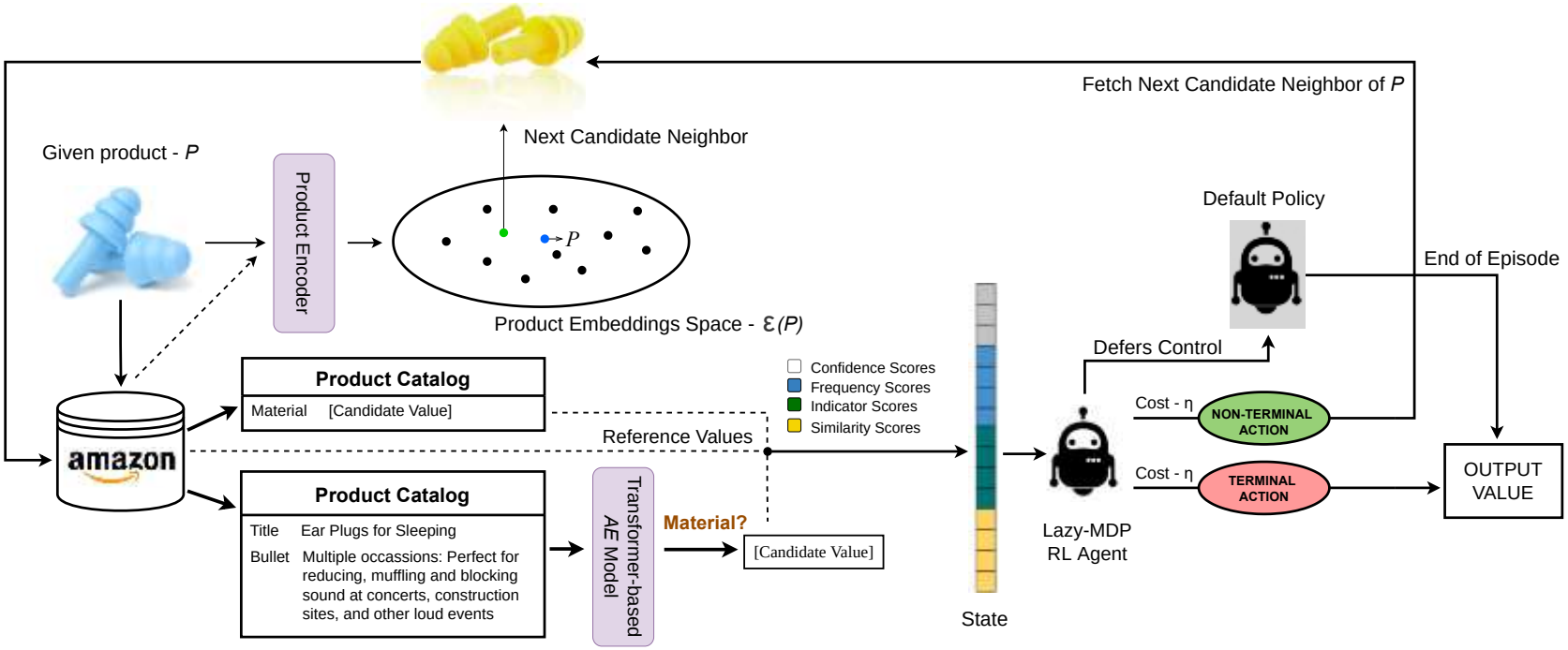
Why Should We Ensemble Values?



RL based Ensemble

Model Description	Precision	Recall	F1 score
Traditional AE models	67.1%	38.9%	47%
First neighbor ensemble	42.8%	57.7%	48.4%
Majority vote ensemble	43.5%	57.3%	48.7%
Most confident ensemble	40.4%	54.7%	45.8%
Model Based ensemble	58.0%	39.3%	46.7%
Oracle score (20 neighbors)	-	82.9%	-

PAVE – Product Attribute Value Ensemble



Liu et. al. Knowledge-guided Open Attribute Value Extraction with Reinforcement Learning

PAVE

$M_+ = (M, \bar{a}, \bar{\pi}, \eta)$

$\bar{\pi}$ - Ignores current candidate

Handles noise, stable and interpretable policy

Confidence scores

Frequency scores

Indicator scores

Similarity scores

Non-Terminal Actions:
1 (replace BV with AV)
2 (replace BV with VV)
3 (replace BV with CV)

Terminal Actions:
4 (stop)
5 (replace BV with blank & stop)

Episode Reward:
-α, blank
+/-1, match with ground truth

Intermediate Reward

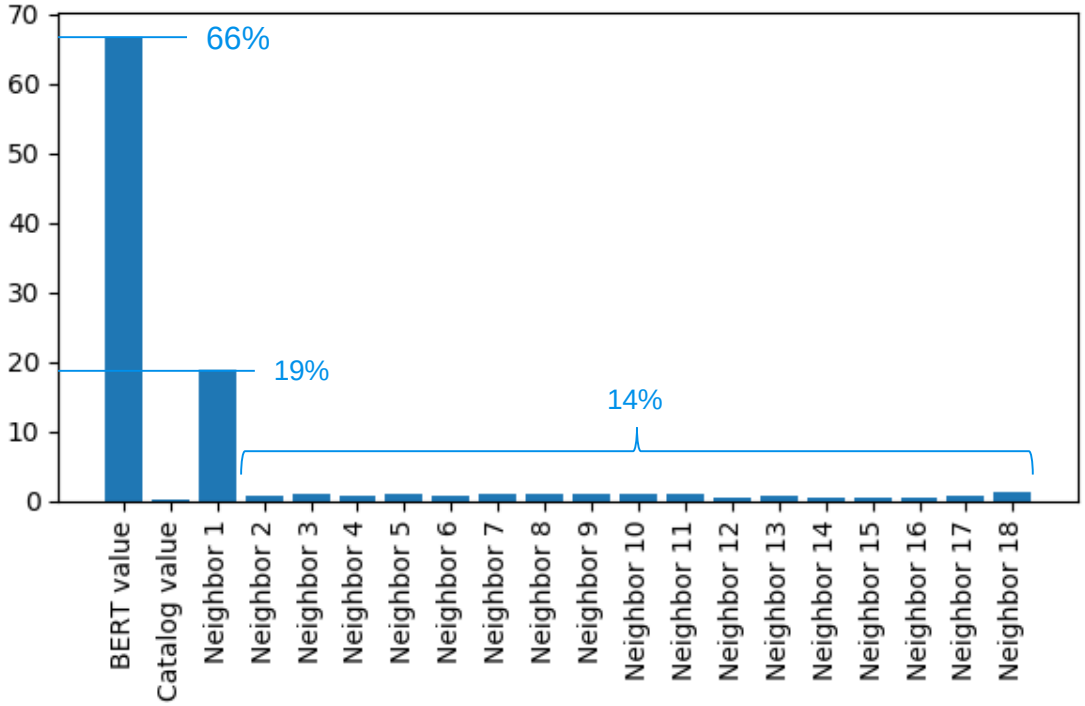
Penalty:
-ν, switch to same value

RESULTS

Comparison with Baselines

	Color			Flavor			Scent			Material			Container			All (macro)		
	Pr%	Re%	F1%	Pr%	Re%	F1%	Pr%	Re%	F1%	Pr%	Re%	F1%	Pr%	Re%	F1%	Pr%	Re%	F1%
Baseline AE model.																		
BERT AE	75.7	39.0	51.5	50.9	56.3	53.5	70.7	41.8	52.5	74.3	43.4	54.8	64.1	13.8	22.7	67.1	38.9	47.0
Baseline ensemble models.																		
First ensemble	54.6	56.3	55.4	41.8	58.0	48.6	28.5	59.6	38.6	43.8	66.9	52.9	45.4	47.8	46.6	42.8	57.7	48.4
Confidence ensemble	50.3	51.9	55.1	42.0	58.7	49.0	27.3	56.7	36.9	41.3	63.2	50.0	40.9	43.0	41.9	40.4	54.7	45.8
Majority ensemble	55.0	56.2	55.6	42.4	58.1	49.0	29.9	59.4	39.8	44.1	66.6	53.1	46.1	46.4	43.9	43.5	57.3	48.7
Model ensemble	70.0	39.0	50.1	55.0	40.7	46.8	48.2	29.7	36.8	64.9	43.8	52.3	52.1	43.4	47.4	58.0	39.3	46.7
PAVE model and its variants.																		
PAVE	**67.8**	**50.5**	**57.9**	**51.8**	**59.8**	**55.5**	51.6	61.9	56.3	69.7	69.4	69.5	57.1	42.1	48.5	59.8	56.7	57.5
BERT AE + PAVE ensemble	67.4	50.2	57.5	49.4	58.9	53.7	51.4	61.3	55.9	69.1	65.0	67.0	56.7	39.6	46.6	58.8	55.0	56.1
PAVE-DQN	52.1	44.2	47.8	53.9	50.3	52.0	51.5	40.5	45.3	61.4	60.4	60.9	56.0	26.4	35.9	55.0	44.4	48.4
BERT AE + PAVE-DQN ensemble	54.8	53.8	54.3	48.6	58.8	53.2	54.4	54.0	54.2	61.5	61.6	61.5	54.1	28.4	37.2	54.7	51.3	52.1
Confidence thresholding on PAVE models.																		
PAVE-maxF1	67.8	50.5	57.9	51.8	59.8	55.5	**64.4**	**52.9**	**58.1**	**72.7**	**67.4**	**69.9**	**61.9**	**40.3**	**48.8**	**63.7**	**54.2**	**58.1**
PAVE-maxPr	76.7	38.4	51.2	66.7	0.9	1.8	71.6	41.0	52.1	78.1	52.5	62.8	73.2	26.8	39.2	73.3	31.9	41.4
PAVE-samePr	75.8	39.9	52.3	51.8	59.8	55.5	70.5	42.8	53.3	74.3	65.2	69.5	63.8	38.5	39.4	67.2	49.2	55.7

Breaking Up Recall Lift



Thanks!

Questions?

You can find me at:

kushlku@amazon.com

Link to the paper