Knowledge-aware Zero-shot Learning (K-ZSL): Concepts, Methods and Resources

Part II – KG-aware ZSL Methods

Jiaoyan Chen[1], Yuxia Geng[2], Yufeng Huang[2] and Huajun Chen[2]

1. Department of Computer Science, University of Oxford, UK
2. College of Computer Science and Technology, Zhejiang University, China

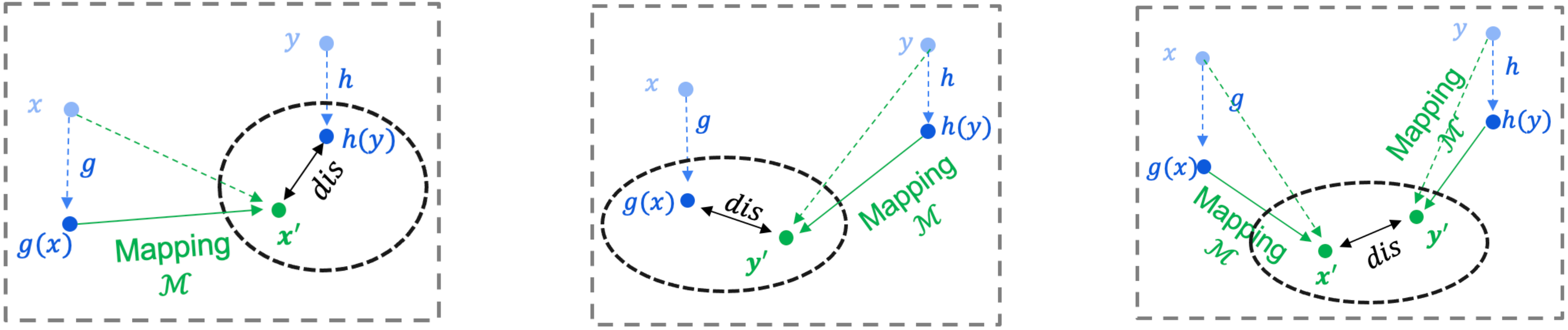
https://china-uk-zsl.github.io/kg-zsl-tutorial-iswc-2022/

Tutorial of The 21th International Semantic Web Conference (October 23, 2022, Virtual)

&1

Mapping-based Paradigm

General Frameworks



(a) Input Mapping

- The mapping function $\mathcal{M}$ maps the input x or $g(x)$ to its class's vector encoding $h(y)$

(b) Class Mapping

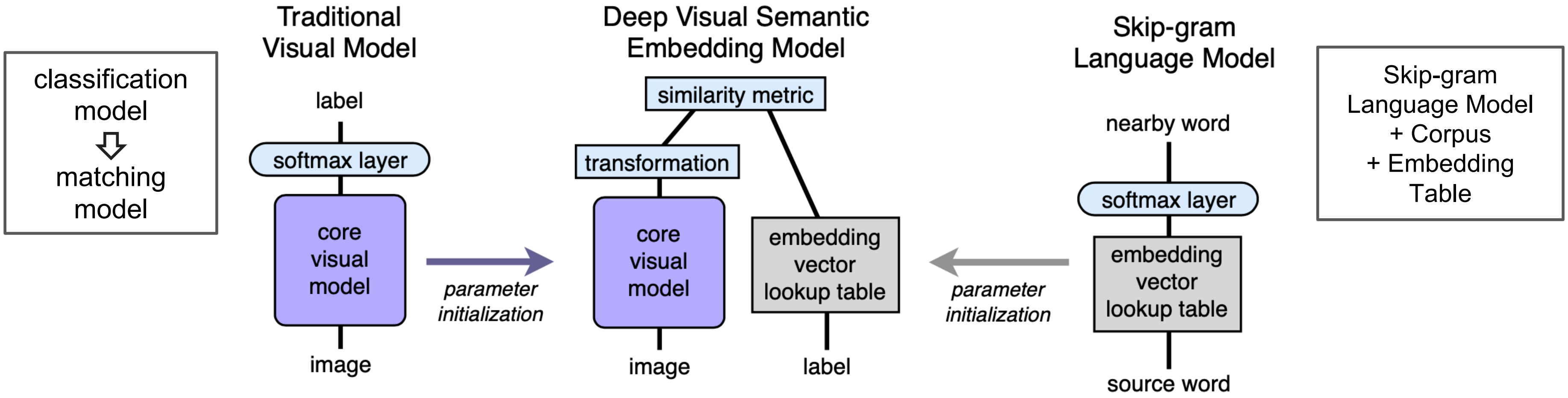
- The class (external knowledge) y or its encoding $h(y)$ is mapped to the vector of its sample $g(x)$

(c) Joint Mapping

- Both the class label and the input are mapped by two different functions to an intermediate space

DeViSE

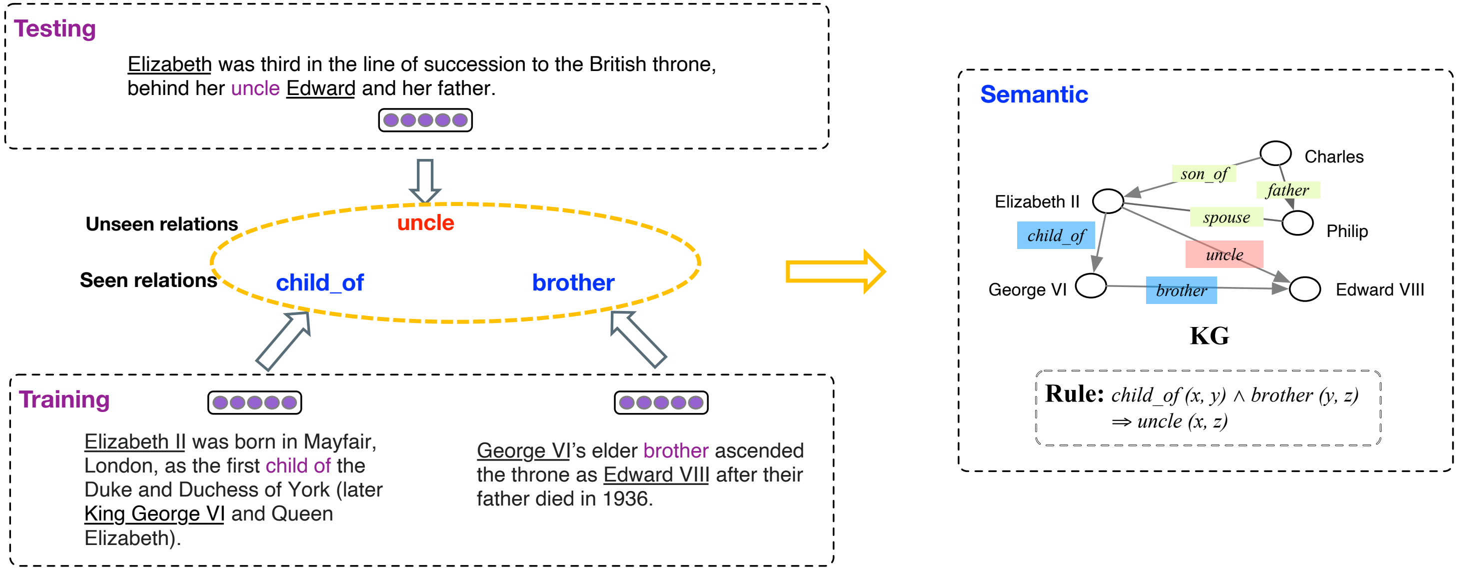
- Zero-shot Image Classification & Word embedding-based External Knowledge
- mapping images into the word vector space, where the correlations between classes are implied



Frome, Andrea, et al. "Devise: A deep visual-semantic embedding model." NIPS 2013.

Logic-guided Zero-shot Relation Extraction

- **Zero-shot Relation Extraction:** given a context and annotated entity pairs, predicting the target (unseen) relation
- **KG-based External Knowledge:** KG embeddings + Rule

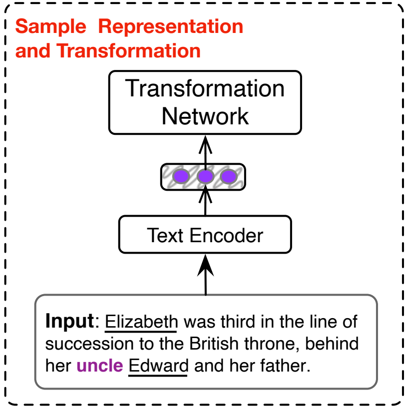


Li, Juan, et al. "Logic-guided semantic representation learning for zero-shot relation classification." COLING 2020.

Logic-guided Zero-shot Relation Extraction

- Zero-shot Relation Extraction
- KG-based External Knowledge: KG embeddings + Rule

1. Sentence Feature Extraction: PCNN (Position-aware Convolutional Network)

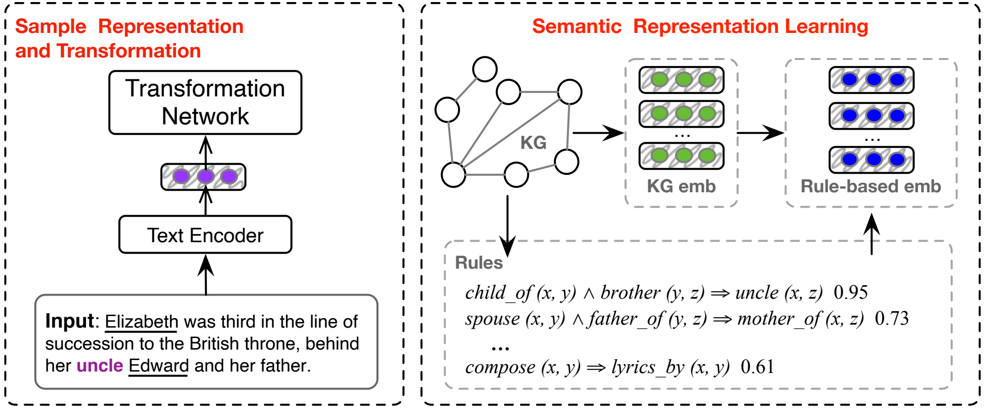


Li, Juan, et al. "Logic-guided semantic representation learning for zero-shot relation classification." COLING 2020.

Logic-guided Zero-shot Relation Extraction

- Zero-shot Relation Extraction
- KG-based External Knowledge: KG embeddings + Rule

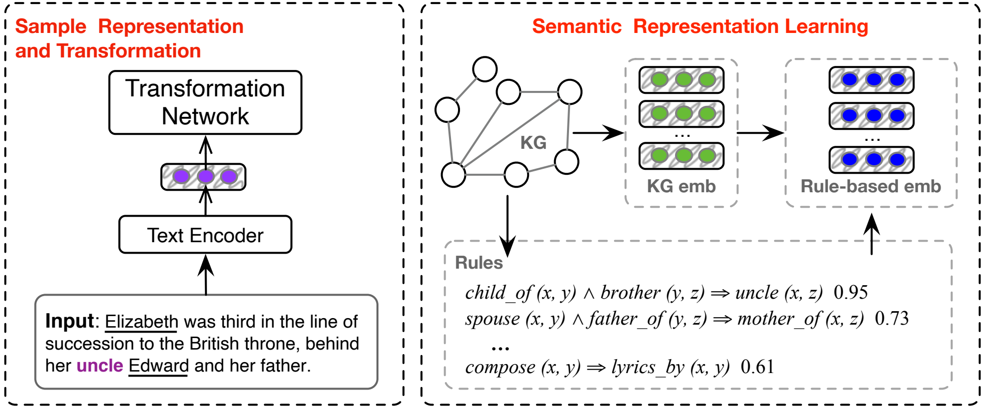
1. Sentence Feature Extraction: PCNN
2. KG embedding: TransE



Li, Juan, et al. "Logic-guided semantic representation learning for zero-shot relation classification." COLING 2020.

Logic-guided Zero-shot Relation Extraction

- Zero-shot Relation Extraction
- KG-based External Knowledge: KG embeddings + Rule



1. Sentence Feature Extraction: PCNN
2. KG embedding: Wikidata dump + TransE pre-train
3. Rule-augmented KG embedding

For an unseen relation r_u with its three rules, $R1$, $R2$ and $R3$, its rule-guided embedding is calculated as:

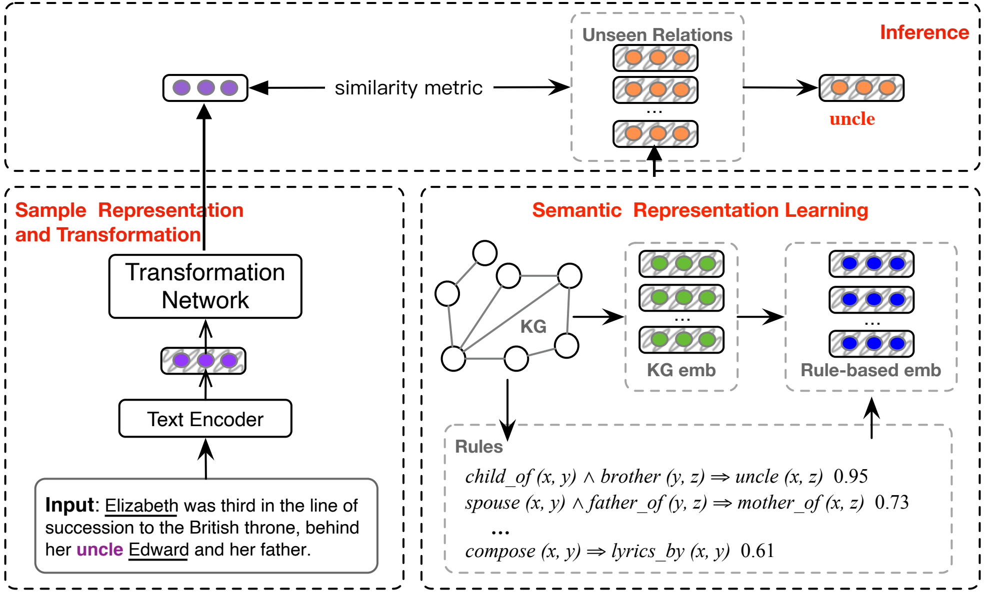
rule	equation
$R1: r_A \Rightarrow r_u$	$\vec{r_A} \approx \vec{r_u}$
$R2: r_B \wedge r_C \Rightarrow r_u$	$\vec{r_B} + \vec{r_C} \approx \vec{r_u}$
$R3: r_D \wedge r_u \Rightarrow r_E$	$\vec{r_D} + \vec{r_u} \approx \vec{r_E}$

$$E_{rl}(r_u) = \frac{s_1 * \vec{r_A} + s_2 * (\vec{r_B} + \vec{r_C}) + s_3 * (\vec{r_E} - \vec{r_D})}{s_1 + s_2 + s_3}$$

Li, Juan, et al. “Logic-guided semantic representation learning for zero-shot relation classification.” COLING 2020.

Logic-guided Zero-shot Relation Extraction

- Zero-shot Relation Extraction
- KG-based External Knowledge: KG embeddings + Rule



1. Sentence Feature Extraction: PCNN
2. KG embedding: TransE
3. Rule-augmented KG embedding
4. Mapping-based ZSL model

   DeViSE

Li, Juan, et al. "Logic-guided semantic representation learning for zero-shot relation classification." COLING 2020.

Ontology-guided Semantic Composition

- Zero-shot Image Classification & Zero-shot Visual Question Answering
- OWL ontology-based External Knowledge: compositional class semantics

logical constructors in DL

- *inclusion:* $\sqsubseteq$
- *intersection:* $\sqcap$
- *union:* $\sqcup$
- *existential restriction:* $\exists$
- *universal restriction:* $\forall$
- …

define a new class/concept **A**

$$A \sqsubseteq \text{Female} \sqcap \exists\, \text{likes.Movie} \sqcap \exists \text{hasSon.(Student} \sqcap \exists \text{attends.CSCourse)}$$

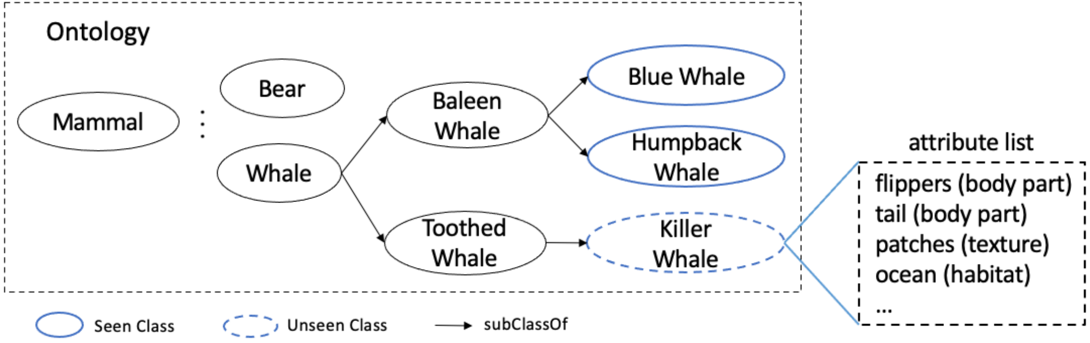
Krötzsch, Markus, Frantisek Simancik, and Ian Horrocks. "A description logic primer."
Chen, Jiaoyan, et al. "Ontology-guided semantic composition for zero-shot learning." KR 2020.

Ontology-guided Semantic Composition

- Zero-shot Image Classification & Zero-shot Visual Question Answering
- OWL ontology-based External Knowledge: compositional class semantics

logical constructors in DL

- *inclusion:* ⊑
- *intersection:* ⊓
- *union:* ⊔
- *existential restriction:* ∃
- *universal restriction:* ∀
- …

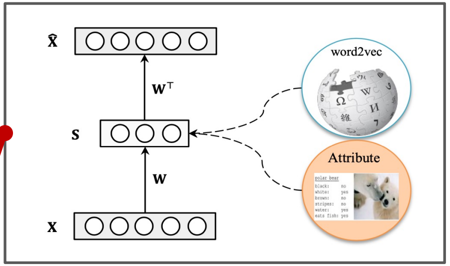


Killer_Whale ⊑ Toothed_Whale ⊓ ∃hasTexture.Patches ⊓ ∃hasHabitat.Ocean …

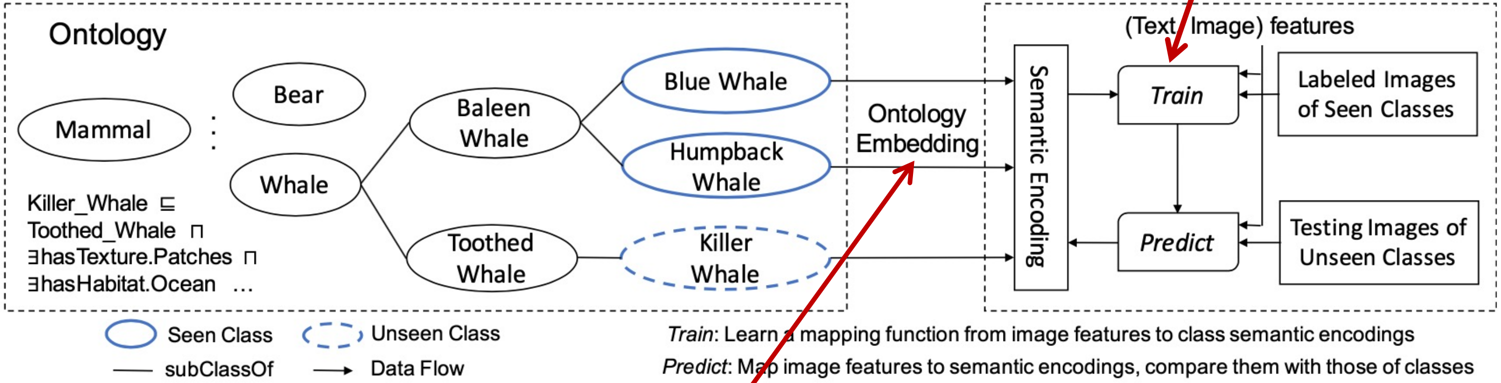
Chen, Jiaoyan, et al. "Ontology-guided semantic composition for zero-shot learning." KR 2020.

Ontology-guided Semantic Composition

- Zero-shot Image Classification & Zero-shot visual question answering
- OWL ontology-based External Knowledge: compositional class semantics



Semantic Autoencoder

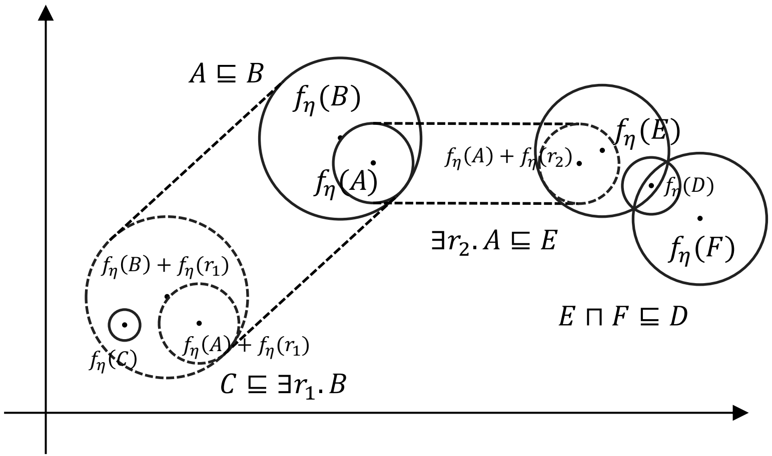


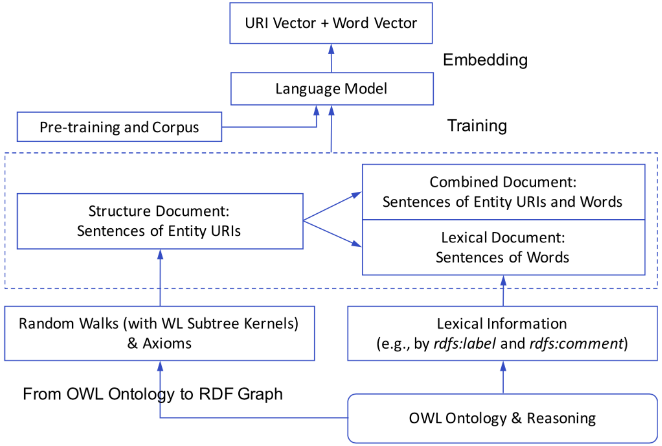
1. Embedding of Logical Axioms e.g., by geometric learning (Kulmanov et al. 2019)
2. Embedding of text and paths e.g., by neural language model (Chen et al. 2021)

Chen, Jiaoyan, et al. "Ontology-guided semantic composition for zero-shot learning." KR 2020.

Ontology-guided Semantic Composition

- Zero-shot Image Classification & Zero-shot visual question answering
- OWL ontology-based External Knowledge: compositional class semantics





1. Embedding of Logical Axioms e.g., by geometric learning (Kulmanov et al., EL Embeddings: Geometric Construction of Models for the Description Logic EL++. IJCAI 2019)
2. Embedding of text and paths e.g., by neural language model (Chen et al., OWL2Vec*: embedding of OWL ontologies. Machine Learning 2021)

Chen, Jiaoyan, et al. "Ontology-guided semantic composition for zero-shot learning." KR 2020.

More Reading

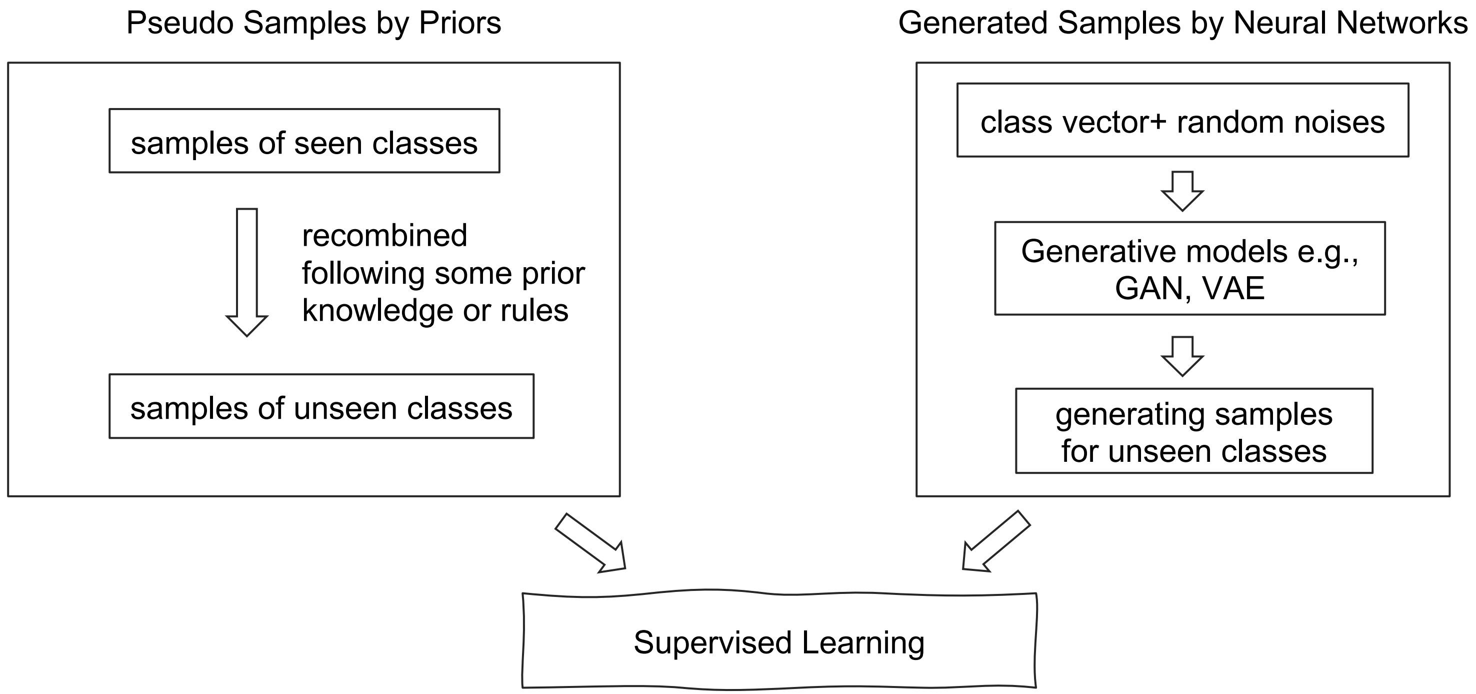
- Chen et al., Zero-shot Text Classification via Knowledge Graph Embedding for Social Media Data, The IEEE IoT Journal
- Nayak et al., Zero-Shot Learning with Common Sense Knowledge Graphs, arXiv 2021
- Shah et al., An Open-World Extension to Knowledge Graph Completion Models, AAAI 2019
- Rios et al., Few-Shot and Zero-Shot Multi-Label Learning for Structured Label Spaces, EMNLP 2018
- Roy et al., Improving Zero Shot Learning Baselines with Commonsense Knowledge, arXiv 2020
- Huang et al., Zero-Shot Transfer Learning for Event Extraction, ACL 2018
- ...

&2

Data augmentation-based Paradigm

(generative models)

Data Augmentation Frameworks

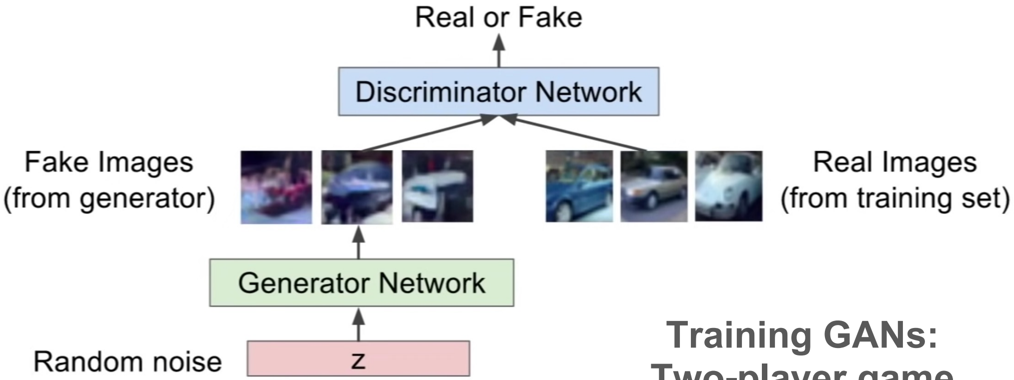


Generative Adversarial Networks (GANs)

- Generative models: Generative Adversarial Networks (GANs)

A human face synthesized by Deepfake

Generator Network: try to fool the discriminator by generating real-looking images;
Discriminator Network: try to distinguish between real and fake images

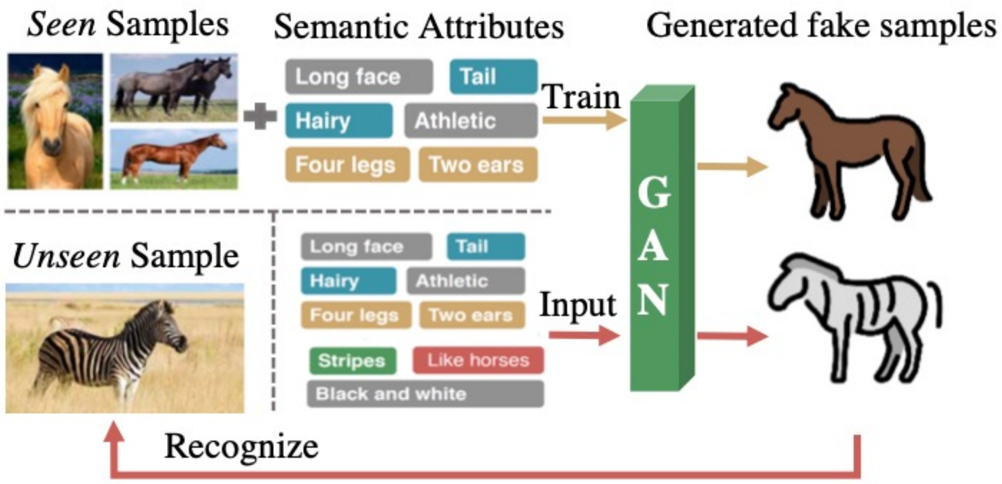


Training GANs:
Two-player game

Ian Goodfellow et al., Generative Adversarial Nets, NIPS 2014

GANs for ZSL

- Generative Adversarial Networks (GANs) & Zero-shot Learning

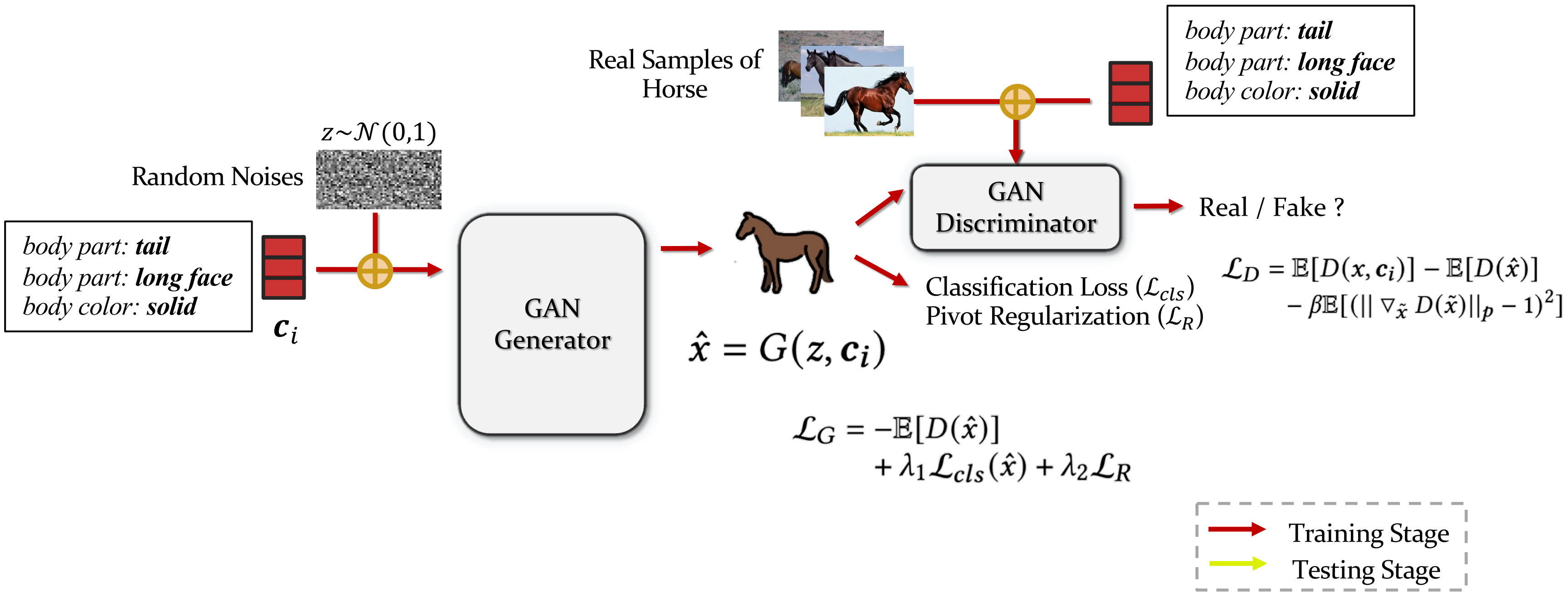


Conditional GANs

Xian, Yongqin, et al. "Feature generating networks for zero-shot learning." CVPR 2018.
Li, Jingjing, et al. "Leveraging the invariant side of generative zero-shot learning." CVPR 2019.

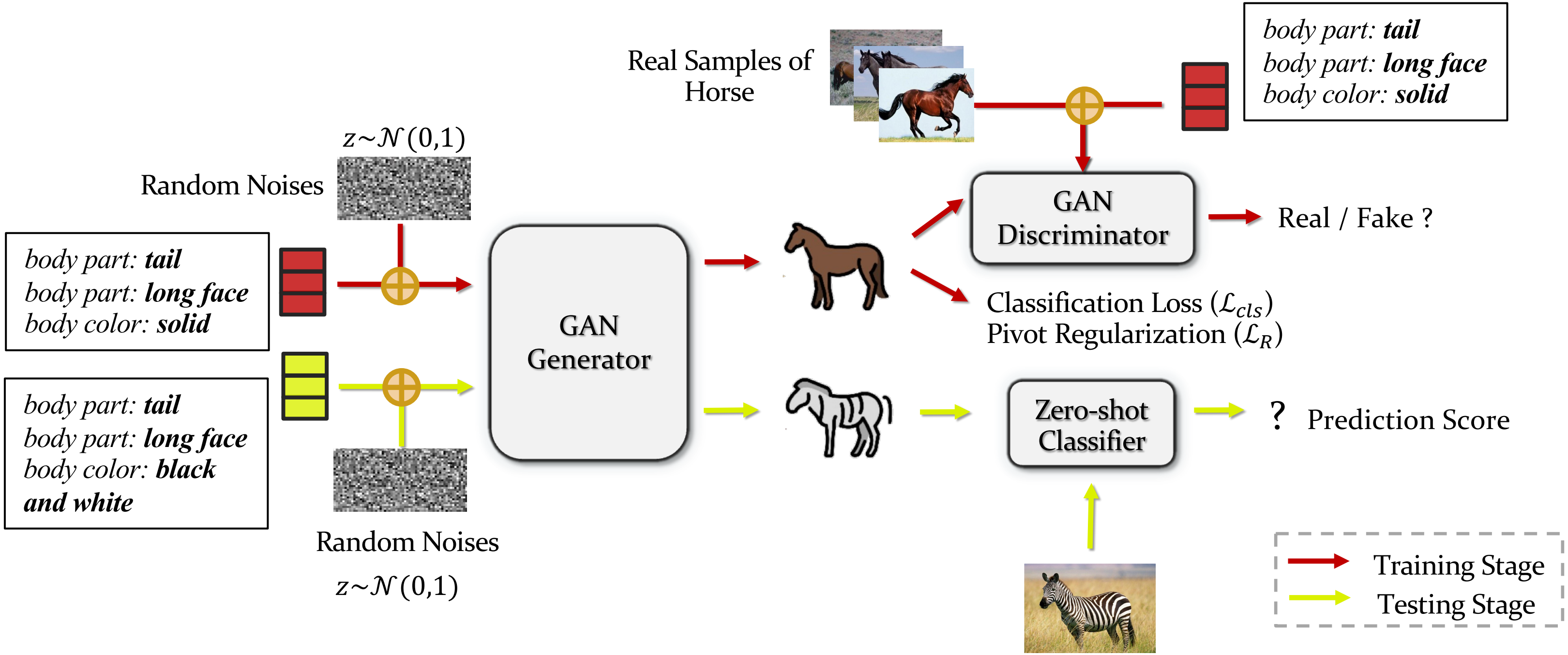
GANs for ZSL

- A running example with zero-shot animal classification



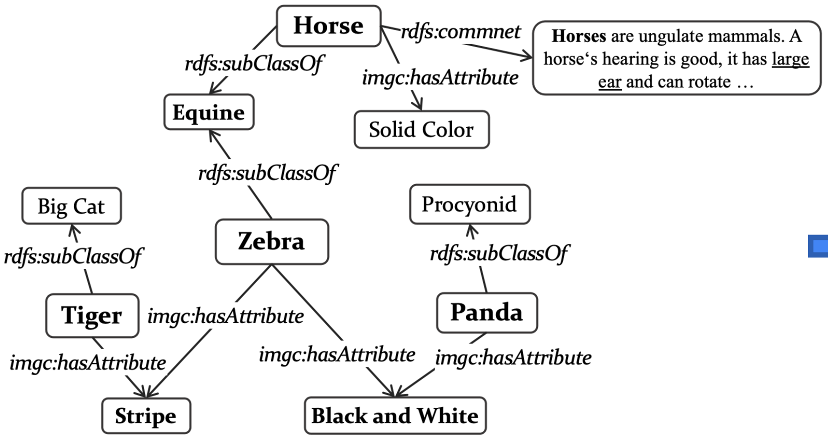
GANs for ZSL

- A running example with zero-shot animal classification



GANs for ZSL

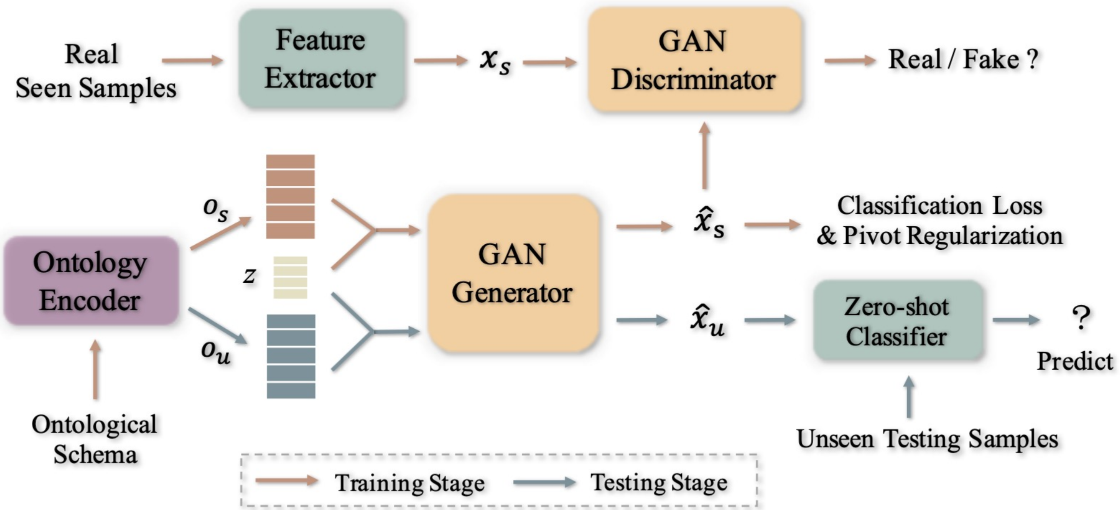
- How do we incorporate a KG or an ontology with GAN for ZSL problem?



+ GAN

OntoZSL

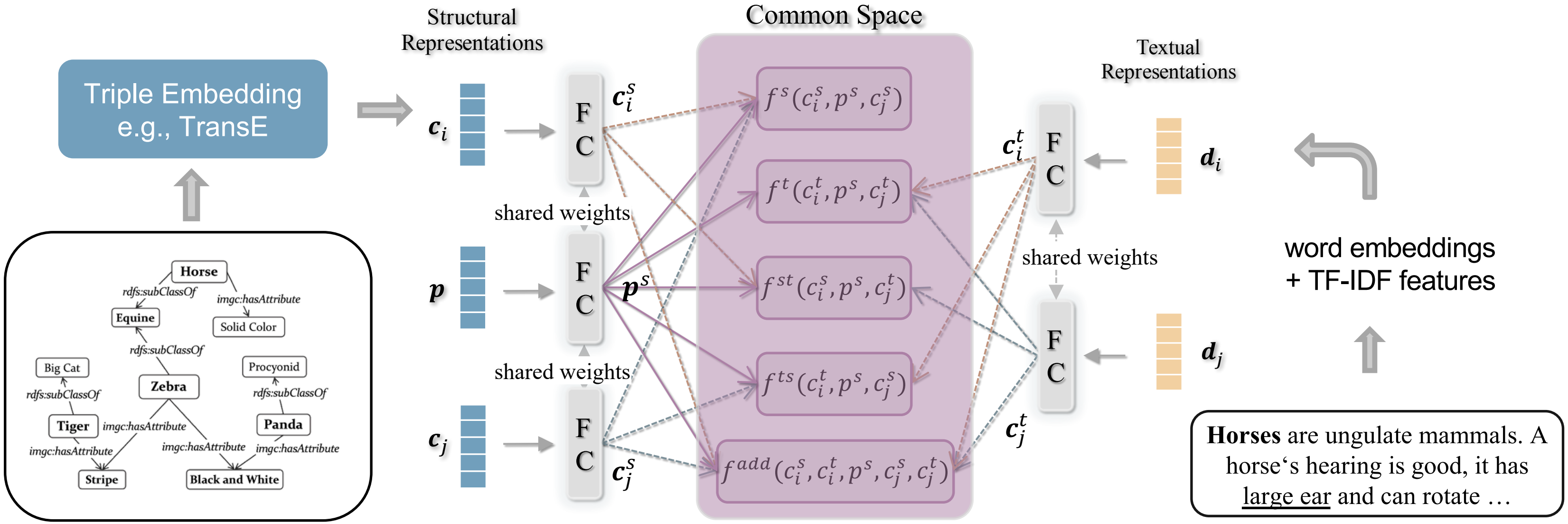
- How do we incorporate a KG or ontology with GAN for ZSL problem?
  - pre-training the KG or ontology using an ontology encoder module



Geng, Yuxia, et al. "OntoZSL: Ontology-enhanced zero-shot learning." WWW 2021.

OntoZSL

- KG-based side information & pre-training the KG
  - text-aware pre-training

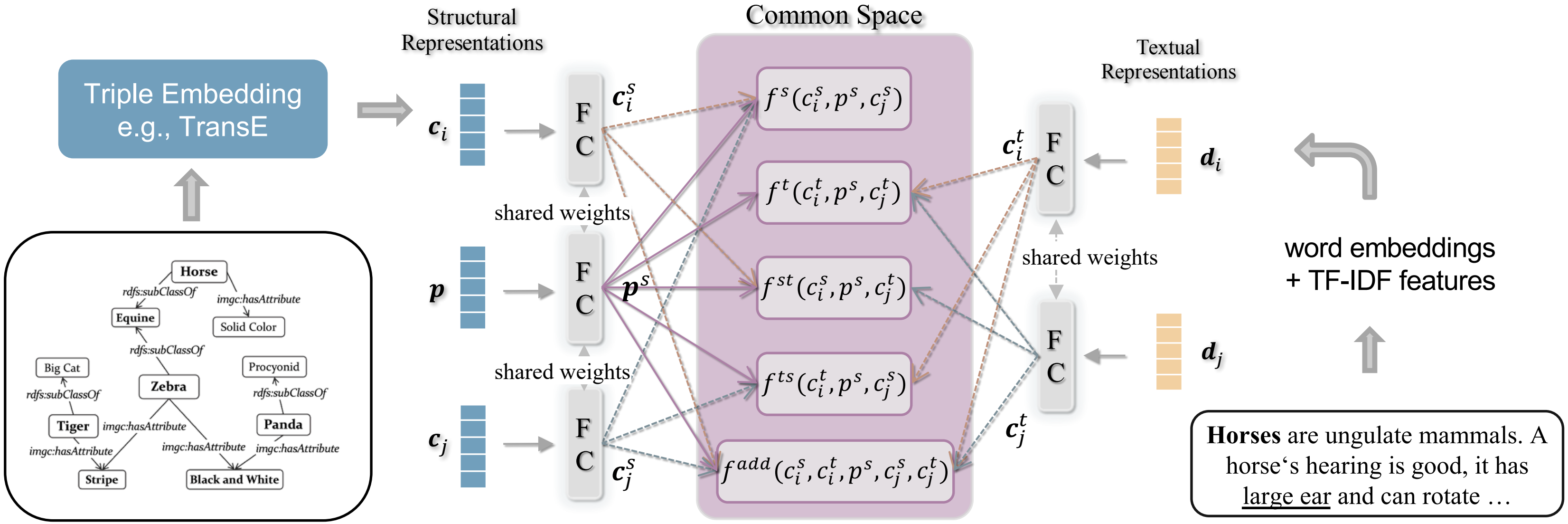


Geng, Yuxia, et al. "OntoZSL: Ontology-enhanced zero-shot learning." WWW 2021.

OntoZSL

- KG-based side information & pre-training the KG
    - text-aware pre-training

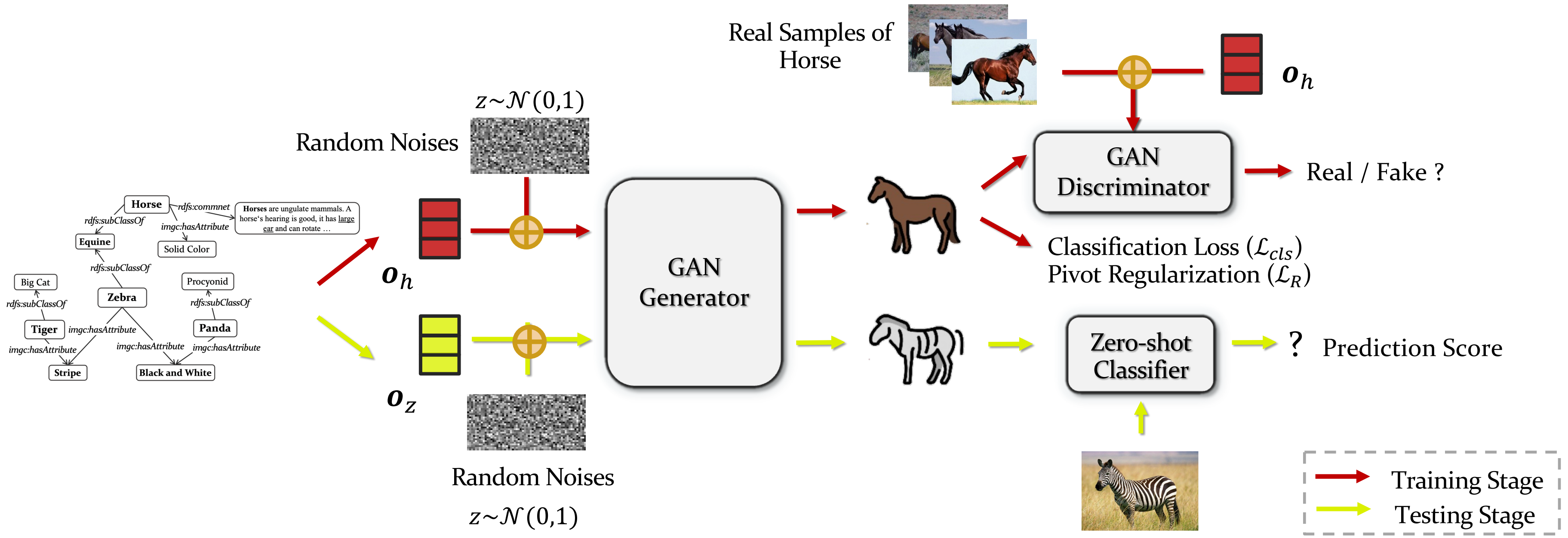
$$o_i = [c_i^s ; c_i^t]$$



Geng, Yuxia, et al. "OntoZSL: Ontology-enhanced zero-shot learning." WWW 2021.

OntoZSL

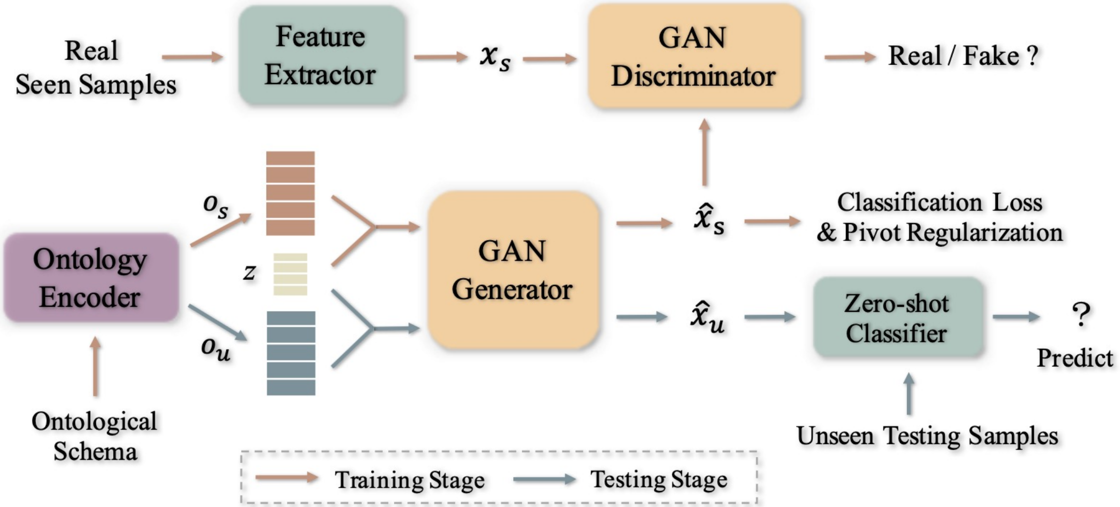
- How do we incorporate a KG or ontology with GAN for ZSL problem?



Geng, Yuxia, et al. "OntoZSL: Ontology-enhanced zero-shot learning." WWW 2021.

OntoZSL

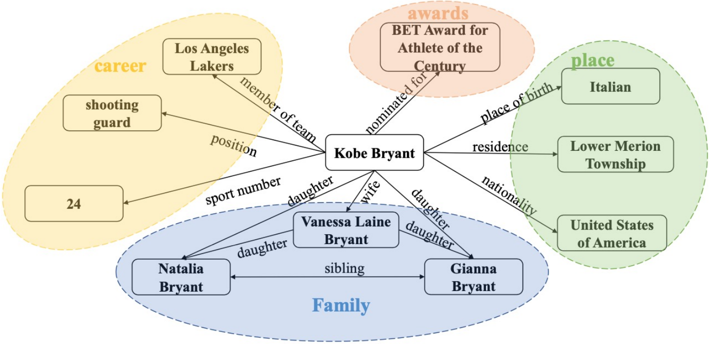
- How do we incorporate a KG or ontology with GAN for ZSL problem?
  - pre-training the KG or ontology using an ontology encoder module

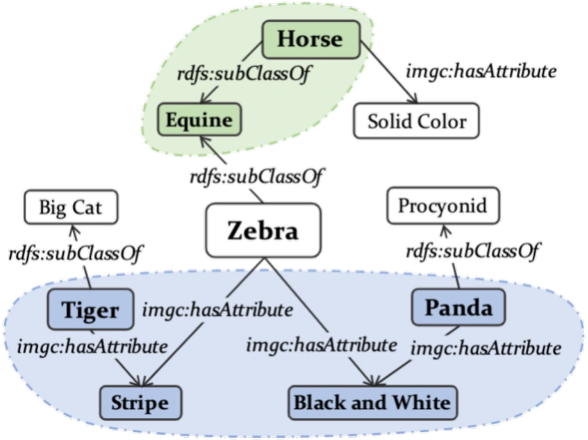


Geng, Yuxia, et al. "OntoZSL: Ontology-enhanced zero-shot learning." WWW 2021.
Geng, Yuxia, et al. “K-ZSL: Benchmarking knowledge-driven zero-shot learning.” Journal of Web Semantics 2022.

OntoZSL with Disentangled Ontology Embedding

- KG-based side information & pre-training the KG
    - disentanglement-aware pre-training

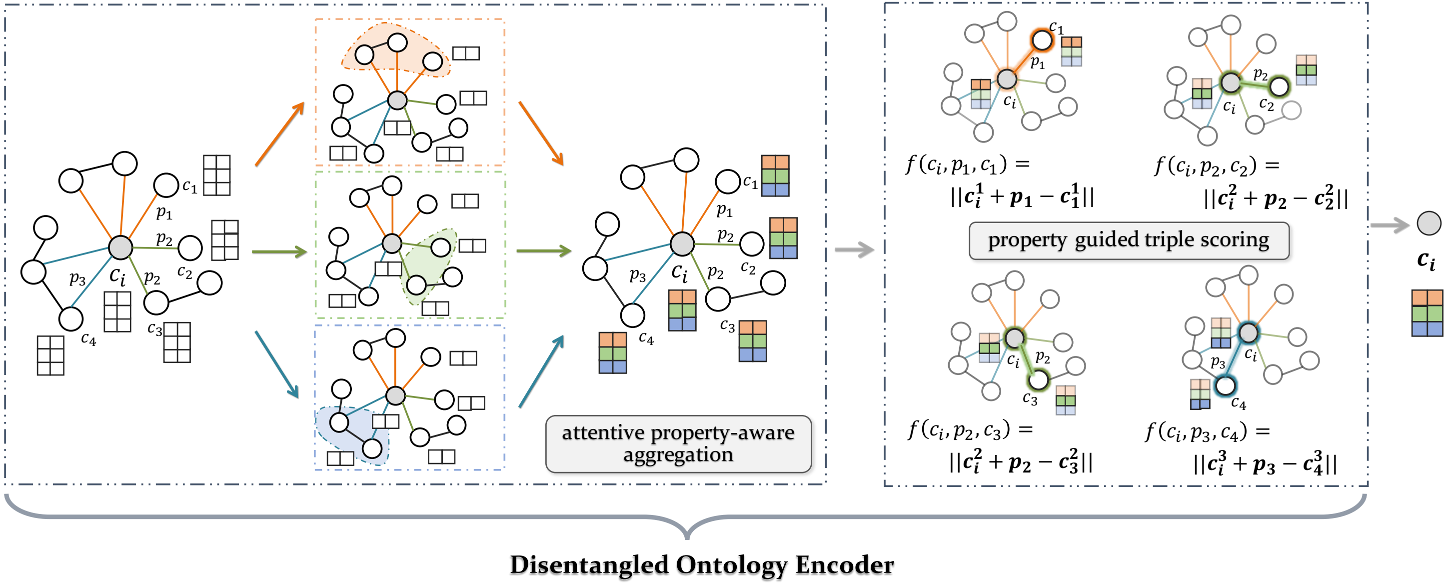




The entanglement characteristics of KGs and ontologies: a class (entity) is often related to other classes (entities) in different semantic aspects.

OntoZSL with Disentangled Ontology Embedding

- KG-based side information & pre-training the KG
  - disentanglement-aware pre-training



Geng, Yuxia, et al. "Disentangled Ontology Embedding for Zero-shot Learning." KDD 2022.

OntoZSL with Disentangled Ontology Embedding

- Learning disentangled embeddings for each class according to its semantics of different aspects
  a) split the initial embedding of class i into K components, $\boldsymbol{c}_i = [\boldsymbol{c}_i^1, \boldsymbol{c}_i^2, \ldots, \boldsymbol{c}_i^K]$;

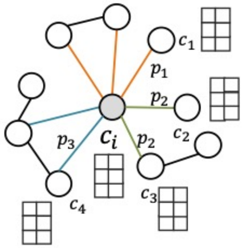


Illustration of DOZSL with K = 3, d = 6.

Geng, Yuxia, et al. "Disentangled Ontology Embedding for Zero-shot Learning." KDD 2022.

OntoZSL with Disentangled Ontology Embedding

- Learning disentangled embeddings for each class according to its semantics of different aspects
  a) split the initial embedding of class i into K components, $\boldsymbol{c}_i = [\boldsymbol{c}_i^1, \boldsymbol{c}_i^2, \dots, \boldsymbol{c}_i^K]$;
  b) attentively aggregate the graph neighborhood information for each component;

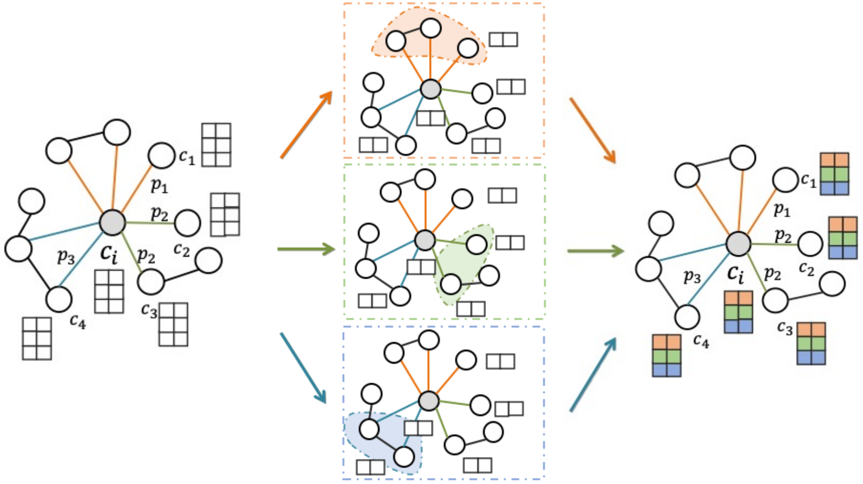


Illustration of DOZSL with K = 3, d = 6.

attention weight computation:

$$\alpha^{k,l}_{(c_i,p,c_j)} = softmax((h^l_{i,k,p})^T \cdot h^l_{j,k,p}) = \frac{exp((h^l_{i,k,p})^T \cdot h^l_{j,k,p})}{\sum_{(c_{j'},p') \in \mathcal{N}(i)} exp((h^l_{i,k,p'})^T \cdot h^l_{j',k,p'})}$$

$$h^l_{i,k,p} = h^l_{i,k} \circ W_p, \quad h^l_{j,k,p} = h^l_{j,k} \circ W_p$$

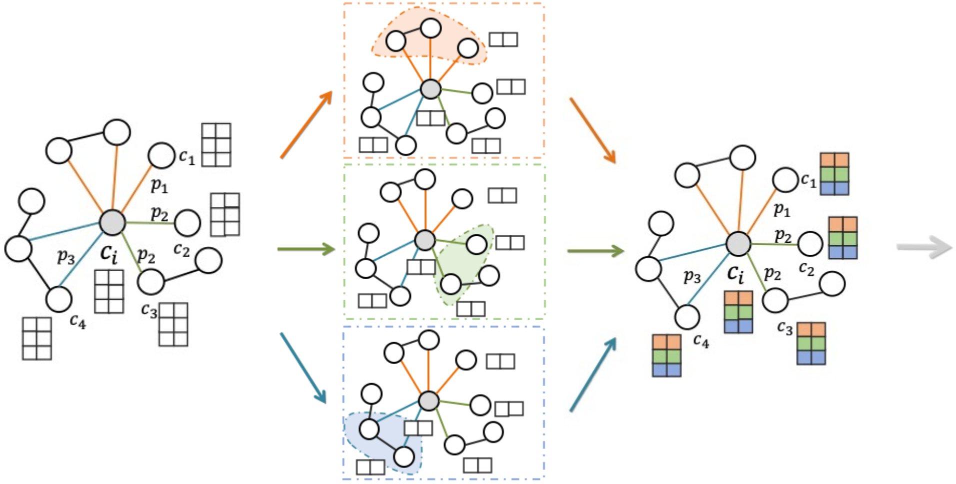
neighborhood aggregation:

$$h^{l+1}_{i,k} = \sigma(\sum_{(c_j,p) \in \mathcal{N}(i)} \alpha^{k,l}_{(c_i,p,c_j)} \phi(h^l_{j,k}, h^l_p, W_p)), \quad h^{l+1}_p = h^l_p \cdot \Theta^l_p$$

Geng, Yuxia, et al. "Disentangled Ontology Embedding for Zero-shot Learning." KDD 2022.

OntoZSL with Disentangled Ontology Embedding

- Learning disentangled embeddings for each class according to its semantics of different aspects
  - a) split the initial embedding of class i into K components, $\boldsymbol{c}_i = [\boldsymbol{c}_i^1, \boldsymbol{c}_i^2, \ldots, \boldsymbol{c}_i^K]$;
  - b) attentively aggregate the graph neighborhood information for each component;
  - c) extract property-specific components to compute the triple scores to further refine the semantics of each component



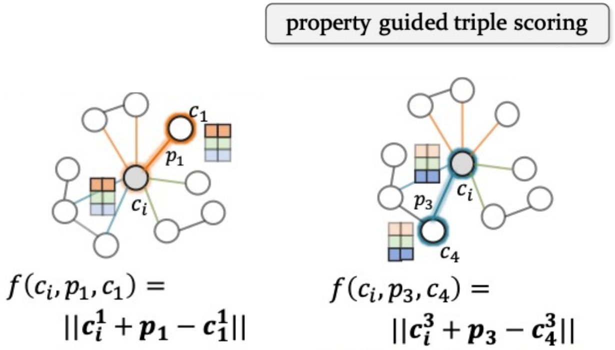


Illustration of DOZSL with K = 3, d = 6.

triple scoring function:

$$q_{(c_i, p_k, c_j)} = f(-||\boldsymbol{c}_i^k + \boldsymbol{p}_k - \boldsymbol{c}_j^k||)$$

Geng, Yuxia, et al. "Disentangled Ontology Embedding for Zero-shot Learning." KDD 2022.

OntoZSL with Disentangled Ontology Embedding

- Learning disentangled embeddings for each class according to its semantics of different aspects
  a) split the initial embedding of class i into K components, $\boldsymbol{c}_i = [\boldsymbol{c}_i^1, \boldsymbol{c}_i^2, \ldots, \boldsymbol{c}_i^K]$;
  b) attentively aggregate the graph neighborhood information for each component;
  c) extract property-specific components to compute the triple scores to further refine the semantics of each component

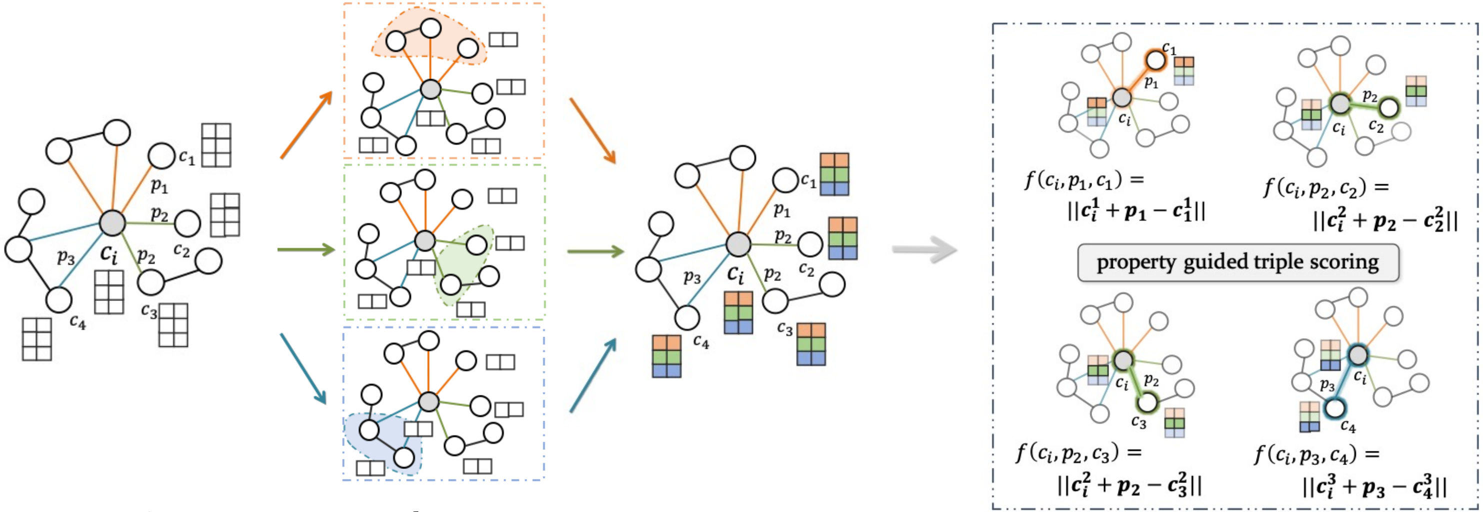


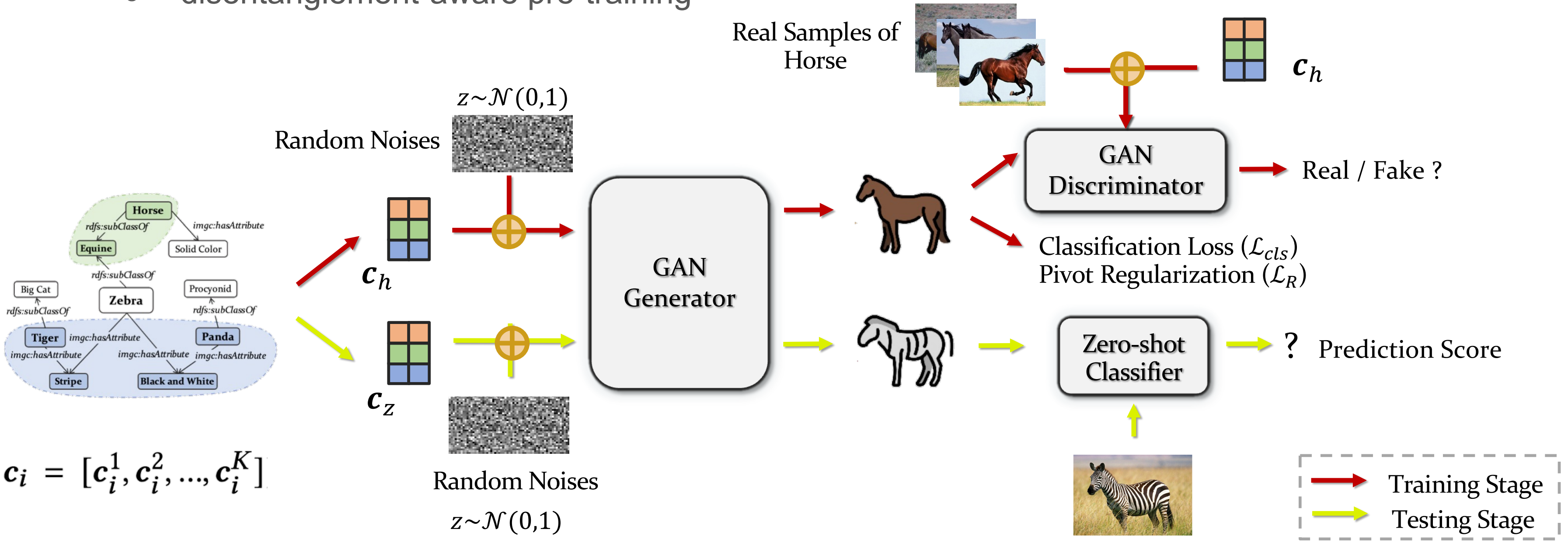
Illustration of DOZSL with K = 3, d = 6.

a standard cross entropy loss

Geng, Yuxia, et al. "Disentangled Ontology Embedding for Zero-shot Learning." KDD 2022.

OntoZSL with Disentangled Ontology Embedding

- KG-based side information & pre-training the KG
  - disentanglement-aware pre-training



More Reading

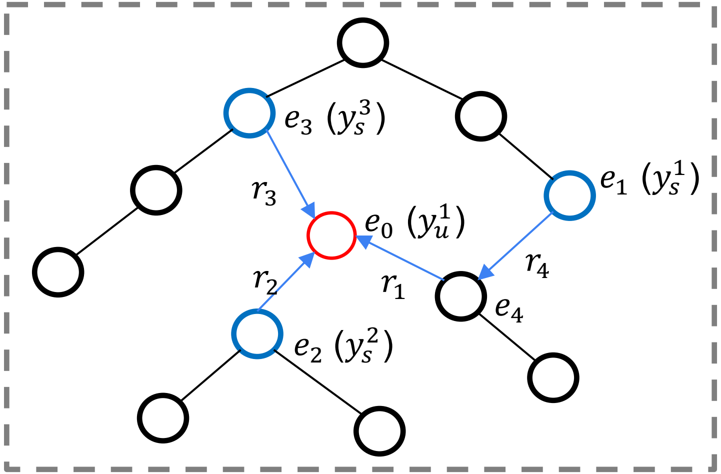
- VAE-based
  - Mishra et al., A Generative Model For Zero Shot Learning Using Conditional Variational Autoencoders, CVPR 2018
  - Wang et al., Zero-Shot Learning via Class-Conditioned Deep Generative Models, AAAI 2018
- Prior-based
  - Zero and Few Shot Learning with Semantic Feature Synthesis and Competitive Learning. TPAMI 2020
  - Iteratively learning embeddings and rules for knowledge graph reasoning. WWW 2019
  - Injecting Logical Background Knowledge into Embeddings for Relation Extraction. NAACL 2015

&3

Propagation-based Paradigm

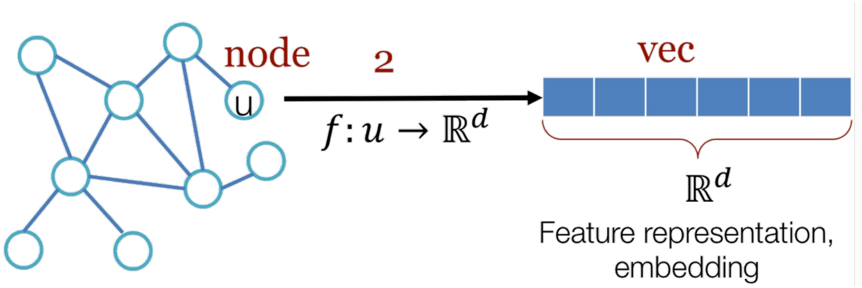
Feature Propagation Framework

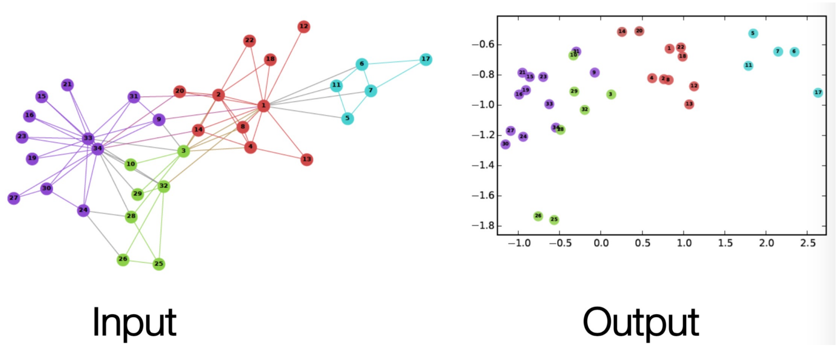
- Regarding the graph structure of KGs and ontologies, how about directly preforming computation (feature transfer) on graph?
- For example, propagate model parameter or features from seen class nodes to unseen class nodes



Graph Neural Networks (GNNs)

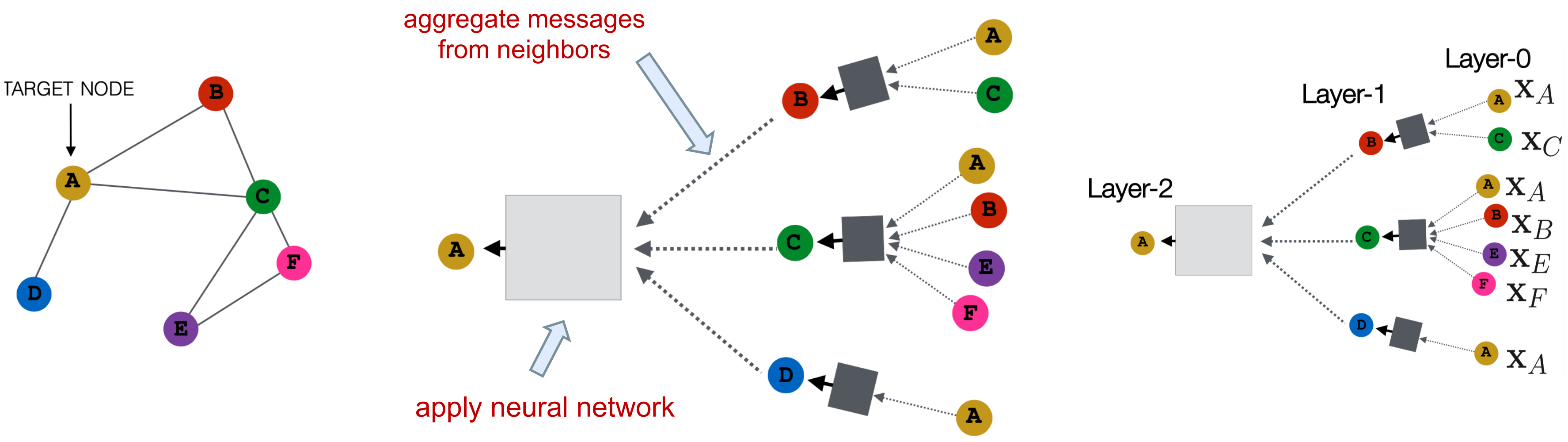
- Deep learning based methods that operate on graph domain
- Objective
  - Input: a graph G =(V, E)
  - Output: $Z \in \mathbb{R}^{|V| \times d}$, $d \ll |V|$, a d-dim vector Z_u for each node u.





Graph Neural Networks (GNNs)

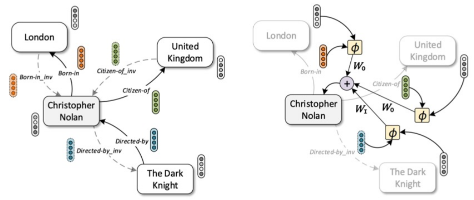
- Basic idea
  - neighbourhood aggregation: aggregating neighbour information and pass into a neural network
  - self updating: update the node's hidden states using another network



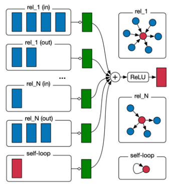
Graph Neural Networks (GNNs)

- Representative GNNs

Composition-based Multi-Relational Graph Convolutional Networks (CompGCN)	Vashishth et al., ICLR'20
Modeling Relational Data with Graph Convolutional Networks (RGCN)	Schlichtkrull et al., ESWC'18
Graph Attention Networks (GAT)	Velickovic et al., ICLR'18
Inductive Representation Learning on Large Graphs (GraphSAGE)	Hamilton et al., NIPS'17
Semi-Supervised Classification with Graph Convolutional Networks (GCN)	Kipf and Welling, ICLR'17

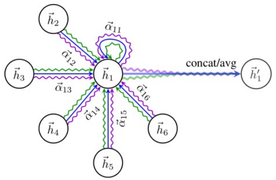


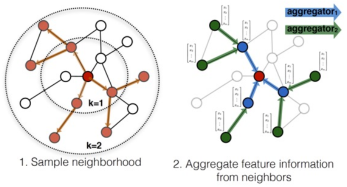
CompGCN



RGCN

GAT





GraphSAGE

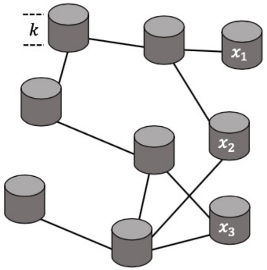
Zhou, Jie, et al. "Graph neural networks: A review of methods and applications." AI Open 1 (2020): 57-81.

GCNZ (Zero-shot Image Classification with KG)

- Input: a graph $G = (V, E)$ with $|V|$ class nodes and hierarchical relationship edges from WordNet
  - Adjacency matrix

$$A_{i,j} = \begin{cases} a_{i,j} = 1 & (i,j) \in E \\ 0 & (i,j) \notin E \end{cases}$$

  - Node feature matrix: $X \in \mathbb{R}^{n \times k}$ for k-dim word embeddings of n class nodes



Inputs: Word Embeddings $\mathcal{X}$
(k dimensions)

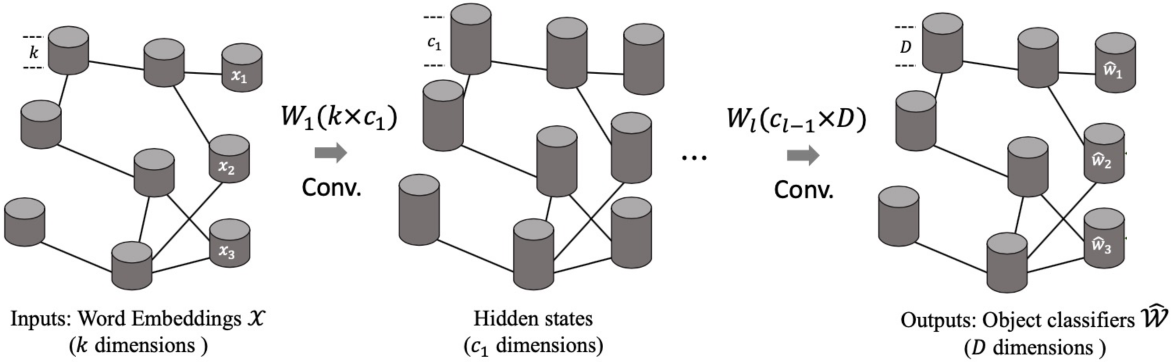
Wang, Xiaolong, Yufei Ye, and Abhinav Gupta. "GCNZ: Zero-shot recognition via semantic embeddings and knowledge graphs." CVPR 2018.

GCNZ (Zero-shot Image Classification with KG)

- Input: a graph $G = (V, E)$ with $|V|$ class nodes and hierarchical relationship edges from WordNet
  - Adjacency matrix

$$A_{i,j} = \begin{cases} a_{i,j} = 1 & (i,j) \in E \\ 0 & (i,j) \notin E \end{cases}$$

  - Node feature matrix: $X \in \mathbb{R}^{n \times k}$ for k-dim word embeddings of n class nodes

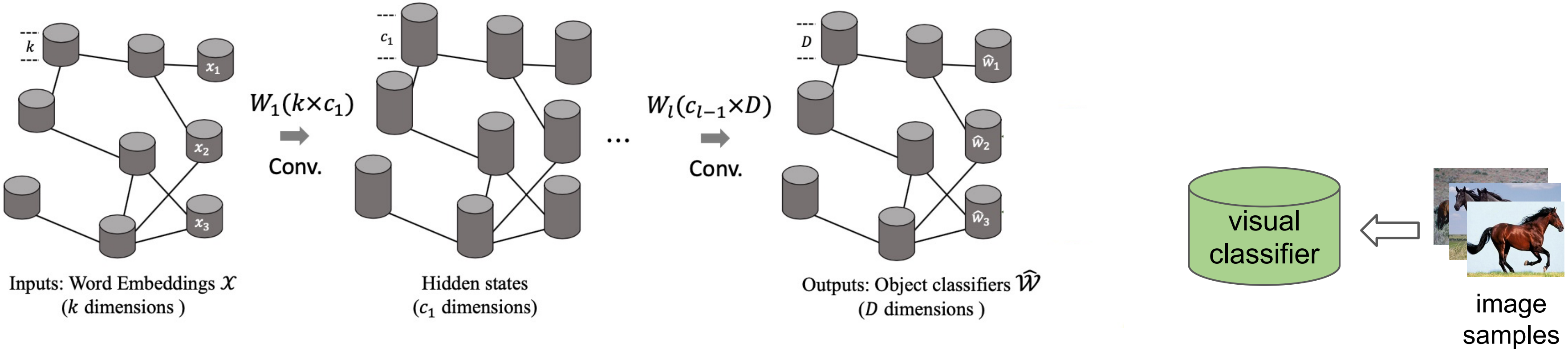


GCN: propagate features among class nodes

Wang, Xiaolong, Yufei Ye, and Abhinav Gupta. "GCNZ: Zero-shot recognition via semantic embeddings and knowledge graphs." CVPR 2018.

GCNZ (Zero-shot Image Classification with KG)

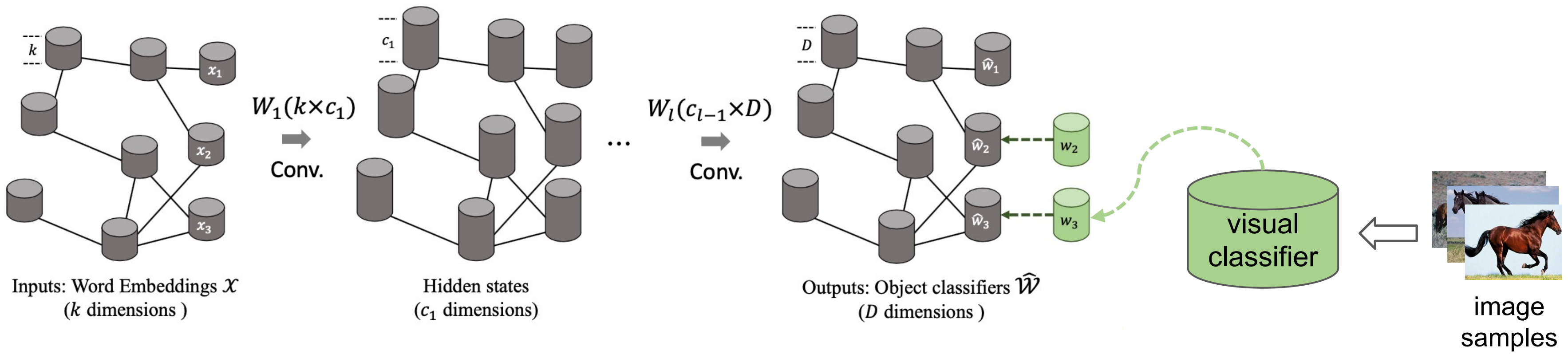
- Output: sample classifiers
  - each seen class node is supervised by a classifier, which is often a D-dim vector for class-specific sample features
  - each unseen class node is inferred to output its corresponding classifier for classification



Wang, Xiaolong, Yufei Ye, and Abhinav Gupta. "GCNZ: Zero-shot recognition via semantic embeddings and knowledge graphs." CVPR 2018.

GCNZ (Zero-shot Image Classification with KG)

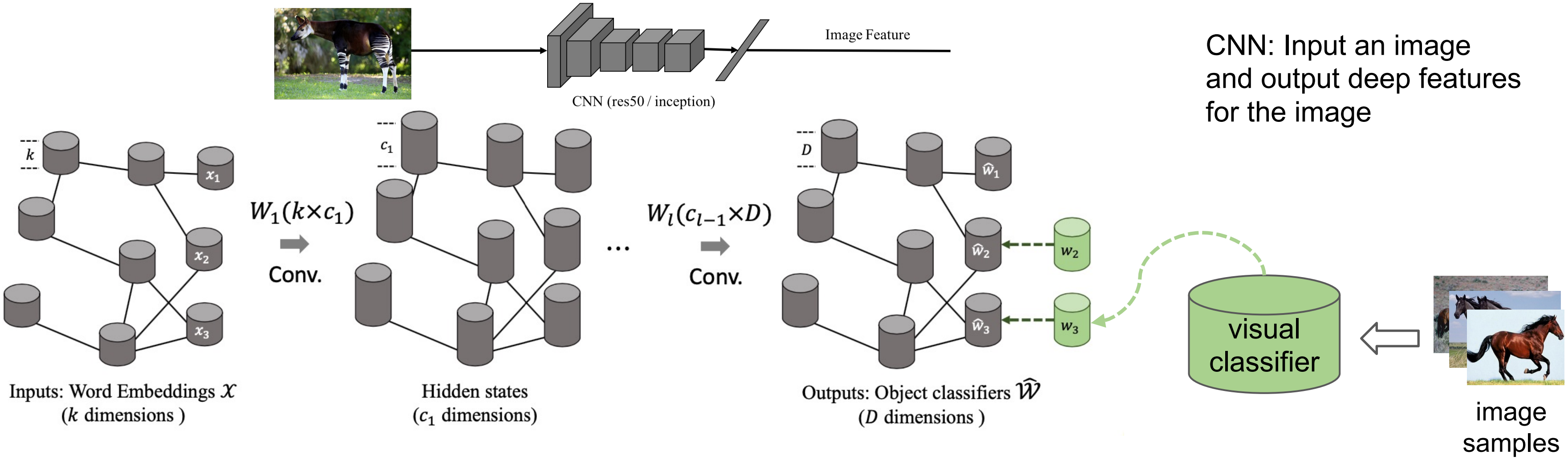
- Output: sample classifiers
  - each seen class node is supervised by a classifier, which is often a D-dim vector for class-specific sample features
  - each unseen class node is inferred to output its corresponding classifier for classification



Wang, Xiaolong, Yufei Ye, and Abhinav Gupta. "GCNZ: Zero-shot recognition via semantic embeddings and knowledge graphs." CVPR 2018.

GCNZ (Zero-shot Image Classification with KG)

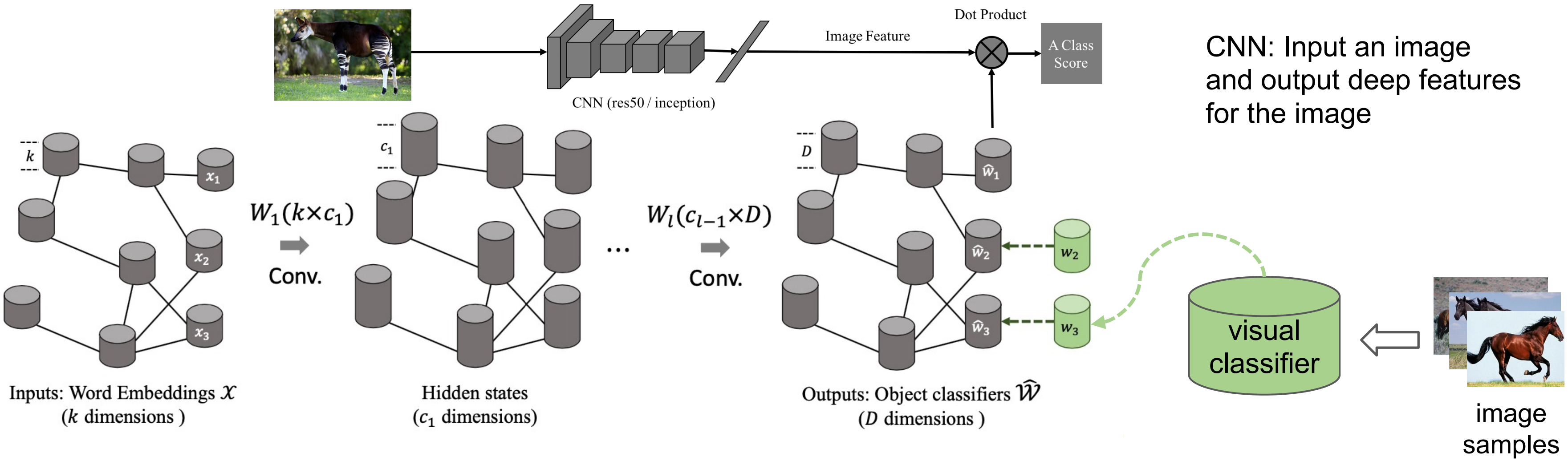
- Output: sample classifiers
  - each seen class node is supervised by a classifier, which is often a D-dim vector for class-specific sample features
  - each unseen class node is inferred to output its corresponding classifier for classification



Wang, Xiaolong, Yufei Ye, and Abhinav Gupta. "GCNZ: Zero-shot recognition via semantic embeddings and knowledge graphs." CVPR 2018.

GCNZ (Zero-shot Image Classification with KG)

- Output: sample classifiers
  - each seen class node is supervised by a classifier, which is often a D-dim vector for class-specific sample features
  - each unseen class node is inferred to output its corresponding classifier for classification

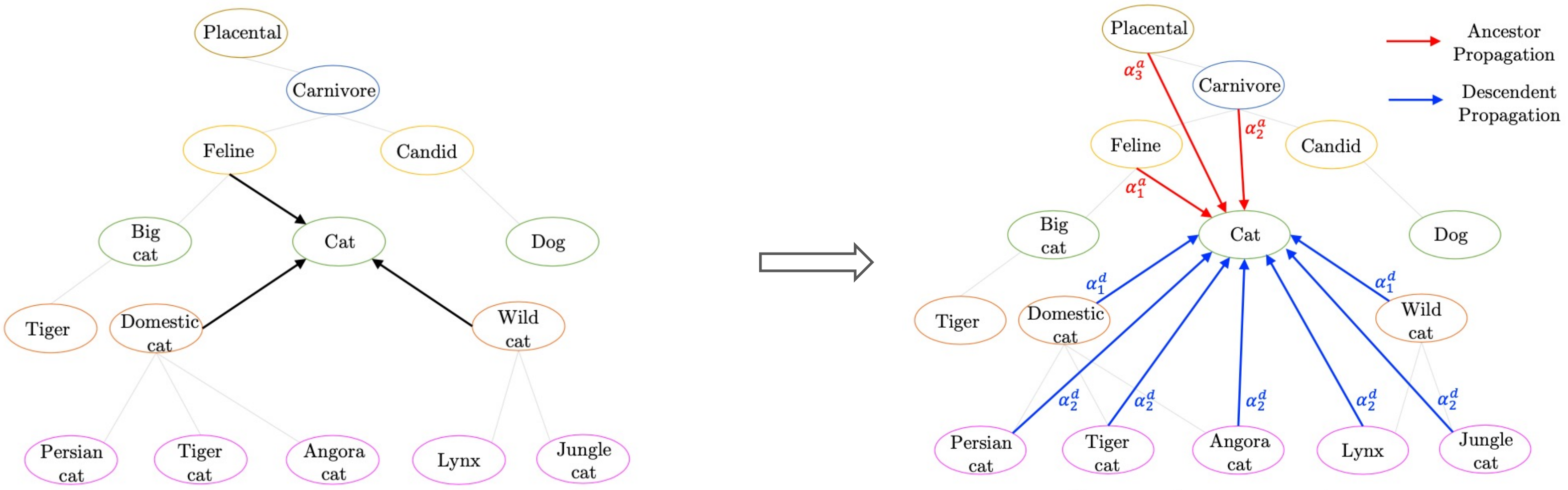


CNN: Input an image and output deep features for the image

Wang, Xiaolong, Yufei Ye, and Abhinav Gupta. "GCNZ: Zero-shot recognition via semantic embeddings and knowledge graphs." CVPR 2018.

DGP (Zero-shot Image Classification with KG)

- Optimized Graph Propagation with **denser connections**



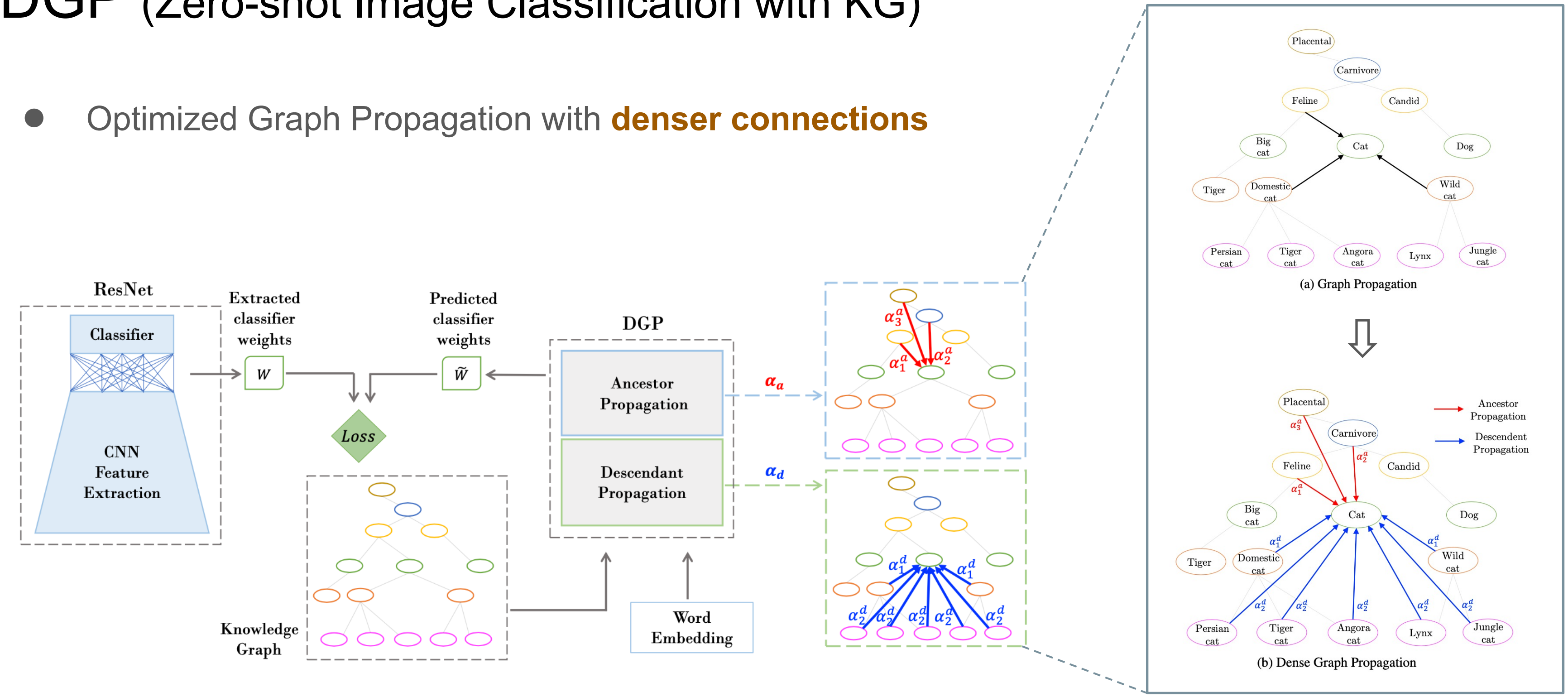
(a) Graph Propagation

(b) Dense Graph Propagation

Kampffmeyer, Michael, et al. "Rethinking knowledge graph propagation for zero-shot learning." CVPR 2019.

DGP (Zero-shot Image Classification with KG)

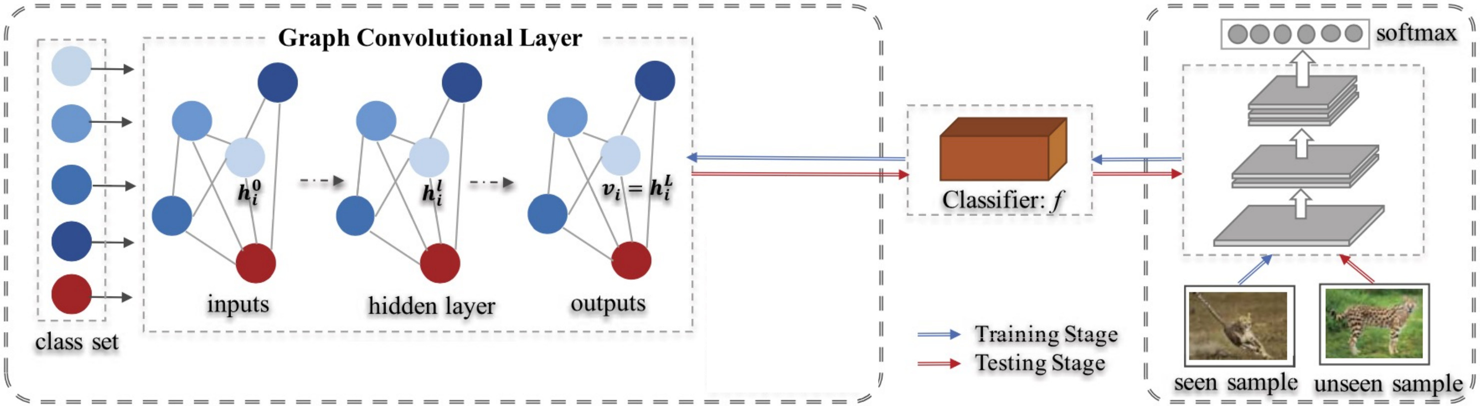
- Optimized Graph Propagation with **denser connections**



Kampffmeyer, Michael, et al. "Rethinking knowledge graph propagation for zero-shot learning." CVPR 2019.

Attentive GCN (Zero-shot Image Classification with KG)

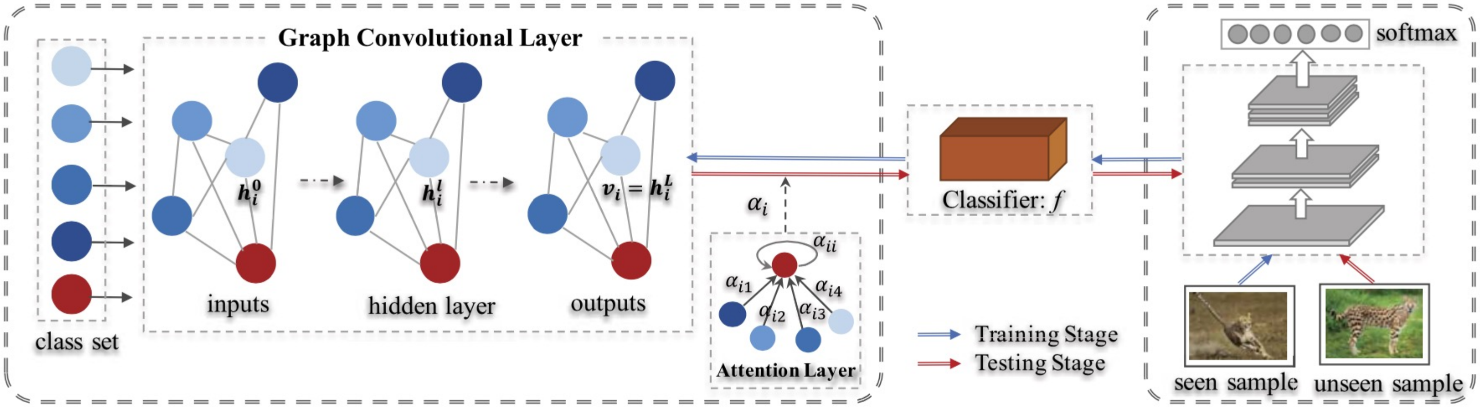
- Optimized Graph Propagation with **attention, to highlight significant neighbors**



Geng, Yuxia, et al. "Explainable zero-shot learning via attentive graph convolutional network and knowledge graphs." Semantic Web Journal 2021.

Attentive GCN (Zero-shot Image Classification with KG)

- Optimized Graph Propagation with **attention, to highlight significant neighbors**

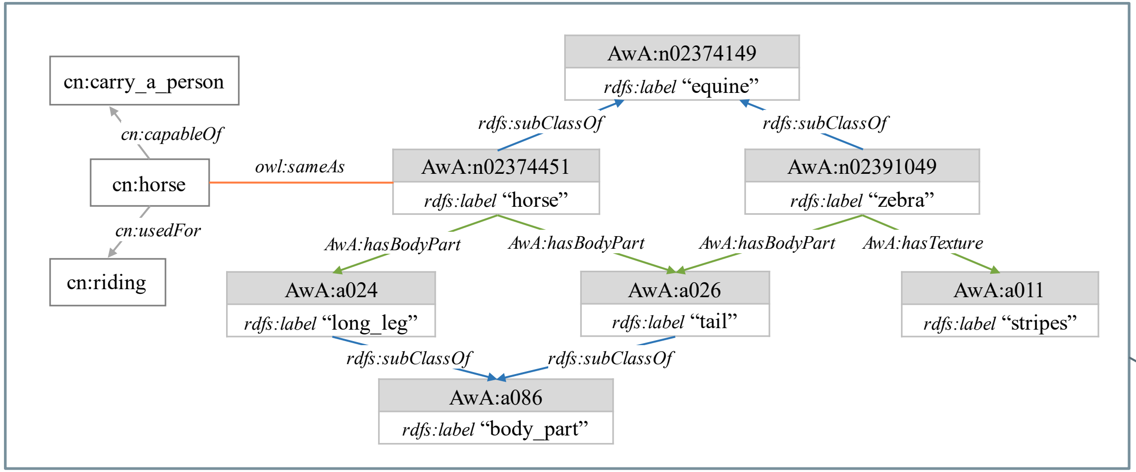


$$\alpha_{ij} = \frac{\exp(\cos(v_i, v_j))}{\sum_{k \in \mathcal{N}_i} \exp(\cos(v_i, v_k))} \qquad \bar{v}_i = \sum_{j \in \mathcal{N}_i} \alpha_{ij} \cdot v_j$$

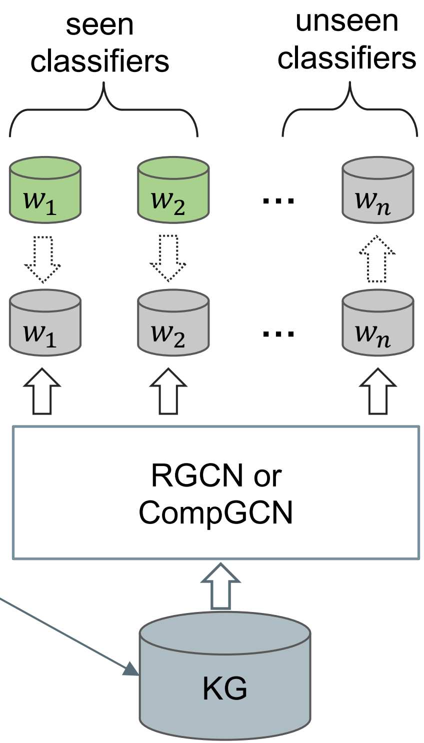
Geng, Yuxia, et al. "Explainable zero-shot learning via attentive graph convolutional network and knowledge graphs." Semantic Web Journal 2021.

Graph Propagation with Disentangled Ontology Embedding

- Optimized Graph Propagation with **relation-aware GCN**



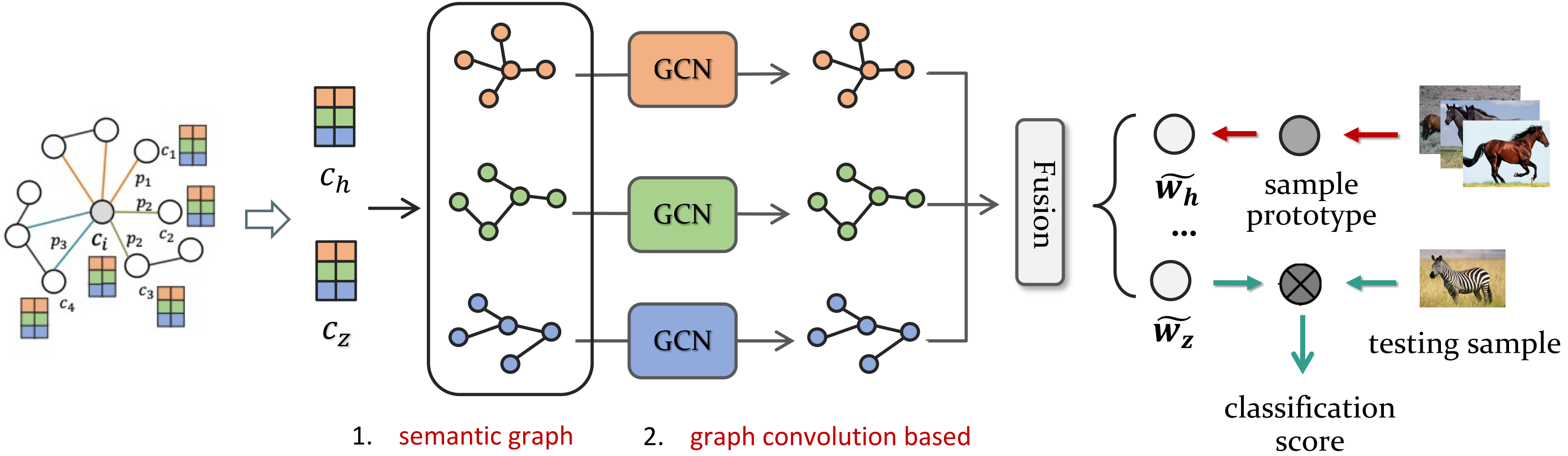
A KG fragment for animal classes, including 6 relations for different semantics



Geng, Yuxia, et al. "Disentangled Ontology Embedding for Zero-shot Learning." KDD 2022.

Graph Propagation with Disentangled Ontology Embedding

- Optimized Graph Propagation with **disentangled semantic graphs**



1. semantic graph construction

$$A_k(i,j) = \begin{cases} 1 & \text{if} \quad sim(\boldsymbol{c}_i^k, \boldsymbol{c}_j^k) > \tau \\ 0 & \text{otherwise} \end{cases}$$

2. graph convolution based feature propagation

$$H_k^{l+1} = \sigma(\widehat{A}_k H_k^l \Phi_k^l)$$

3. classifier fusion

$$\widetilde{W} = \frac{1}{K} \sum_k Z_k$$

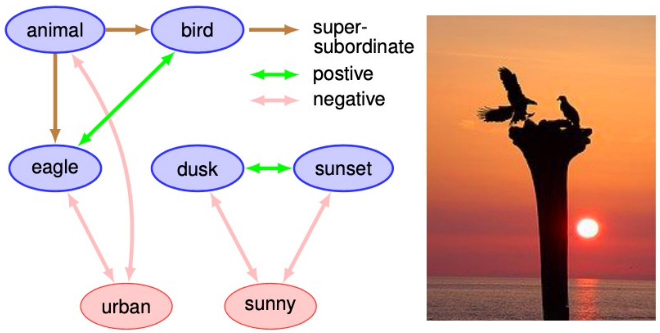
$$\widetilde{W} = W_1([Z_1; Z_2; \dots; Z_K])$$

Training Stage
Testing Stage

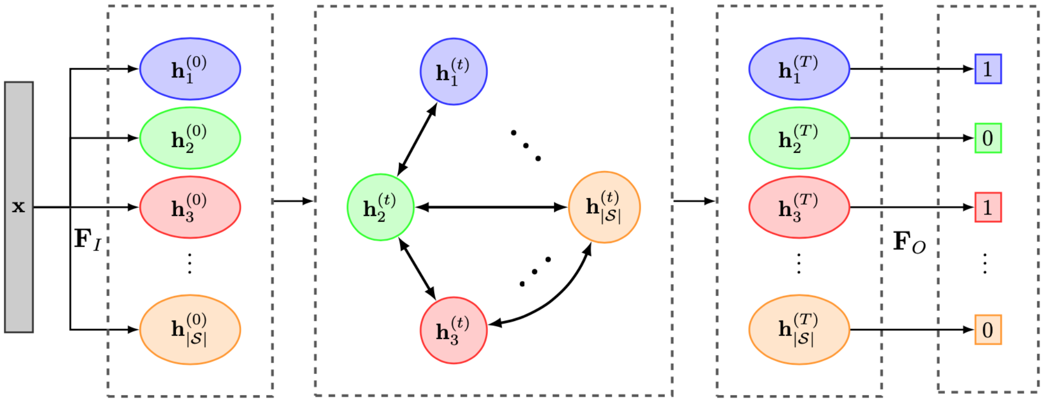
Geng, Yuxia, et al. "Disentangled Ontology Embedding for Zero-shot Learning." KDD 2022.

Label Score Propagation

- Zero-shot Multi-label Image Classification & Knowledge Graph



Structured Label Relationship

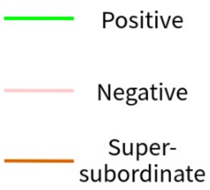


Graph Gated Neural Networks

Lee, Chung-Wei, et al. "Multi-label zero-shot learning with structured knowledge graphs." CVPR 2018.

Label Score Propagation

- Zero-shot Multi-label Image Classification & Knowledge Graph

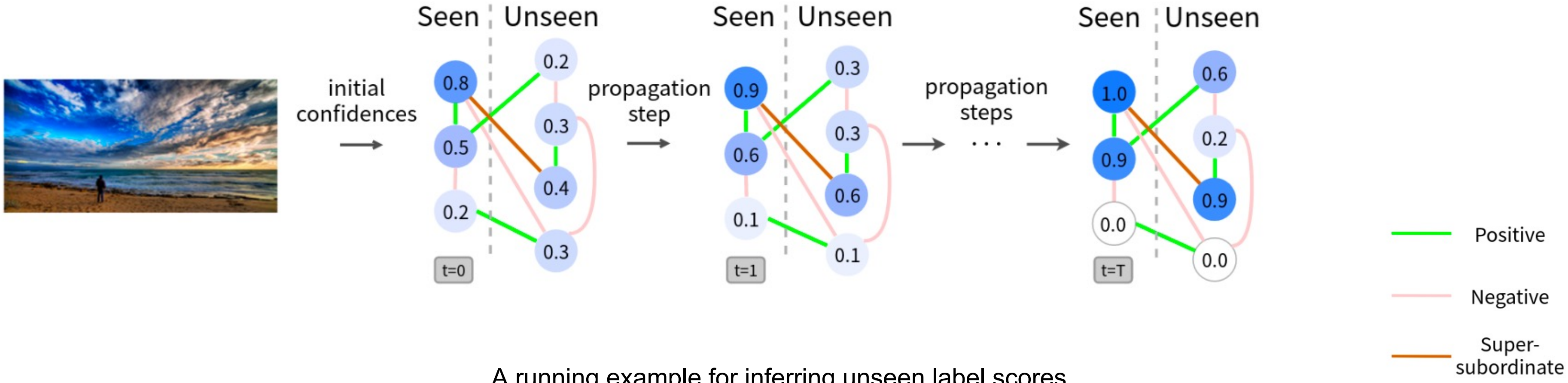


A running example for inferring unseen label scores

Lee, Chung-Wei, et al. "Multi-label zero-shot learning with structured knowledge graphs." CVPR 2018.

Label Score Propagation

- Zero-shot Multi-label Image Classification & Knowledge Graph



A running example for inferring unseen label scores

Lee, Chung-Wei, et al. "Multi-label zero-shot learning with structured knowledge graphs." CVPR 2018.

More Reading

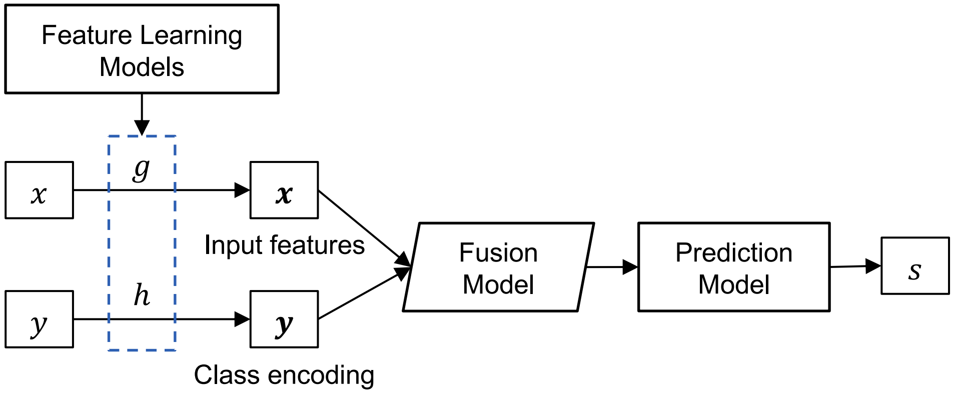
- Wang et al., Zero-Shot Learning via Contrastive Learning on Dual Knowledge Graphs, ICCV 2021
- Chen et al., Zero-shot Ingredient Recognition by Multi-Relational Graph Convolutional Network, AAAI 2020
- Wei et al., Residual Graph Convolutional Networks for Zero-Shot Learning, ACMMM Asia 2019
- Bosselut et al., Dynamic Neuro-Symbolic Knowledge Graph Construction for Zero-Shot Commonsense Question Answering, AAAI 2021

&4

Class Feature Paradigm

General Framework

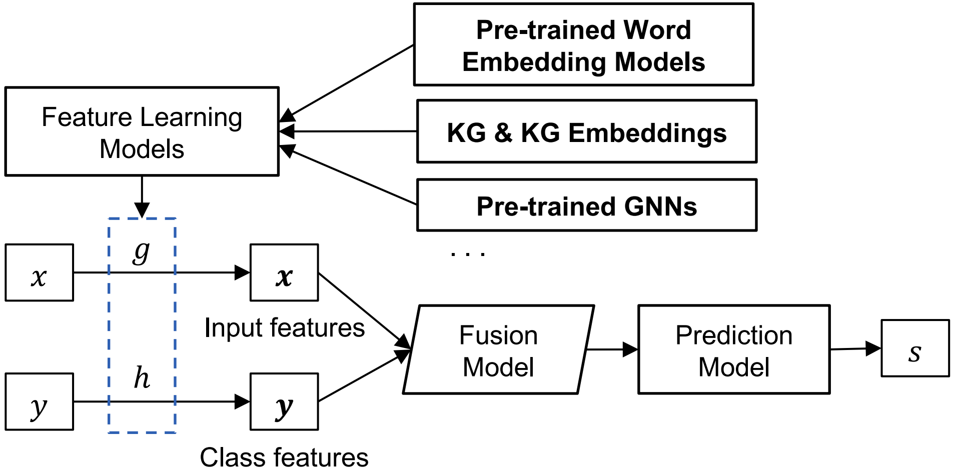
- Class encoding $g(x)$ and input feature $h(y)$ are fed into one model
- The model predicts a score which indicates whether the input x matches the class y
- The model is usually learned from D_{tr}



- In prediction, the encoding of an unseen class is input; and the original ZSL problem in transformed into a domain adaption problem

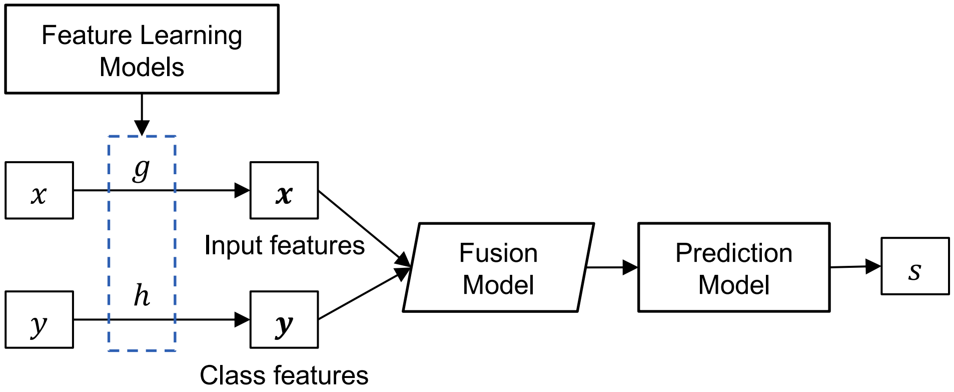
General Framework

- External knowledge are incorporated and utilized mainly via the the class encoding (h):
  - Text features e.g., TF-IDF
  - Text embeddings via e.g., pre-trained word embedding models and language models
  - External features extracted from KGs e.g., common sense knowledge of image contents
  - (Pre-trained) KG embeddings and Graph Neural Networks
  - Etc.



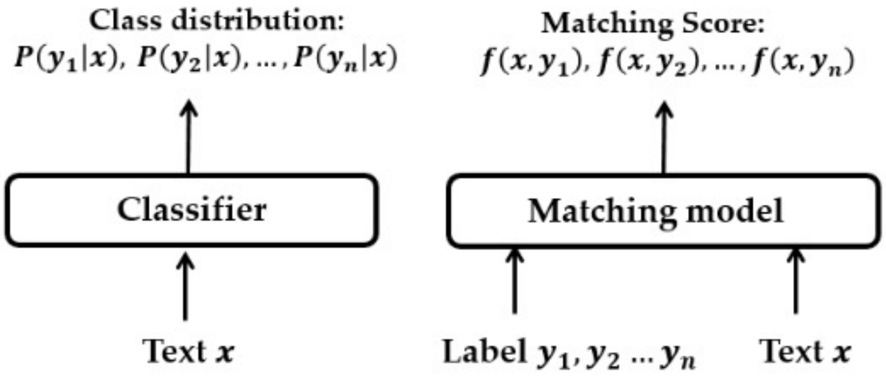
Text Feature vs Multi-modal Feature Fusion

- Two categories
  - Text Feature Fusion: both $\boldsymbol{x}$ and $\boldsymbol{y}$ are features learned from text
  - Multi-modal Feature Fusion: $\boldsymbol{x}$ and $\boldsymbol{y}$ are features of different modalities (e.g., text feature vs KG embedding, image feature vs text embedding)



Text Feature Fusion

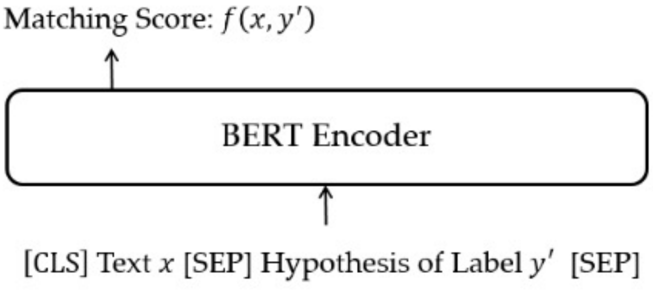
- Typical ZSL tasks
  - Sentiment (text) classification where both input and output (class label) are text



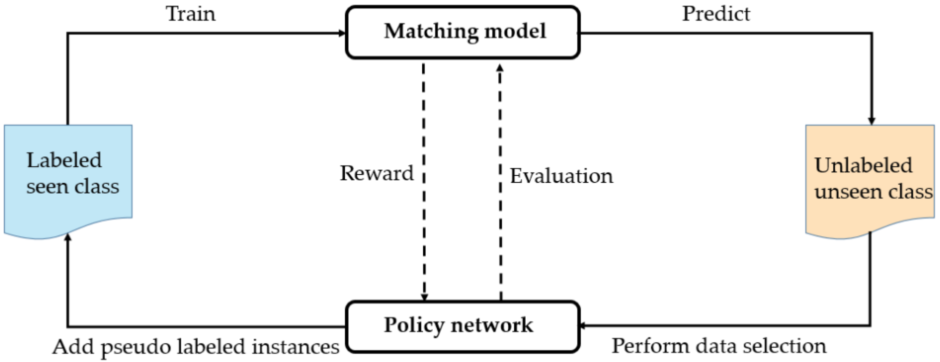
(a) Traditional Classifier (b) Standard ZSL Model

Text Feature Fusion

- Typical ZSL tasks
  - Sentiment (text) classification where both input and output (class label) are text
    - See bellow for a zero-shot text classification method (ACL'2020)



1. Pre-trained BERT as the base matching model



2. Reinforced self-training framework (MLP for Policy Network)

Ye, Zhiquan, et al. "Zero-shot text classification via reinforced self-training." Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020.

Text Feature Fusion

- Typical ZSL tasks
  - Sentiment (text) classification
  - Question answering
  - KG completion where entities and relations are described by name phrases and/or textual descriptions

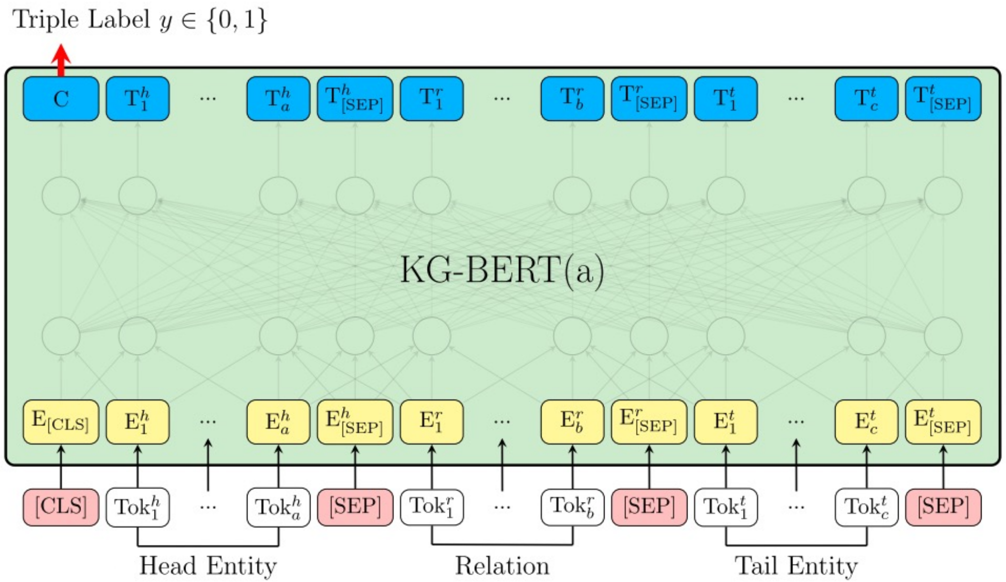
- Next, we will introduce a typical KG completion method named KG-BERT which can utilize the entity and relation labels for addressing unseen entities and/or relations

Yao, Liang, Chengsheng Mao, and Yuan Luo. "KG-BERT: BERT for knowledge graph completion." arXiv preprint arXiv:1909.03193 (2019).

KG-BERT

- Represent entities and relations by their names or descriptions
- Loss for fine-tuning (negative triples are generated by corrupting tail entities and head entities):

$$\mathcal{L} = -\sum_{\tau \in \mathbb{D}^+ \cup \mathbb{D}^-} (y_\tau \log(s_{\tau 0}) + (1 - y_\tau) \log(s_{\tau 1}))$$



Predict the likelihood of a triple

$$\mathbf{s}_\tau = f(h, r, t) = \text{sigmoid}(CW^T)$$

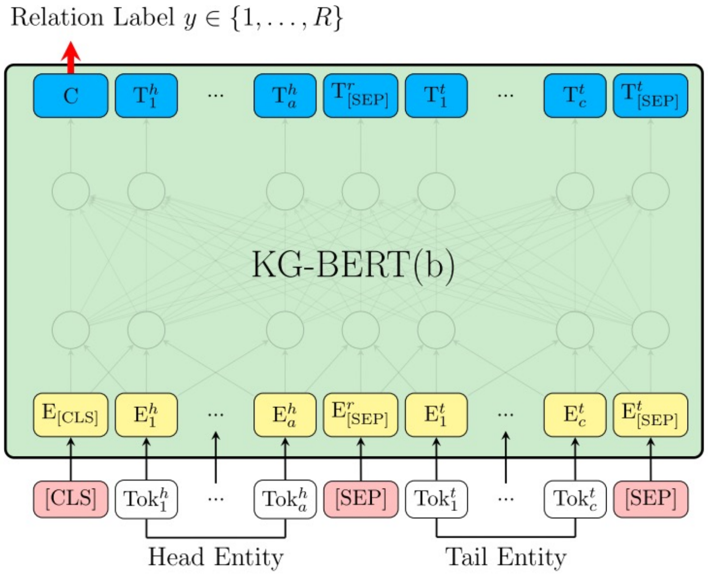
$$\mathbf{s}_\tau \in \mathbb{R}^2$$

KG-BERT

- Cross-entropy for fine-tuning with negative triples generated by corrupting the relations

$$\mathcal{L}' = -\sum_{\tau \in \mathbb{D}^+} \sum_{i=1}^{R} y'_{\tau i} \log(s'_{\tau i})$$

- Better performance than traditional KG embedding methods e.g., TransE and DistMult on typical benchmarks e.g., FB15K and WN18RR
- KG-BERT (b) performs better than KG-BERT (a)



Predict the relation of two entities

$$\mathbf{s}'_\tau = f(h, t) = \mathrm{softmax}(CW'^T)$$

$$\mathbf{s}'_\tau \in \mathbb{R}^R$$

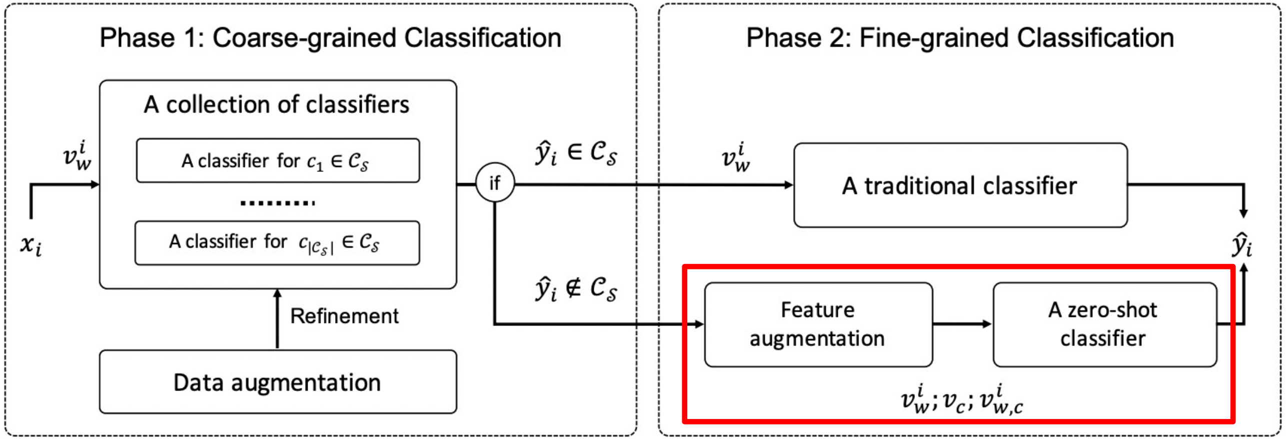
Multi-modal Feature Fusion

- Input features x and class features y are of different modalities
- In some tasks with text input, KG is used as side information (the class labels are aligned with KG entities), and KG features are extracted and fused
    - Case study #1: zero-shot text classification [Zhang et al. 2019]

Zhang, Jingqing, Piyawat Lertvittayakumjorn, and Yike Guo. "Integrating Semantic Knowledge to Tackle Zero-shot Text Classification." ACL 2019.

Multi-modal Feature Fusion

- Case study #1: zero-shot text classification [Zhang et al. 2019]



v_w^i: the embedding of each word in the document x_i (the input feature)

v_c: the embedding of each class label

$v_{w,c}^i$: additional features indicating how the word w and the class c are related considering the relations in a general KG such as ConceptNet

Multi-modal Feature Fusion

- Case study #1: zero-shot text classification [Zhang et al. 2019]
  - Assume the class c is “Educational Institution”, three sets of ConceptNet nodes are extracted:
    - the_class_nodes: educational_institution, educational, institution
    - superclass_nodes: organization, node
    - description_nodes: place, people, ages, education
  - To calculate $v_{w,c}^{i}$, the word w is compared with each node set; for example, if w is inside a node set, its corresponding vector slot is set to 1, otherwise, it is set to 0

Multi-modal Feature Fusion

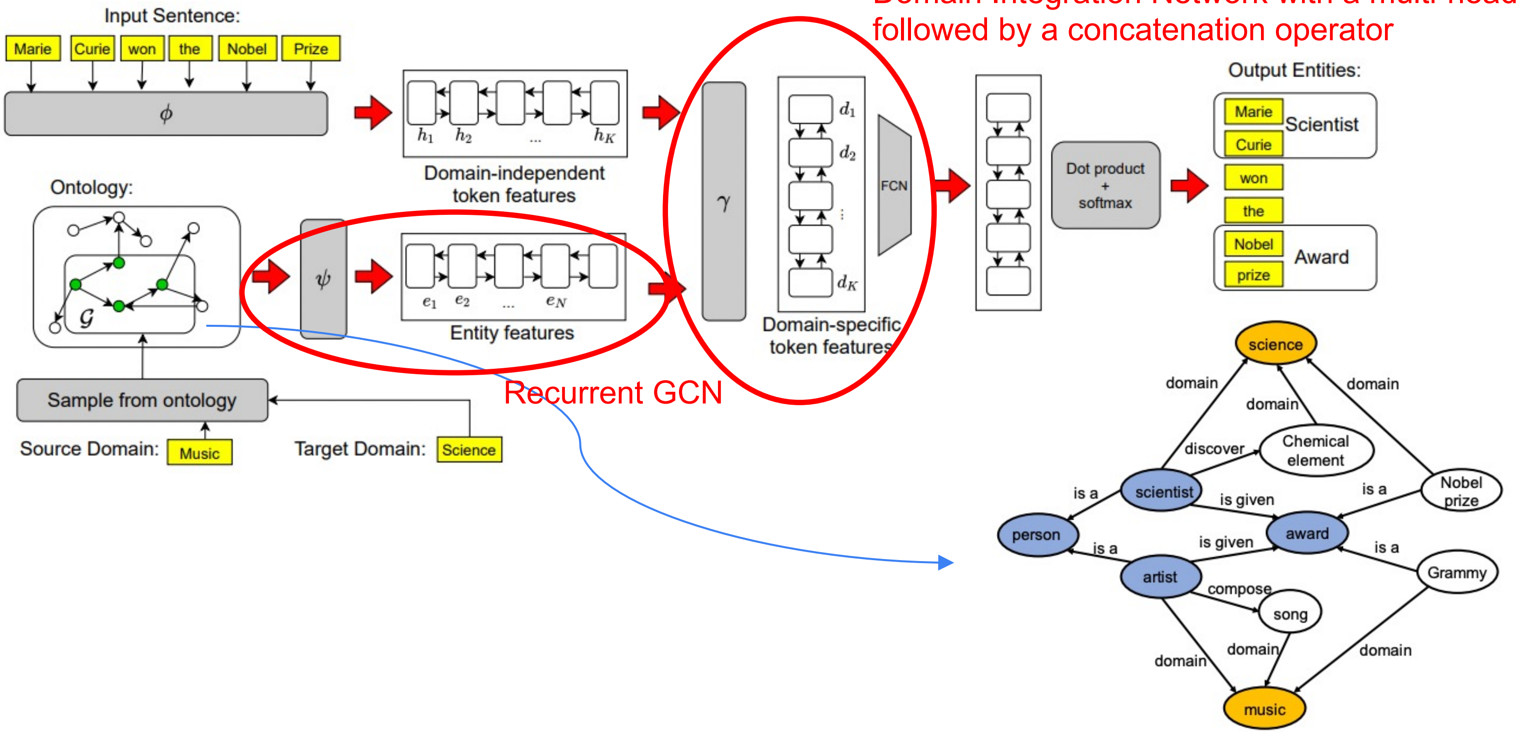
- Case study #2: **DOZEN** -- zero-shot named entity recognition
  - The entities to recognize are not only unseen in training, but also from a different domain as the training entities (domains e.g., Music, Natural Science, Politics)
  - $\boldsymbol{x}$: the input sentence is encoded by a pre-trained BERT
  - $\boldsymbol{y}$: the entity (class label) is encoded as graph features learned by a Recurrent GCN over an ontology (as external knowledge)

Nguyen, Hoang-Van, Francesco Gelli, and Soujanya Poria. "DOZEN: cross-domain zero shot named entity recognition with knowledge graph." SIGIR 2021.

Multi-modal Feature Fusion

- Case study #2: **DOZEN** -- zero-shot named entity recognition

Domain Integration Network with a multi-head attention followed by a concatenation operator



Comparison

	Basic Idea	Pros	Cons
Mapping-based	Map the input and/or class into the same space for matching	Simple, widely investigated	Often biased to seen classes in generalized ZSL
Generative Models	Generate samples of unseen classes conditioned on their semantic embeddings	Avoid prediction bias towards seen classes, flexible	Hard to train generative models
Propagation-based	Propagate model parameters or class scores over a graph	Directly and well utilize the graph structure	Inflexible to incorporate other kinds of knowledge
Class Feature based	Move the class encoding as new input	Flexible to use well-developed class encoding models	The domain adaption problem remains and is hard to be addressed

Explainability is also important!

&5

Explainable Zero-shot Learning

Explainable AI (X-AI)

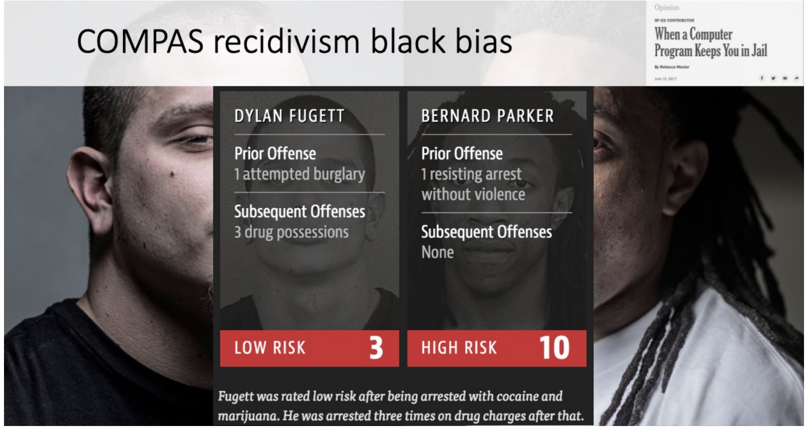
- Motivation
  - Create a suite of techniques that produce more explainable models while maintaining a high level of searching, learning, planning and reasoning performance (accuracy, precision and optimization)
  - Enable human users to understand, appropriately trust, and efficiently manage the emerging generation of AI systems
  - Improve the safety and fairness of (critical) AI systems
  - Etc.

Explainable AI (X-AI)

- Examples

Finance:

- Credit scoring, loan approval
- Insurance quotes



The Big Read Artificial intelligence + Add to myFT

Insurance: Robots learn the business of covering risk

Artificial intelligence could revolutionise the industry but may also allow clients to calculate if they need protection

Save

Oliver Ralph MAY 16, 2017 24

https://www.ft.com/content/e07cee0c-3949-11e7-821a-6027b8a20f23

Freddy Lécué, "On the Role of Knowledge Graphs in Explainable AI -- A Machine Learning Perspective", October 27th, 2019, International Semantic Web Conference – Workshop on Semantic Explainability, Auckland, New Zealand

Explainability in Machine Learning

- Categories/targets
  - Model interpretation (how does a model work?)
  - Prediction justification (why the prediction is good or bad?)
- Methods
  - Inherently interpretable models, e.g., rule set and sparse linear model
  - Black-model approximation
  - Visualization
  - Attention
  - …
- Human-centric explanation
  - People without machine learning expertise can understand
    - E.g., Generate captions to explain image classification [Hendricks et al. ECCV'16]
    - E.g., Describe features for stock price prediction with articles of Wikipedia [Biran and McKeown IJCAI'17]
    - E.g., Explain data clusters with entities of Linked Data [Tiddi et al. ESWC'14]

KG in Machine Learning Explanation

- External knowledge for human-centric machine learning explanation
- Post-explanation of prediction results
- Feature and model understanding with external knowledge

Case study -- "How to explain feature transferability from domain to domain? Which domains to transfer?"

Chen, Jiaoyan, et al. "Knowledge-based transfer learning explanation." Sixteenth International Conference on Principles of Knowledge Representation and Reasoning. 2018.

Knowledge-based Transfer Learning Explanation

D1: Flight of Delta Airlines from ORD to LAX

Positive transfer

D2: Flight of American Airlines from ORD to SFO

Negative Transfer

D3: Flight of JetBlue from LAX to JFK

$$Dep \sqcap \exists hasDelMin.\{Pos\} \sqsubseteq DelayedDep \tag{1}$$
$$Dep \sqcap \exists hasDelMin.\{Neg\} \sqsubseteq OnTimeDep \tag{2}$$
$$hasCarrier \circ hasCarHub \sqsubseteq hasDepHub \tag{3}$$
$$hasNebApt \circ hasRecDep \sqsubseteq hasRecNebDep \tag{4}$$
$$Dep \sqcap \exists hasOri.\{CA\} \sqcap \exists hasDes.\{CA\} \sqsubseteq \exists withIn.\{CA\} \tag{5}$$
$$\exists withIn.\top \sqsubseteq InStateDep \tag{6}$$

$Airport(LAX)$	(7)	$locatedIn(LAX, CA)$	(8)
$Carrier(DL)$	(9)	$Departure(d)$	(10)
$hasDelMin(d, Pos)$	(11)	$hasWea(d, wea)$	(12)
$hasOri(d, LAX)$	(13)	$hasCarrier(d, DL)$	(14)
$Airport(JFK)$	(15)	$hasDes(d, JFK)$	(16)
$LAX = ori$	(17)	$DL = car$	(18)
$hasRecDep(d, d_1)$	(19)	$hasCarrier(d_1, MU)$	(20)
$hasRecDep(d, d_2)$	(21)	$hasCarrier(d_2, AA)$	(22)
$DelayedDep(d)$	(23)	$HeavySnow(wea)$	(24)

Case: CNN features learned from historical records for flight delay forecasting

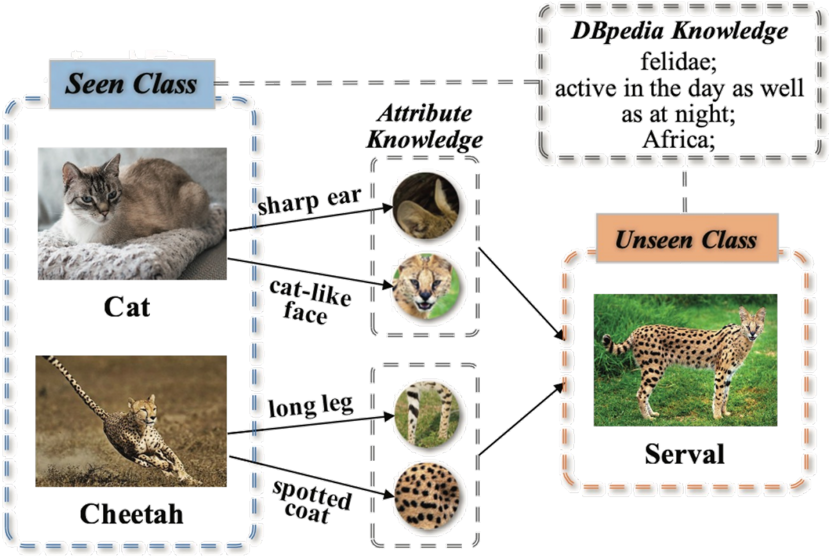
Why?

An ontology/KG with knowledge on flights, airports, cities …

Explanations:
1. The original airports of D1 and D2 are the same
2. The target airports of D1 and D2 are both in California
3. The carrier of D1 is a big airline company while that of D3 is a small airline company
4. …

A KG-based Explainable ZSL Method

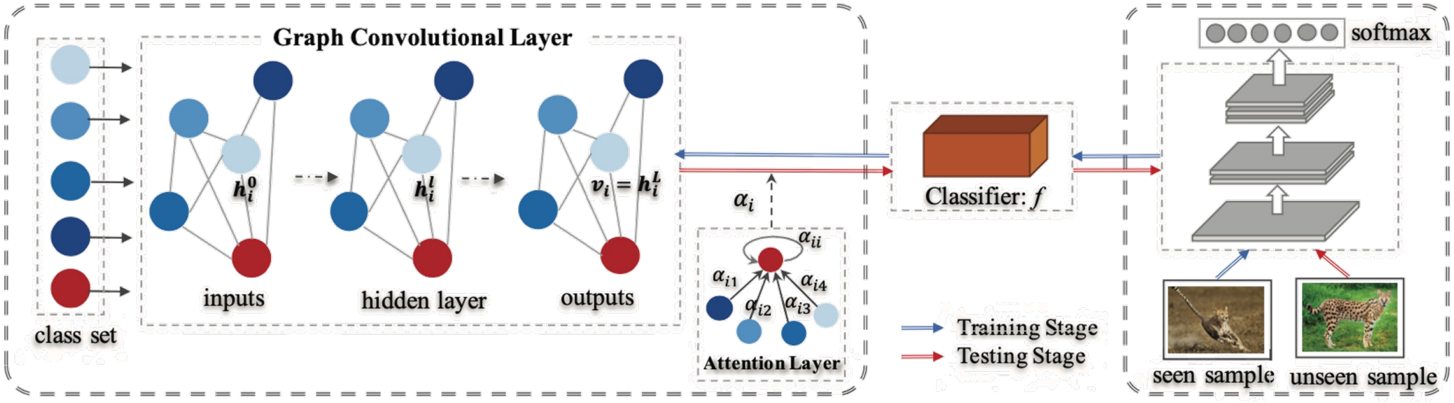
"Why does a model transfer features of seen classes *Cat* and *Cheetah* to unseen *Serval*? Why the prediction for *Serval* is trustful?"



Geng, Yuxia, et al. "Explainable zero-shot learning via attentive graph convolutional network and knowledge graphs." Semantic Web Preprint (2021): 1-28.

A KG-based Explainable ZSL Method

- Module #1: a **ZSL Learner** of the propagation-based paradigm)
  - Classes are aligned with WordNet nodes
  - Attentive Graph Convolutional Network
    - Attentions: which seen classes play a more important role to an unseen classes, i.e., impressive seen classes

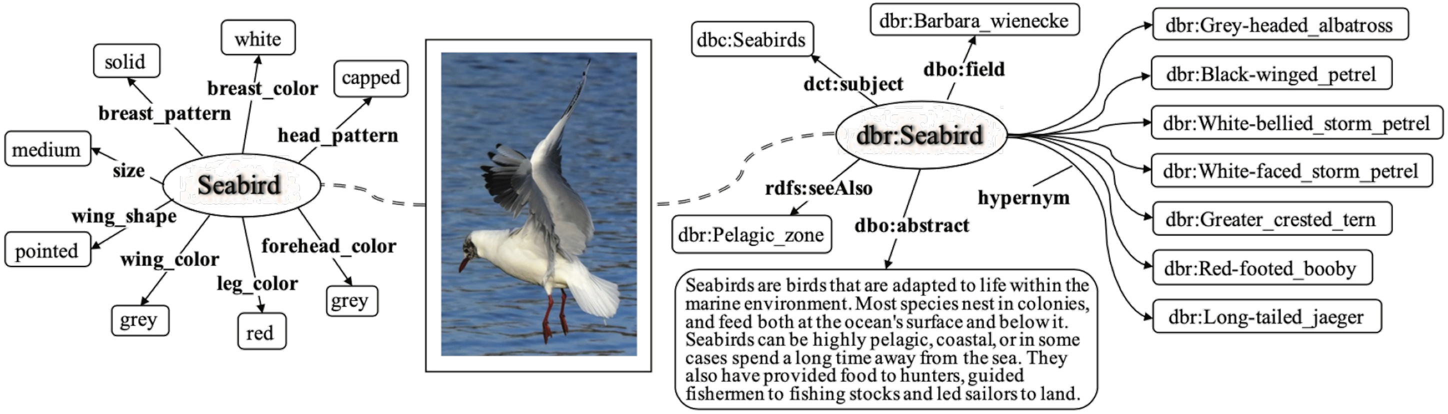


A KG-based Explainable ZSL Method

- Module #2: **Explanation Generator** with impressive seen classes
  - KG: constructed by attributes and DBpedia knowledge

A KG-based Explainable ZSL Method

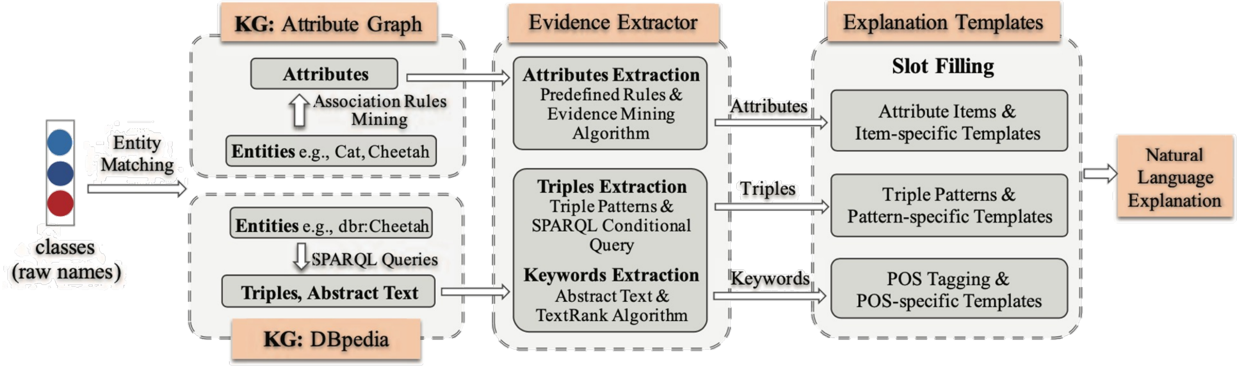
- An example of the KG resources for the animal class *Seabird*



(Attribute Knowledge) (DBpedia Knowledge)

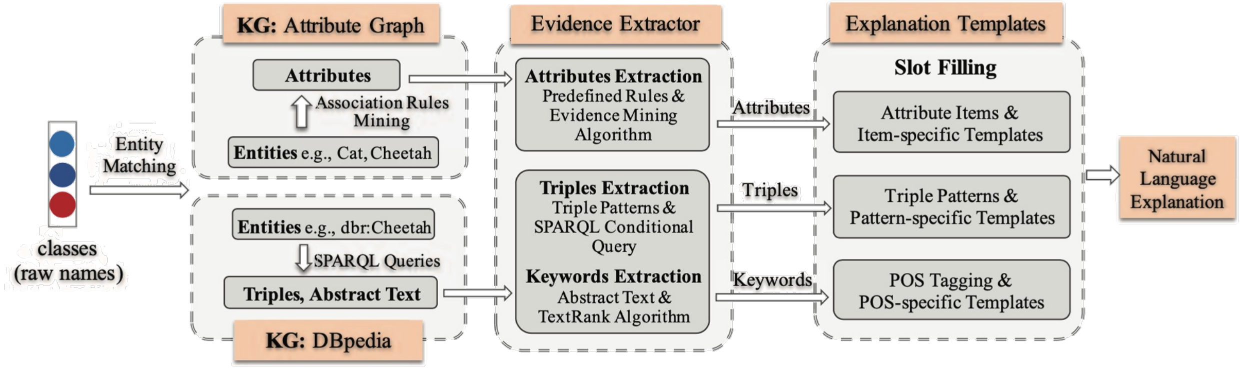
A KG-based Explainable ZSL Method

- Module #2: **Explanation Generator** with impressive seen classes
  - KG: constructed by attributes and DBpedia knowledge
  - Evidence extractor:
    - Rules for **common attributes**
    - **Common triples** and **abstract text** from DBpedia by
      - Its lookup service for relevant DBpedia entities
      - SPARQL query and heuristic triple patterns for common triples



A KG-based Explainable ZSL Method

- Module #2: Explanation Generator based on impressive seen classes
  - KG: constructed by attributes and DBpedia knowledge
  - Evidence extractor
  - Human understandable explanations by templates for different evidences (see explanation examples)



A KG-based Explainable ZSL Method

- Two explanation examples

		Dolphin (Unseen Class)	IMSCs of Dolphin
			Killer whale (w: 1.0)
Image			
DBpedia Entity		dbr:Dolphin	dbr:Killer_whale
Knowledge from Attribute Graph		hairless, toughskin, flippers, swims, tail, ocean, grouped, smart, fast, active	
		For rule $\{killer\ whale\} \Rightarrow \{dolphin\}$: sup. = 48.7%; con. = 62.1%	
Knowledge from DBpedia	**DBpedia Property/ Relation**	(dbr:Dolphin, dct:subject, dbc:Animals_that_use_echolocation)	(dbr:Killer_whale, dct:subject, dbc:Animals_that_use_echolocation)
	Keywords of Abstract Text	aquatic, shaped **teeth**, well-developed **hearing**, widespread, **blubber** under skin	oceanic, apex predators, **toothed** whale, a layer of **blubber**, excellent **hearing**, diverse diet
Generated Explanation & Score		The prediction for samples of **dolphin** is supported by **killer whale**. *They are both hairless, grouped, both have tough skin, flippers, tail, both swim, live in ocean, and behave smart, fast and active.* They are both **animals that use echolocation**. They both have **teeth**, **hearing** and **blubber**.	
		Readability [G/M/B]: M **Rationality [G/M/B]: M**	

Stork (Unseen Class)	IMSCs of Stork	
	White stork (w: 0.51)	**Black stork (w: 0.49)**
		
dbr:Ciconiiformes	dbr:White_stork	dbr:Black_stork
white, black, water, wild, fish		
For rule $\{white\ stork, black\ stork\} \Rightarrow \{stork\}$: sup. = 83.3%; con. = 100.0%		
(dbr:White_stork, dbo:order, dbr:Ciconiiformes) (dbr:Black_stork, dbo:order, dbr:Ciconiiformes)		
large, **long-legged**, long-necked, **wading**, **birds**, soaring, long, stout bills, **migratory**	**large bird**, **long** red **legs**, **wading**, family, Ciconiidae, **migrant**, carnivore	**large bird**, **wading**, family, Ciconiidae, black plumage, **long** red **legs**, red beak
The prediction for samples of **stork** is supported by **white stork** and **black stork**. *They are both white, black, wild, both live in water and eat fish.* **White stork** and **black stork** both belong to **stork** biologically. They are both **large wadding birds**, and are similar in **long legs**.		
Readability [G/M/B]: G **Rationality [G/M/B]: G**		

Conclusion

- Mapping-based paradigm
  - DeViSE [Frome et al. NeurIPS'13], Logic-guided Relation Extraction [Li et al. COLING'20], Ontology-guised Sematic Composition [Chen et al. KR'20]
- Data augmentation-based paradigm
  - OntoZSL [Geng et al. WWW'21], DOZSL [Geng et al. KDD'22]
- Propagation-based paradigm
  - GCNZ [Wang et al. CVPR'18], DGP [Kampffmeyer et al. CVPR'19], ML-ZSL [Lee et al. CVPR'18], Attentive GCN [Geng et al. Semantic Web Journal 2021], DOZSL [Geng et al. KDD'22]
- Class feature paradigm
  - KG-BERT [Yao et al. 2019], text classification [Zhang et al. ACL'19] [Ye et al. ACL'20], DOZEN [Nguyen et al, SIGIR'21]
- Explainable ZSL
  - Attentive GCN [Geng et al. Semantic Web Journal 2021], Transfer Learning Explanation [Chen et al. KR'18]

See our hands-on session (Part III) for more details of experimenting with representative KG-aware ZSL methods

Thanks!

Contacts:

Jiaoyan Chen (jiaoyan.chen@cs.ox.ac.uk)
Yuxia Geng (gengyx@zju.edu.cn)