The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21)

# Embracing Domain Differences in Fake News: Cross-domain Fake News Detection using Multi-modal Data

**Amila Silva, Ling Luo, Shanika Karunasekera, Christopher Leckie**
School of Computing and Information Systems
The University of Melbourne
Parkville, Victoria, Australia
{amila.silva@student., ling.luo@, karus@, caleckie@}unimelb.edu.au

**AAAI'21**

**220331 Chia-Chun Ho**

# Outline

Introduction

Related Works

Methodology

Experiments

Conclusion

Comments

# Introduction
## Motivation

- Social media is considered as one of the leading and fastest media to seek news information online.
  - Thus provide an ideal environment to spread fake news.
- Many times the cost and damage due to fake news are high and early detection to stop spreading such information is of importance.
  - Like the US president election, COVID-19 pandemic etc.
- Due to the high volumes of news generated on a daily basis, it's not practical to identify fake news using manual fact checking.

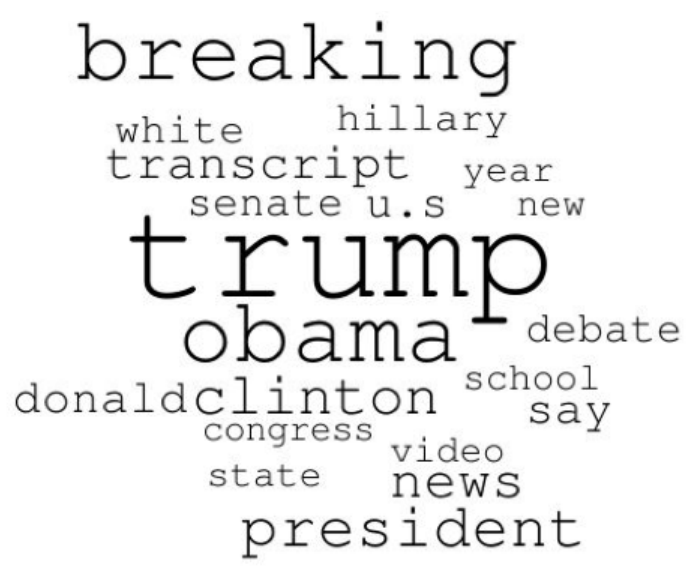
# Introduction

## Challenges (1/2)

| Feature | Weiner Index | Network Depth | Maximum Outdegree | Propagation Speed |
|---|---|---|---|---|
| p-value | 1.81e-2 | 5.81e-19 | 4.11e-4 | 3.42e-29 |

T-test result conducted using 2 datasets

- Most existing techniques are trained and evaluated using datasets that are limited to a single domain such as politics, entertainment, healthcare.
  - Real news stream typically covers a wide variety of domains.
  - Existing approaches perform poorly for such a cross-domain news dataset.
- Due to two reasons:
  - Domain-specific word usage
  - Domain-specific propagation patterns




PolitiFact


GossipCop

Word cloud for the top 20 words in two dataset

# Introduction
## Challenges (1/2)

- To address this challenge, some previous works learned models to overlook such domain-specific information and only rely on cross-domain information.
  - E.g., web-markup and readability features
- Domain-specific knowledge could be useful for accurate identification of fake news.
- As a solution, this work aims to address
  - preserve domain-specific and cross-domain knowledge in news at the same time
  - detect fake news in cross-domain datasets.

# Introduction

## Challenges (2/2)

- Studies show that most approach are not good at identifying fake news from unseen or rarely-seen domains during training.
- Models can be learned using a dataset that covers as many domains as possible.
- Here assume that FND model requires supervision as supervised techniques are known to be substantially better at identifying fake news compared to unsupervised methods.
- Due to sheer volume of unlabeled news available, there is a need to identify information news to annotate such that the labelled dataset ultimately covers many domains while avoiding any selection biases.

# Introduction
## Contributions

- Propose a multimodal fake news detection technique for cross-domain news datasets
    - that learns domain-specific & cross-domain information of news using two independent embedding spaces.
- Propose an unsupervised technique to select a given number of news from a large data pool
    - such that the selected dataset maximizes that domain coverage.

# Related Works

## Limitations

- Domain-agnostic FND has two limitations:
    - It assumes that the news records from different domains arrive sequentially, though this is not always true for real-world streams.
    - It requires the domain of news records to be known, which is not generally available.
- Active Learning for FND has two limitations:
    - It requires a pre-trained model to select instances.
    - it is known to be highly vulnerable to the biases introduced by the initial model.

# Problem Statement
## Notations

- Each news record $r$ in set of news $R$ ($r \in R$) is represented as a tuple $< t^r, W^r, G^r >$
  - $t^r$: timestamp when $r$ is published online
  - $W^r$: text content of $r$
  - $G^r(V^r, E^r, X^r)$: attributed directed graph represent the propagation network of $r$ for time bound $\Delta T$ ($\Delta T$ set 5 hours for evaluate early detection performance)
    - $V^r$: nodes represent the tweets/retweets of $r$
    - $E^r$: edges represent the retweet relationships among them.
    - $X^r$: set of attributes of the nodes in $G^r$.

# Problem Statement

## Sub-Tasks

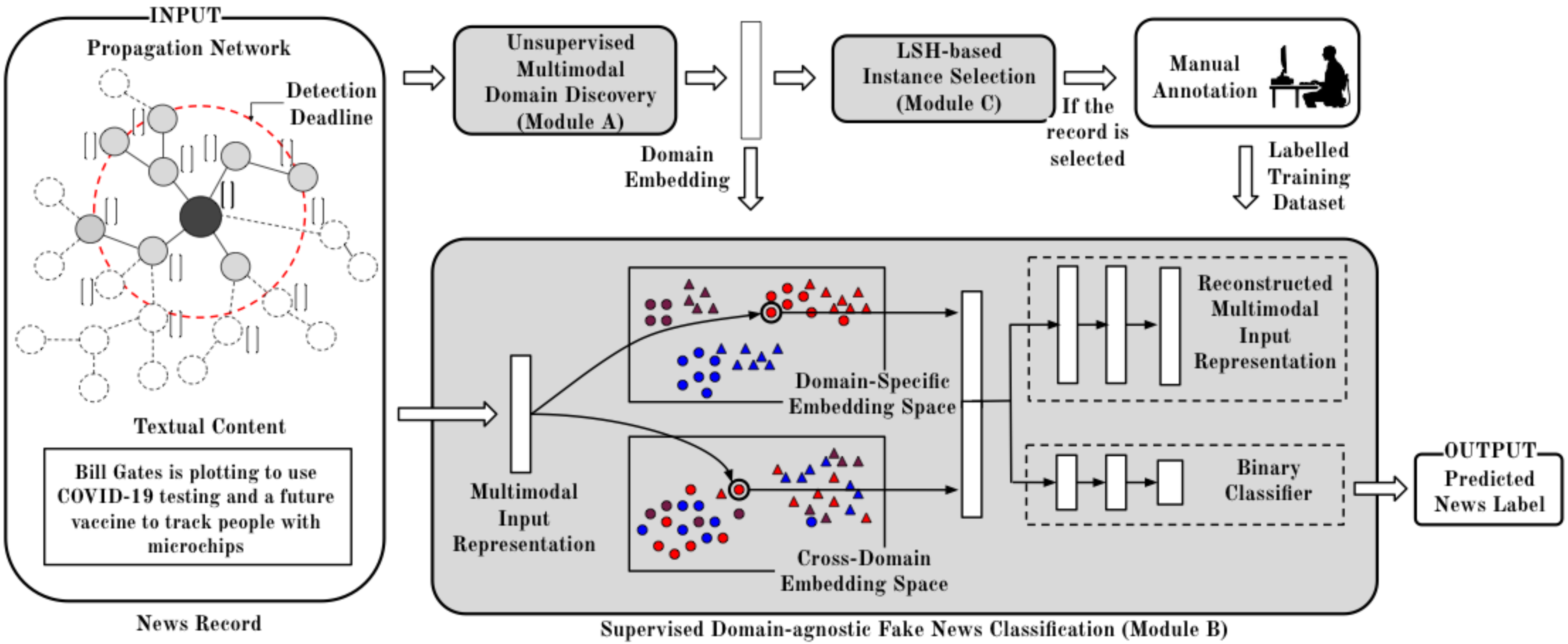
| Dataset | PolitiFact | GossipCop | CoAID |
|---|---|---|---|
| # Fake News | 269 | 1269 | 135 |
| # Real News | 230 | 2466 | 1568 |

Statics of 3 datasets

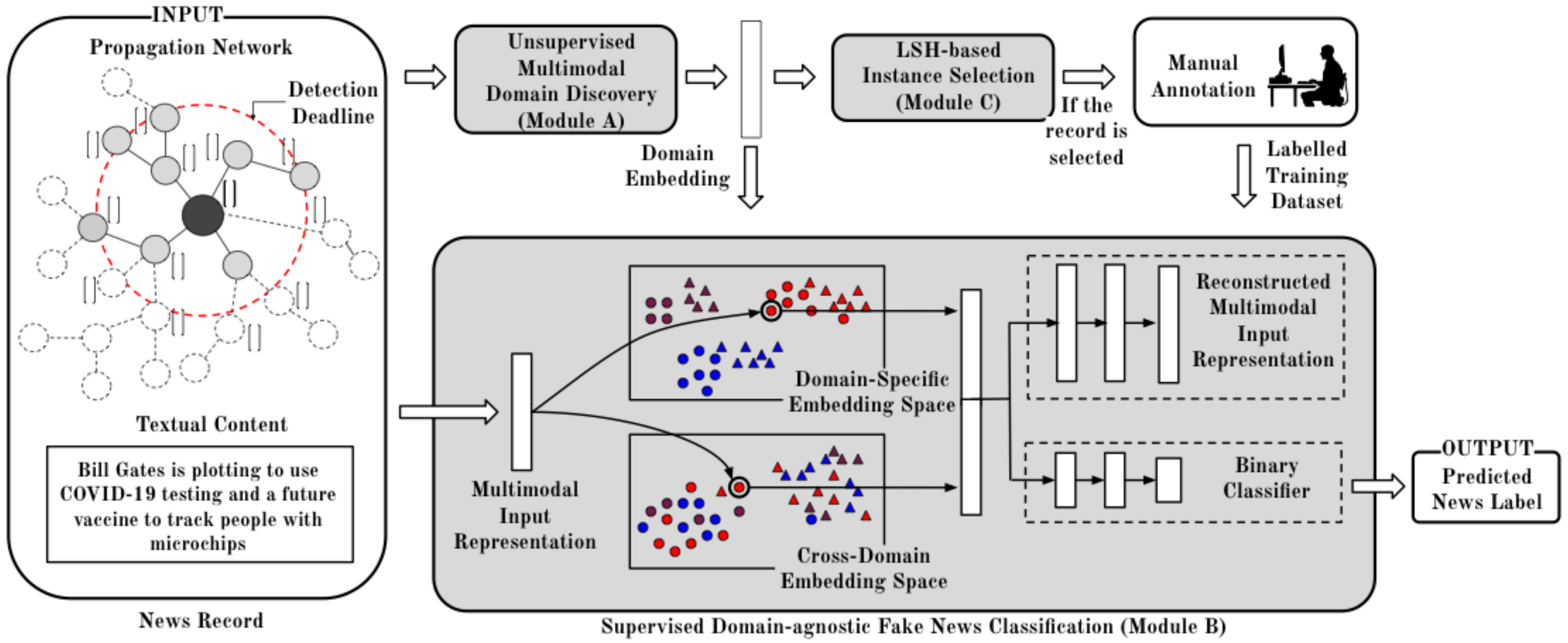
- Select a set of instances $R^L$ from $R$ to label while give labeling budget $B$ (constrains the # of instances in $R^L$).
  - Assign a binary label $y^r$ (1: fake 0: real)
- Learn an effective model using $R^L$ to predict the label $y^r$ for unlabeled news $r \in R^U$ as false or real news.
  - $R(R^L \cup R^U)$ is not constrained to a specific domain.
  - To emulate such a domain-agnostic dataset, combine PolitiFact, GossipCop, CoAID.

# Methodology

## Embracing Domain Differences in Fake News (EDDFN*)



# Methodology
## EDDFN




- EDDFN includes 2 main components for FND.
  - Unsupervised embedding learning
  - Supervised domain-agnostic news classification
  - These components are integrated to identify fake news while exploiting domain-specific and cross-domain knowledge in news.
- In addition, the proposed instance selection approach
  - Adopts the same domain embedding learning component to select informative news for labeling, which eventually yields a labeled dataset that maximizes the domain-coverage.

# Methodology

## Unsupervised Domain Discovery


Word cloud for the top 20 words in two dataset

- For a give news $r$, assume that its domain label is not available.
- Proposed unsupervised domain embedding learning technique exploits multimodal content of $r$ to represent the domain of $r$ as low-dimensional vector $f_{domain}(r)$.
- Approach is motivated by:
  - Tendency of users to form groups containing people with similar interests (homophily), which results in different domain have distinct user bases.
  - Significant differences in domain-specific word usage as shown in previous figure.

# Methodology

## Network Construction

- Create set for each news by adding all users and all words appearing in the news title.
- For each pair of items, build a weighted edge linking the two items in the graph.
- Repeat above steps for all news to obtain the final network.

**Algorithm 1:** Domain Embedding Learning

**Input:** A collection of news records $R$
**Output:** Domain embeddings $f_{domain}(r)$ of $r \in R$

`// Network construction`
Initialize an empty graph $G$;
**for** $r \in R$ **do**
  $S^r \leftarrow X^r \cup U^r$
  **for** *each pair* $(s_1, s_2) \in S$ **do**
    $e \leftarrow (\{s_1, s_2\}, 1)$;
    **if** *edge* $e$ *exists in graph* $G$ **then**
      Increment edge $e$ in graph $G$ by 1;
    **else**
      Add edge $e$ to graph $G$;
`// Community Detection`
$C \leftarrow$ Find communities in $G$ using Louvain;
`// Embedding Learning`
**for** $r \in R$ **do**
  Compute $f_{domain}(r)$ using Eq. 2
Return $f_{domain}(r)$ of $r \in R$.

# Methodology

## Community Detection

- Adopt Louvain algorithm* to identify communities in graph.
    - Best performing parameter-free community detection algorithm.
- Obtain a set of communities/clusters, each having either a highly connected set of users or words.
- Assume each community belongs to a single domain.

**Algorithm 1:** Domain Embedding Learning

**Input:** A collection of news records $R$
**Output:** Domain embeddings $f_{domain}(r)$ of $r \in R$
// Network construction
Initialize an empty graph $G$;
**for** $r \in R$ **do**
  $S^r \leftarrow X^r \cup U^r$
  **for** *each pair* $(s_1, s_2) \in S$ **do**
    $e \leftarrow (\{s_1, s_2\}, 1)$;
    **if** *edge* $e$ *exists in graph* $G$ **then**
      Increment edge $e$ in graph $G$ by 1;
    **else**
      Add edge $e$ to graph $G$;
// Community Detection
$C \leftarrow$ Find communities in $G$ using Louvain;
// Embedding Learning
**for** $r \in R$ **do**
  Compute $f_{domain}(r)$ using Eq. 2
Return $f_{domain}(r)$ of $r \in R$.

# Methodology

## Embedding Learning

- Compute the soft membership $p(r \in c)$ of $r$ in a cluster $c$:
  - $p(r \in c) = \sum_{v \in c \cap r} v_{deg} / \sum_{c \in C} \sum_{v \in r} v_{deg}$
- $p(r \in c)$ is proportional to # of common users or words that $r$ and $c$ have.
- Each node $v$ is weighted using the degree $v_{deg}$ in G to reflect their caring importance for the corresponding community.

**Algorithm 1:** Domain Embedding Learning

**Input:** A collection of news records $R$
**Output:** Domain embeddings $f_{domain}(r)$ of $r \in R$

```
// Network construction
Initialize an empty graph G;
for r ∈ R do
    S^r ← X^r ∪ U^r
    for each pair (s_1, s_2) ∈ S do
        e ← ({s_1, s_2}, 1);
        if edge e exists in graph G then
            Increment edge e in graph G by 1;
        else
            Add edge e to graph G;
// Community Detection
C ← Find communities in G using Louvain;
// Embedding Learning
for r ∈ R do
    Compute f_domain(r) using Eq. 2
Return f_domain(r) of r ∈ R.
```

# Methodology

## Embedding Learning

- Finally, produce the domain embedding $f_{domain}(r) \in \mathbb{R}^{|C|}$ of $r$ as concatenation of $r$'s likelihood belonging to communities in $C$:

- $f_{domain}(r) = p(r \in c_1) \oplus p(r \in c_2) \oplus \ldots p(r \in c_{|C|})$

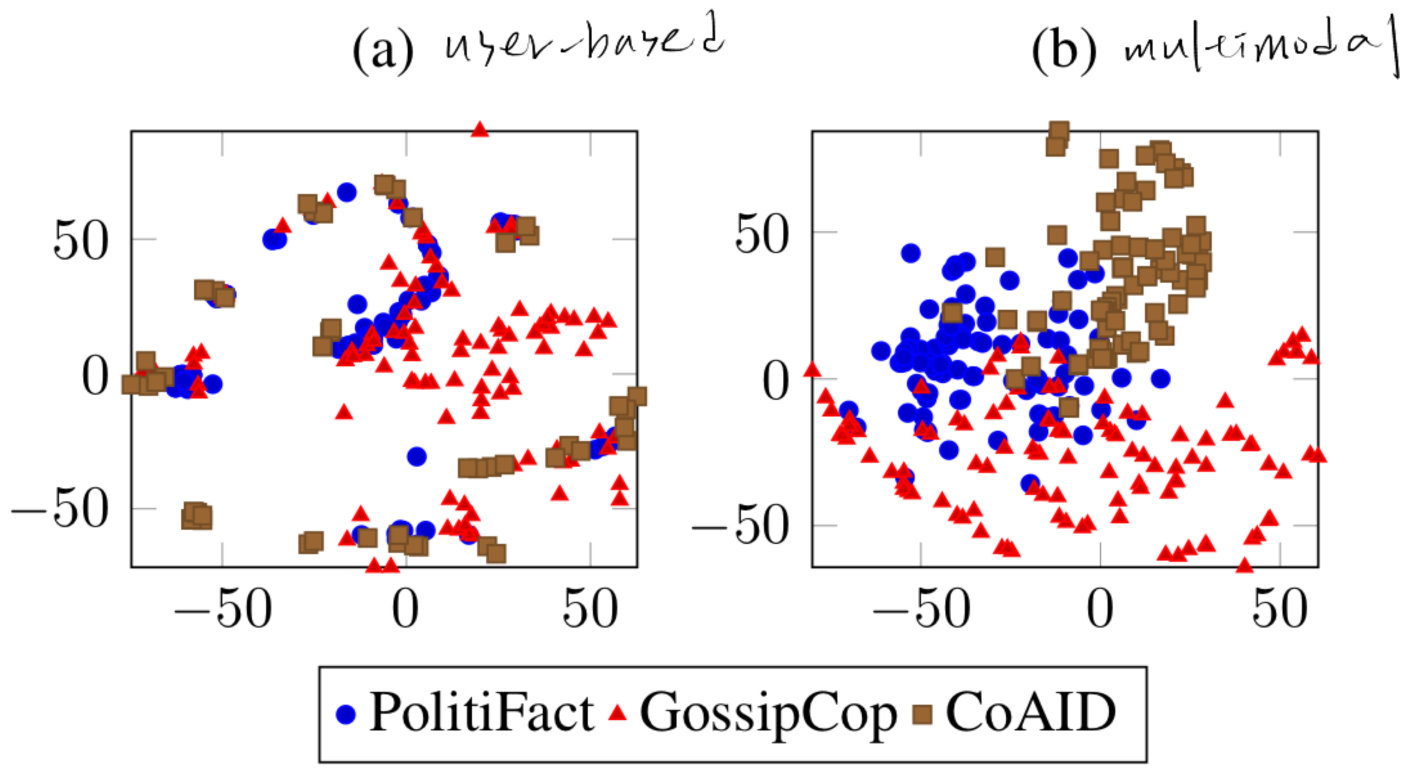
**Algorithm 1:** Domain Embedding Learning

**Input:** A collection of news records $R$
**Output:** Domain embeddings $f_{domain}(r)$ of $r \in R$

```
// Network construction
```
Initialize an empty graph $G$;
**for** $r \in R$ **do**
  $S^r \leftarrow X^r \cup U^r$
  **for** *each pair* $(s_1, s_2) \in S$ **do**
    $e \leftarrow (\{s_1, s_2\}, 1)$;
    **if** *edge* $e$ *exists in graph* $G$ **then**
      Increment edge $e$ in graph $G$ by 1;
    **else**
      Add edge $e$ to graph $G$;
```
// Community Detection
```
$C \leftarrow$ Find communities in $G$ using Louvain;
```
// Embedding Learning
```
**for** $r \in R$ **do**
  Compute $f_{domain}(r)$ using Eq. 2
Return $f_{domain}(r)$ of $r \in R$.

# Methodology

## Comparison of domain embedding




t-SNE result (user-based/multimodal)

- Proposed approach (fig.b) yields a clear separation between the domains compared to user-based domain discovery algorithm (fig.a).
- May mainly due to ability of proposed approach to jointly exploit multi-modalities, both users and text of news to discover their domain.
- Most previous works on domain discovery ultimately assign hard domain labels for news, but some news may belong to multiple domains.
  - Hence, proposed method use low-dimensional vector to represent domain embedding can preserved such knowledge related to the domains of news.

# Methodology

## Domain-agnostic News Classification

- In classification model, each news $r$ is represented as a vector $f_{input}(r)$ using textual content $W^r$ and propagation network $G^r$ of $r$.
- Then, model maps $f_{input}(r)$ into 2 subspace:
    - Preserves domain-specific knowledge: $f_{specific} : f_{input}(r) \to \mathbb{R}^d$
    - Preserves cross-domain knowledge: $f_{shared} : f_{input}(r) \to \mathbb{R}^d$

# Methodology

## Domain-agnostic News Classification

- Then, $f_{specific}(r) \oplus f_{shared}(r)$ is used to recover the label $y^r$ and input representation $f_{input}(r)$ of $r$ during training via two decoder function $g_{pred}$ and $g_{recon}$ respectively.
- BCE: binary cross-entropy loss function
- Minimize $L_{pred}$ & $L_{recon}$ to find optimal parameters of $(f_{specific}, f_{shared}, g_{pred}, g_{recon})$.

- $\overline{y^r} = g_{pred}(f_{specific}(r) \oplus f_{shared}(r))$
- $\overline{f_{input}(r)} = g_{recon}(f_{specific}(r) \oplus f_{shared}(r))$
- $L_{pred} = BCE(y^r, \overline{y^r})$
- $L_{recon} = \|f_{input}(r) - \overline{f_{input}(r)}\|^2$

# Methodology

## Domain-agnostic News Classification

- However, $L_{pred}$ & $L_{recon}$ do not leverage domain difference in news.
- Hence, now discuss how the mapping function for subspaces, $f_{specific}$ & $f_{shared}$.
    - Further learned to preserve the domain-specific and cross-domain knowledge in news.

# Methodology

## Leveraging Domain-specific Knowledge

- To preserve domain-specific knowledge, introduce an auxiliary loss term $L_{specific}$ to learn a news decoder function $g_{specific}$ to recover the domain embedding $f_{domain}(r)$ using the domain-specific representation $f_{specific}(r)$.

- Minimize $L_{specific}$ to find optimal parameters for $(f_{specific}, g_{specific})$ to capture the domain-specific knowledge by $f_{specific}$.

  - $L_{specific} = \| f_{domain}(r) - g_{specific}(f_{specific}(r)) \|^2$

  - $(\hat{g}_{specific}, \hat{f}_{specific}) = \underset{(g_{specific}, f_{specific})}{\operatorname{argmin}} (L_{specific})$

# Methodology

## Leveraging Cross-domain Knowledge

- In contrast, learn $f_{shared}$ to overlook domain-specific knowledge of news.
- Train a decoder function $g_{shared}$ to accurately predict domain of $r$ using $f_{shared}(r)$.
- Meanwhile, learn $f_{shared}$ to fool decoder $g_{shared}$ by maximizing loss if $g_{shared}$.
  - Such a formulation forces $f_{shared}$ to only rely in cross-domain knowledge.
  - Useful to transfer the knowledge across domains.

# Methodology

## Leveraging Cross-domain Knowledge

- Can be defined as a minimax game between $g_{shared}$ & $f_{shared}$.
  - $L_{shared} = \|g_{shared}(f_{shared}(r)) - f_{domain}(r)\|^2$
  - $(\hat{g}_{shared}, \hat{f}_{shared}) = \underset{f_{shared}}{\operatorname{argmin}} \, \underset{g_{shared}}{\operatorname{argmax}}(-L_{shared})$

# Methodology

## Integrated model

- Then final loss function of model is formulated as:
  - $L_{final} = L_{pred} + \lambda_1 L_{recon} + \lambda_2 L_{specific} - \lambda_3 L_{shared}$
  - where $\lambda$ controls the importance given to each loss term compared to $L_{pred}$ (main).
- To learn minimax game in $L_{shared}$, the final loss function $L_{final}$:
  - $(\widehat{\theta_1}) = \underset{\theta_1}{\operatorname{argmin}} L_{final}(\theta_1, \theta_2)$, $\quad \theta_1 : (f_{specific}, f_{shared}, g_{specific}, g_{pred}, g_{recon})$
  - $(\widehat{\theta_2}) = \underset{\theta_2}{\operatorname{argmax}} L_{final}(\widehat{\theta_1}, \theta_2) \quad \theta_2 : g_{shared}$

# Methodology

## LHS-based Instance Selection

- Aforementioned model is able to exploit the domain-specific and cross-domain knowledge in news to identify their veracity.
- Empirically observe that the performance of the model substantially drops when identify unseen or rarely appearing domain news during training.
- Proposed an unsupervised technique to come up with a labeled training dataset for a given labeling budget $B$ such that it covers as many domains as possible.
- Ultimate objective of this technique is to learn a model using such a dataset that performs well for many domains.

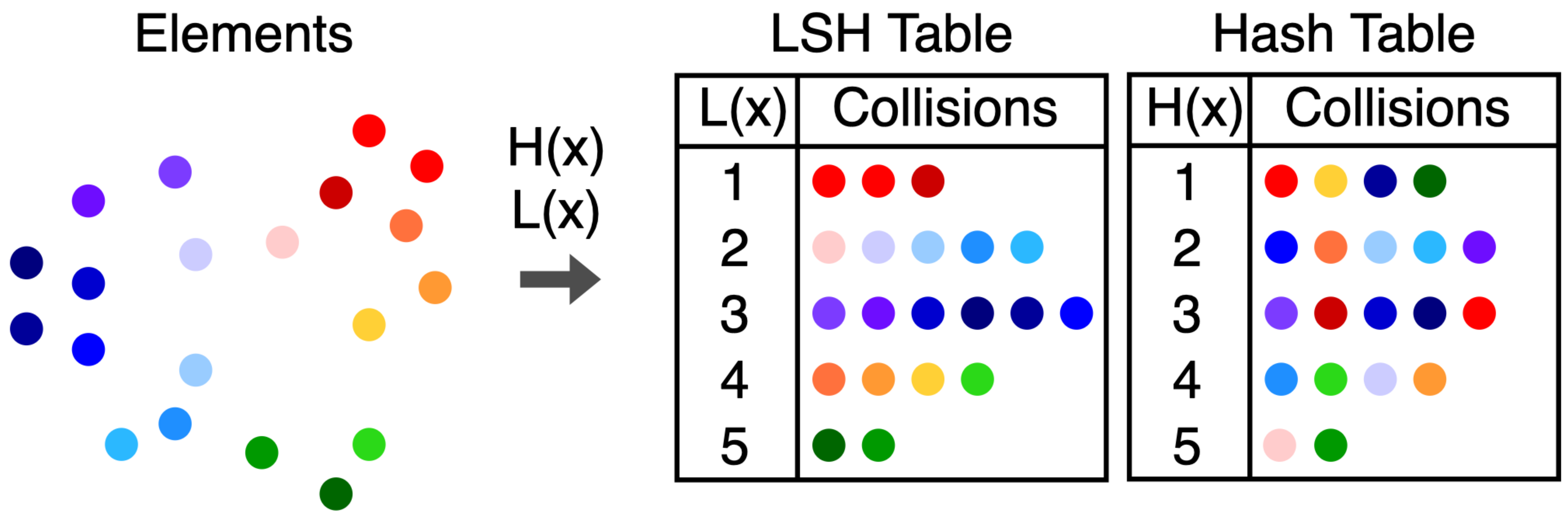
# Methodology

## LHS-based Instance Selection

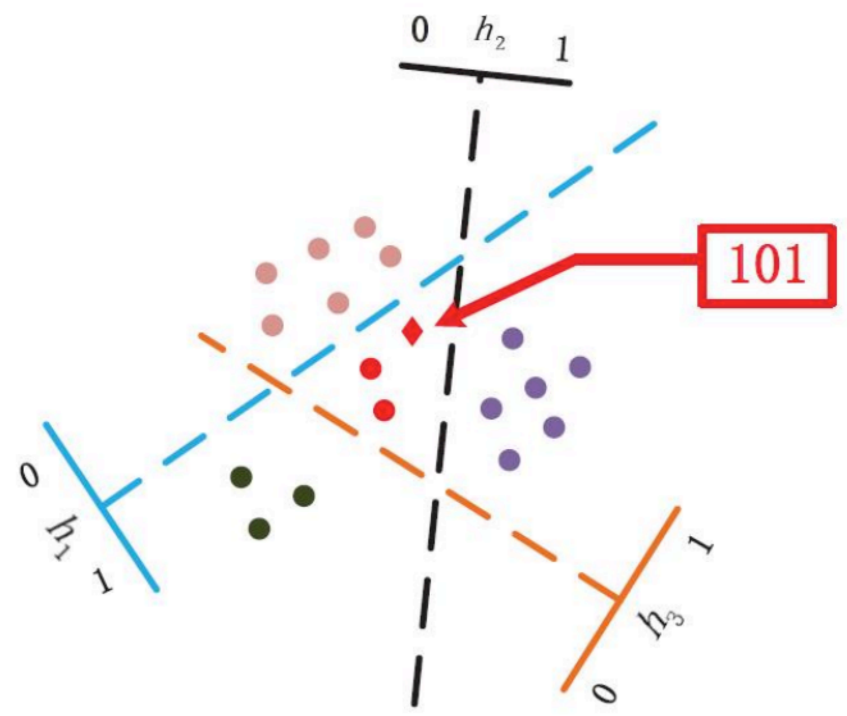
- Initially represents each news $r \in R$ using its domain embedding $f_{domain}(r)$.
- Proposed* a Locality-Sensitive Hashing (LSH) algorithm based on random projection to select a set of news in $R$ that are distant in the domain embedding space.

# Methodology

## LSH algorithm




credit: https://randorithms.com/2019/09/19/Visual-LSH.html




credit: http://nnw.cz/doi/2017/NNW.2017.27.005.pdf

- LSH is an algorithmic technique that hashes similar input items into the same "buckets" with high probability.

- Since similar items end up in the same buckets, this technique can be used for data clustering and nearest neighbor search.

- It differs from conventional hashing techniques in that hash collisions are maximized, not minimized.

- Alternatively, the technique can be seen as a way to reduce the dimensionality of high-dimensional data; high-dimensional input items can be reduced to low-dimensional versions while preserving relative distances between items.

# Methodology

## Step of Instance Selection

$$h_{i,j} = \sqrt{3} \times \begin{cases} +1 & \text{with probability } 1/6 \\ 0 & \text{with probability } 2/3 \\ -1 & \text{with probability } 1/6 \end{cases}$$

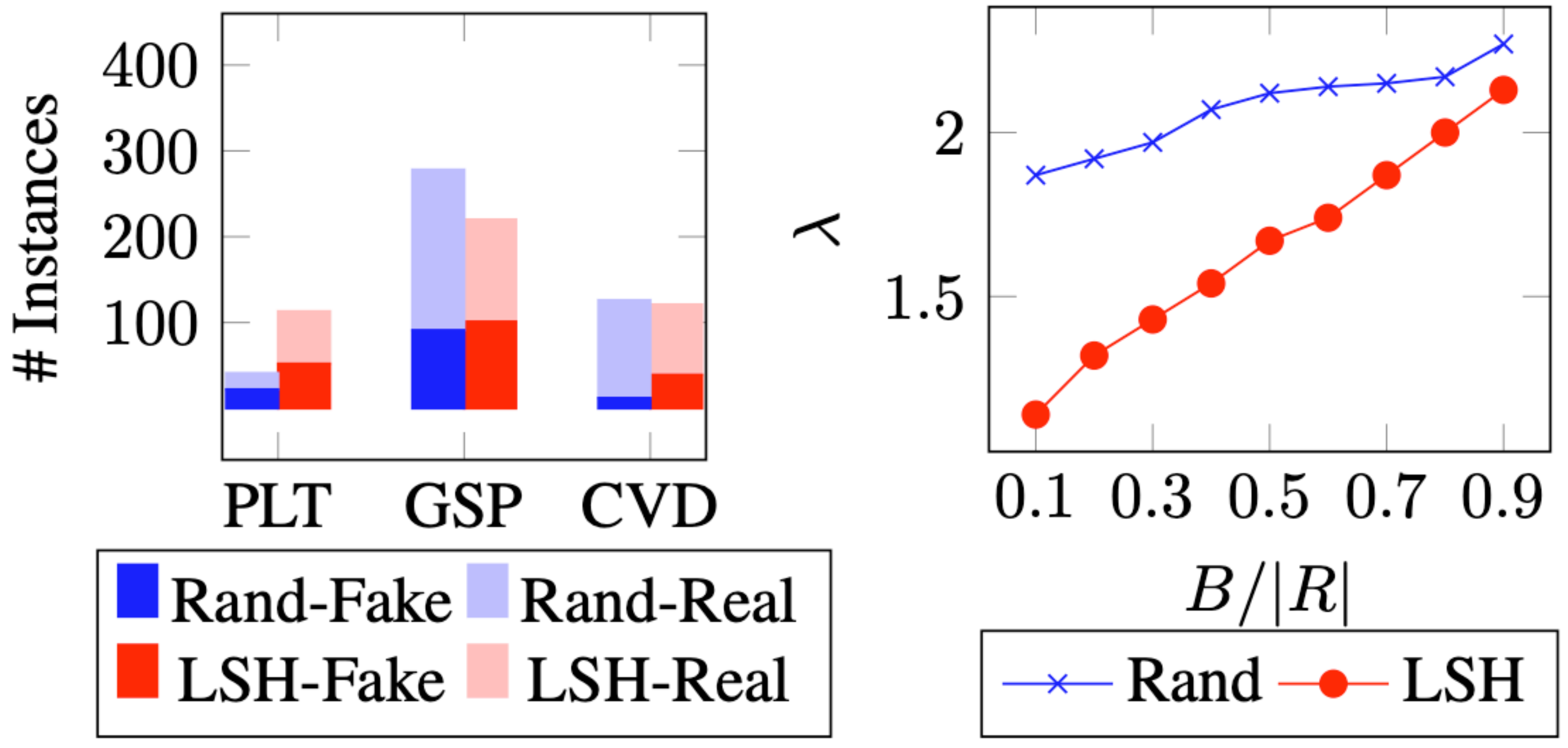
probability distribution

- Create $|H|$ different hash functions such as $H_i(r) = sgn(h_i \cdot f_{domain}(r))$, $i \in [0, |H| - 1]$ .
- Construct an H-dimensional hash vector for each news as $H_0(r) \oplus \ldots \oplus H_{|H|-1}(r)$.
- Group the news with similar hash values to construct a hash table.
- Randomly pick a news from each bin in the hash table and add to the selected dataset pool.
- Repeat above steps until the size of the dataset pool reaches the labelling budget $B$.

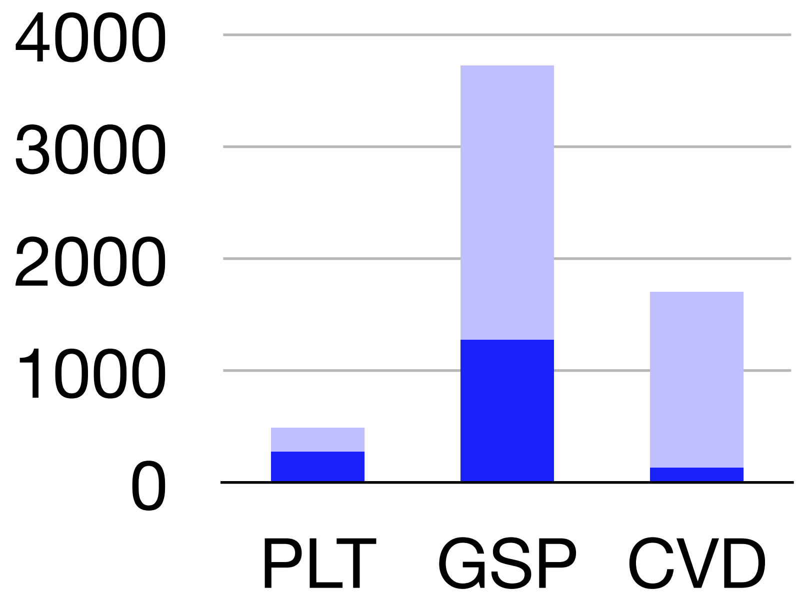
# Methodology

## Instance Selection Performance

- As can be seen, random selection follows empirical distribution of dataset and pick few instances from rarely domain.
- In contrast, the proposed approach provides a significant # of samples from even rarely occurring domains.
- Also mentioned complexity is efficient…




(a) (b)

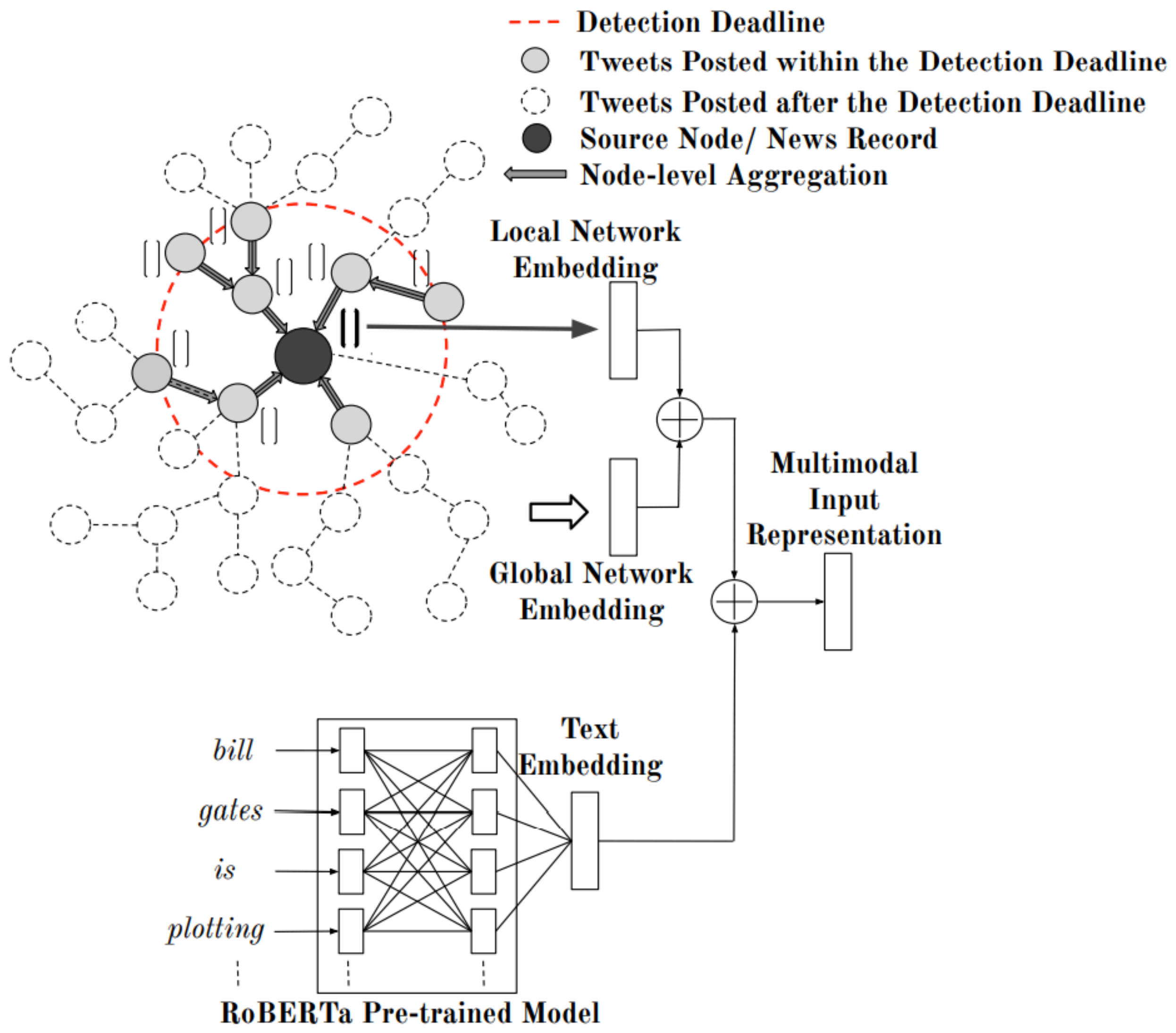



| Dataset | PolitiFact | GossipCop | CoAID |
|---|---|---|---|
| # Fake News | 269 | 1269 | 135 |
| # Real News | 230 | 2466 | 1568 |

Statics of 3 datasets

# Experiments

## Encoding & Decoding Functions




- Adopt RoBERTa-base to learn text-based representation $f_{text}(r)$.

- Propagation network-based representation $f_{network}(r)$ is using unsupervised network representation learning technique proposed by author.

- Then final input representation $f_{input}(r)$ is constructed as $f_{text}(r) \oplus f_{network}(r)$.

- All other encoding/decoding functions ($f_{specific}, f_{shared}, g_{specific}, g_{shared}, g_{pred}, g_{recon}$) are modeled as 2-layer feed-forward network with sigmoid activation.

# Experiments

## Datasets

- Combine PolitiFact, GossipCop, CoAID to produce a cross-domain news dataset.
- 75% candidate data pool for training : 25% for testing
- For given a budget $B$, select $B$ instances from pool to train the model.

# Experiments
## Baselines

- LIWC: learns feature vectors from the text content of news by counting the #of lexicons falling into different psycho-linguistic categories.
- HAN: adopts a hierarchical attention neural network framework to model the text content of news.
- EANN: use EANN-Unimodal (text) & EANN-multimodal (text, network).
- HPNF: extracts various features from the propagation network of news to generate its feature representation.
- AE: adopts an Auto-encoder architecture to learn latent representation for each news based on its propagation network.
- SAFE: propose modality-similarity method by caption news image compare with news text content.

# Experiments

## Performance Comparison

| Method | Type | | | Politifact | | | | Gossipcop | | | | CoAID | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | T | S | M | Acc | Prec | Rec | F1 | Acc | Prec | Rec | F1 | Acc | Prec | Rec | F1 |
| LIWC (Pennebaker et al. 2015) | ✓ | | | 0.488 | 0.680 | 0.565 | 0.432 | 0.662 | 0.550 | 0.516 | 0.472 | 0.903 | 0.586 | 0.531 | 0.538 |
| text-CNN (Kim 2014) | ✓ | | | 0.608 | 0.621 | 0.623 | 0.608 | 0.733 | 0.698 | 0.703 | 0.701 | 0.903 | 0.679 | 0.674 | 0.677 |
| HAN (Yang et al. 2016) | ✓ | | | 0.632 | 0.672 | 0.651 | 0.648 | 0.716 | 0.703 | 0.709 | 0.706 | 0.919 | 0.698 | 0.682 | 0.688 |
| EANN-Unimodal (Wang et al. 2018) | ✓ | | | 0.794 | 0.811 | 0.790 | 0.791 | 0.765 | 0.732 | 0.738 | 0.734 | 0.925 | 0.842 | 0.763 | 0.792 |
| HPNF (Shu et al. 2020b) | | ✓ | | 0.697 | 0.692 | 0.683 | 0.687 | 0.721 | 0.703 | 0.689 | 0.695 | 0.902 | 0.652 | 0.693 | 0.672 |
| AE (Silva et al. 2020) | | ✓ | | 0.784 | 0.783 | 0.774 | 0.779 | 0.834 | 0.828 | 0.802 | 0.812 | 0.928 | 0.686 | 0.673 | 0.677 |
| HPNF + LIWC (Shu et al. 2020b) | | | ✓ | 0.704 | 0.723 | 0.708 | 0.716 | 0.734 | 0.715 | 0.706 | 0.708 | 0.911 | 0.682 | 0.709 | 0.690 |
| SAFE (Zhou et al. 2020) | | | ✓ | 0.793 | 0.782 | 0.771 | 0.775 | 0.831 | 0.822 | 0.798 | 0.806 | 0.931 | 0.754 | 0.744 | 0.748 |
| EANN-Multimodal (Wang et al. 2018) | | | ✓ | 0.804 | 0.808 | 0.794 | 0.798 | 0.836 | 0.812 | 0.815 | 0.813 | 0.944 | 0.849 | 0.803 | 0.808 |
| Our Approach ($B = 100\%\|R_{pool}\|$) | | | ✓ | **0.840** | **0.836** | **0.831** | **0.835** | **0.877** | **0.840** | **0.832** | **0.836** | **0.970** | **0.876** | **0.863** | **0.869** |
| Our Approach ($B = 50\%\|R_{pool}\|$) | | | ✓ | **0.838** | **0.836** | **0.828** | **0.833** | **0.848** | **0.822** | **0.797** | **0.808** | **0.963** | **0.870** | **0.854** | **0.862** |
| **Ablation Study** ($B = 100\%\|R_{pool}\|$) | | | | | | | | | | | | | | | |
| (-) *Domain-shared loss* | | | | 0.823 | 0.821 | 0.812 | 0.815 | 0.864 | 0.832 | 0.828 | 0.829 | 0.956 | 0.857 | 0.861 | 0.858 |
| (-) *Domain-specific loss* | | | | 0.792 | 0.800 | 0.783 | 0.786 | 0.858 | 0.832 | 0.821 | 0.828 | 0.934 | 0.850 | 0.857 | 0.853 |
| (-) *Network modality* | | | | 0.816 | 0.815 | 0.817 | 0.815 | 0.765 | 0.749 | 0.745 | 0.746 | 0.945 | 0.803 | 0.855 | 0.827 |
| (-) *Text modality* | | | | 0.804 | 0.798 | 0.793 | 0.795 | 0.837 | 0.835 | 0.815 | 0.817 | 0.932 | 0.711 | 0.704 | 0.707 |

# Experiments

## Performance Comparison

| Method | Type | | | Politifact | | | | Gossipcop | | | | CoAID | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | T | S | M | Acc | Prec | Rec | F1 | Acc | Prec | Rec | F1 | Acc | Prec | Rec | F1 |
| LIWC (Pennebaker et al. 2015) | ✓ | | | 0.488 | 0.680 | 0.565 | 0.432 | 0.662 | 0.550 | 0.516 | 0.472 | 0.903 | 0.586 | 0.531 | 0.538 |
| text-CNN (Kim 2014) | ✓ | | | 0.608 | 0.621 | 0.623 | 0.608 | 0.733 | 0.698 | 0.703 | 0.701 | 0.903 | 0.679 | 0.674 | 0.677 |
| HAN (Yang et al. 2016) | ✓ | | | 0.632 | 0.672 | 0.651 | 0.648 | 0.716 | 0.703 | 0.709 | 0.706 | 0.919 | 0.698 | 0.682 | 0.688 |
| EANN-Unimodal (Wang et al. 2018) | ✓ | | | 0.794 | 0.811 | 0.790 | 0.791 | 0.765 | 0.732 | 0.738 | 0.734 | 0.925 | 0.842 | 0.763 | 0.792 |
| HPNF (Shu et al. 2020b) | | ✓ | | 0.697 | 0.692 | 0.683 | 0.687 | 0.721 | 0.703 | 0.689 | 0.695 | 0.902 | 0.652 | 0.693 | 0.672 |
| AE (Silva et al. 2020) | | ✓ | | 0.784 | 0.783 | 0.774 | 0.779 | 0.834 | 0.828 | 0.802 | 0.812 | 0.928 | 0.686 | 0.673 | 0.677 |
| HPNF + LIWC (Shu et al. 2020b) | | | ✓ | 0.704 | 0.723 | 0.708 | 0.716 | 0.734 | 0.715 | 0.706 | 0.708 | 0.911 | 0.682 | 0.709 | 0.690 |
| SAFE (Zhou et al. 2020) | | | ✓ | 0.793 | 0.782 | 0.771 | 0.775 | 0.831 | 0.822 | 0.798 | 0.806 | 0.931 | 0.754 | 0.744 | 0.748 |
| EANN-Multimodal (Wang et al. 2018) | | | ✓ | 0.804 | 0.808 | 0.794 | 0.798 | 0.836 | 0.812 | 0.815 | 0.813 | 0.944 | 0.849 | 0.803 | 0.808 |
| Our Approach ($B = 100\%\lvert R_{pool}\rvert$) | | | ✓ | **0.840** | **0.836** | **0.831** | **0.835** | **0.877** | **0.840** | **0.832** | **0.836** | **0.970** | **0.876** | **0.863** | **0.869** |
| Our Approach ($B = 50\%\lvert R_{pool}\rvert$) | | | ✓ | **0.838** | **0.836** | **0.828** | **0.833** | **0.848** | **0.822** | **0.797** | **0.808** | **0.963** | **0.870** | **0.854** | **0.862** |
| **Ablation Study** ($B = 100\%\lvert R_{pool}\rvert$) | | | | | | | | | | | | | | | |
| (-) *Domain-shared loss* | | | | 0.823 | 0.821 | 0.812 | 0.815 | 0.864 | 0.832 | 0.828 | 0.829 | 0.956 | 0.857 | 0.861 | 0.858 |
| (-) *Domain-specific loss* | | | | 0.792 | 0.800 | 0.783 | 0.786 | 0.858 | 0.832 | 0.821 | 0.828 | 0.934 | 0.850 | 0.857 | 0.853 |
| (-) *Network modality* | | | | 0.816 | 0.815 | 0.817 | 0.815 | 0.765 | 0.749 | 0.745 | 0.746 | 0.945 | 0.803 | 0.855 | 0.827 |
| (-) *Text modality* | | | | 0.804 | 0.798 | 0.793 | 0.795 | 0.837 | 0.835 | 0.815 | 0.817 | 0.932 | 0.711 | 0.704 | 0.707 |

# Experiments

## Performance Comparison

| Method | Type | | | Politifact | | | | Gossipcop | | | | CoAID | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | T | S | M | Acc | Prec | Rec | F1 | Acc | Prec | Rec | F1 | Acc | Prec | Rec | F1 |
| LIWC (Pennebaker et al. 2015) | ✓ | | | 0.488 | 0.680 | 0.565 | 0.432 | 0.662 | 0.550 | 0.516 | 0.472 | 0.903 | 0.586 | 0.531 | 0.538 |
| text-CNN (Kim 2014) | ✓ | | | 0.608 | 0.621 | 0.623 | 0.608 | 0.733 | 0.698 | 0.703 | 0.701 | 0.903 | 0.679 | 0.674 | 0.677 |
| HAN (Yang et al. 2016) | ✓ | | | 0.632 | 0.672 | 0.651 | 0.648 | 0.716 | 0.703 | 0.709 | 0.706 | 0.919 | 0.698 | 0.682 | 0.688 |
| EANN-Unimodal (Wang et al. 2018) | ✓ | | | 0.794 | 0.811 | 0.790 | 0.791 | 0.765 | 0.732 | 0.738 | 0.734 | 0.925 | 0.842 | 0.763 | 0.792 |
| HPNF (Shu et al. 2020b) | | ✓ | | 0.697 | 0.692 | 0.683 | 0.687 | 0.721 | 0.703 | 0.689 | 0.695 | 0.902 | 0.652 | 0.693 | 0.672 |
| AE (Silva et al. 2020) | | ✓ | | 0.784 | 0.783 | 0.774 | 0.779 | 0.834 | 0.828 | 0.802 | 0.812 | 0.928 | 0.686 | 0.673 | 0.677 |
| HPNF + LIWC (Shu et al. 2020b) | | | ✓ | 0.704 | 0.723 | 0.708 | 0.716 | 0.734 | 0.715 | 0.706 | 0.708 | 0.911 | 0.682 | 0.709 | 0.690 |
| SAFE (Zhou et al. 2020) | | | ✓ | 0.793 | 0.782 | 0.771 | 0.775 | 0.831 | 0.822 | 0.798 | 0.806 | 0.931 | 0.754 | 0.744 | 0.748 |
| EANN-Multimodal (Wang et al. 2018) | | | ✓ | 0.804 | 0.808 | 0.794 | 0.798 | 0.836 | 0.812 | 0.815 | 0.813 | 0.944 | 0.849 | 0.803 | 0.808 |
| Our Approach ($B = 100\%\lvert R_{pool}\rvert$) | | | ✓ | **0.840** | **0.836** | **0.831** | **0.835** | **0.877** | **0.840** | **0.832** | **0.836** | **0.970** | **0.876** | **0.863** | **0.869** |
| Our Approach ($B = 50\%\lvert R_{pool}\rvert$) | | | ✓ | **0.838** | **0.836** | **0.828** | **0.833** | **0.848** | **0.822** | **0.797** | **0.808** | **0.963** | **0.870** | **0.854** | **0.862** |
| **Ablation Study** ($B = 100\%\lvert R_{pool}\rvert$) | | | | | | | | | | | | | | | |
| (-) *Domain-shared loss* | | | | 0.823 | 0.821 | 0.812 | 0.815 | 0.864 | 0.832 | 0.828 | 0.829 | 0.956 | 0.857 | 0.861 | 0.858 |
| (-) *Domain-specific loss* | | | | 0.792 | 0.800 | 0.783 | 0.786 | 0.858 | 0.832 | 0.821 | 0.828 | 0.934 | 0.850 | 0.857 | 0.853 |
| (-) *Network modality* | | | | 0.816 | 0.815 | 0.817 | 0.815 | 0.765 | 0.749 | 0.745 | 0.746 | 0.945 | 0.803 | 0.855 | 0.827 |
| (-) *Text modality* | | | | 0.804 | 0.798 | 0.793 | 0.795 | 0.837 | 0.835 | 0.815 | 0.817 | 0.932 | 0.711 | 0.704 | 0.707 |

# Experiments
## Ablation Study (1/2)

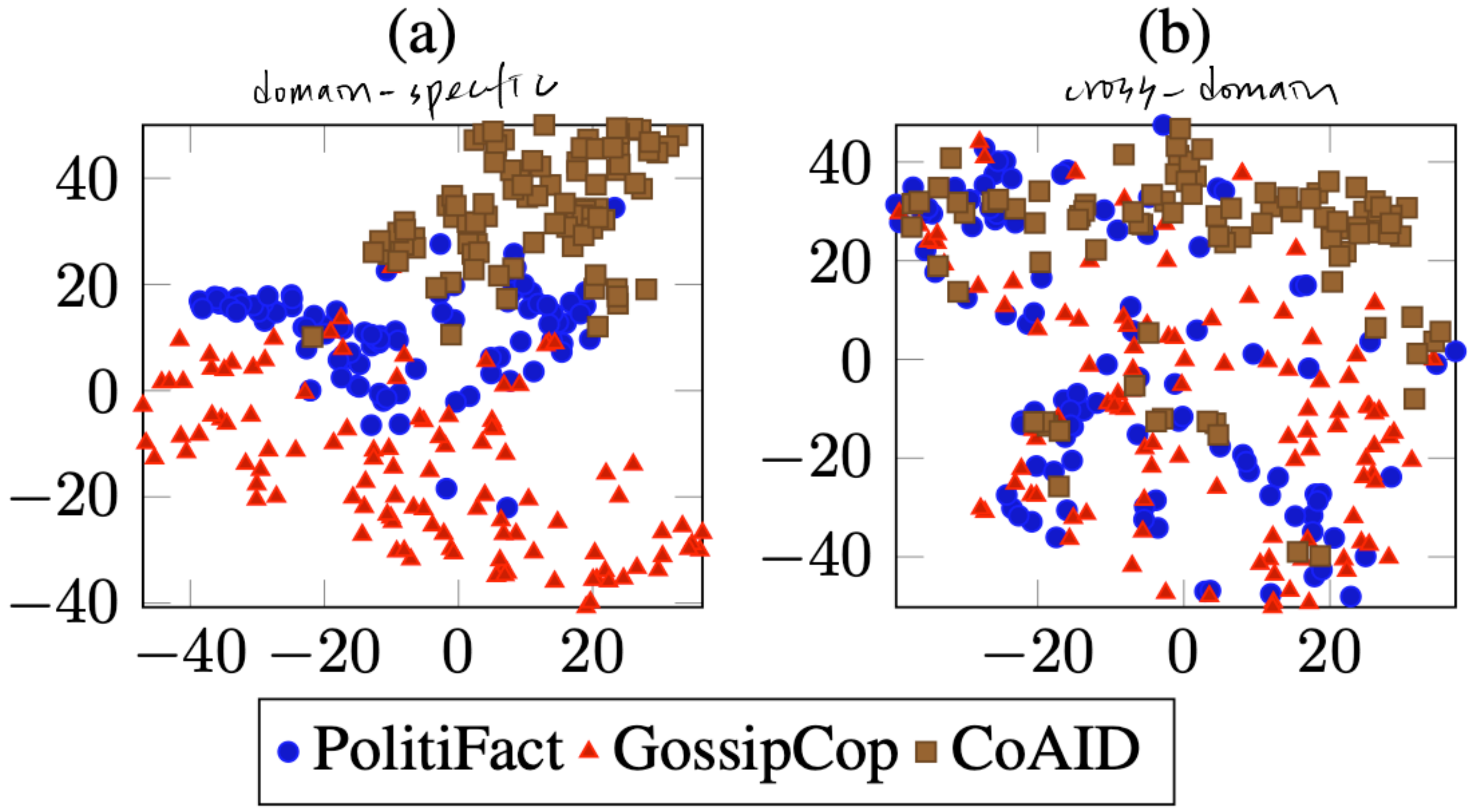
| Method | Type | | | Politifact | | | | Gossipcop | | | | CoAID | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | T | S | M | Acc | Prec | Rec | F1 | Acc | Prec | Rec | F1 | Acc | Prec | Rec | F1 |
| LIWC (Pennebaker et al. 2015) | ✓ | | | 0.488 | 0.680 | 0.565 | 0.432 | 0.662 | 0.550 | 0.516 | 0.472 | 0.903 | 0.586 | 0.531 | 0.538 |
| text-CNN (Kim 2014) | ✓ | | | 0.608 | 0.621 | 0.623 | 0.608 | 0.733 | 0.698 | 0.703 | 0.701 | 0.903 | 0.679 | 0.674 | 0.677 |
| HAN (Yang et al. 2016) | ✓ | | | 0.632 | 0.672 | 0.651 | 0.648 | 0.716 | 0.703 | 0.709 | 0.706 | 0.919 | 0.698 | 0.682 | 0.688 |
| EANN-Unimodal (Wang et al. 2018) | ✓ | | | 0.794 | 0.811 | 0.790 | 0.791 | 0.765 | 0.732 | 0.738 | 0.734 | 0.925 | 0.842 | 0.763 | 0.792 |
| HPNF (Shu et al. 2020b) | | ✓ | | 0.697 | 0.692 | 0.683 | 0.687 | 0.721 | 0.703 | 0.689 | 0.695 | 0.902 | 0.652 | 0.693 | 0.672 |
| AE (Silva et al. 2020) | | ✓ | | 0.784 | 0.783 | 0.774 | 0.779 | 0.834 | 0.828 | 0.802 | 0.812 | 0.928 | 0.686 | 0.673 | 0.677 |
| HPNF + LIWC (Shu et al. 2020b) | | | ✓ | 0.704 | 0.723 | 0.708 | 0.716 | 0.734 | 0.715 | 0.706 | 0.708 | 0.911 | 0.682 | 0.709 | 0.690 |
| SAFE (Zhou et al. 2020) | | | ✓ | 0.793 | 0.782 | 0.771 | 0.775 | 0.831 | 0.822 | 0.798 | 0.806 | 0.931 | 0.754 | 0.744 | 0.748 |
| EANN-Multimodal (Wang et al. 2018) | | | ✓ | 0.804 | 0.808 | 0.794 | 0.798 | 0.836 | 0.812 | 0.815 | 0.813 | 0.944 | 0.849 | 0.803 | 0.808 |
| Our Approach ($B = 100\%\|R_{pool}\|$) | | | ✓ | **0.840** | **0.836** | **0.831** | **0.835** | **0.877** | **0.840** | **0.832** | **0.836** | **0.970** | **0.876** | **0.863** | **0.869** |
| Our Approach ($B = 50\%\|R_{pool}\|$) | | | ✓ | **0.838** | **0.836** | **0.828** | **0.833** | **0.848** | **0.822** | **0.797** | **0.808** | **0.963** | **0.870** | **0.854** | **0.862** |
| **Ablation Study** ($B = 100\%\|R_{pool}\|$) | | | | | | | | | | | | | | | |
| (-) *Domain-shared loss* | | | | 0.823 | 0.821 | 0.812 | 0.815 | 0.864 | 0.832 | 0.828 | 0.829 | 0.956 | 0.857 | 0.861 | 0.858 |
| (-) *Domain-specific loss* | | | | 0.792 | 0.800 | 0.783 | 0.786 | 0.858 | 0.832 | 0.821 | 0.828 | 0.934 | 0.850 | 0.857 | 0.853 |
| (-) *Network modality* | | | | 0.816 | 0.815 | 0.817 | 0.815 | 0.765 | 0.749 | 0.745 | 0.746 | 0.945 | 0.803 | 0.855 | 0.827 |
| (-) *Text modality* | | | | 0.804 | 0.798 | 0.793 | 0.795 | 0.837 | 0.835 | 0.815 | 0.817 | 0.932 | 0.711 | 0.704 | 0.707 |

# Experiments
## Ablation Study (1/2)

| Method | Type | | | Politifact | | | | Gossipcop | | | | CoAID | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | T | S | M | Acc | Prec | Rec | F1 | Acc | Prec | Rec | F1 | Acc | Prec | Rec | F1 |
| LIWC (Pennebaker et al. 2015) | ✓ | | | 0.488 | 0.680 | 0.565 | 0.432 | 0.662 | 0.550 | 0.516 | 0.472 | 0.903 | 0.586 | 0.531 | 0.538 |
| text-CNN (Kim 2014) | ✓ | | | 0.608 | 0.621 | 0.623 | 0.608 | 0.733 | 0.698 | 0.703 | 0.701 | 0.903 | 0.679 | 0.674 | 0.677 |
| HAN (Yang et al. 2016) | ✓ | | | 0.632 | 0.672 | 0.651 | 0.648 | 0.716 | 0.703 | 0.709 | 0.706 | 0.919 | 0.698 | 0.682 | 0.688 |
| EANN-Unimodal (Wang et al. 2018) | ✓ | | | 0.794 | 0.811 | 0.790 | 0.791 | 0.765 | 0.732 | 0.738 | 0.734 | 0.925 | 0.842 | 0.763 | 0.792 |
| HPNF (Shu et al. 2020b) | | ✓ | | 0.697 | 0.692 | 0.683 | 0.687 | 0.721 | 0.703 | 0.689 | 0.695 | 0.902 | 0.652 | 0.693 | 0.672 |
| AE (Silva et al. 2020) | | ✓ | | 0.784 | 0.783 | 0.774 | 0.779 | 0.834 | 0.828 | 0.802 | 0.812 | 0.928 | 0.686 | 0.673 | 0.677 |
| HPNF + LIWC (Shu et al. 2020b) | | | ✓ | 0.704 | 0.723 | 0.708 | 0.716 | 0.734 | 0.715 | 0.706 | 0.708 | 0.911 | 0.682 | 0.709 | 0.690 |
| SAFE (Zhou et al. 2020) | | | ✓ | 0.793 | 0.782 | 0.771 | 0.775 | 0.831 | 0.822 | 0.798 | 0.806 | 0.931 | 0.754 | 0.744 | 0.748 |
| EANN-Multimodal (Wang et al. 2018) | | | ✓ | 0.804 | 0.808 | 0.794 | 0.798 | 0.836 | 0.812 | 0.815 | 0.813 | 0.944 | 0.849 | 0.803 | 0.808 |
| Our Approach ($B = 100\%\lvert R_{pool}\rvert$) | | | ✓ | **0.840** | **0.836** | **0.831** | **0.835** | **0.877** | **0.840** | **0.832** | **0.836** | **0.970** | **0.876** | **0.863** | **0.869** |
| Our Approach ($B = 50\%\lvert R_{pool}\rvert$) | | | ✓ | **0.838** | **0.836** | **0.828** | **0.833** | **0.848** | **0.822** | **0.797** | **0.808** | **0.963** | **0.870** | **0.854** | **0.862** |
| **Ablation Study** ($B = 100\%\lvert R_{pool}\rvert$) | | | | | | | | | | | | | | | |
| (-) *Domain-shared loss* | | | | 0.823 | 0.821 | 0.812 | 0.815 | 0.864 | 0.832 | 0.828 | 0.829 | 0.956 | 0.857 | 0.861 | 0.858 |
| (-) *Domain-specific loss* | | | | 0.792 | 0.800 | 0.783 | 0.786 | 0.858 | 0.832 | 0.821 | 0.828 | 0.934 | 0.850 | 0.857 | 0.853 |
| (-) *Network modality* | | | | 0.816 | 0.815 | 0.817 | 0.815 | 0.765 | 0.749 | 0.745 | 0.746 | 0.945 | 0.803 | 0.855 | 0.827 |
| (-) *Text modality* | | | | 0.804 | 0.798 | 0.793 | 0.795 | 0.837 | 0.835 | 0.815 | 0.817 | 0.932 | 0.711 | 0.704 | 0.707 |

# Experiments

## Ablation Study (2/2)

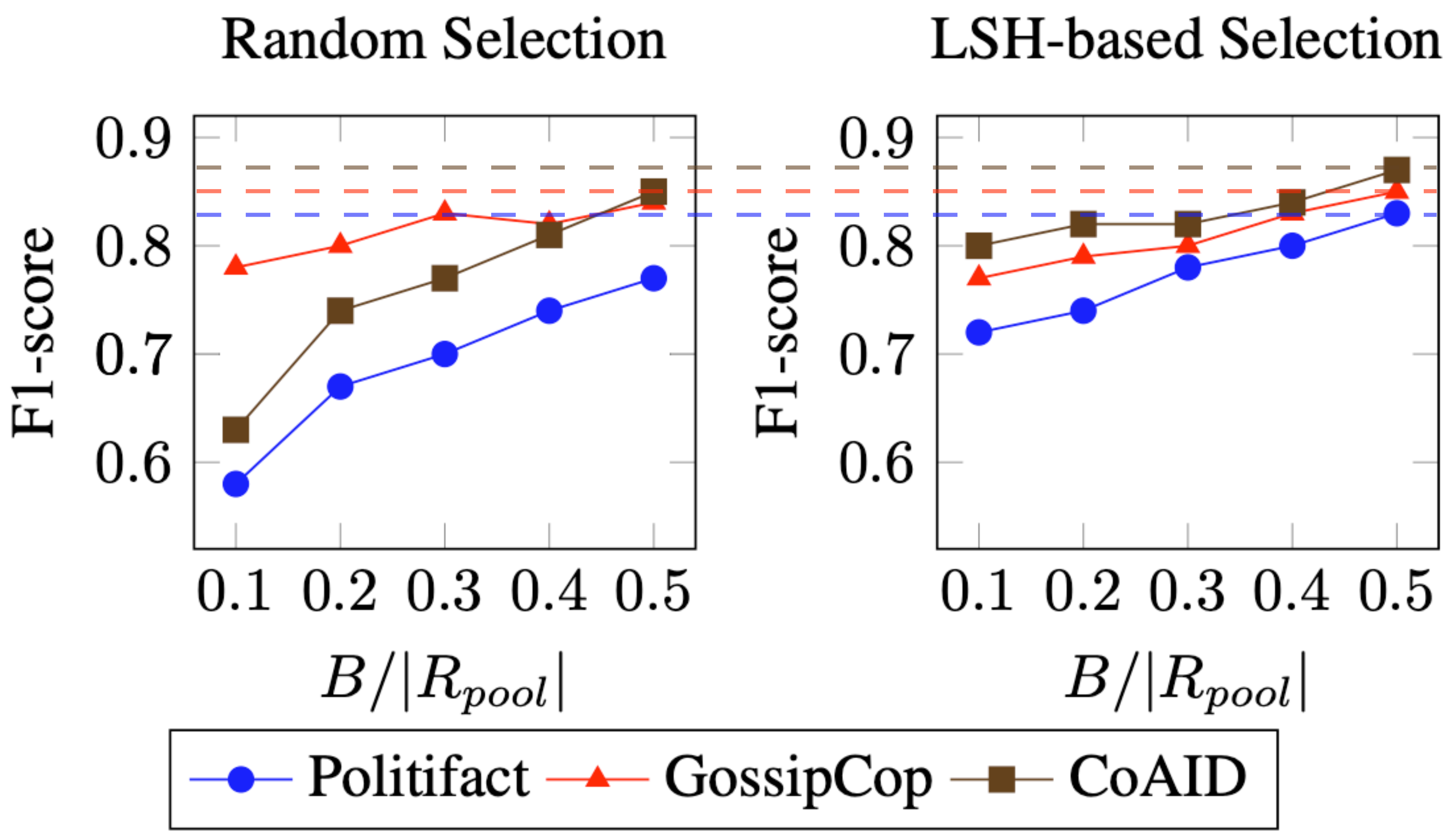



- As can be seen, the domain-specific embedding layer preserves the domain of the news by mapping different domains into different clusters (fig.a).

- In contrast, cannot identify the domain labels of news from the cross-domain embedding space (fig.b).
  - Hence, this embedding space is useful to share common knowledge between news from different domains.

# Experiments

## Evaluation of LSH-based Instance Selection




- Proposed approach substantially outperforms the random instance selection for the rarely-appearing or highly imbalance domains.
- It increases F1-score by 14% for PolitiFact and 17% for CoAID, when $B/|R_{pool}| = 0.1$.
  - May due to the ability of approach to maximize the coverage of domains when selecting instances.
  - Instead of biasing towards a domain with larger # of records.

# Conclusion

- Proposed a novel FND framework, which exploits domain-specific & cross-domain knowledge in news to determine fake news from different domains.
- Also introduced novel unsupervised approach to select informative instances for manual labelling from a large pool of unlabelled news.
  - Selected data pool is subsequently used to train a model that can perform equally for different domains.
- For future work, authors intend to extend model as an online learning framework to determine fake news in a real-world news stream, which typically covers a large number of domains.

# Comments
## of EDDFN

- Multimodal setting in this paper is use text & network information.
  - Different w/ text & image as usual.
- Preserve domain-invariant & domain-specific feature at the same time.
  - Curious about the domain-specific detail performance (only cancel loss in ablation).
- Notation in this paper is uncleared. 😠 ($f_{shared}$ vs. $f_{shared}(r)$)
- Network feature not detailed explained cause use authors' previous work.