Ontology Mining

Taxonomy Enrichment & Relation Discovery

Section Structure

- Problem Definition

  What is needed beyond techniques for building generic KGs?
- Short answer -- key intuition

  What are key intuitions for ontology mining?
- Long answer -- details

  What are practical tips?
- Reflection/short-answer

  Can we apply the techniques to other domains?

Why Ontology Mining?

- Living in a world that is constantly changing...
    - Emerging Product Categories

Anthony's Organic Coconut Flour, 4lbs, Batch Tested Gluten Free, Non GMO, Vegan, Keto Friendly

NEW

Coconut Flour

Why Ontology Mining?

- Living in a world that is constantly changing...
  - Emerging Product Categories

Anthony's Organic Coconut Flour, 4lbs, Batch Tested Gluten Free, Non GMO, Vegan, Keto Friendly

NEW

Coconut Flour

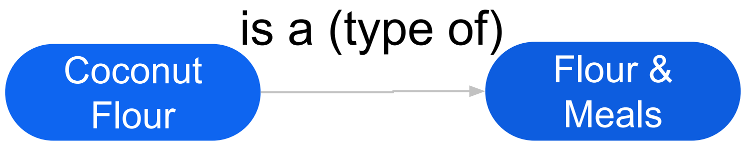
  - Understand the relation between products

Why Ontology Mining?

- Living in a world that is constantly changing...
  - Emerging Product Categories

Anthony's Organic Coconut Flour, 4lbs, Batch Tested Gluten Free, Non GMO, Vegan, Keto Friendly

  - Understand the relation between products



Why Ontology Mining?

- Living in a world that is constantly changing...
    - Emerging Product Categories

Anthony's Organic Coconut Flour, 4lbs, Batch Tested Gluten Free, Non GMO, Vegan, Keto Friendly

NEW

Coconut Flour

- Understand the relation between products

Coconut Flour → is a (type of) → Flour & Meals

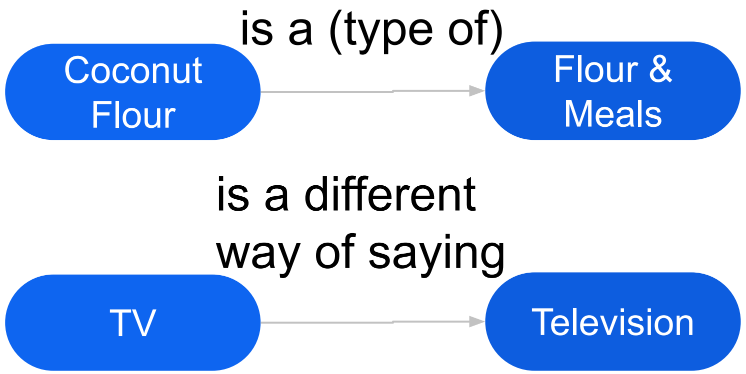
TV → is a different way of saying → Television

Why Ontology Mining?

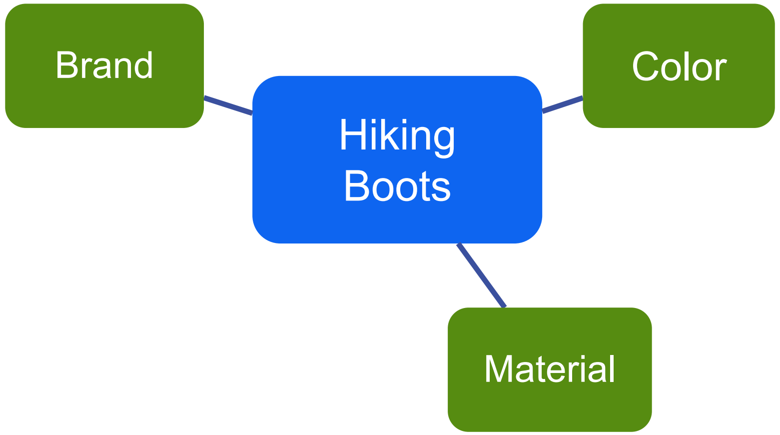
- Living in a world that is constantly changing...
    - Emerging Product Categories

Anthony's Organic Coconut Flour, 4lbs, Batch Tested Gluten Free, Non GMO, Vegan, Keto Friendly

    - Understand the relation between products



    - Discover relations between product category and attributes



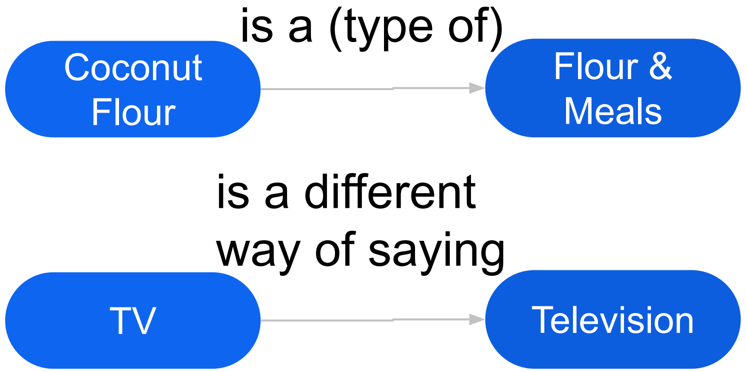
Why Ontology Mining?

- Living in a world that is constantly changing...
  - Emerging Product Categories

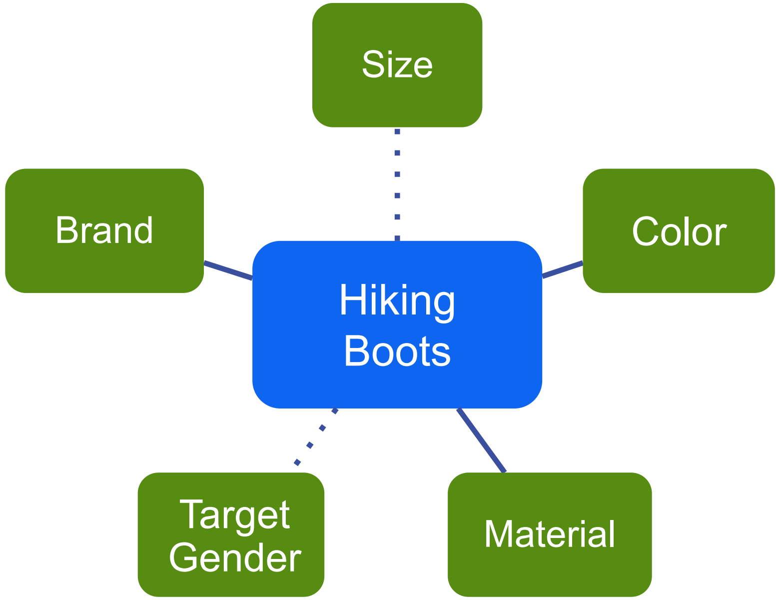
Anthony's Organic Coconut Flour, 4lbs, Batch Tested Gluten Free, Non GMO, Vegan, Keto Friendly

NEW
Coconut Flour

  - Understand the relation between products



  - Discover relations between product category and attributes



Why Ontology Mining?

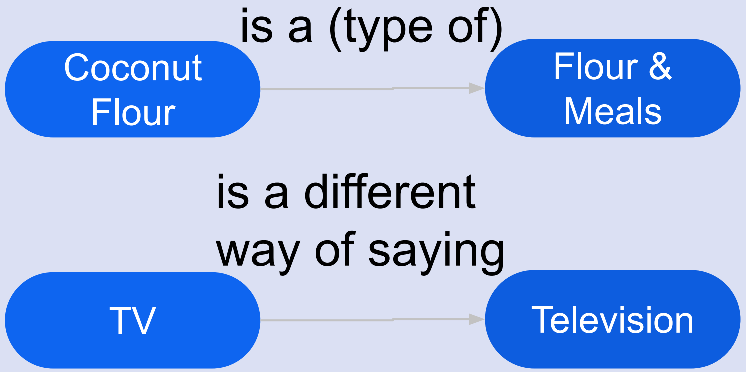
- Living in a world that is constantly changing...
  - Emerging Product Categories

Anthony's Organic Coconut Flour, 4lbs, Batch Tested Gluten Free, Non GMO, Vegan, Keto Friendly

NEW

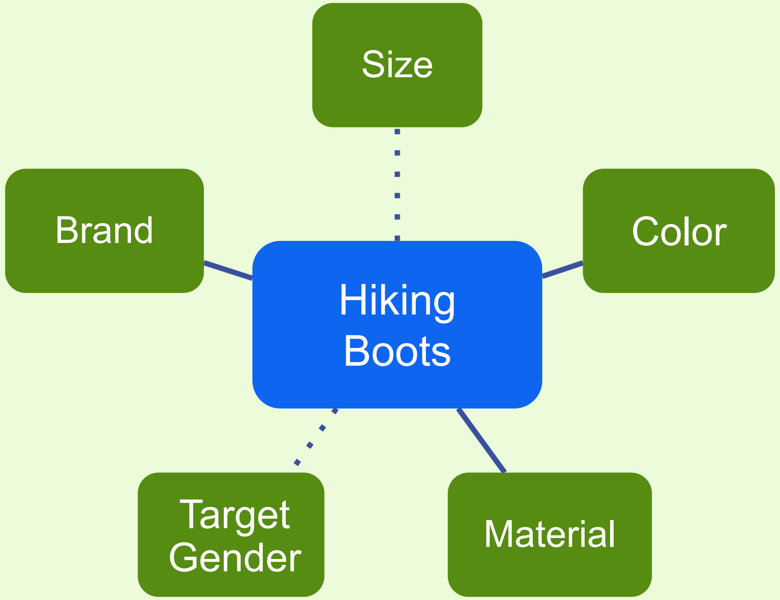
Coconut Flour

  - Understand the relation between products



Taxonomy Enrichment

  - Discover relations between product category and attributes



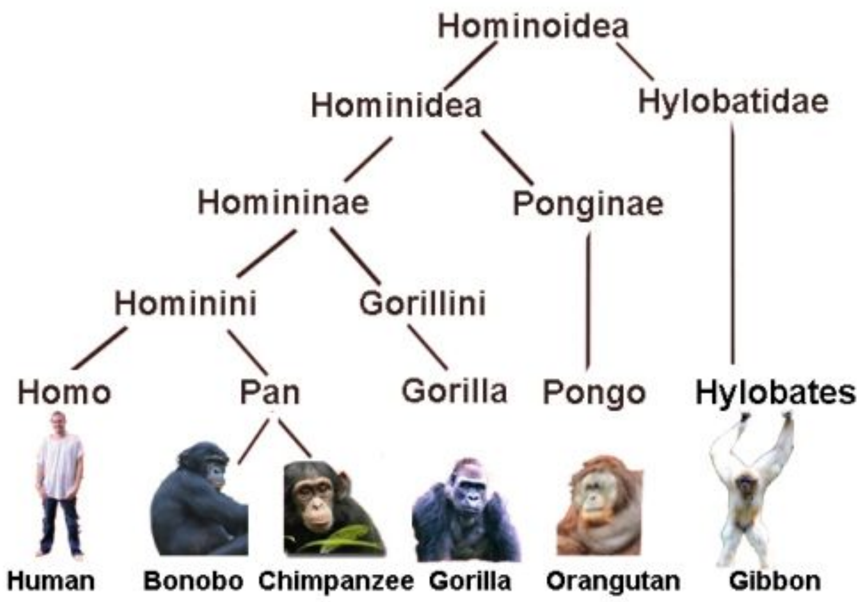
Relation Discovery

Taxonomy Enrichment

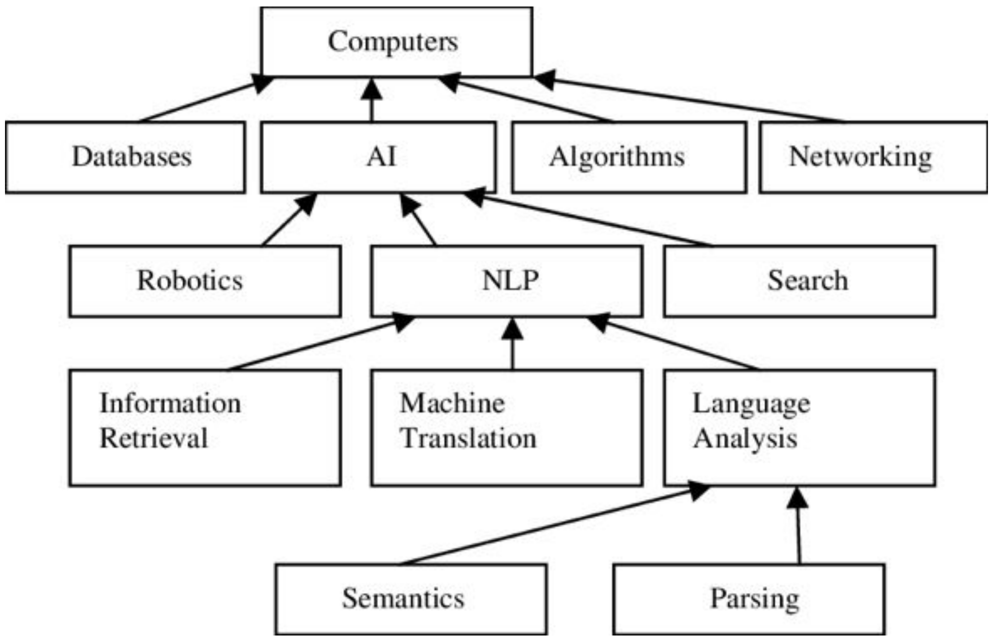
Discover new product categories &
Attach them to the existing taxonomy

What is a Taxonomy?

- **A taxonomy of biological organisms**

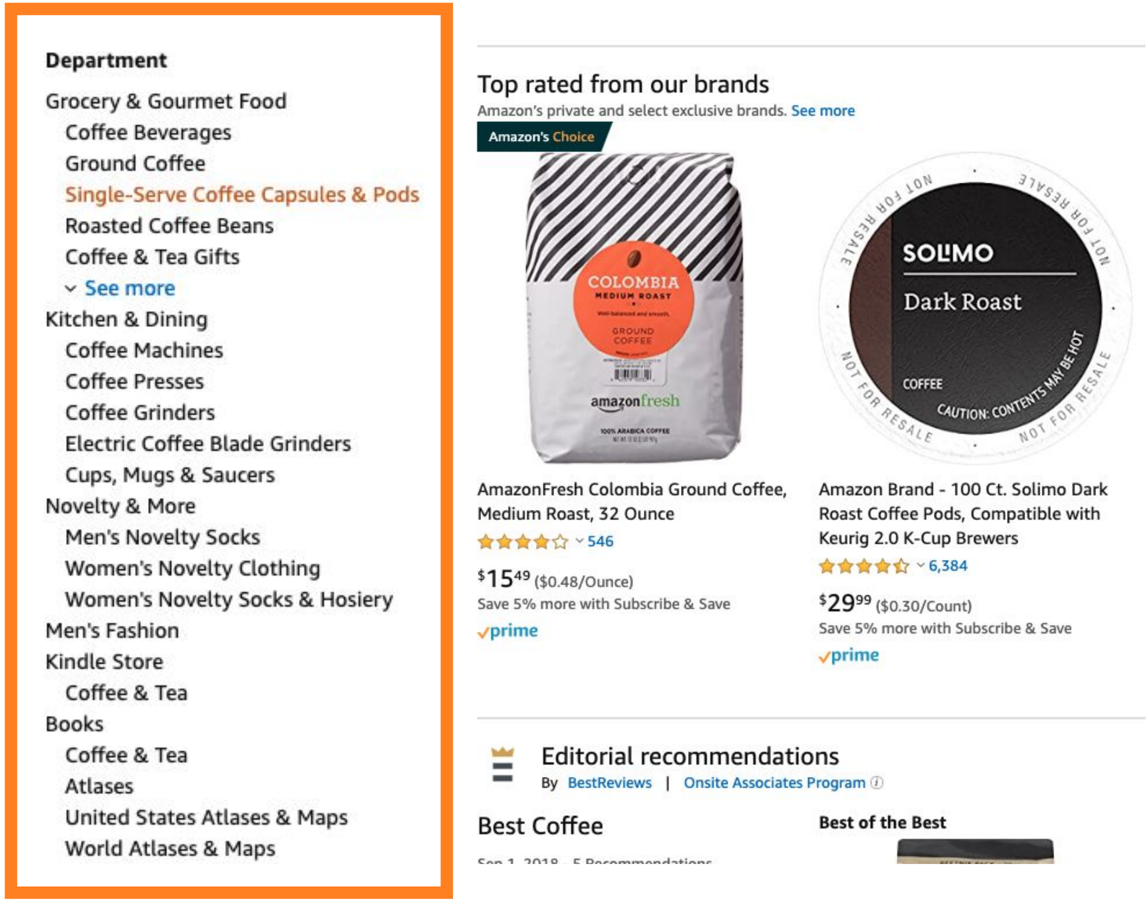


- **A taxonomy for computer science**



What is a Product Taxonomy?

From the eyes of customers



What is a Product Taxonomy?

From the eyes of customers

Department

Grocery & Gourmet Food
- Coffee Beverages
- Ground Coffee
- Single-Serve Coffee Capsules & Pods
- Roasted Coffee Beans
- Coffee & Tea Gifts
- See more

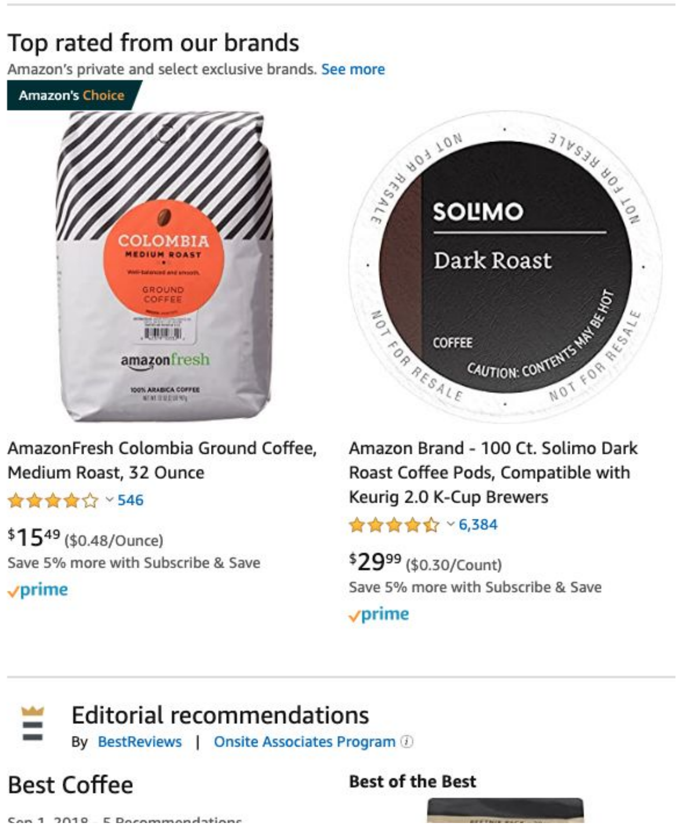
Kitchen & Dining
- Coffee Machines
- Coffee Presses
- Coffee Grinders
- Electric Coffee Blade Grinders
- Cups, Mugs & Saucers

Novelty & More
- Men's Novelty Socks
- Women's Novelty Clothing
- Women's Novelty Socks & Hosiery

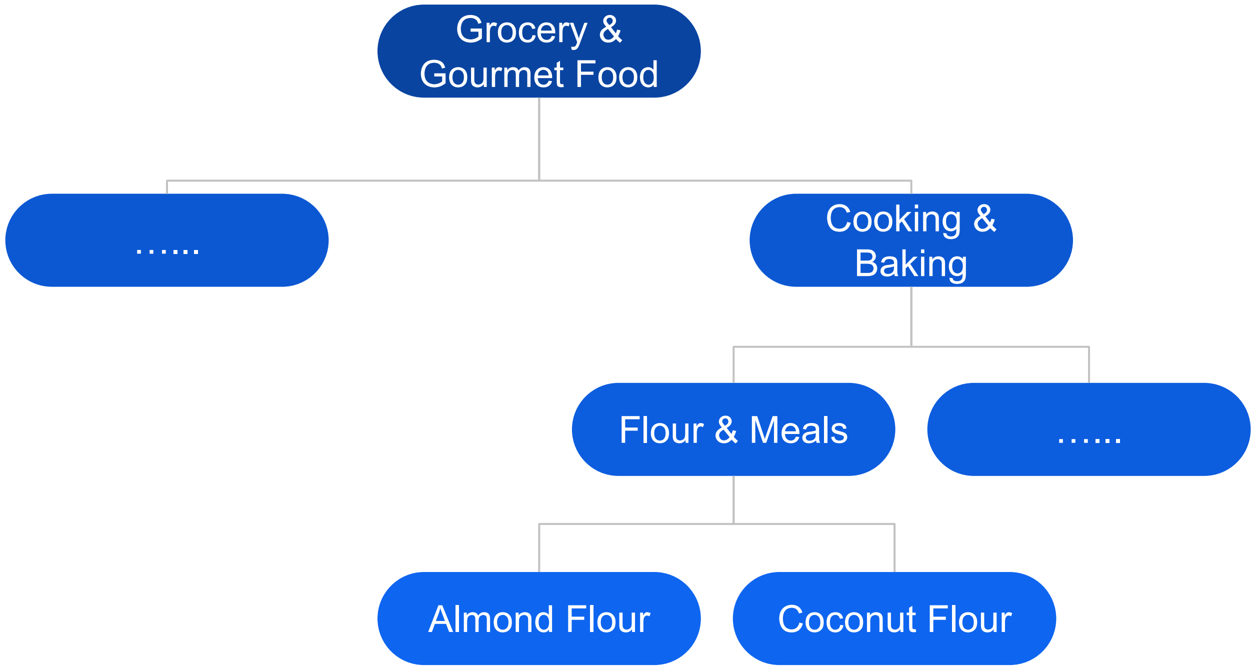
Men's Fashion

Kindle Store
- Coffee & Tea

Books
- Coffee & Tea
- Atlases
- United States Atlases & Maps
- World Atlases & Maps



Backend taxonomy structure



What is Taxonomy Enrichment?

- Given a base Taxonomy T=<V, R>, where
  - v ∈ V is the product category node
  - R is the relationship between product categories, selected from {‘hypernym’, ‘synonym’}

What is Taxonomy Enrichment?

- Given a base Taxonomy T=<V, R>, where
  - v ∈ V is the product category node
  - R is the relationship between product categories, selected from {‘hypernym’, ‘synonym’}
- Taxonomy Enrichment tries to
  - Identify new product category v’
  - And attach the new product category v’ to v ∈ V with certain r ∈ R

Generic Solution

- Treat Taxonomy Enrichment as a two-stage approach

Term Extraction → **Taxonomy Expansion**

- “Which term to attach”
- Input: texts, images
- Output: candidate terms extracted from texts/images

Generic Solution

- Treat Taxonomy Enrichment as a two-stage approach

Term Extraction → **Taxonomy Expansion**

Term Extraction
- "Which term to attach"
- Input: texts, images
- Output: candidate terms extracted from texts/images

Taxonomy Expansion
- "Where to attach to"
- Input: terms to add to the taxonomy, existing taxonomy
- Output: enriched taxonomy

Unique Challenges for Product Taxonomy Enrichment

Recognize fine-grained product category in an open-world setting

Fine-grained categorization space for e-commerce products

Minimize manual curation efforts & Ensure freshness

Manually curated core taxonomy over years + constantly emerging categories

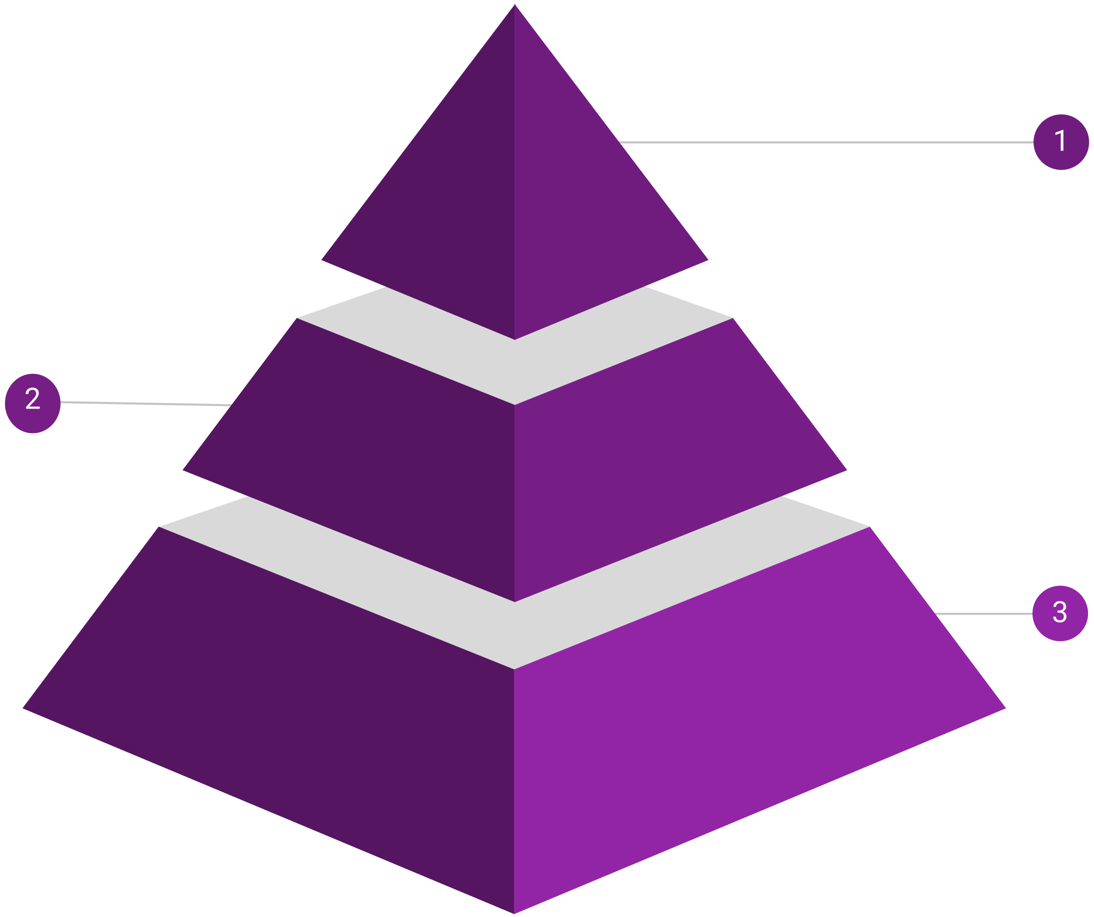
Leverage rich heterogeneous information sources

Limited supervisions from taxonomy nodes themselves

Why Taxonomy Enrichment?

- Why enrichment, instead of construction from scratch?
  - Already have a decent taxonomy built by experts and used for years
  - Most common terms are covered
- What is missing?
  - Emerging terms (product type) take time for domain-experts to discover
  - Long-tail / fine-grained terms are likely to be neglected
- What is needed?
  - Discover emerging product categories
  - Organize product categories in a tree-like structure

Short answer -- Key Intuitions



Start with the core Taxonomy

Built by experts and accumulated for years

Most common terms are covered

Add long-tail nodes

Learn from existing hypernum pairs in the core Taxonomy

Automatically discover long-tail categories from heterogeneous sources

Understand super long-tail, ad-hoc terms

For specific downstream applications using NLP, Text Mining, and etc.

Long answer -- Term Extraction

- Leverage information from various modalities for term extraction
  - **Extracted/Summarized from seller-provided texts**
  - From product images
  - From user search queries

Sellers mention product category in product titles

NESCAFE CLASICO Dark Roast Instant Coffee 7 Ounce

Long answer -- Term Extraction

- Leverage information from various modalities for term extraction
  - Extracted/Summarized from seller-provided texts
  - **From product images**
  - From user search queries

Product Image helps determine the product category

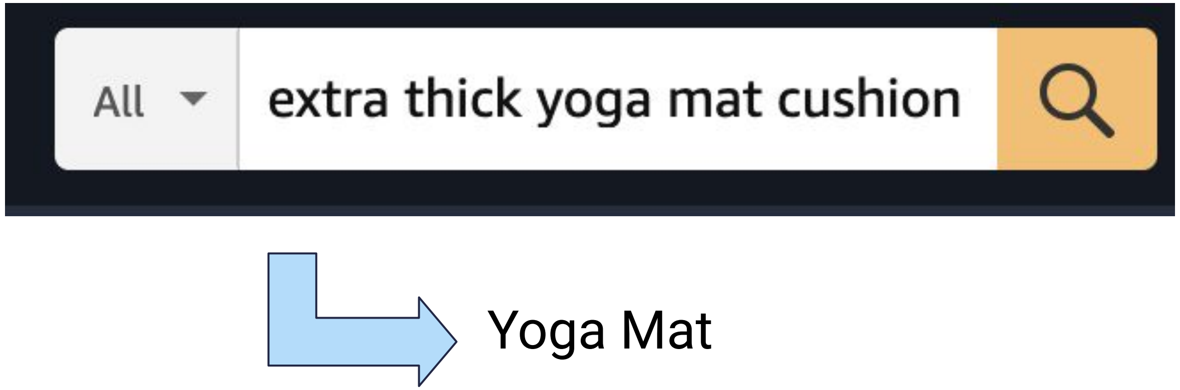
Bucket Hat

300N Unisex 100% Cotton Packable Summer Travel Bucket Beach Sun Hat

Long answer -- Term Extraction

- Leverage information from various modalities for term extraction
  - Extracted/Summarized from seller-provided texts
  - From product images
  - **From user search queries**

Customers search with product category keywords



Long answer -- Term Extraction/Summarization from Seller-Provided Texts

- Consider term extraction as a sequence labeling task [Mao+ 2020]

NESCAFE	CLASICO	Dark	Roast	Instant	Coffee	7	Ounce
O	O	O	O	B	E	O	O

Mao et al., Octet: Online Catalog Taxonomy Enrichment with Self-Supervision, SigKDD, 2020.
Li et al., Don't Classify, Translate: Multi-Level E-Commerce Product Categorization Via Machine Translation, TMIS, 2018.

Long answer -- Term Extraction/Summarization from Seller-Provided Texts

- Consider term extraction as a sequence labeling task [Mao+ 2020]

NESCAFE	CLASICO	Dark	Roast	Instant	Coffee	7	Ounce
O	O	O	O	B	E	O	O

Term may not be mentioned as a consecutive phrase in the text

Mao et al., Octet: Online Catalog Taxonomy Enrichment with Self-Supervision, SigKDD, 2020.
Li et al., Don't Classify, Translate: Multi-Level E-Commerce Product Categorization Via Machine Translation, TMIS, 2018.

Long answer -- Term Extraction/Summarization from Seller-Provided Texts

- Consider term extraction as a sequence labeling task [Mao+ 2020]

NESCAFE	CLASICO	Dark	Roast	Instant	Coffee	7	Ounce
O	O	O	O	B	E	O	O

Term may not be mentioned as a consecutive phrase in the text

Ariel Detergent Liquid Color Power Detergent

Liquid Detergent

Mao et al., Octet: Online Catalog Taxonomy Enrichment with Self-Supervision, SigKDD, 2020.
Li et al., Don't Classify, Translate: Multi-Level E-Commerce Product Categorization Via Machine Translation, TMIS, 2018.

Long answer -- Term Extraction/Summarization from Seller-Provided Texts

- Consider term extraction as a sequence labeling task [Mao+ 2020]

NESCAFE	CLASICO	Dark	Roast	Instant	Coffee	7	Ounce
O	O	O	O	B	E	O	O

Term may not be mentioned as a consecutive phrase in the text

- Term summarization as a sequence-to-sequence prediction task [Li+ 2018]

Ariel Detergent Liquid Color Power Detergent

Liquid Detergent

Mao et al., Octet: Online Catalog Taxonomy Enrichment with Self-Supervision, SigKDD, 2020.
Li et al., Don't Classify, Translate: Multi-Level E-Commerce Product Categorization Via Machine Translation, TMIS, 2018.

Long answer -- Term Extraction/Summarization from Seller-Provided Texts

- Consider term extraction as a sequence labeling task [Mao+ 2020]

NESCAFE	CLASICO	Dark	Roast	Instant	Coffee	7	Ounce
O	O	O	O	B	E	O	O

Term may not be mentioned as a consecutive phrase in the text

- Term summarization as a sequence-to-sequence prediction task [Li+ 2018]

Can't derive categorical information from the text at all?

Mao et al., Octet: Online Catalog Taxonomy Enrichment with Self-Supervision, SigKDD, 2020.
Li et al., Don't Classify, Translate: Multi-Level E-Commerce Product Categorization Via Machine Translation, TMIS, 2018.

Long answer -- Term Extraction/Summarization from Seller-Provided Texts

- Consider term extraction as a sequence labeling task [Mao+ 2020]

NESCAFE	CLASICO	Dark	Roast	Instant	Coffee	7	Ounce
O	O	O	O	B	E	O	O

Term may not be mentioned as a consecutive phrase in the text

- Term summarization as a sequence-to-sequence prediction task [Li+ 2018]

Can't derive categorical information from the text at all?

Mao et al., Octet: Online Catalog Taxonomy Enrichment with Self-Supervision, SigKDD, 2020.
Li et al., Don't Classify, Translate: Multi-Level E-Commerce Product Categorization Via Machine Translation, TMIS, 2018.

Long answer -- Term Extraction/Summarization from Seller-Provided Texts

- Consider term extraction as a sequence labeling task [Mao+ 2020]

Term may not be mentioned as a consecutive phrase in th

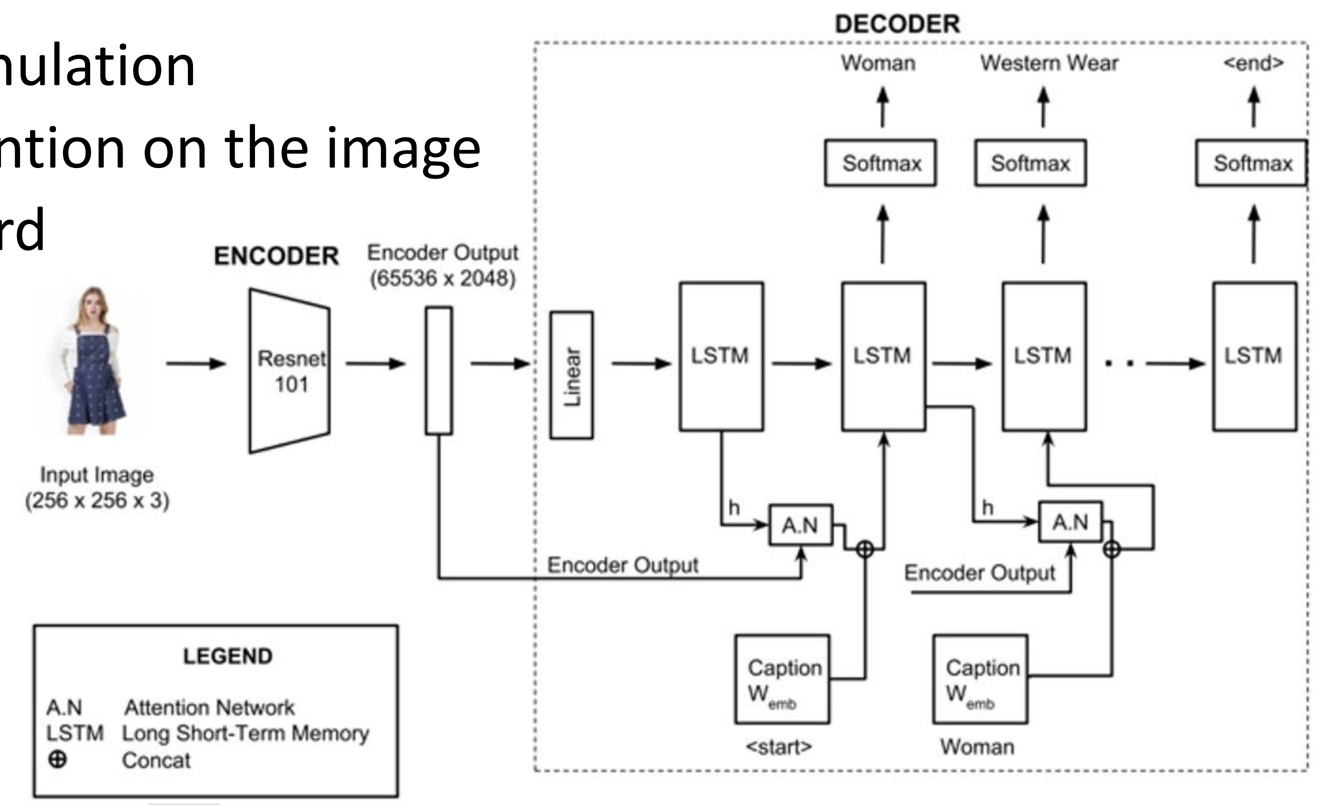
- Term summarization as a sequence-to-sequence prediction

Can't derive categorical information from the text at all?

Mao et al., Octet: Online Catalog Taxonomy Enrichment with Self-Supervision, SigKDD, 2020.
Li et al., Don't Classify, Translate: Multi-Level E-Commerce Product Categorization Via Machine Translation, TMIS, 2018.

Long Answer -- Term Extraction from Images

- From product images [Umaashankar+ 2020]
  - Encode image into a vector
  - Use an Encoder-Decoder formulation
  - Decoder is an LSTM with attention on the image
  - Decode the term word by word



Umaashankar et al., Atlas: A Dataset and Benchmark for E-commerce Clothing Product Categorization. SigIR, 2020.

Long Answer -- Term Extraction from Images

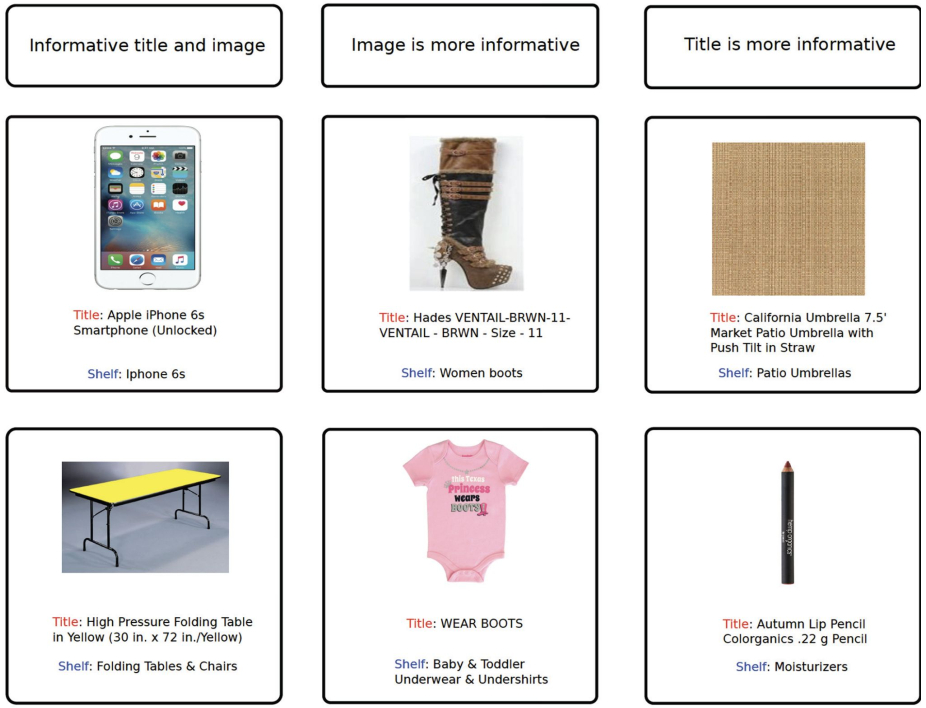
- From product images [Umaashankar+ 2020]
  - Encode image into a vector
  - Use an Encoder-Decoder formulatio
  - Decoder is an LSTM with attention c
  - Decode the term word by word

When both text and image are available, which modality is more helpful?

Umaashankar et al., Atlas: A Dataset and Benchmark for E-commerce Clothing Product Categorization. SigIR, 2020.

Long Answer -- Term Extraction from Images

- Study the impact from each modality [Zahavy+ 2018]

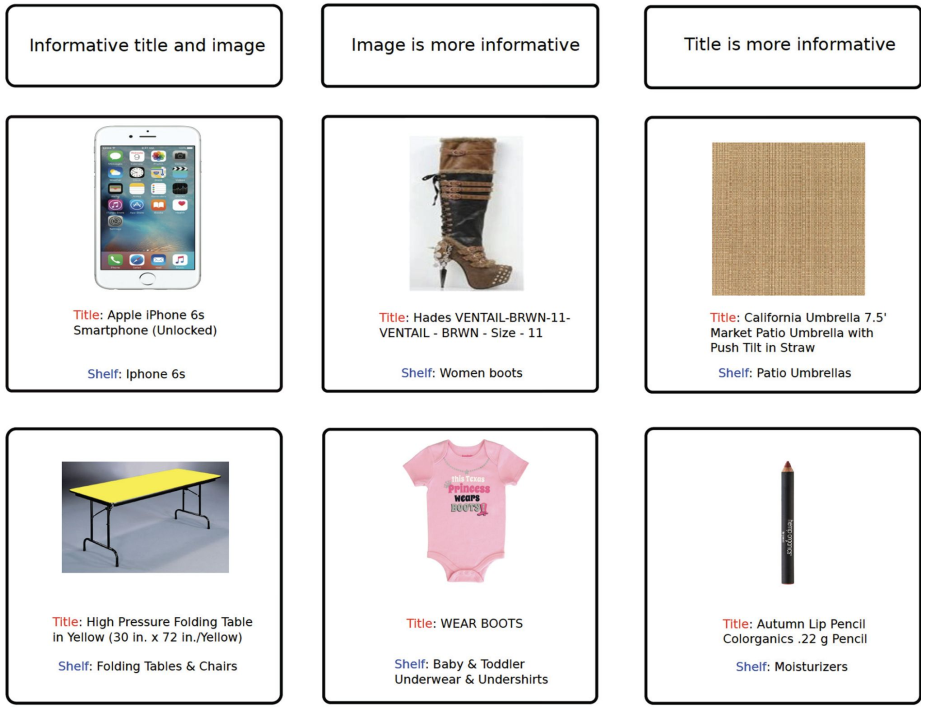


Zahavy et al., Is a picture worth a thousand words? a deep multi-modal fusion architecture for product classification in e-commerce, AAAI, 2018.

Long Answer -- Term Extraction from Images

- Study the impact from each modality [Zahavy+ 2018]
  - Encode both textual and visual features
  - Learns a policy (decision fusion rule) using a deep network
  - Multimodal architecture improves accuracy over both input-specific networks

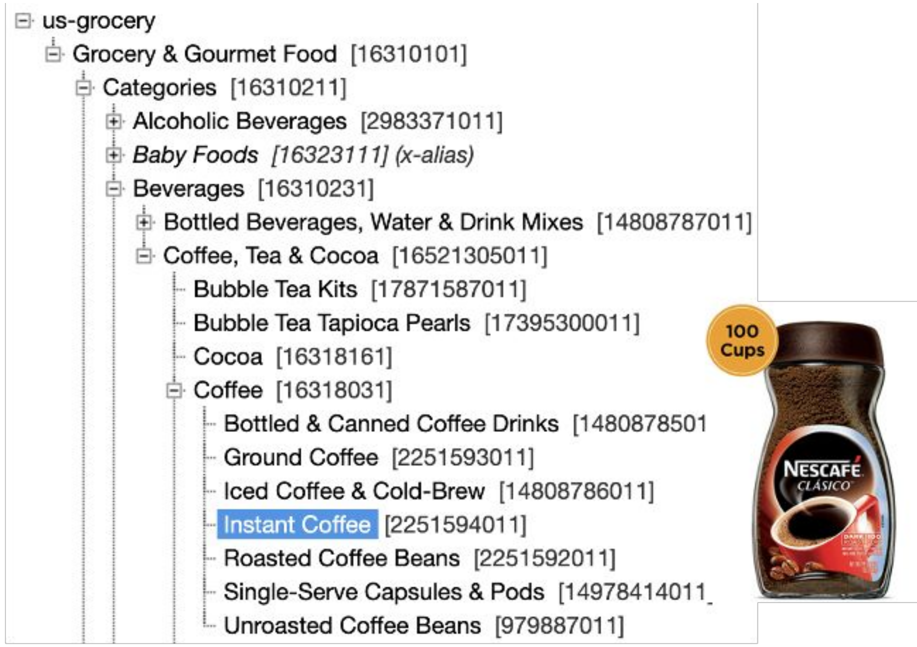
Policy network input	Policy	Optimal Policy	Policy accuracy
Top-3 class probabilities	**71.8 (+1.7)**	**77.7 (+7.6)**	**84.2**
Image	68.5(-1.6)	77.6 (+7.5)	80.3
Text	69.0 (-1.1)	77.6 (+7.5)	83.7
Both	66.1 (-4)	77.6 (+7.5)	73.7



Zahavy et al., Is a picture worth a thousand words? a deep multi-modal fusion architecture for product classification in e-commerce, AAAI, 2018.

Long Answer -- Minimum Manual Efforts

- Collecting Training Data using Distant-Supervision



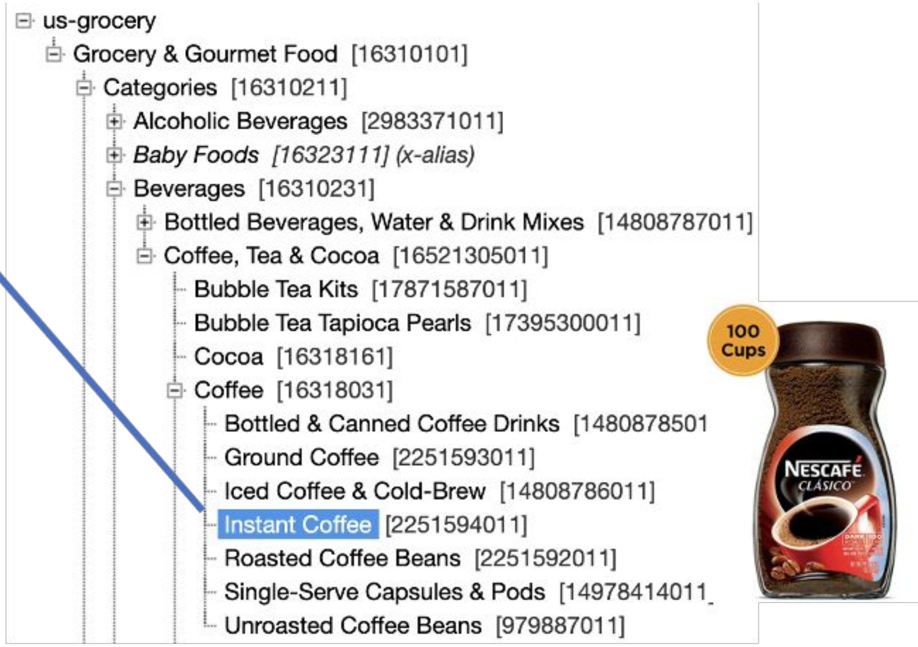
Mao et al., Octet: Online Catalog Taxonomy Enrichment with Self-Supervision, SigKDD, 2020.

Long Answer -- Minimum Manual Efforts

- Collecting Training Data using Distant-Supervision
  - Leveraging existing product category assignments & category surface names

Product's category surface name as desired terms to be extracted

"NESCAFE CLASICO Dark Roast **Instant Coffee** 7 Ounce"



Mao et al., Octet: Online Catalog Taxonomy Enrichment with Self-Supervision, SigKDD, 2020.

Long Answer -- Minimum Manual Efforts

- Collecting Training Data using Distant-Supervision
    - Leveraging existing product category assignments & category surface names

      Product's category surface name as desired terms to be extracted

      "NESCAFE CLASICO Dark Roast **Instant Coffee** 7 Ounce"
    - Leveraging query & category annotations

Mao et al., Octet: Online Catalog Taxonomy Enrichment with Self-Supervision, SigKDD, 2020.

Long Answer -- Minimum Manual Efforts

- Collecting Training Data using Distant-Supervision
  - Leveraging existing product category assignments & category surface names

    Product's category surface name as desired terms to be extracted

    "NESCAFE CLASICO Dark Roast **Instant Coffee** 7 Ounce"
  - Leveraging query & category annotations

- Such distant-supervised approach matches 62% of products and generates training data **without additional labeling efforts**

Mao et al., Octet: Online Catalog Taxonomy Enrichment with Self-Supervision, SigKDD, 2020.

Long Answer -- Learning under an open-world setting

- Rely on the generalization ability of the product category extraction/summarization model [Dong+ 2020]

Dong et al., AutoKnow: Self-driving knowledge collection for products of thousands of types, SigKDD, 2020.

Long Answer -- Learning under an open-world setting

- Rely on the generalization ability of the product category extraction/summarization model [Dong+ 2020]
  - Achieved **87.7% precision** across four major categories: Grocery, Health, Beauty, and Baby
  - Increased the number of categories by **2.9X**

Dong et al., AutoKnow: Self-driving knowledge collection for products of thousands of types, SigKDD, 2020.

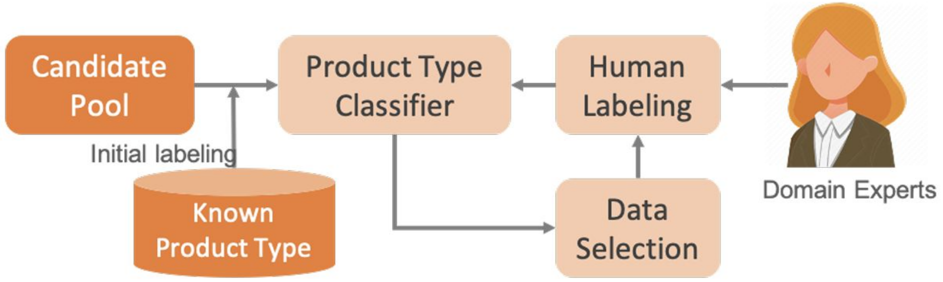
Long Answer -- Learning under an open-world setting

- Rely on the generalization ability of the product category extraction/summarization model
- Classification model with the capability to handle new classes [Xu+ 2019]
  - Ability to reject examples from unseen classes
  - Incrementally learn new classes to expand the existing model

Xu et al, Open-world learning and application to product classification, WWW, 2019.

Long Answer -- Learning under an open-world setting

- Rely on the generalization ability of the product category extraction/summarization model
- Classification model with the capability to handle new classes [Xu+ 2019]
  - Ability to reject examples from unseen classes
  - Incrementally learn new classes to expand the existing model
- Utilize domain experts' knowledge for Active Learning [Zhu+ 2020]



Zhu et al., Active Learning for Product Type Ontology Enhancement in E-commerce, SigKDD, 2020.

Generic Solution

- Treat Taxonomy Enrichment as a two-stage approach

Term Extraction → **Taxonomy Expansion**

Term Extraction

- “Which term to attach?”
- Input: texts, images
- **Output: candidate terms extracted from texts/images**

Taxonomy Expansion

- “Where to attach to?”
- Input: terms to add to the taxonomy, existing taxonomy
- Output: enriched taxonomy

Generic Solution

- Treat Taxonomy Enrichment as a two-stage approach

Term Extraction → **Taxonomy Expansion**

Term Extraction

- "Which term to attach?"
- Input: texts, images
- Output: candidate terms extracted from texts/images

Taxonomy Expansion

- **"Where to attach to?"**
- Input: terms to add to the taxonomy, existing taxonomy
- Output: enriched taxonomy

Short Answer -- Key Intuitions

- Generic
  - Hearst Patterns
  - Embeddings
- Product Specific
  - User Behaviors

Short Answer -- Key Intuitions

- Generic
    - **Hearst Patterns**
    - Embeddings
- Product Specific
    - User Behaviors

Rely on corpuses where parent-child terms co-occur

“The definition of a <u>sandal</u> ***is a type of*** <u>shoe</u> with straps that wrap around various parts of the foot and attach to a sole under the foot”

Short Answer -- Key Intuitions

- Generic
  - Hearst Patterns
  - **Embeddings**
- Product Specific
  - User Behaviors

Based on the *Distributional Inclusion Hypothesis*

- Assume that more general words like "animal" appear in a variety of different contexts, while more specific words like "cat" appear in a few specific contexts.
- When the contexts of "animal" include all the contexts of "cat", we can assume that "animal" is a hypernym of "cat".

Learn a better embedding for each term

Short Answer -- Key Intuitions

- Generic
  - Hearst Patterns
  - Embeddings
- **Product Specific**
  - **User Behaviors**
    - Search
    - Click
    - Add to cart
    - Purchase

Short Answer -- Key Intuitions

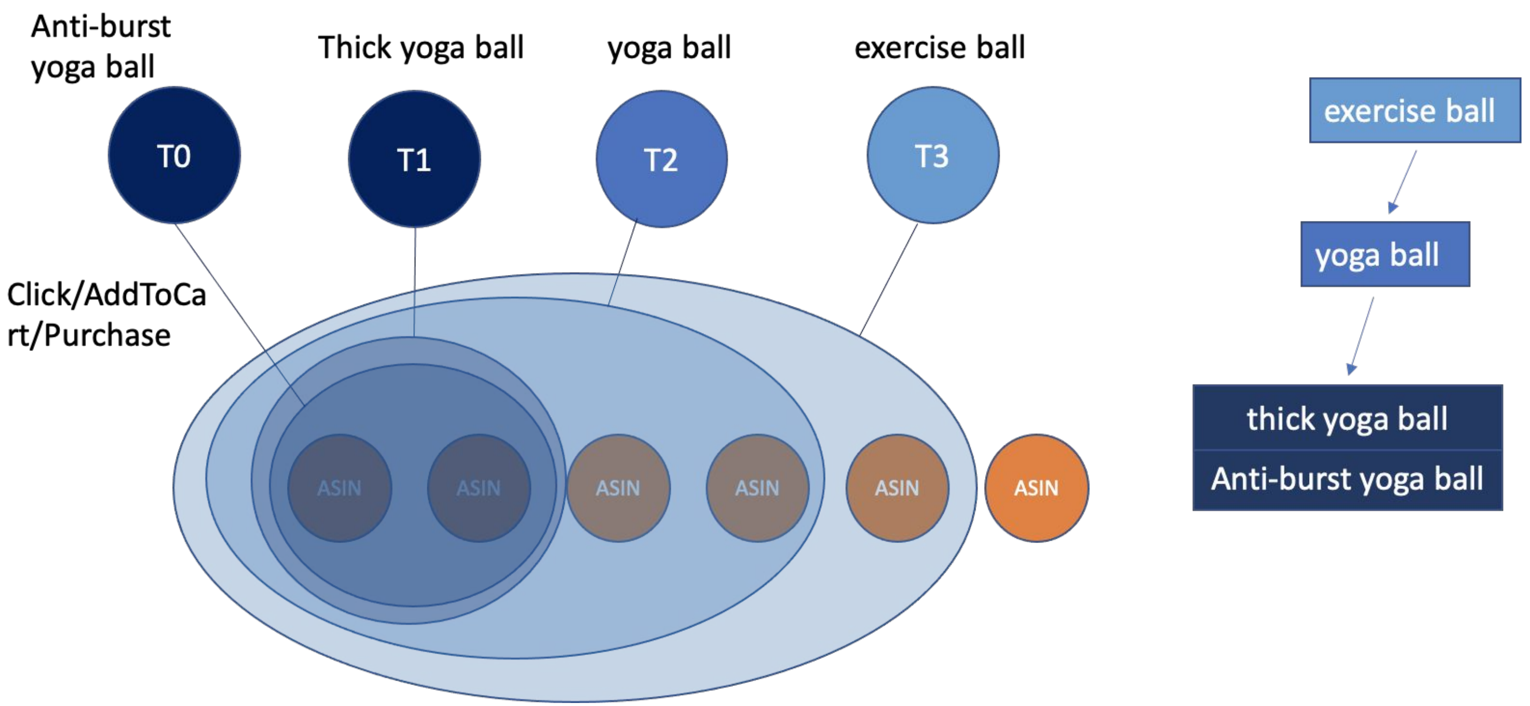
- Generic
  - Hearst Patterns
  - Embeddings
- **Product Specific**
  - **User Behaviors**
    - Search
    - Click
    - Add to cart
    - Purchase

Generic terms in search query result in diverse purchasing signals

Short Answer -- Key Intuitions

- Generic
  - Hearst Patterns
  - Embeddings
- **Product Specific**
  - **User Behaviors**
    - Search
    - Click
    - Add to cart
    - Purchase

Generic terms in search query result in diverse purchasing signals



Long Answer -- Hearst Patterns [Hearst 1992]

- Leveraging lexical patterns such as

Pattern
X which is a (example|class|kind|...) of Y
X (and|or) (any|some) other Y
X which is called Y
X is `JJS` (most)? Y
X a special case of Y
X is an Y that
X is a !(member|part|given) Y
!(features|properties) Y such as X_1, X_2, ...
(Unlike|like) (most|all|any|other) Y, X
Y including X_1, X_2, ...

Hearst, Automatic acquisition of hyponyms from large text corpora, COLING, 1992.
Roller et al., Hearst Patterns Revisited: Automatic Hypernym Detection from Large Text Corpora, ACL, 2018.

Long Answer -- Hearst Patterns [Hearst 1992]

- Leveraging lexical patterns such as

- Provide **high-quality** and **robust** predictions on large corpora by capturing important contextual constraints [Roller+ 2018]

Pattern
X which is a (example|class|kind|...) of Y
X (and|or) (any|some) other Y
X which is called Y
X is JJS (most)? Y
X a special case of Y
X is an Y that
X is a !(member|part|given) Y
!(features|properties) Y such as X_1, X_2, ...
(Unlike|like) (most|all|any|other) Y, X
Y including X_1, X_2, ...

Hearst, Automatic acquisition of hyponyms from large text corpora, COLING, 1992.
Roller et al., Hearst Patterns Revisited: Automatic Hypernym Detection from Large Text Corpora, ACL, 2018.

Long Answer -- Hearst Patterns [Hearst 1992]

- Leveraging lexical patterns such as

- Provide **high-quality** and **robust** predictions on large corpora by capturing important contextual constraints [Roller+ 2018]

May not able to find co-occurrence terms in e-commerce corpus, therefore low recall

Pattern
X which is a (example|class|kind|...) of Y
X (and|or) (any|some) other Y
X which is called Y
X is JJS (most)? Y
X a special case of Y
X is an Y that
X is a !(member|part|given) Y
!(features|properties) Y such as X_1, X_2, ...
(Unlike|like) (most|all|any|other) Y, X
Y including X_1, X_2, ...

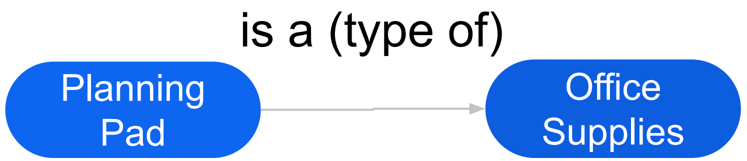
Hearst, Automatic acquisition of hyponyms from large text corpora, COLING, 1992.
Roller et al., Hearst Patterns Revisited: Automatic Hypernym Detection from Large Text Corpora, ACL, 2018.

Long Answer -- Hearst Patterns [Hearst 1992]

- Leveraging lexical patterns such as

- Provide **high-quality** and **robust** predictions on large corpora by capturing important contextual constraints [Roller+ 2018]

May not able to find co-occurrence terms in e-commerce corpus, therefore low recall



"Office Supplies" rarely appears in text profiles for planning pad products

Pattern
X which is a (example|class|kind|...) of Y
X (and|or) (any|some) other Y
X which is called Y
X is JJS (most)? Y
X a special case of Y
X is an Y that
X is a !(member|part|given) Y
!(features|properties) Y such as X_1, X_2, ...
(Unlike|like) (most|all|any|other) Y, X
Y including X_1, X_2, ...

Hearst, Automatic acquisition of hyponyms from large text corpora, COLING, 1992.
Roller et al., Hearst Patterns Revisited: Automatic Hypernym Detection from Large Text Corpora, ACL, 2018.

Long Answer -- Embeddings

- The founding blocks...

Hearst, Automatic acquisition of hyponyms from large text corpora, COLING, 1992.
Roller et al., Hearst Patterns Revisited: Automatic Hypernym Detection from Large Text Corpora, ACL, 2018.

Long Answer -- Embeddings

- The founding blocks...
- *Distributional Hypothesis* [Harris 1968]
    - Words that occur within similar contexts are semantically similar

Hearst, Automatic acquisition of hyponyms from large text corpora, COLING, 1992.
Roller et al., Hearst Patterns Revisited: Automatic Hypernym Detection from Large Text Corpora, ACL, 2018.

Long Answer -- Embeddings

- The founding blocks...
- *Distributional Hypothesis* [Harris 1968]
  - Words that occur within similar contexts are semantically similar
- *Distributional Inclusion Hypothesis* [Geffet & Dagan 2005]
  - We assume that more general words like "animal" appear in a variety of different contexts, while more specific words like "cat" appear in a few specific contexts.
  - When the contexts of "animal" include all the contexts of "cat", we can assume that "animal" is a hypernym of "cat".

Hearst, Automatic acquisition of hyponyms from large text corpora, COLING, 1992.
Roller et al., Hearst Patterns Revisited: Automatic Hypernym Detection from Large Text Corpora, ACL, 2018.

Long Answer -- Contexts

- Unstructured: neighboring words

“hiking shoes”

“This waterproof hiking shoe is ready for off-road adventure.”
“Leave no trail untrekked with this all-weather hiking shoe.”
“This brilliantly versatile hiking shoe cushions the foot well.”

Long Answer -- Contexts

- Unstructured: neighboring words

“hiking shoes”

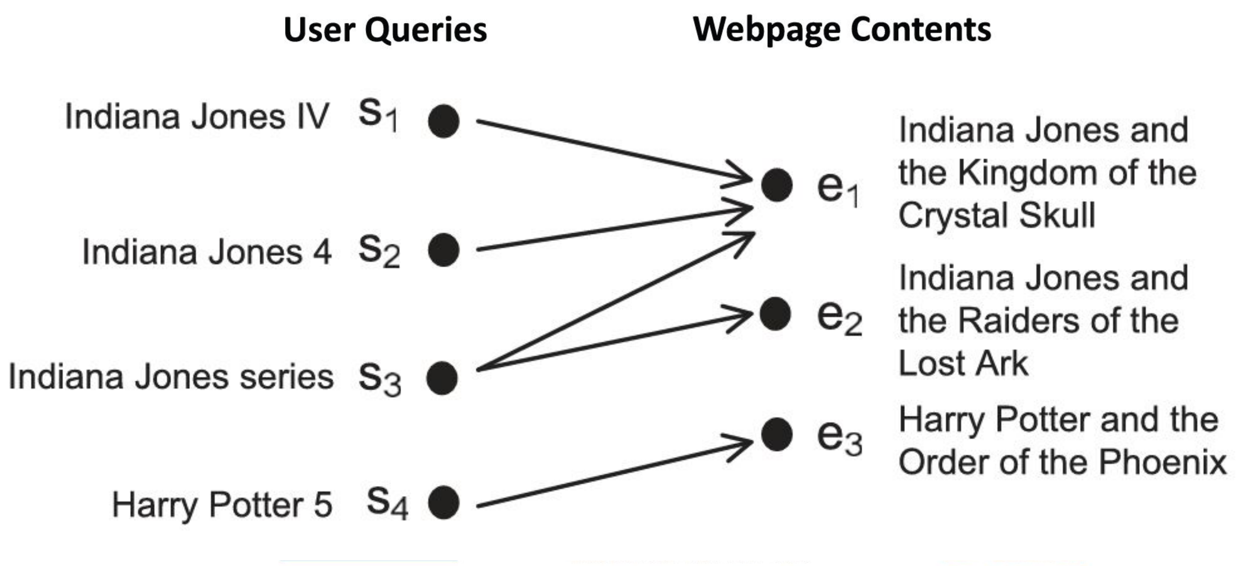
“This waterproof hiking shoe is ready for off-road adventure.”
“Leave no trail untrekked with this all-weather hiking shoe.”
“This brilliantly versatile hiking shoe cushions the foot well.”

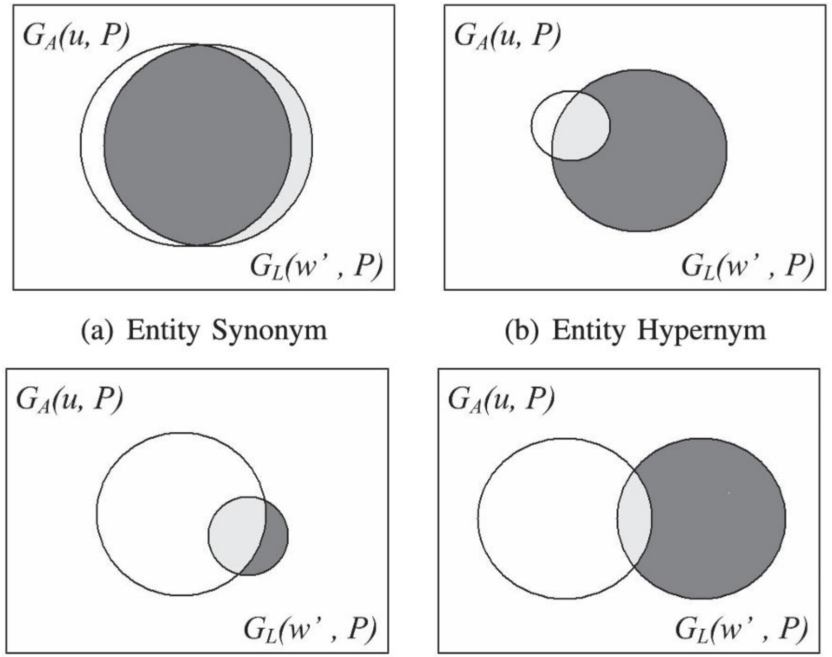
- Structured: clicked/purchased products [Cheng + 2011]
  - A broad term in the query has diverse results across multiple subcategories (“clothes for women”, "TV").
  - A specific term in the query has concentrated results over few products (“OLED 4K TV”).

Cheng et al., Entity synonyms for structured web search, TKDE, 2011

Long Answer -- Structured Contexts

- From term mentioned in search queries to clicked/purchased products



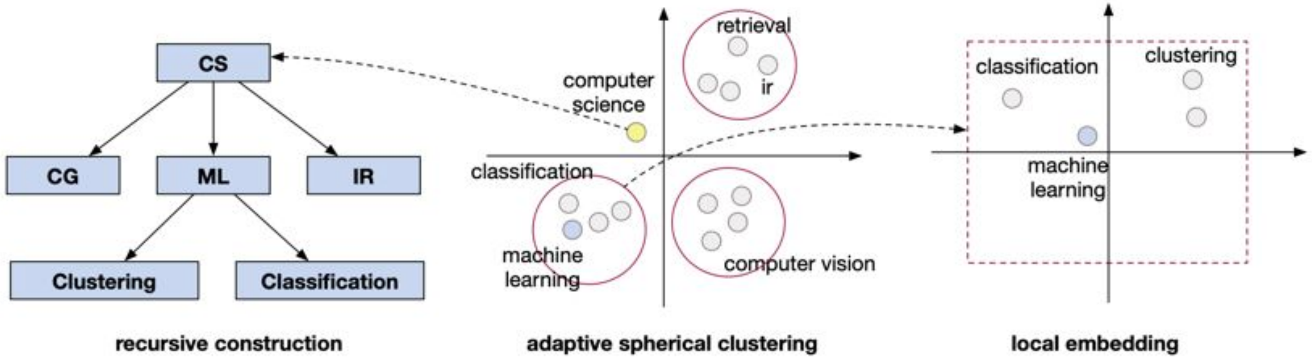


(a) Entity Synonym (b) Entity Hypernym (c) Entity Hyponym (d) Not Equivalent

Cheng et al., Entity synonyms for structured web search, TKDE, 2011

Long Answer -- Unstructured Contexts

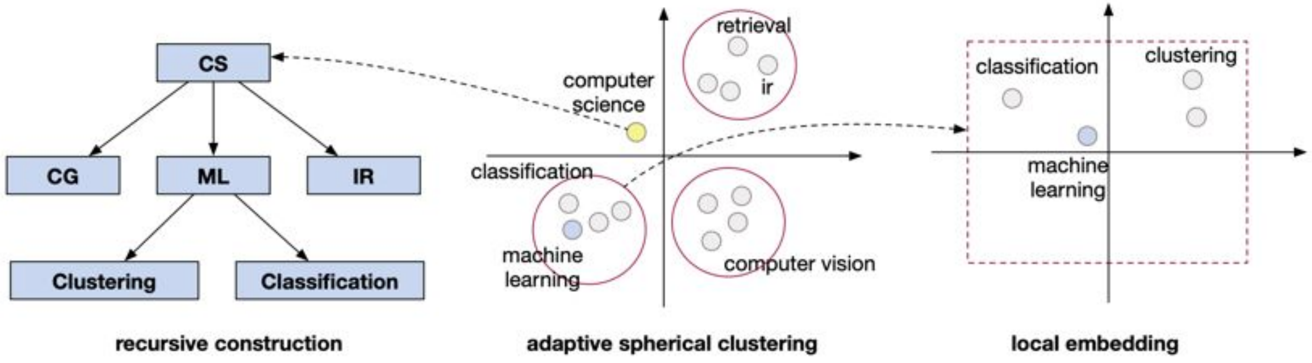
- For Unsupervised Hypernym Detection [Zhang+ 2018]
  - Recursive clustering process
  - An adaptive spherical clustering module to split coarse topics
  - A local embedding module to learn topic-specific embeddings



Zhang et al., Taxogen: Unsupervised topic taxonomy construction by adaptive term embedding and clustering. SigKDD, 2018.

Long Answer -- Unstructured Contexts

- For Unsupervised Hypernym Detection [Zhang+ 2018]
  - Recursive clustering process
  - An adaptive spherical clustering module to split coarse topics
  - A local embedding module to learn topic-specific embeddings



Totally unsupervised? The existing taxonomy is incomplete, but still gives meaningful supervisions.

Zhang et al., Taxogen: Unsupervised topic taxonomy construction by adaptive term embedding and clustering. SigKDD, 2018.

Long Answer -- Lexical Features for Supervised Hypernym Detection

- Supervised methods formulate hypernym detection as a term pair classification task

Boteanu et al., Synonym expansion for large shopping taxonomies, AKBC, 2018

Long Answer -- Lexical Features for Supervised Hypernym Detection

- Supervised methods formulate hypernym detection as a term pair classification task
- Based on Lexical Features [Boteanu+ 2019]

Boteanu et al., Synonym expansion for large shopping taxonomies, AKBC, 2018

Long Answer -- Lexical Features for Supervised Hypernym Detection

- Supervised methods formulate hypernym detection as a term pair classification task
- Based on Lexical Features [Boteanu+ 2019]
  - Local features such as
    - Character and word edit distance

Boteanu et al., Synonym expansion for large shopping taxonomies, AKBC, 2018

Long Answer -- Lexical Features for Supervised Hypernym Detection

- Supervised methods formulate hypernym detection as a term pair classification task
- Based on Lexical Features [Boteanu+ 2019]
  - Local features such as
    - Character and word edit distance
    - Embedding cosine similarity

Boteanu et al., Synonym expansion for large shopping taxonomies, AKBC, 2018

Long Answer -- Lexical Features for Supervised Hypernym Detection

- Supervised methods formulate hypernym detection as a term pair classification task
- Based on Lexical Features [Boteanu+ 2019]
  - Local features such as
    - Character and word edit distance
    - Embedding cosine similarity
  - Structure features such as
    - The node's parent name
    - The name of the taxonomy root (Indicate the department, such as "fashion")

Boteanu et al., Synonym expansion for large shopping taxonomies, AKBC, 2018

Long Answer -- Lexical Features for Supervised Hypernym Detection

- Supervised methods formulate hypernym detection as a term pair classification task
- Based on Lexical Features [Boteanu+ 2019]
  - Local features such as
    - Character and word edit distance
    - Embedding cosine similarity
  - Structure features such as
    - The node's parent name
    - The name of the taxonomy root (Indicate the department, such as "fashion")
- Each parent/child term is represented as a concatenation of lexical features

Boteanu et al., Synonym expansion for large shopping taxonomies, AKBC, 2018

Long Answer -- Lexical Features for Supervised Hypernym Detection

- Supervised methods formulate hypernym detection as a term pair classification task
- Based on Lexical Features [Boteanu+ 2019]
  - Local features such as
    - Character and word edit distance
    - Embedding cosine similarity
  - Structure features such as
    - The node's parent name
    - The name of the taxonomy root (Indicate the department, such as "fashion")
- Each parent/child term is represented as a concatenation of lexical features

Not all hypernym pairs have shared lexical features

("shoe", "sandal") ("Fish", "Salmon")

Boteanu et al., Synonym expansion for large shopping taxonomies, AKBC, 2018

Long Answer -- Lexical Features for Supervised Hypernym Detection

- Supervised methods formulate hypernym detection as a term pair classification task
- Based on Lexical Features [Boteanu+ 2019]
  - Local features such as
    - Character and word edit distance
    - Embedding cosine similarity
  - Structure features such as
    - The node's parent name
    - The name of the taxonomy root (Indicate the department, such as "fashion")
- Each parent/child term is represented as a concatenation of lexical features

Not all hypernym pairs have shared lexical features

("shoe", "sandal") ("Fish", "Salmon")

Does not model interactions between parent-child terms explicitly

Boteanu et al., Synonym expansion for large shopping taxonomies, AKBC, 2018

Long Answer -- Supervised Hypernym Detection using Graph Neural Networks

- Learning better term representations via Graph Neural Networks [Mao+ 2020][Shen+ 2020][Zeng+ 2021]

Shen et al., TaxoExpan: Self-supervised taxonomy expansion with position-enhanced graph neural network, WWW, 2020.

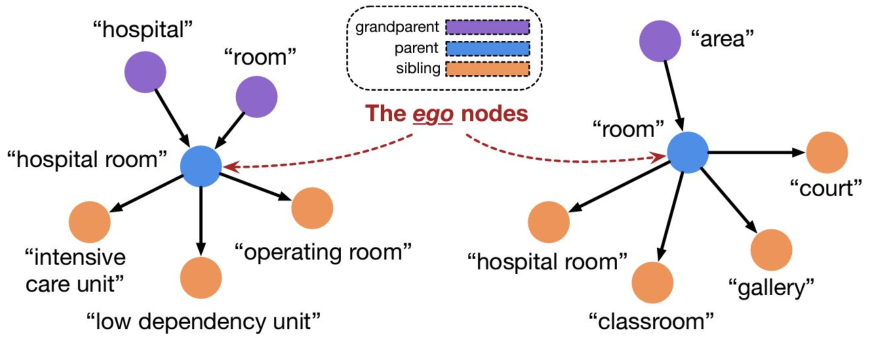
Long Answer -- Supervised Hypernym Detection using Graph Neural Networks (1)

- Learning better term representations via Graph Neural Networks [Mao+ 2020][Shen+ 2020][Zeng+ 2021]
- Better encoding of the taxonomy itself using Position-enhanced GNN [Shen+ 2020]

Shen et al., TaxoExpan: Self-supervised taxonomy expansion with position-enhanced graph neural network, WWW, 2020.

Long Answer -- Supervised Hypernym Detection using Graph Neural Networks (1)

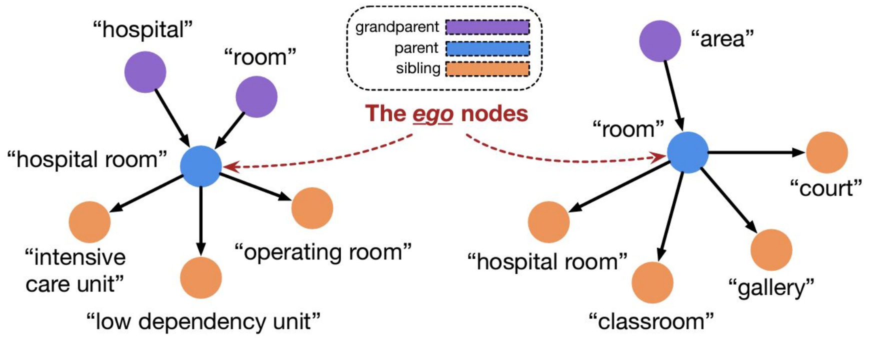
- Learning better term representations via Graph Neural Networks [Mao+ 2020][Shen+ 2020][Zeng+ 2021]
- Better encoding of the taxonomy itself using Position-enhanced GNN [Shen+ 2020]
  - Create an ego net for each term



Shen et al., TaxoExpan: Self-supervised taxonomy expansion with position-enhanced graph neural network, WWW, 2020.

Long Answer -- Supervised Hypernym Detection using Graph Neural Networks (1)

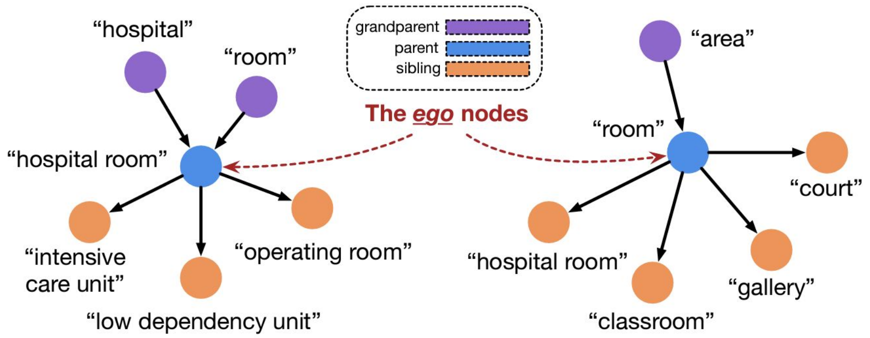
- Learning better term representations via Graph Neural Networks [Mao+ 2020][Shen+ 2020][Zeng+ 2021]
- Better encoding of the taxonomy itself using Position-enhanced GNN [Shen+ 2020]
    - Create an ego net for each term
    - Attach a learnable positional embedding {‘grandparent’,’parent’, ‘sibling’} to each node

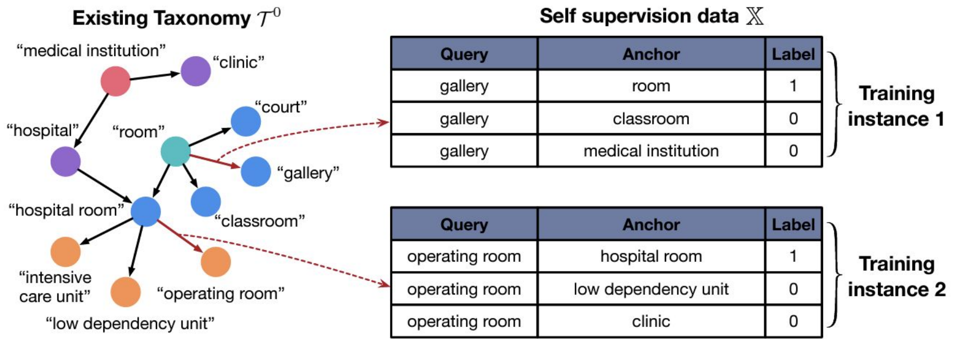


Shen et al., TaxoExpan: Self-supervised taxonomy expansion with position-enhanced graph neural network, WWW, 2020.

Long Answer -- Supervised Hypernym Detection using Graph Neural Networks (1)

- Learning better term representations via Graph Neural Networks [Mao+ 2020][Shen+ 2020][Zeng+ 2021]
- Better encoding of the taxonomy itself using Position-enhanced GNN [Shen+ 2020]
  - Create an ego net for each term
  - Attach a learnable positional embedding {‘grandparent’,’parent’, ‘sibling’} to each node
  - Training data generated from the existing taxonomy





Query	Anchor	Label
gallery	room	1
gallery	classroom	0
gallery	medical institution	0

Query	Anchor	Label
operating room	hospital room	1
operating room	low dependency unit	0
operating room	clinic	0

Shen et al., TaxoExpan: Self-supervised taxonomy expansion with position-enhanced graph neural network, WWW, 2020.

Long Answer -- Supervised Hypernym Detection using Graph Neural Networks (2)

- Besides interactions on taxonomy nodes……
- Construct a graph that connects terms through products/queries/taxonomy nodes [Mao+ 2020]

Mao et al., Octet: Online Catalog Taxonomy Enrichment with Self-Supervision, SigKDD, 2020.

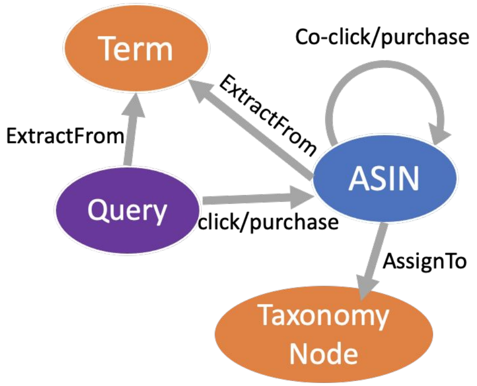
Long Answer -- Supervised Hypernym Detection using Graph Neural Networks (2)

- Besides interactions on taxonomy nodes……
- Construct a graph that connects terms through products/queries/taxonomy nodes [Mao+ 2020]
  - Term ⇨ Product — Remember we extract terms from products
  - Product ⇨ Taxonomy Node — Leverage existing taxonomy-node assignment to help
  - Query ⇨ Product ⇨ Term — Search log also tells us how term are linked with queries and corresponding clicked/purchased products
  - Product ⇨ Product — Co-click/co-purchase relationship between products

Mao et al., Octet: Online Catalog Taxonomy Enrichment with Self-Supervision, SigKDD, 2020.

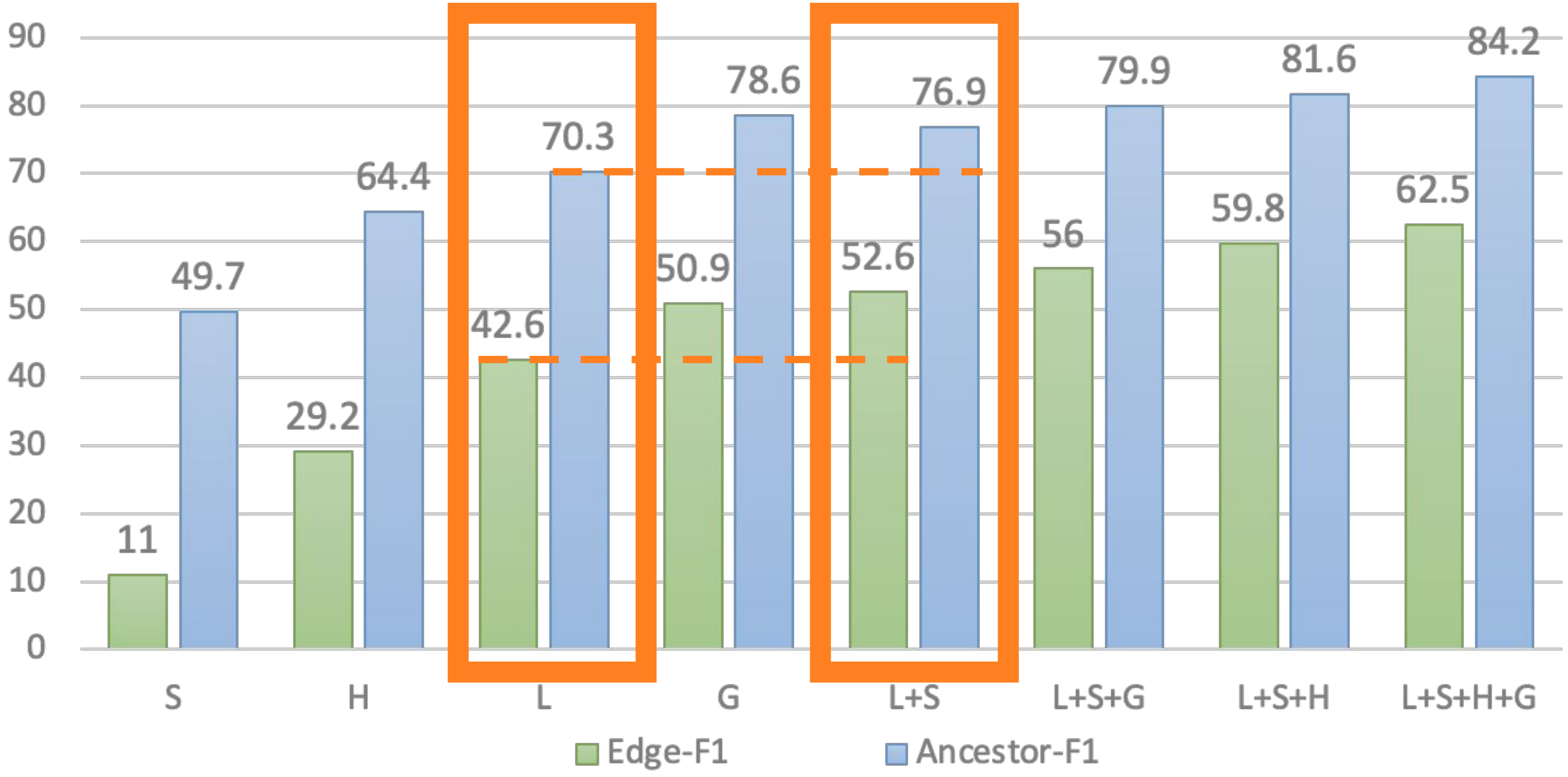
Long Answer -- Supervised Hypernym Detection using Graph Neural Networks (2)

- Besides interactions on taxonomy nodes……
- Construct a graph that connects terms through products/queries/taxonomy nodes [Mao+ 2020]
  - Term ⇒ Product — Remember we extract terms from products
  - Product ⇒ Taxonomy Node — Leverage existing taxonomy-node assignment to help
  - Query ⇒ Product ⇒ Term — Search log also tells us how term are linked with queries and corresponding clicked/purchased products
  - Product ⇒ Product — Co-click/co-purchase relationship between products
- Refine term embeddings using Relational Graph Convolutional Network (RGCN)



Mao et al., Octet: Online Catalog Taxonomy Enrichment with Self-Supervision, SigKDD, 2020.

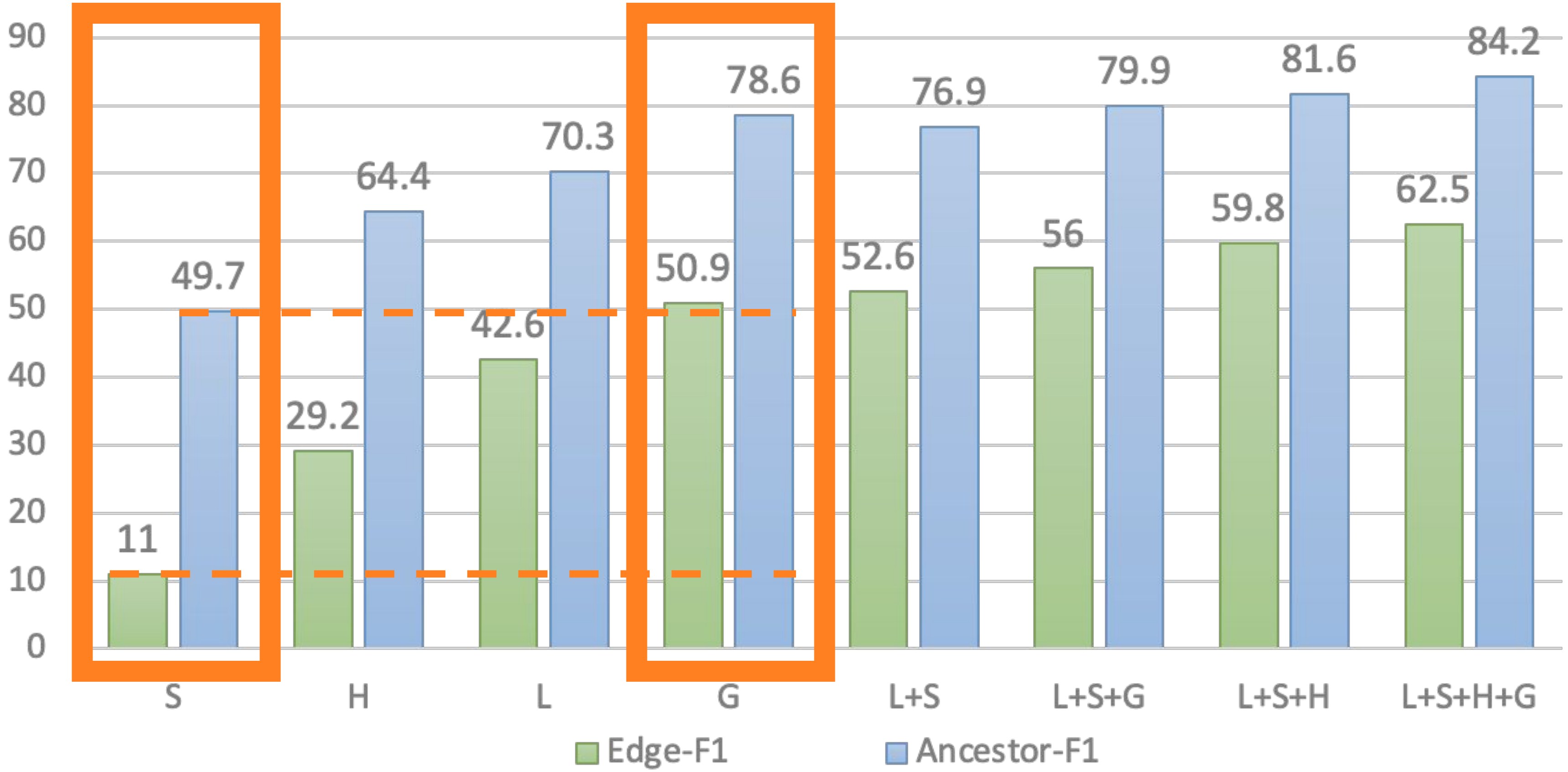
Long Answer -- Supervised Hypernym Detection using Graph Neural Networks (2)



S: Semantic feature
SH: Head word semantic feature
L: Lexical feature
G: Structured graph-based semantic feature

Semantic features add to lexical features

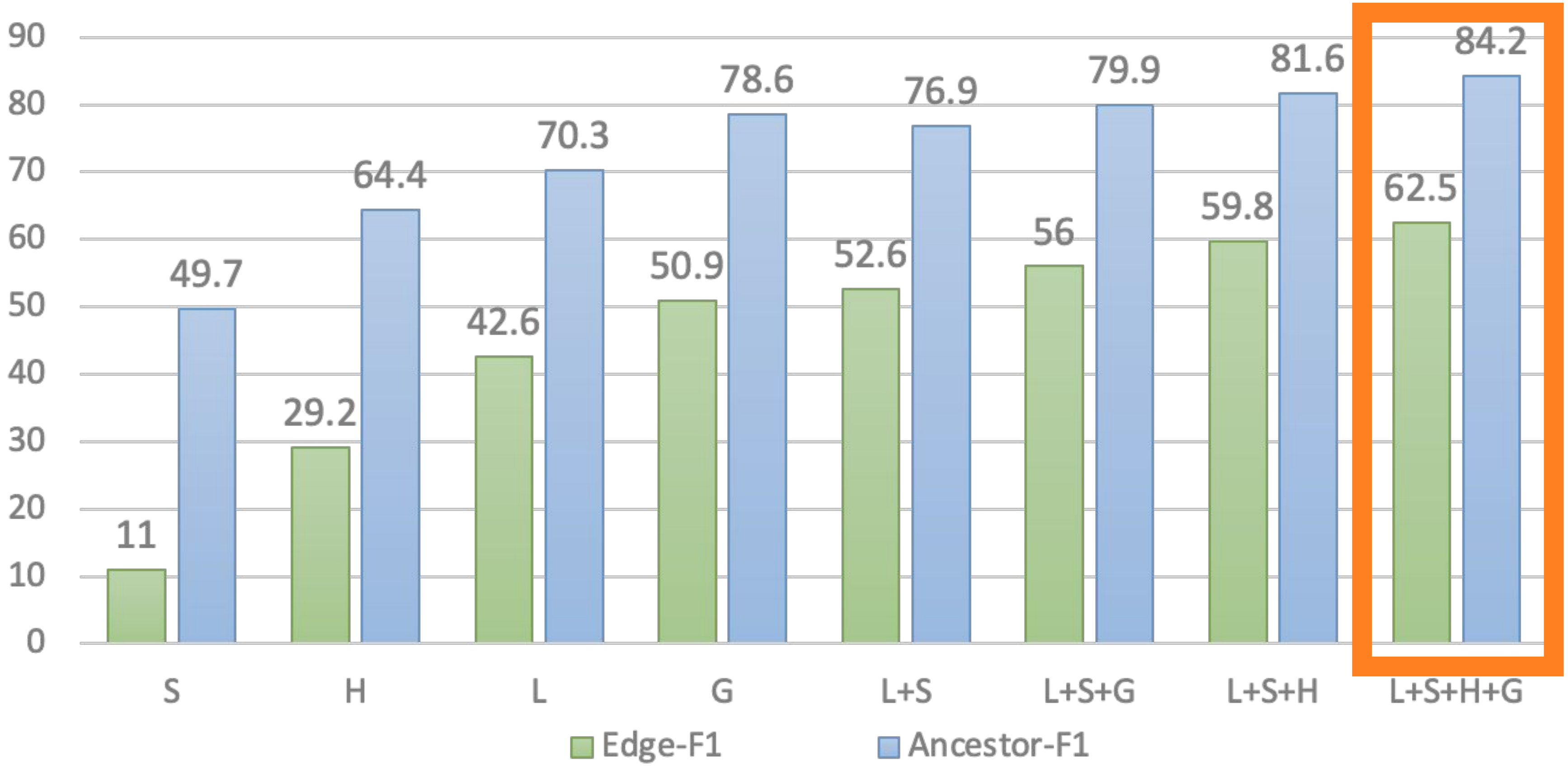
Long Answer -- Supervised Hypernym Detection using Graph Neural Networks (2)



S: Semantic feature
SH: Head word semantic feature
L: Lexical feature
G: Structured graph-based semantic feature

Graph structure significantly improves semantic features

Long Answer -- Supervised Hypernym Detection using Graph Neural Networks (2)



S: Semantic feature
SH: Head word semantic feature
L: Lexical feature
G: Structured graph-based semantic feature

Best performance achieved using features from all sources

Reflection/short-answer

- **Definition**: Taxonomy enrichment organizes product categories in a neat hierarchy

Reflection/short-answer

- **Definition**: Taxonomy enrichment organizes product categories in a neat hierarchy
- **Recipe**:
    - Term Extraction: which new term to attach
    - Term Attachment: where to attach

Reflection/short-answer

- **Definition**: Taxonomy enrichment organizes product categories in a neat hierarchy
- **Recipe**:
  - Term Extraction: which new term to attach
  - Term Attachment: where to attach
- **Key to Success for Products**: Leverage customer behavior

Reflection/short-answer

- **Definition**: Taxonomy enrichment organizes product categories in a neat hierarchy
- **Recipe**:
  - Term Extraction: which new term to attach
  - Term Attachment: where to attach
- **Key to Success for Products**: Leverage customer behavior
- **Applicable to other domains**:

Reflection/short-answer

- **Definition**: Taxonomy enrichment organizes product categories in a neat hierarchy
- **Recipe**:
  - Term Extraction: which new term to attach
  - Term Attachment: where to attach
- **Key to Success for Products**: Leverage customer behavior
- **Applicable to other domains**:
  - Fine-grained categorization space

Reflection/short-answer

- **Definition**: Taxonomy enrichment organizes product categories in a neat hierarchy
- **Recipe**:
  - Term Extraction: which new term to attach
  - Term Attachment: where to attach
- **Key to Success for Products**: Leverage customer behavior
- **Applicable to other domains**:
  - Fine-grained categorization space
  - Emerging categories when freshness is crucial

Reflection/short-answer

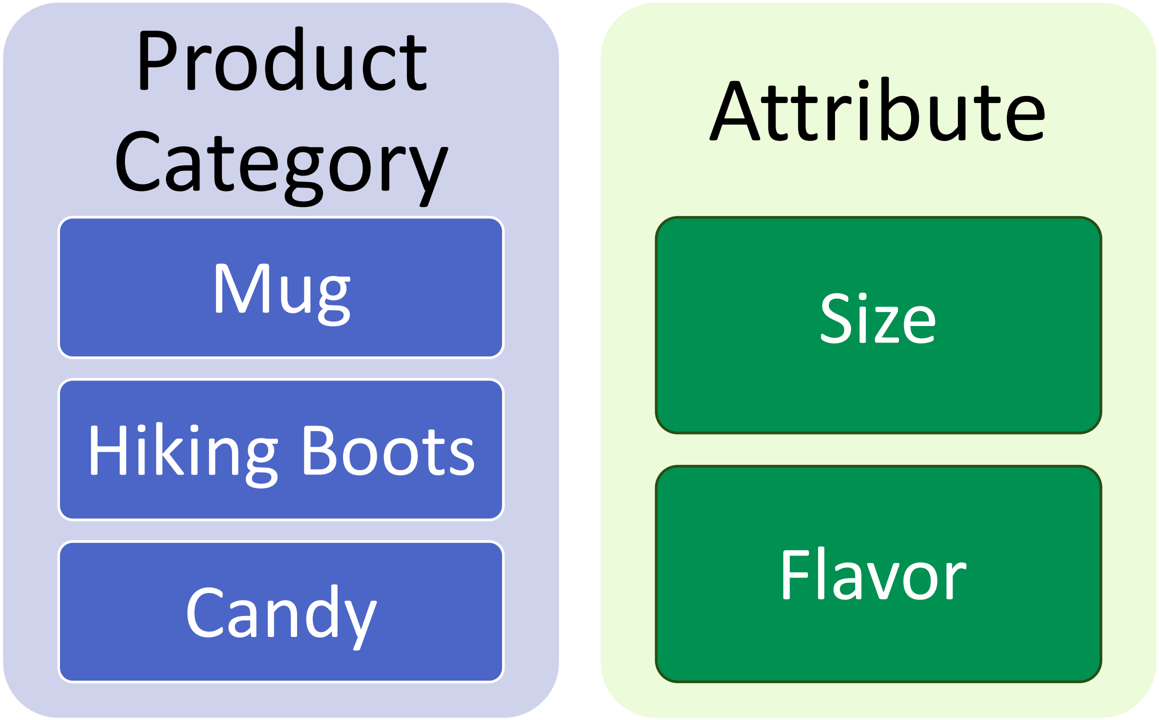
- **Definition**: Taxonomy enrichment organizes product categories in a neat hierarchy
- **Recipe**:
    - Term Extraction: which new term to attach
    - Term Attachment: where to attach
- **Key to Success for Products**: Leverage customer behavior
- **Applicable to other domains**:
    - Fine-grained categorization space
    - Emerging categories when freshness is crucial
    - Scarce direct supervision but rich structured signals

Relation Discovery

Attribute Applicability, Attribute Importance

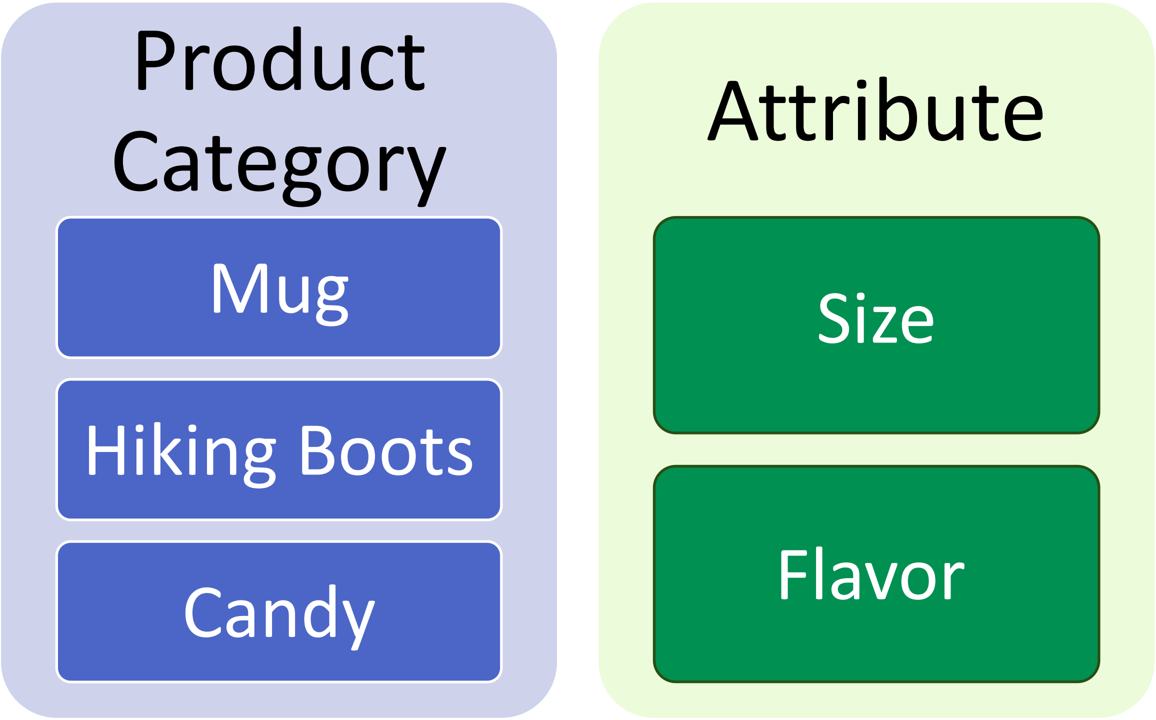
What is Relation Discovery ?

- Given
  - A set of product categories V
  - A set of attributes A
- Relation Discovery aims to solve:



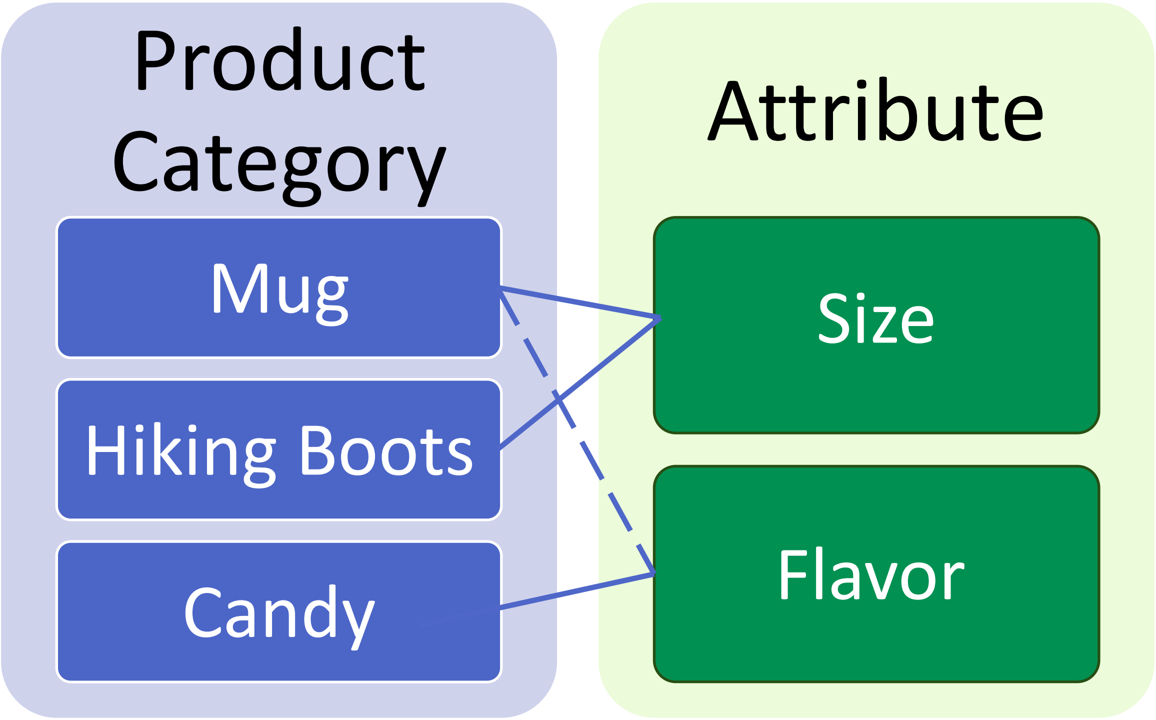
What is Relation Discovery ?

- Given
  - A set of product categories V
  - A set of attributes A
- Relation Discovery aims to solve:
  - Attribute Applicability
    - "Is an attribute applicable to one product category?"
    - Maps ($v \in V$, $a \in A$) to {0,1}



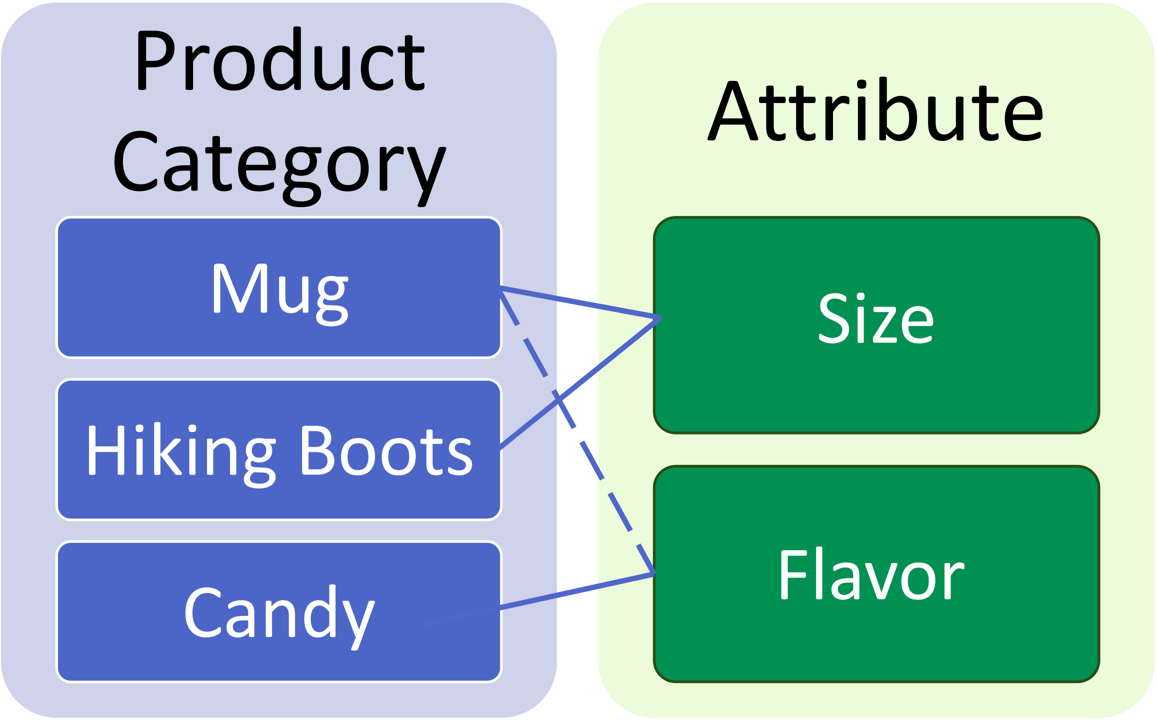
What is Relation Discovery ?

- Given
  - A set of product categories V
  - A set of attributes A
- Relation Discovery aims to solve:
  - Attribute Applicability
    - “Is an attribute applicable to one product category?”



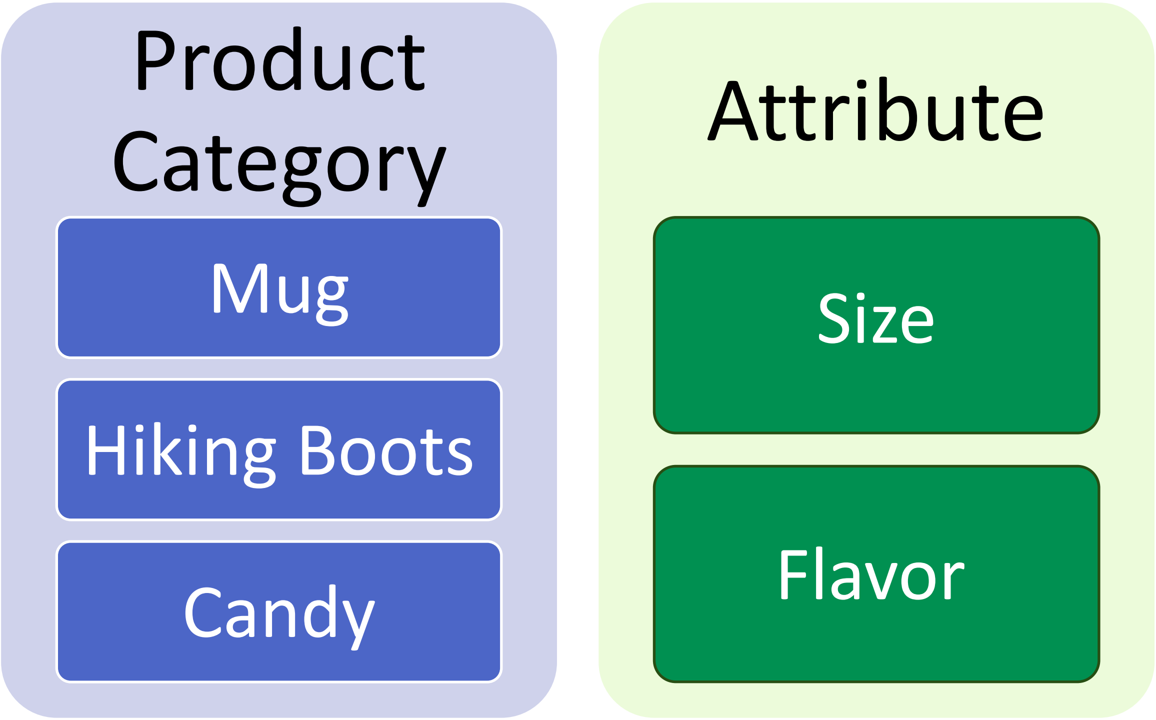
What is Relation Discovery ?

- Given
  - A set of product categories V
  - A set of attributes A
- Relation Discovery aims to solve:
  - Attribute Applicability
    - "Is an attribute applicable to one product category?"
    - Maps (v ∈ V, a ∈ A) to {0,1}



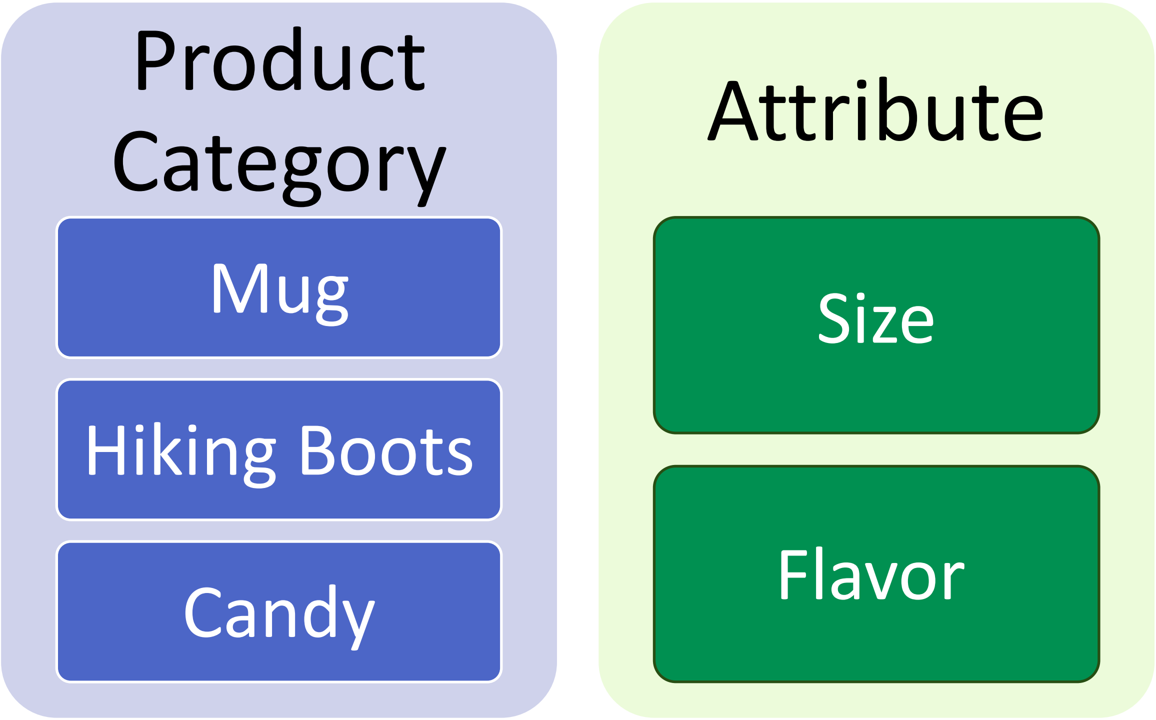
What is Relation Discovery ?

- Given
  - A set of product categories V
  - A set of attributes A
- Relation Discovery aims to solve:
  - Attribute Applicability
    - "Is an attribute applicable to one product category?"
    - Maps ($v \in V$, $a \in A$) to {0,1}
  - Attribute Importance
    - "Is an attribute important when people are making their purchase decisions?"



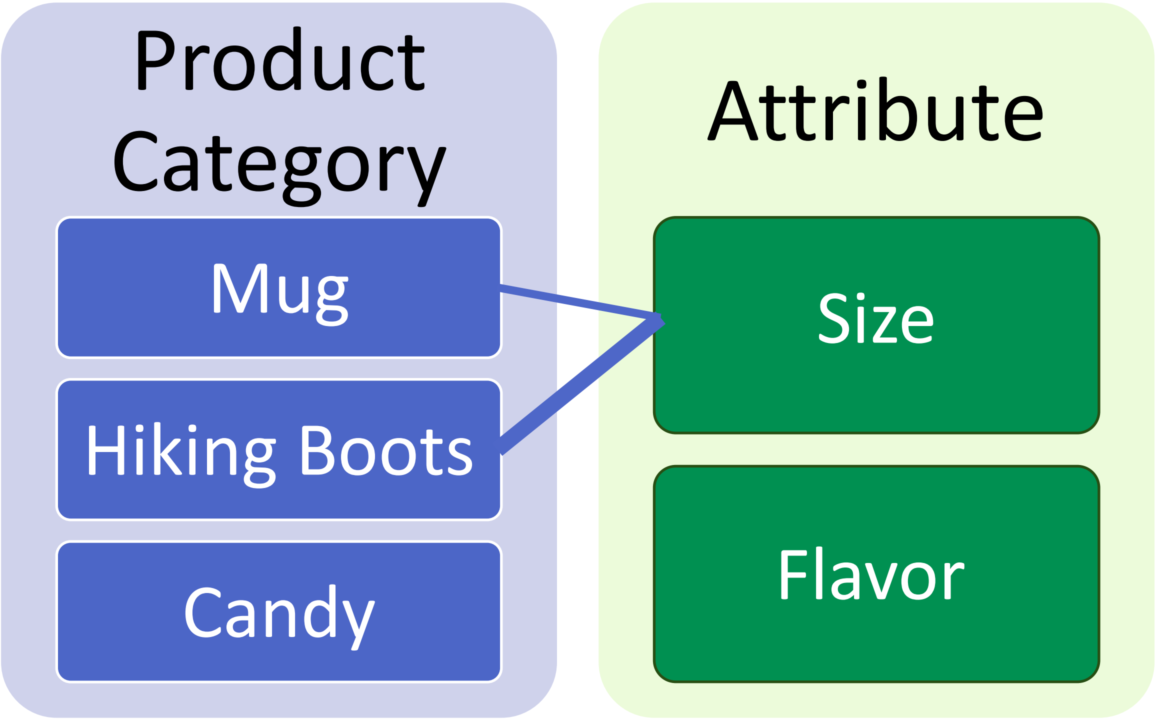
What is Relation Discovery ?

- Given
  - A set of product categories V
  - A set of attributes A
- Relation Discovery aims to solve:
  - Attribute Applicability
    - "Is an attribute applicable to one product category?"
    - Maps ($v \in V$, $a \in A$) to {0,1}
  - Attribute Importance
    - "Is an attribute important when people are making their purchase decisions?"
    - Maps ($v \in V$, $a \in A$) to [0,1]



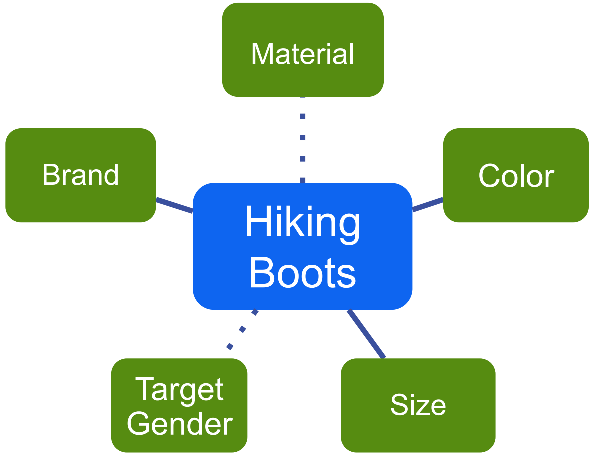
What is Relation Discovery ?

- Given
  - A set of product categories V
  - A set of attributes A
- Relation Discovery aims to solve:
  - Attribute Applicability
    - “Is an attribute applicable to one product category?”
    - Maps (v ∈ V, a ∈ A) to {0,1}
  - Attribute Importance
    - “Is an attribute important when people are making their purchase decisions?”
    - Maps (v ∈ V, a ∈ A) to [0,1]



Why Attribute Applicability?

- Discover an applicable, known aspect of a new product



Why Attribute Applicability?

- Discover an applicable, known aspect of a new product
- Regularize attribute value extraction results

Why Attribute Applicability?

- Discover an applicable, known aspect of a new product
- Regularize attribute value extraction results
  - Whenever attribute value extraction model generates **Flavor** values for **Mug** products, they are most likely incorrect.

Why Attribute Importance?

- Even if an attribute is applicable to one product category, it may not be important to customers

Why Attribute Importance?

- Even if an attribute is applicable to one product category, it may not be important to customers
  - **Scent** is important to **Shampoo**

Roll over image to zoom in

Garnier Fructis Grow Strong Shampoo, 33.8 Ounces
Visit the Garnier Store
7,407 ratings | 8 answered questions
Amazon's Choice for "fructis shampoo"
Climate Pledge Friendly

Price: $6.97 ($0.21 / Fl Oz) Get **Fast, Free Shipping** with Amazon Prime & FREE Returns

Get $50 off instantly: Pay $0.00 ~~$6.97~~ upon approval for the Amazon Rewards Visa Card. No annual fee.

Brand	Garnier
Scent	Apple
Hair Type	Thin
Liquid Volume	33.8 Fluid Ounces
Item Weight	2.36 Pounds

About this item
- Fortifying Shampoo: Our Grow Strong Shampoo features Apple Extract and Ceramide to fortify hair as it grows to bring life back to every inch: stronger, healthier and shinier
- Healthier, Stronger Hair: This paraben free formula features Active Fruit Protein, our exclusive

Why Attribute Importance?

- Even if an attribute is applicable to one product category, it may not be important to customers
  - Scent is **important** to Shampoo
  - Scent is **applicable** to Candy, but Scent of a Candy rarely affect customers' shopping decisions, thus **less important** comparing with e.g. Flavor for Candy

Why Attribute Importance?

- Even if an attribute is applicable to one product category, it may not be important to customers
  - **Scent** is important to **Shampoo**
  - **Scent** is applicable to **Candy**, but **Scent** of a **Candy** rarely affect customers' shopping decisions, thus less important comparing with e.g. **Flavor**

Roll over image to zoom in

Garnier Fructis Grow Strong Shampoo, 33.8 Ounces

Visit the Garnier Store

7,407 ratings | 8 answered questions

Amazon's Choice for "fructis shampoo"

Climate Pledge Friendly

Price: $6.97 ($0.21 / Fl Oz) Get Fast, Free Shipping with Amazon Prime & FREE Returns

Brand	Garnier
Scent	Apple
Hair Type	Thin
Liquid Volume	33.8 Fluid Ounces
Item Weight	2.36 Pounds

Roll over image to zoom in

36 count

Original Candy, 2.17-Ounce 36 Individual Packs

Brand: Skittles

1,452 ratings

Amazon's Choice for "skittles pack"

Price: $32.04

Flavor	Original
Brand	Skittles
Ingredients	Sugar, Corn Syrup, Hydrogenated P Acid, Tapioca Dextrin, Modified Cor Colors (Red 40 Lake, Titanium Dioxi See more
Item Weight	2.17 Ounces
Color	Assorted

Short Answer -- Key Intuitions

- Instance-level: Products fall into the same category have similar applicable attributes

Short Answer -- Key Intuitions

- Instance-level: Products fall into the same category have similar applicable attributes
  - Guardrail is the homogeneity on product category definition

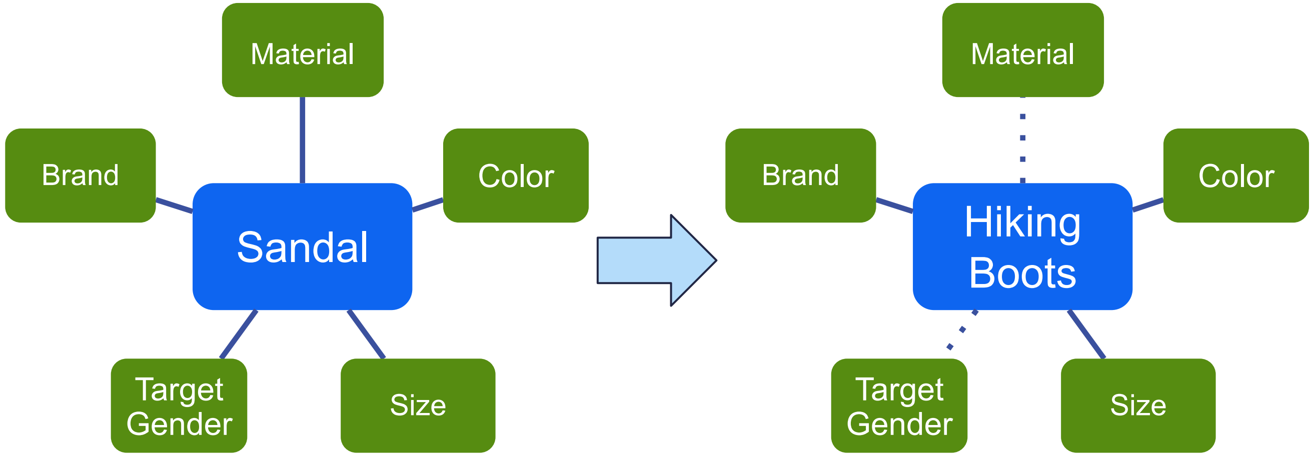
    "A product category identifies a group of real-world products based on similar visible and functional characteristics."

Short Answer -- Key Intuitions

- Instance-level: Products fall into the same category have similar applicable attributes
  - Guardrail is the homogeneity on product category definition

    "A product category identifies a group of real-world products based on similar visible and functional characteristics."

- Category-level: Two related product categories may share similar applicable attributes

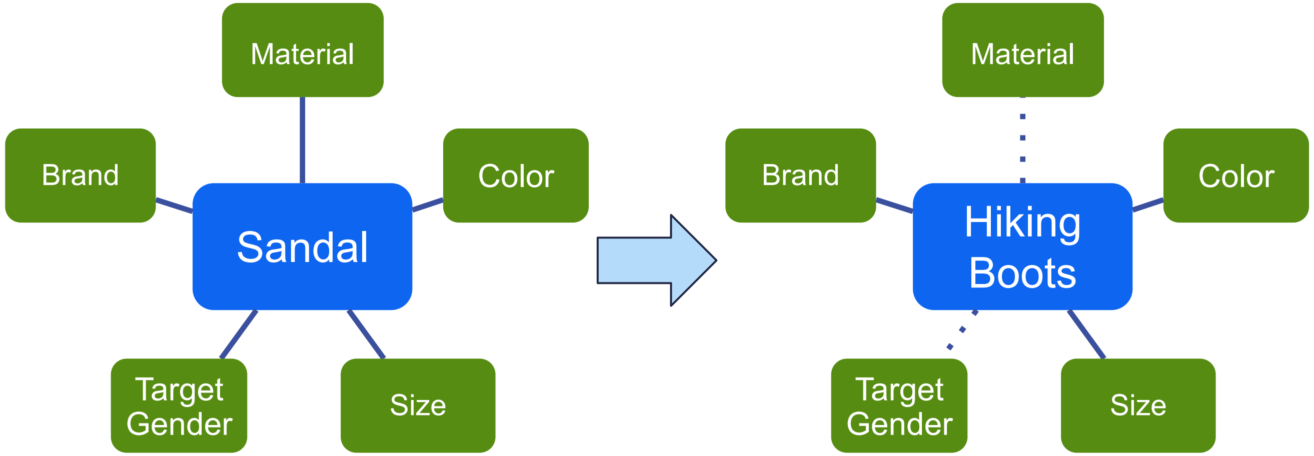


Short Answer -- Key Intuitions

- Instance-level: Products fall into the same category have similar applicable attributes
  - Guardrail is the homogeneity on product category definition

    "A product category identifies a group of real-world products based on similar visible and functional characteristics."

- Category-level: Two related product categories may share similar applicable attributes
  - **Graph mining** on existing links



Short Answer -- Key Intuitions

- Customers talk about attributes in product reviews
  - **Text mining** on customer reviews

    "I will take this to Iceland. Good material. Awesome looking."

    "I would think a half size up would be perfect for those wearing heavy socks but for me the sizing is true."

    "I am a size 12 in women's shoes"

Short Answer -- Key Intuitions

- Customers talk about attributes in product reviews
    - **Text mining** on customer reviews

Material

"I will take this to Iceland. Good material. Awesome looking."

"I would think a half size up would be perfect for those wearing heavy socks but for me the sizing is true."

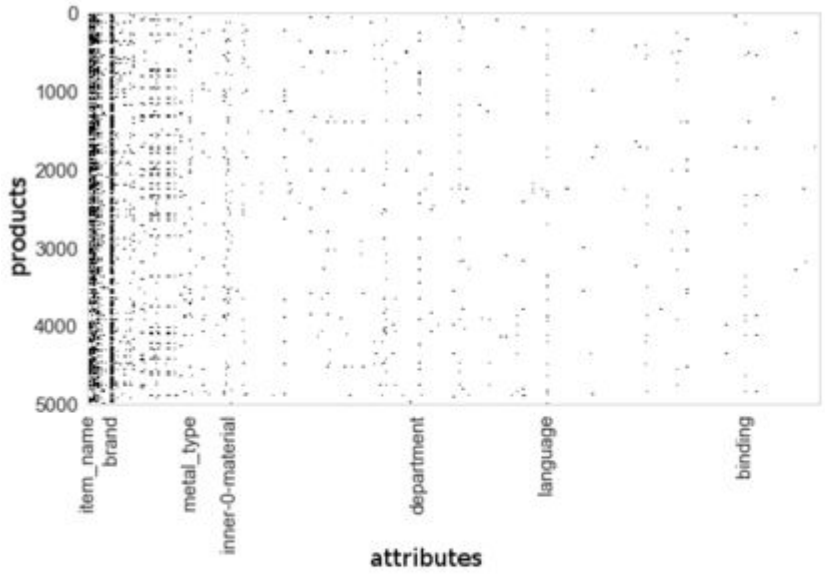
"I am a size 12 in women's shoes"

Size

Target Gender

Long Answer -- Applicability Prediction

- Applicability Prediction methods [Rukat+ 2017]
  - Works on a binary matrix between products and attributes

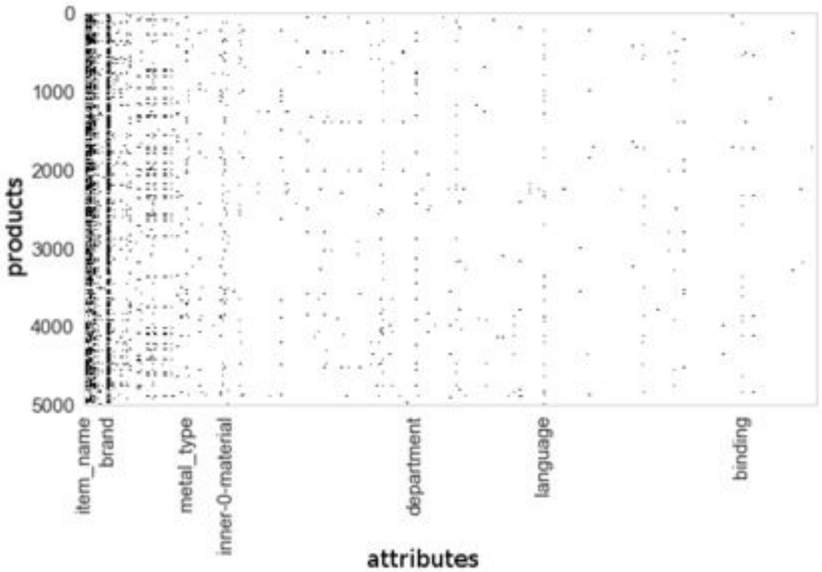


Rukat et al., An interpretable latent variable model for attribute applicability in the amazon catalogue, NIPS, 2017.

Long Answer -- Applicability Prediction

- Applicability Prediction methods [Rukat+ 2017]
  - Works on a binary matrix between products and attributes
  - Using Binary matrix factorization

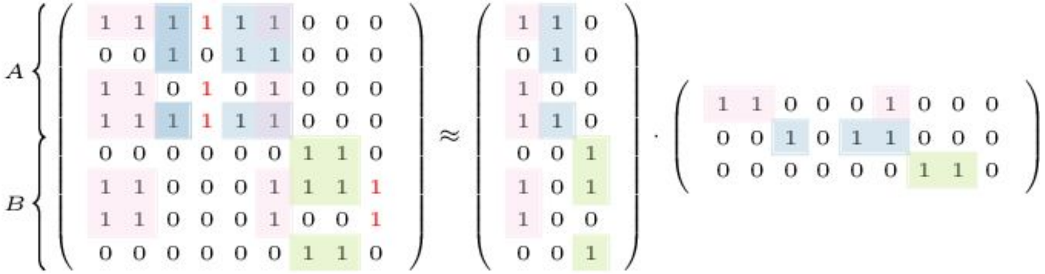
$$\begin{array}{r} A\left\{\begin{array}{c} \\ \\ \\ \\ \end{array}\right. \\ B\left\{\begin{array}{c} \\ \\ \\ \\ \end{array}\right. \end{array}\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix} \approx \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

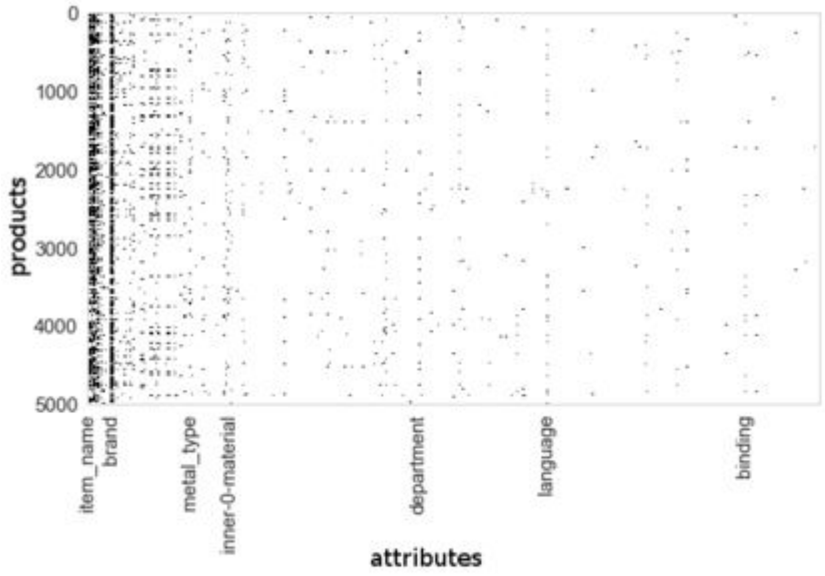


Rukat et al., An interpretable latent variable model for attribute applicability in the amazon catalogue, NIPS, 2017.

Long Answer -- Applicability Prediction

- Applicability Prediction methods [Rukat+ 2017]
  - Works on a binary matrix between products and attributes
  - Using Binary matrix factorization





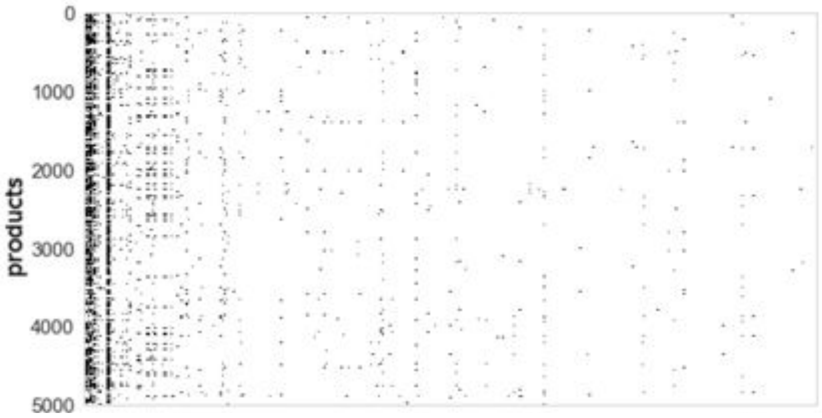
Attribute	cup-size	closure-type	leather-type
Product types with largest p(apply)	Bra 22(10)% Swimwear 3(2)% Underwear 3(2)% Shoes 2(2)% Suit 2(2)%	Shoes 48(18)% Pants 24(10)% Shorts 6(3)% Outerwear 4(2)% Bra 2(2)%	Shoes 48(15)% Outerwear 3(3)% Shorts 2(2)% (< 1%)

Rukat et al., An interpretable latent variable model for attribute applicability in the amazon catalogue, NIPS, 2017.

Long Answer -- Applicability Prediction

- Applicability Prediction methods [Rukat+ 2017]
  - Works on a binary matrix between products and attributes
  - Using Binary matrix factorization

$$
\begin{array}{c}
A \\ \\ \\ \\ B \\ \\ \\ \\
\end{array}
\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0
\end{pmatrix}
\approx
\begin{pmatrix}
1 & 1 & 0 \\
0 & 1 & 0 \\
1 & 0 & 0 \\
1 & 1 & 0 \\
0 & 0 & 1 \\
1 & 0 & 1 \\
1 & 0 & 0 \\
0 & 0 & 1
\end{pmatrix}
\cdot
\begin{pmatrix}
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0
\end{pmatrix}
$$



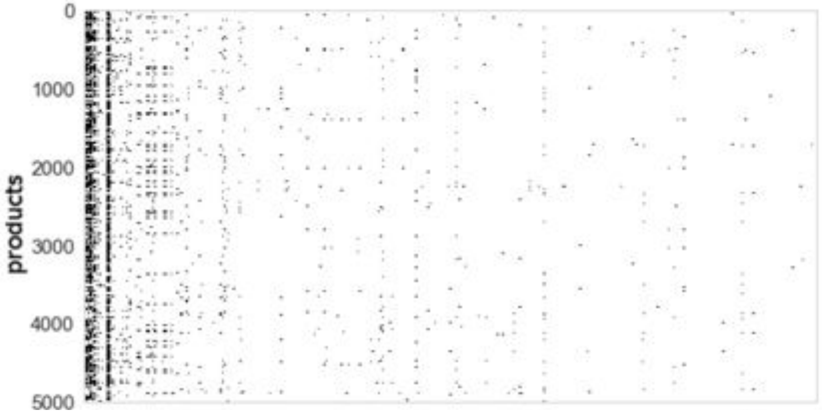
Attribute	cup-size	closure-type	leather-type
Product types with largest p(apply)	Bra 22(10)%	Shoes 48(18)%	Shoes 48(15)%
	Swimwear 3(2)%	Pants 24(10)%	Outerwear 3(3)%
	Underwear 3(2)%	Shorts 6(3)%	Shorts 2(2)%
	Shoes 2(2)%	Outerwear 4(2)%	(< 1%)
	Suit 2(2)%	Bra 2(2)%	

Does not leverage semantics of product categories & typical attribute values

Rukat et al., An interpretable latent variable model for attribute applicability in the amazon catalogue, NIPS, 2017.

Long Answer -- Applicability Prediction

- Applicability Prediction methods [Rukat+ 2017]
  - Works on a binary matrix between products and attributes
  - Using Binary matrix factorization



$$\begin{array}{c} A\left\{\begin{array}{c}\\ \\ \\ \\ \end{array}\right. \\ B\left\{\begin{array}{c}\\ \\ \\ \\ \end{array}\right. \end{array}\begin{pmatrix} 1&1&1&1&1&1&0&0&0\\ 0&0&1&0&1&1&0&0&0\\ 1&1&0&1&0&1&0&0&0\\ 1&1&1&1&1&1&0&0&0\\ 0&0&0&0&0&0&1&1&0\\ 1&1&0&0&0&1&1&1&1\\ 1&1&0&0&0&1&0&0&1\\ 0&0&0&0&0&0&1&1&0 \end{pmatrix} \approx \begin{pmatrix} 1&1&0\\ 0&1&0\\ 1&0&0\\ 1&1&0\\ 0&0&1\\ 1&0&1\\ 1&0&0\\ 0&0&1 \end{pmatrix} \cdot \begin{pmatrix} 1&1&0&0&0&1&0&0&0\\ 0&0&1&0&1&1&0&0&0\\ 0&0&0&0&0&0&1&1&0 \end{pmatrix}$$

Attribute	cup-size	closure-type	leather-type
Product types with largest p(apply)	Bra 22(10)% Swimwear 3(2)% Underwear 3(2)% Shoes 2(2)% Suit 2(2)%	Shoes 48(18)% Pants 24(10)% Shorts 6(3)% Outerwear 4(2)% Bra 2(2)%	Shoes 48(15)% Outerwear 3(3)% Shorts 2(2)% (< 1%)

Does not leverage semantics of product categories & typical attribute values

Discover new attributes?

Rukat et al., An interpretable latent variable model for attribute applicability in the amazon catalogue, NIPS, 2017.

Long Answer -- Attribute Importance

- From seller metadata
  - Manufacturer-provided attributes associated with product catalog

Long Answer -- Attribute Importance

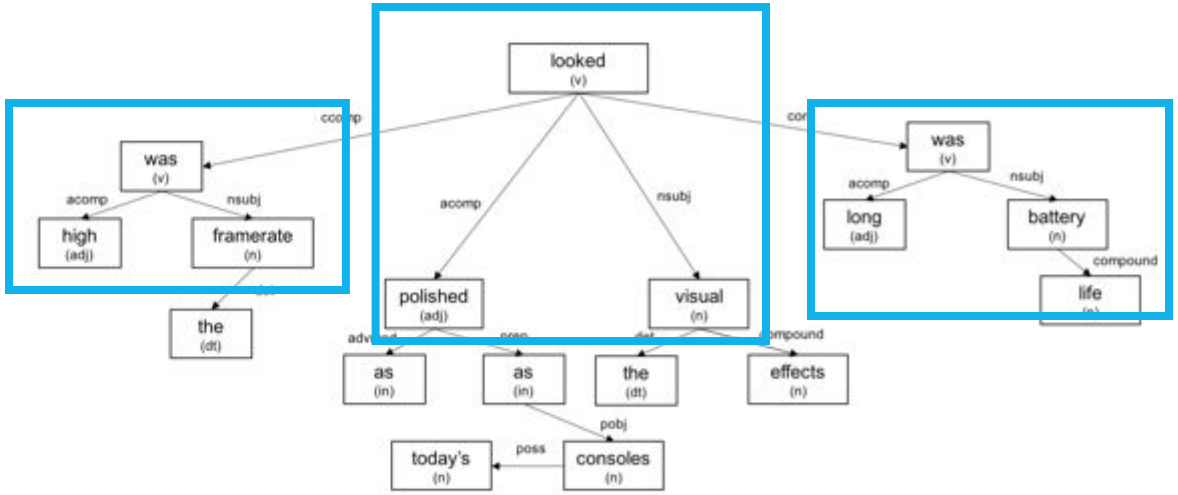
- From seller metadata
  - Manufacturer-provided attributes associated with product catalog
- From customer reviews
  - E.g. for an iPhone “old body style with a larger screen! ”
  - Rule-based [Popescu & Etzioni 2005]
  - Graph-based ranking methods: using PageRank [Yan+ 2015] [Indrakanti & Singh 2018]
  - Fine-grained LDA with keywords [Wang+ 2014]
  - Opinion mining with Fuzzy-c-means clustering [Zimmermann+ 2016]

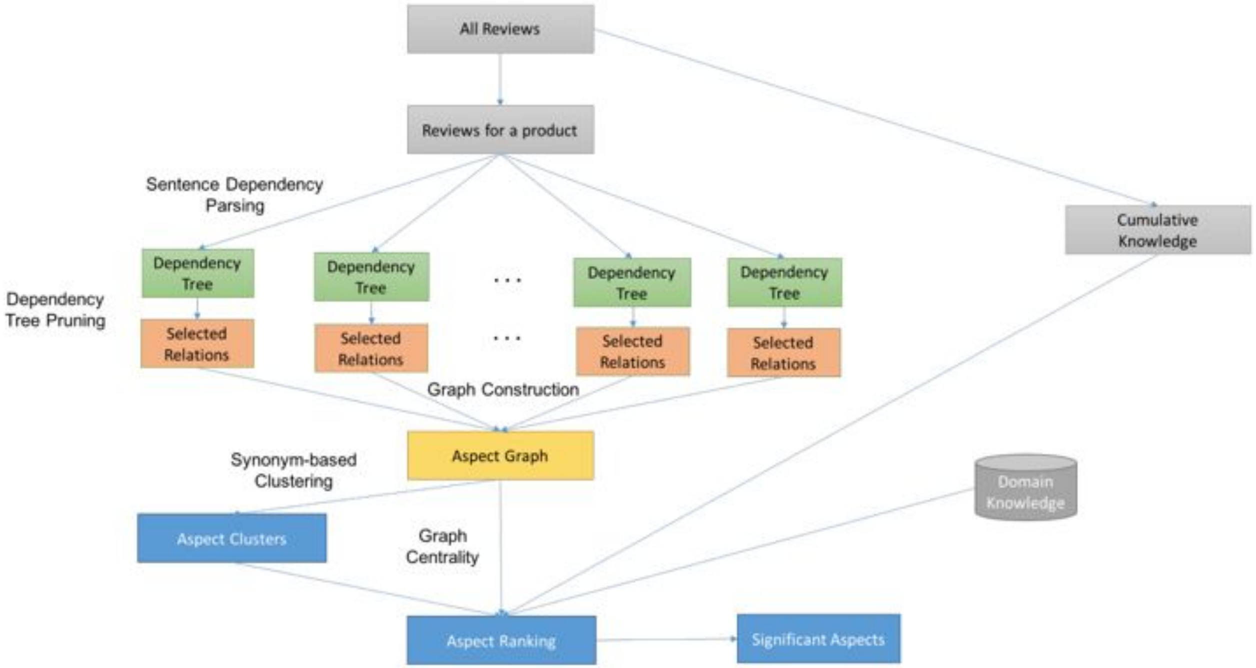
Long Answer -- Attribute Importance

- From seller metadata
  - Manufacturer-provided attributes associated with product catalog
- From customer reviews
  - E.g. for an iPhone "old body style with a larger screen! "
  - Rule-based [Popescu & Etzioni 2005]
  - Graph-based ranking methods: using PageRank [Yan+ 2015] [Indrakanti & Singh 2018]
  - Fine-grained LDA with keywords [Wang+ 2014]
  - Opinion mining with Fuzzy-c-means clustering [Zimmermann+ 2016]
- From customer search query log [Pound+ 2011], and behavior data [Zhou+ 2020]

Long Answer -- Attribute Importance

- Aspect Extraction from reviews
  - Aspect extraction
    via Dependency tree pruning





  - Aspect graph construction
  - Aspect ranking

Reflections/short-answers

- **Definition**: discover relations between product categories and attributes.
  - Attribute Applicability: “Is an attribute applicable to one product category?”
  - Attribute Importance: “Is an attribute important when people are making their purchase decisions?”

Reflections/short-answers

- **Definition**: discover relations between product categories and attributes.
  - Attribute Applicability: “Is an attribute applicable to one product category?”
  - Attribute Importance: “Is an attribute important when people are making their purchase decisions?”
- **Recipe**: Graph Mining and Text Mining

Reflections/short-answers

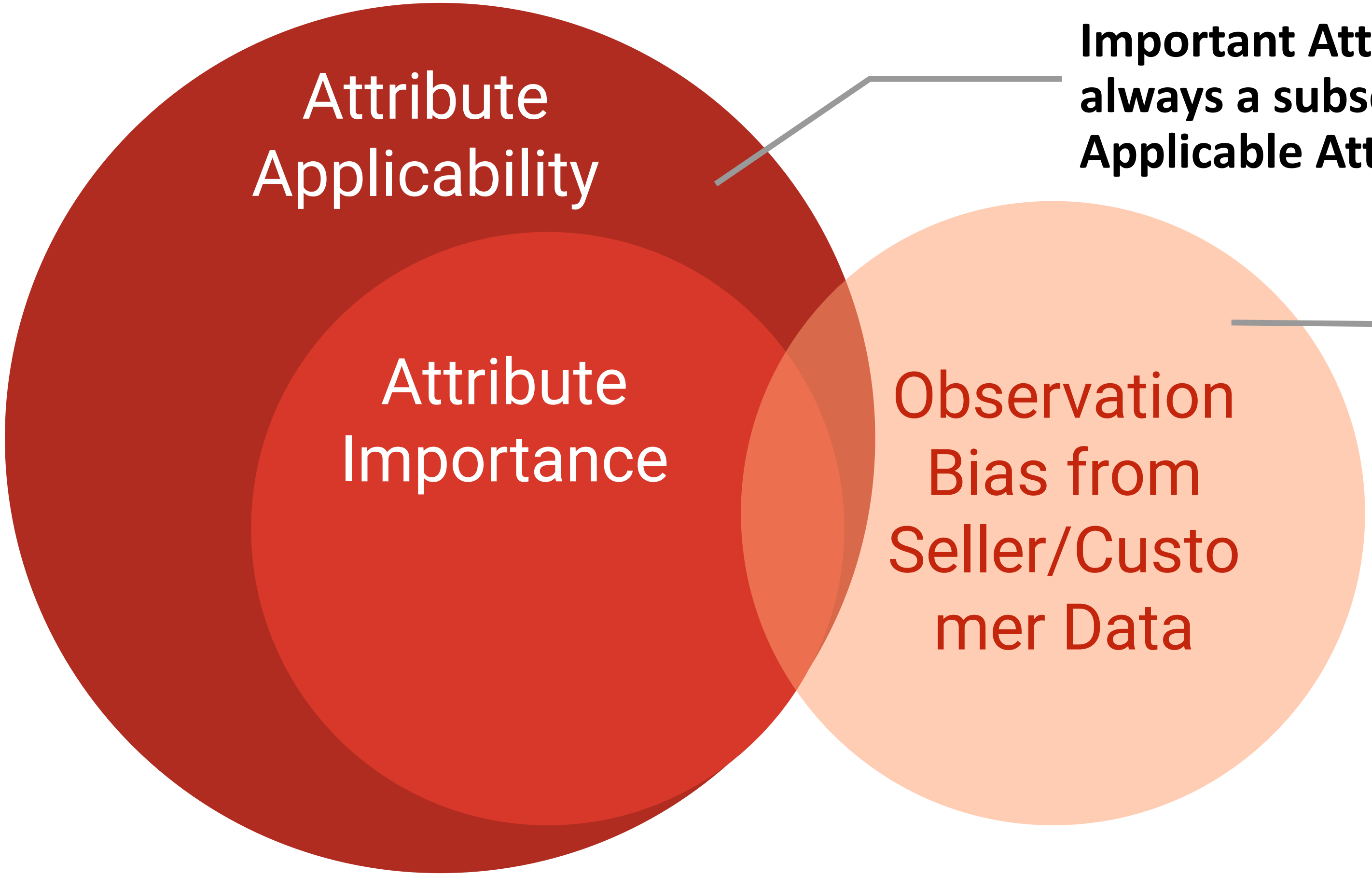
- **Definition**: discover relations between product categories and attributes.
  - Attribute Applicability: “Is an attribute applicable to one product category?”
  - Attribute Importance: “Is an attribute important when people are making their purchase decisions?”
- **Recipe**: Graph Mining and Text Mining
- **Key to Success for Products**: Leverage both seller/customer inputs

Reflections/short-answers

- **Definition**: discover relations between product categories and attributes.
    - Attribute Applicability: “Is an attribute applicable to one product category?”
    - Attribute Importance: “Is an attribute important when people are making their purchase decisions?”
- **Recipe**: Graph Mining and Text Mining
- **Key to Success for Products**: Leverage both seller/customer inputs
- **Applicability to other domains**:
    - An increasing variety of relations or predicate diversity
    - Quantify the relation strength

Reflections/short-answers

- Attribute Applicability, Importance, and Observation Biases



Important Attributes are always a subset of Applicable Attributes

Applicability information can be the "unspoken truth" if attributes are not important

- **Seller may not mention in product profiles**
- **Customer may not talk about it**

The bias gives more negative effects on Applicability than Importance

Recap for Ontology Mining

- Training data can be scarce and noisy. Supervisions from data itself are very useful.
- Leverage signals from heterogeneous sources (text/image/user behavior)
- The presented techniques are applicable for KGs in other domains
- Many other works skipped due to time/space limits

Future Directions

- **Fuse** heterogeneous information sources to achieve a **synergistic effect** on ontology mining

- Better leverage **seed/unlabeled** samples

- Taxonomy Enrichment and Relation Discovery **in one shot**

Resources

- SemEval-2015 Task 17: Taxonomy Extraction Evaluation
- SemEval-2016 Task 13: a Taxonomy Induction Method based on Lexico-Syntactic Patterns, Substrings and Focused Crawling
- SemEval-2016 Task 14: Semantic Taxonomy Enrichment
- SemEval-2018 Task 9: Hypernym Discovery
- Web Data Commons - Gold Standard for Product Matching and Product Feature Extraction
- An important link