Learning representation for Out-of-Vocabulary (OOV) words is challenging.

**Challenge**: OOV in new corpus sometimes have insufficient contexts, and its morphology might not be informative enough.

**Goal**: Can we learn accurate embedding for OOV words by observing their usages in only a few context sentences?

**Our Approach**: 1) Design a model with enough capacity to capture semantics and relations in context sentences. 2) Train the model by simulating an OOV scenario, such that the model learns to generalize well for OOV on a new corpus.

**Our Solution**: Formulate as K-shot Regression Problem

- Generate simulated k-shot data
- K-shot Representation Training Framework
  - (3) Use the embedding learned with all contexts as Oracle Embedding
  - Regression Loss (Cosine Distance)
  - Morphology information
  - Hierarchical Context Encoder (HiCE)
  - Update the model
  - Predict Embedding

**Intrinsic Evaluation (Chimera Dataset)**
- Chimera provides K context sentences for a OOV word and six probe words, with human label of similarity between each probe word and the OOV word.
- The task requires the estimated embedding similarity close to the human label with human label of similarity between each probe word and the OOV word.
- Character sequence (C)
- Intrinsic Evaluation (Chimera Dataset)
- Extrinsic Evaluation (NER, POS Tagging)

**Extrinsic Evaluation (NER, POS Tagging)**
- Word Embedding (T)
- Transformer Encoder Layer
- Concatenation
- Output Layer
- Training Corpus
- New Corpus

**Qualitative Evaluation (Attention Visualization and Case Study)**
- Our proposed few-shot representation learning approach benefits downstream tasks.
- With MAML, the performance improves further.

**MAML**
- Goal: Learn to adapt to new domain with a few examples
- Solution: Model Agnostic Meta-Learning (MAML) [1]
- Procedure:
  1) Run a few steps on training corpus (D_T) to obtain a promising initialization.
  2) Update the model on new corpus (D_N) with the gradient calculated at that initialization point.

**Use the Oracle Embedding**
- Lower Context Encoder encodes a single context sentence into vector E(s)
- Upper Multi-Context Aggregator combines multiple context vectors E(s) with different attention
- The morphological information is incorporated by a Char-CNN

**Innovative World of MAML**
- MAML is helpful when context is rich (4 or 6)
- Morphological information is useful when context is limited, but not helpful when context is rich (4 or 6)
- MAML is helpful when context is rich

**Challenge**
- OOV in new corpus sometimes have insufficient contexts, and its morphology might not be informative enough.

**Goal**
- Can we learn accurate embedding for OOV words by observing their usages in only a few context sentences?

**Our Approach**
- Design a model with enough capacity to capture semantics and relations in context sentences.
- Train the model by simulating an OOV scenario, such that the model learns to generalize well for OOV on a new corpus.

**Learning representation for Out-of-Vocabulary (OOV) words**

**Challenge**
- OOV in new corpus sometimes have insufficient contexts, and its morphology might not be informative enough.

**Goal**
- Can we learn accurate embedding for OOV words by observing their usages in only a few context sentences?

**Our Approach**
- Design a model with enough capacity to capture semantics and relations in context sentences.
- Train the model by simulating an OOV scenario, such that the model learns to generalize well for OOV on a new corpus.

**Learning representation for Out-of-Vocabulary (OOV) words is challenging**

**Challenge**: OOV in new corpus sometimes have insufficient contexts, and its morphology might not be informative enough.

**Goal**: Can we learn accurate embedding for OOV words by observing their usages in only a few context sentences?

**Our Approach**: 1) Design a model with enough capacity to capture semantics and relations in context sentences. 2) Train the model by simulating an OOV scenario, such that the model learns to generalize well for OOV on a new corpus.

**Learning representation for Out-of-Vocabulary (OOV) words is challenging**

**Challenge**: OOV in new corpus sometimes have insufficient contexts, and its morphology might not be informative enough.

**Goal**: Can we learn accurate embedding for OOV words by observing their usages in only a few context sentences?

**Our Approach**: 1) Design a model with enough capacity to capture semantics and relations in context sentences. 2) Train the model by simulating an OOV scenario, such that the model learns to generalize well for OOV on a new corpus.

**Learning representation for Out-of-Vocabulary (OOV) words**

**Challenge**: OOV in new corpus sometimes have insufficient contexts, and its morphology might not be informative enough.

**Goal**: Can we learn accurate embedding for OOV words by observing their usages in only a few context sentences?

**Our Approach**: 1) Design a model with enough capacity to capture semantics and relations in context sentences. 2) Train the model by simulating an OOV scenario, such that the model learns to generalize well for OOV on a new corpus.

**Learning representation for Out-of-Vocabulary (OOV) words is challenging**

**Challenge**: OOV in new corpus sometimes have insufficient contexts, and its morphology might not be informative enough.

**Goal**: Can we learn accurate embedding for OOV words by observing their usages in only a few context sentences?

**Our Approach**: 1) Design a model with enough capacity to capture semantics and relations in context sentences. 2) Train the model by simulating an OOV scenario, such that the model learns to generalize well for OOV on a new corpus.

**Learning representation for Out-of-Vocabulary (OOV) words**

**Challenge**: OOV in new corpus sometimes have insufficient contexts, and its morphology might not be informative enough.

**Goal**: Can we learn accurate embedding for OOV words by observing their usages in only a few context sentences?

**Our Approach**: 1) Design a model with enough capacity to capture semantics and relations in context sentences. 2) Train the model by simulating an OOV scenario, such that the model learns to generalize well for OOV on a new corpus.

**Learning representation for Out-of-Vocabulary (OOV) words is challenging**

**Challenge**: OOV in new corpus sometimes have insufficient contexts, and its morphology might not be informative enough.

**Goal**: Can we learn accurate embedding for OOV words by observing their usages in only a few context sentences?

**Our Approach**: 1) Design a model with enough capacity to capture semantics and relations in context sentences. 2) Train the model by simulating an OOV scenario, such that the model learns to generalize well for OOV on a new corpus.