THE UNIVERSITY OF UTAH

VISA Research

Constrained Non-Affine Alignment of Embeddings

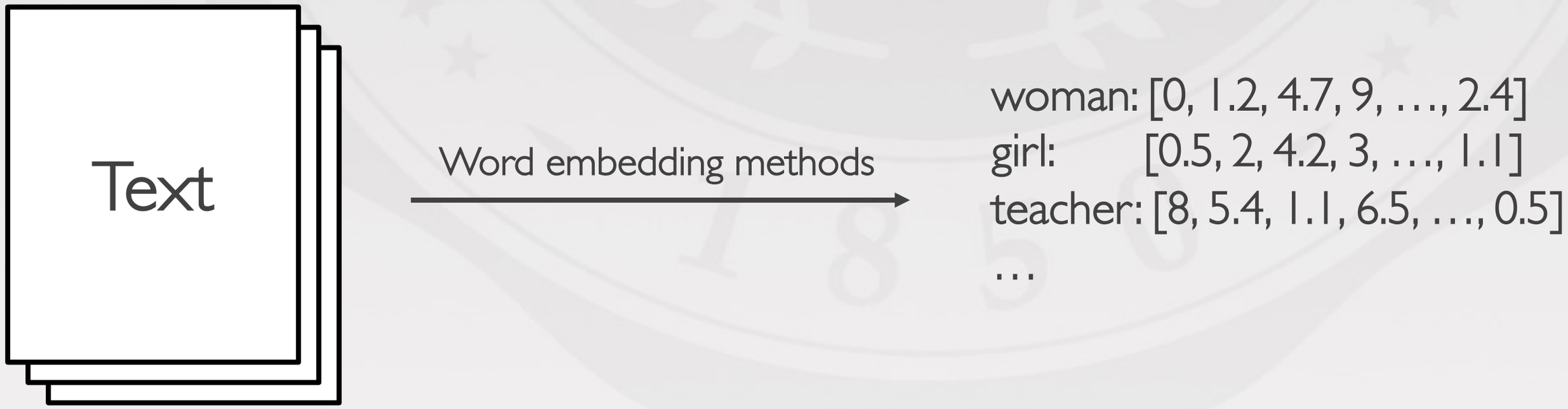
Yuwei Wang, Yan Zheng, Yanqing Peng, Michael Yeh,
Zhongfang Zhuang, Das Mahashweta, Bendre Mangesh,
Feifei Li, Wei Zhang, Jeff M. Phillips

University of Utah, Visa Research

Embeddings

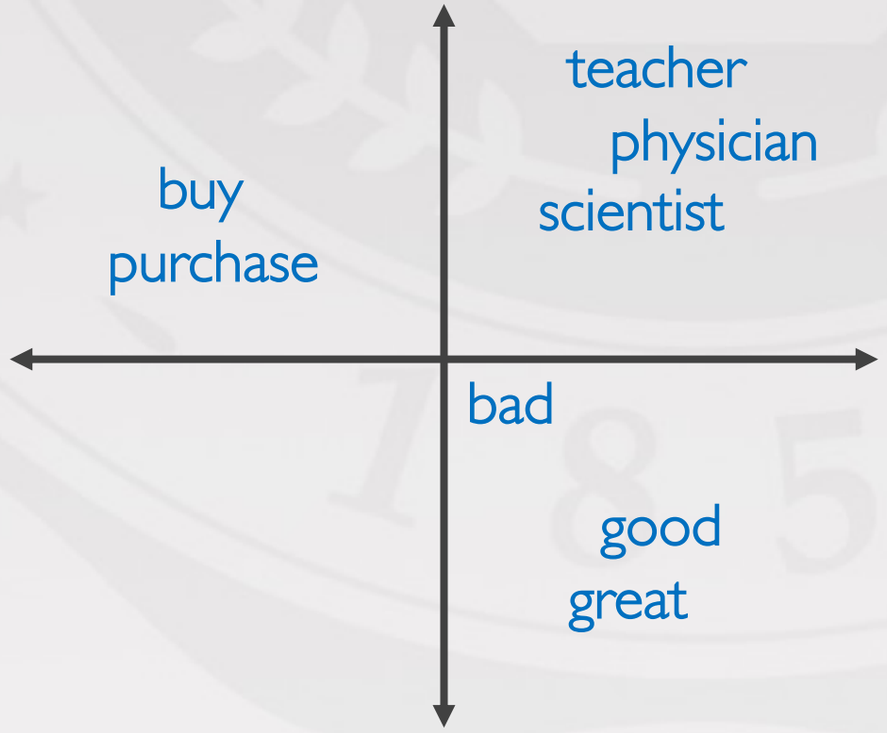
What are embeddings?

An embedding is moderate-dimensional vector representation of an entity where many features can be captured.



Embeddings

Embeddings preserve the information of entities by placing similar entities close together in the embedding space, as measured by cosine distance.



Embeddings

However, it is hard to tease apart features since all the features are entangled together, and there is no simple mapping between the dimensions and the features.

How can we measure the significance of a feature (e.g. polarity)?

THE UNIVERSITY OF UTAH

VISA Research

Feature Measurement

Our scenario

- We can access a subset of features in the original data.
- We measure how significantly an embedding is affected by a known categorical feature.
- We quantify the significance using a classifier.

Feature Measurement

We have dataset D and a balanced feature F with M labels. For fields using numerical values, we set thresholds to label the values into different categories.

Consider an embedding generator $E : \mathcal{D} \to \mathbb{R}^d$ and a balanced feature $F : \mathcal{D} \to \{0, 1, ..., M-1\}$. For a family of classifiers $\mathcal{C}$ on the embedding space, and a positive value ε with the following probability:

$$\max_{C \in \mathcal{C}} \text{Prob}_{x \in \mathcal{D}}[C(E(x)) = F(x)] > \frac{1}{M} + \varepsilon$$

We say that E embeds F with weight ε.

THE UNIVERSITY OF UTAH

VISA Research

Feature Measurement

We have dataset D and a balanced feature F with M labels. For fields using numerical values, we set thresholds to label the values into different categories.

Consider an embedding generator $E : \mathcal{D} \to \mathbb{R}^d$ and a balanced binary feature $F : \mathcal{D} \to \{0, 1\}$. For a family of classifiers $\mathcal{C}$ on the embedding space, and a positive value ε with the following probability:

$$\max_{C \in \mathcal{C}} \text{Prob}_{x \in \mathcal{D}}[C(E(x)) = F(x)] > 50\% + \varepsilon$$

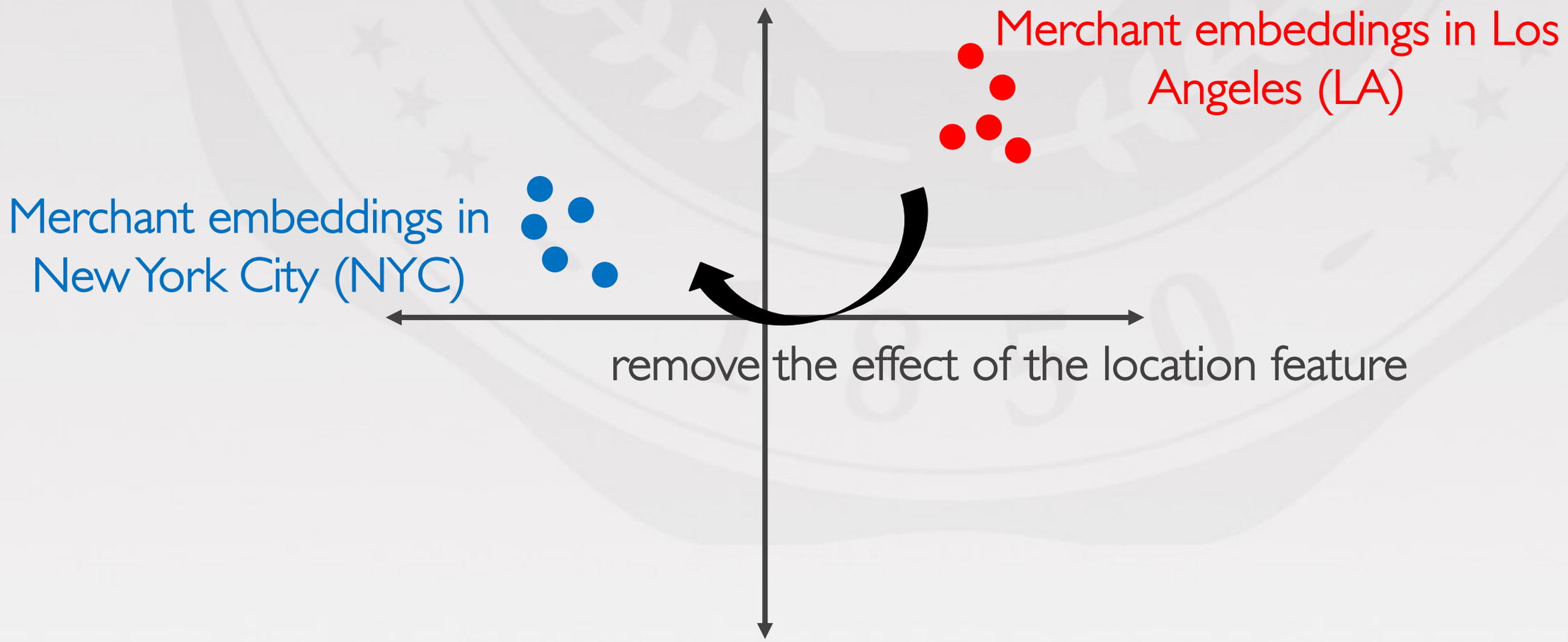
We say that E embeds F with weight ε.

THE UNIVERSITY OF UTAH

VISA Research

Embedding Attenuation and Retention

Given an undesired binary feature F on dataset D, two subsets A and B with corresponding embeddings, how can we align them by removing the effect of the F?

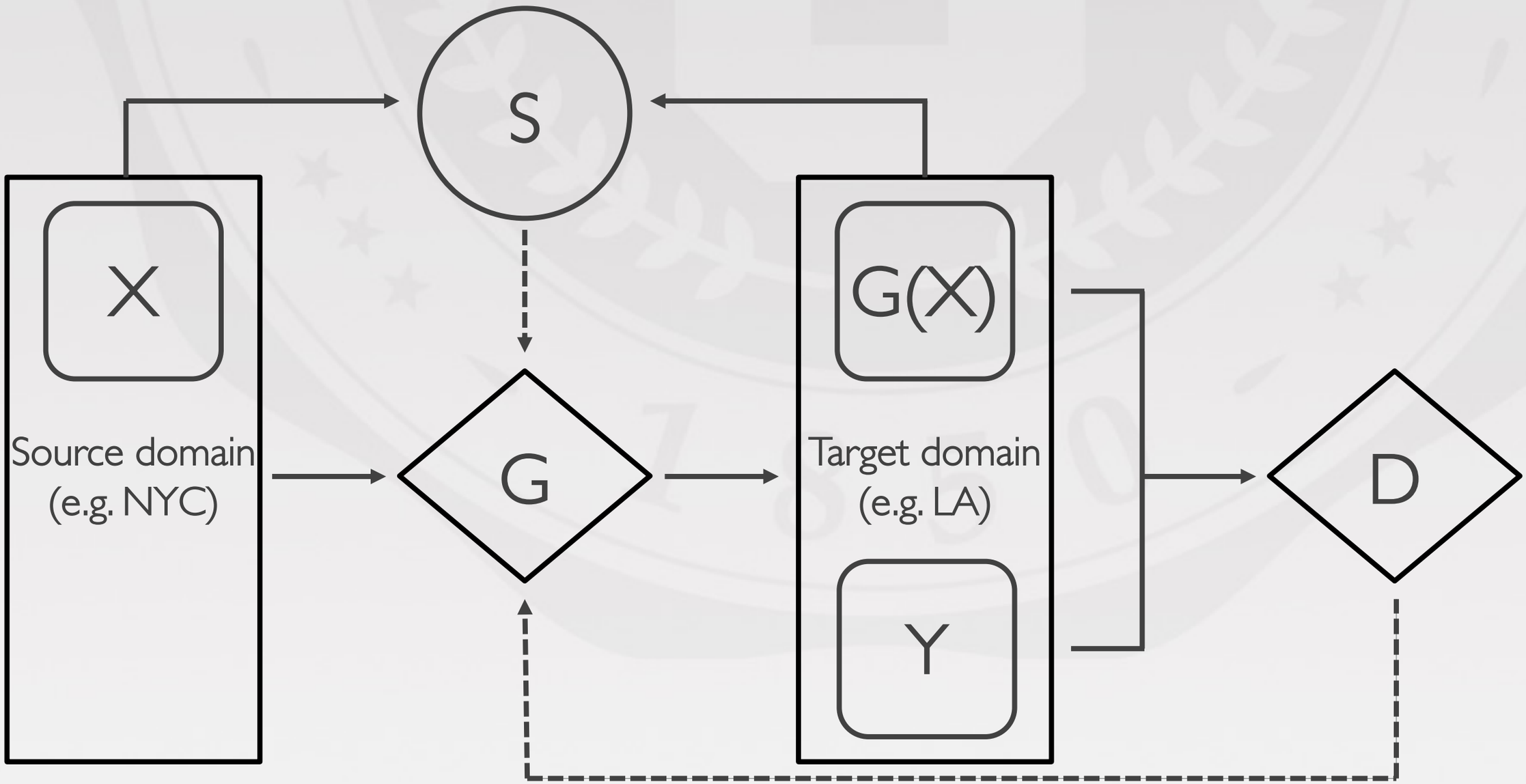


THE UNIVERSITY OF UTAH

VISA Research

Embedding Attenuation and Retention

Our proposed method UCAN (Unsupervised Constrained Alignment that is None-Affine)

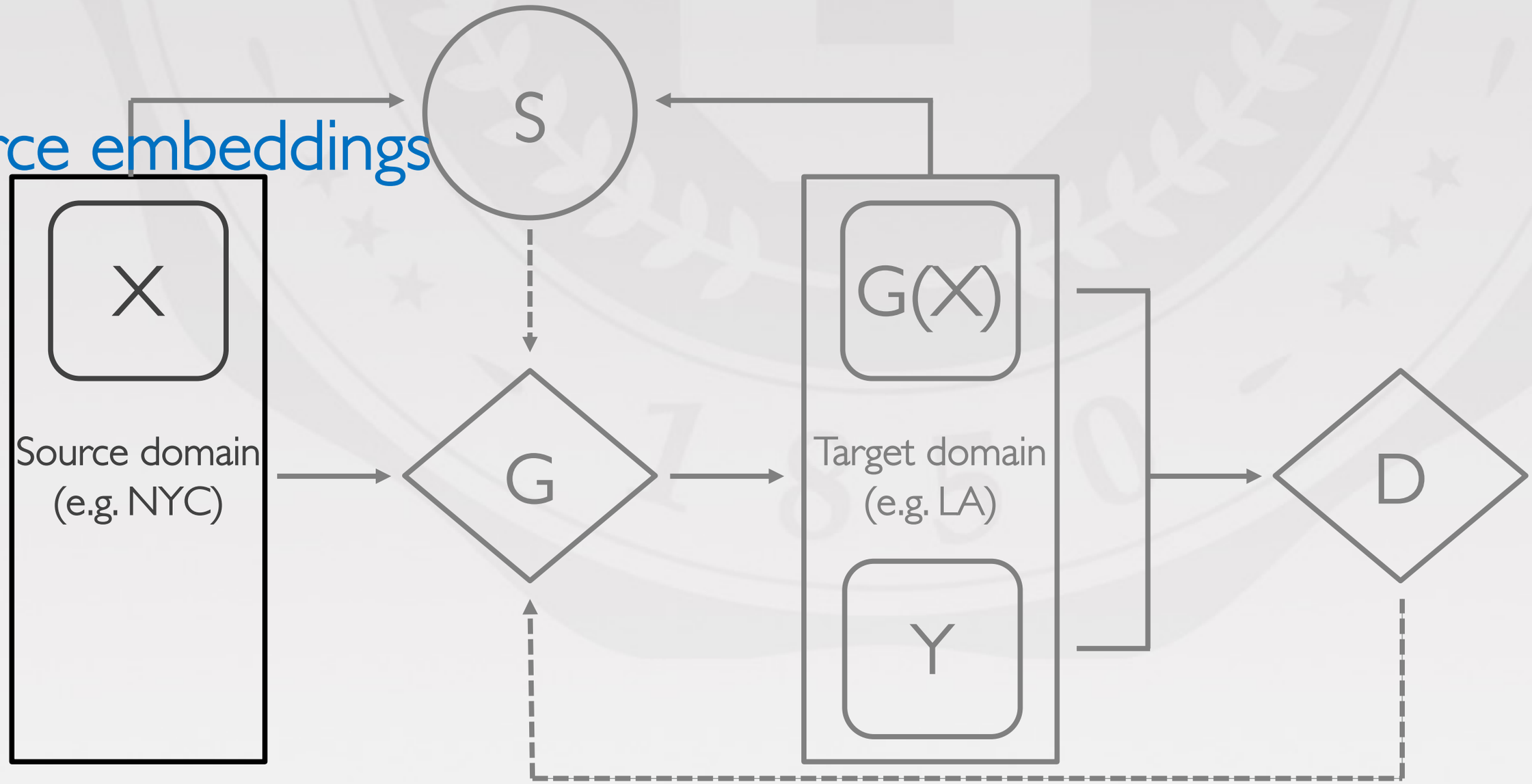


THE UNIVERSITY OF UTAH

VISA Research

Embedding Attenuation and Retention

Our proposed method UCAN (Unsupervised Constrained Alignment that is None-Affine)

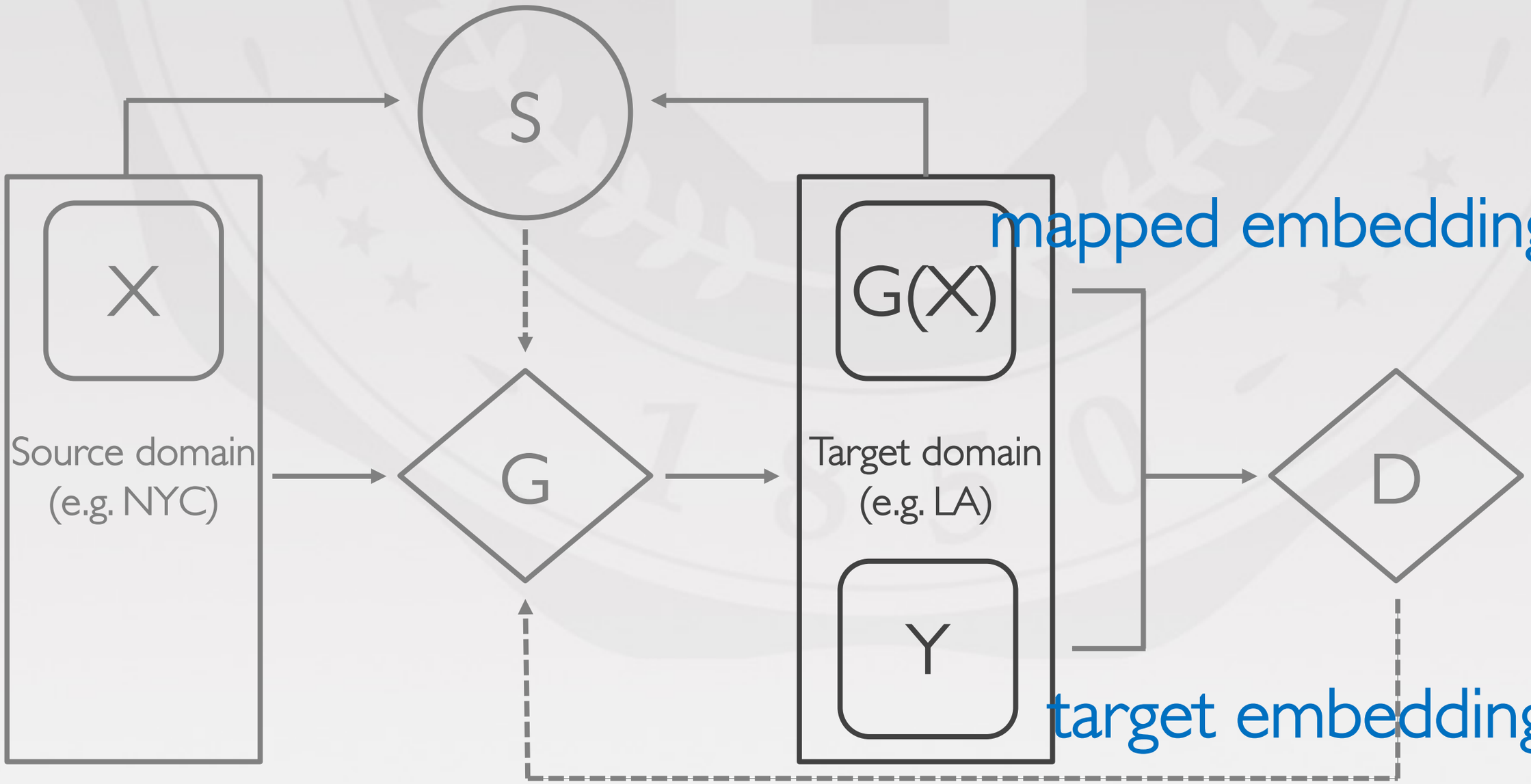


THE UNIVERSITY OF UTAH

VISA Research

Embedding Attenuation and Retention

Our proposed method UCAN (Unsupervised Constrained Alignment that is None-Affine)

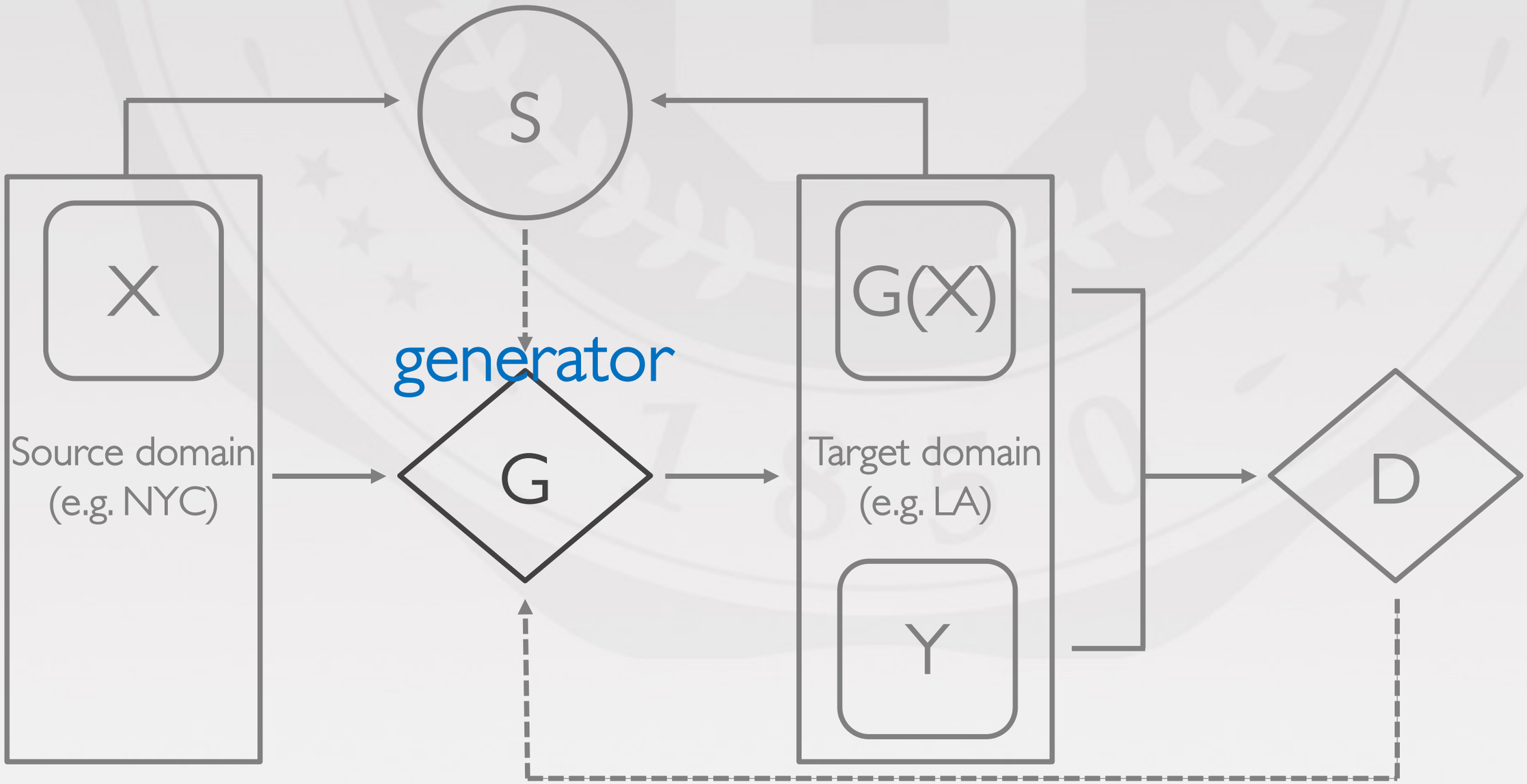


THE UNIVERSITY OF UTAH

VISA Research

Embedding Attenuation and Retention

Our proposed method UCAN (Unsupervised Constrained Alignment that is None-Affine)

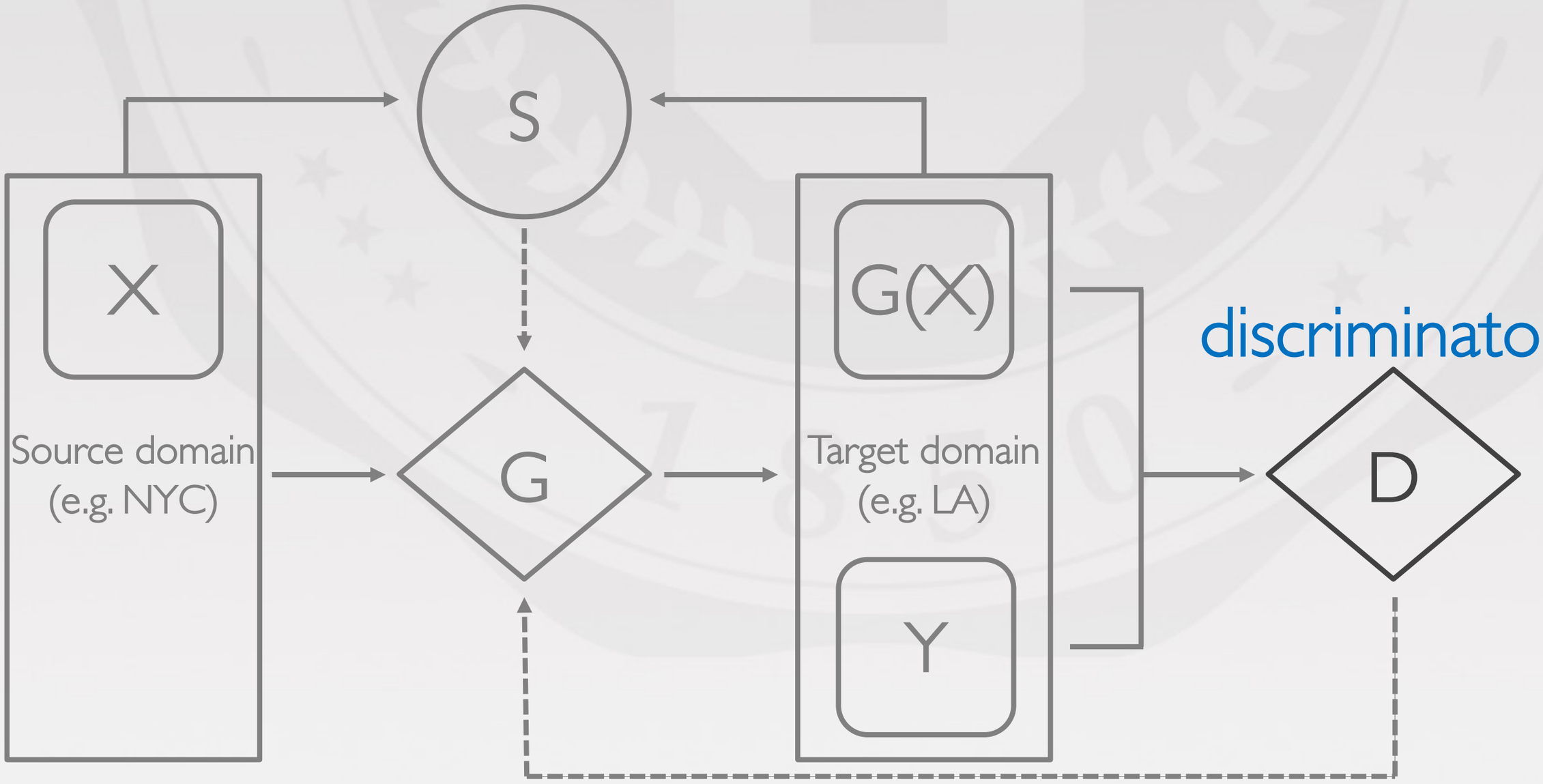


THE UNIVERSITY OF UTAH

VISA Research

Embedding Attenuation and Retention

Our proposed method UCAN (Unsupervised Constrained Alignment that is None-Affine)



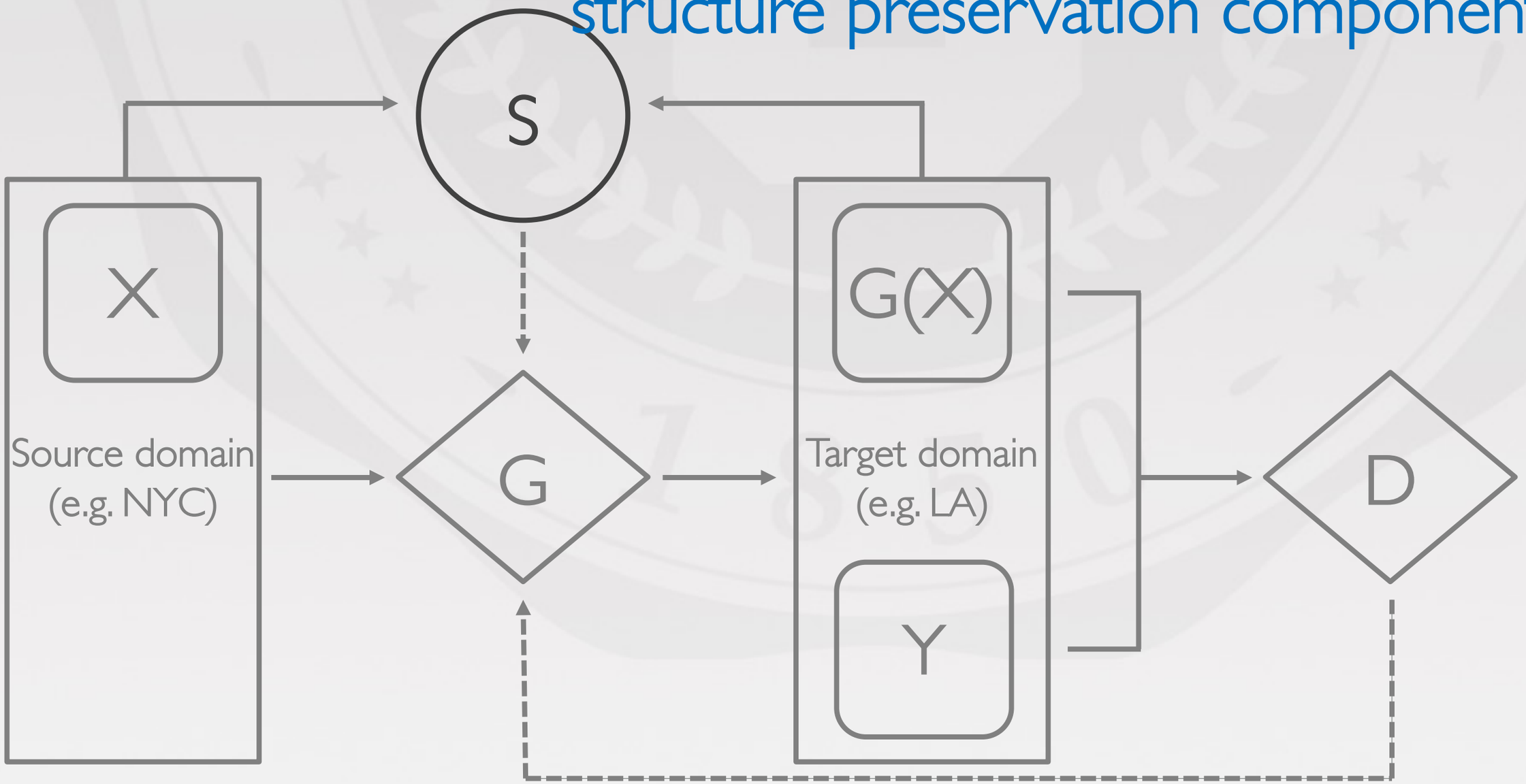
THE UNIVERSITY OF UTAH

VISA Research

Embedding Attenuation and Retention

Our proposed method UCAN (Unsupervised Constrained Alignment that is None-Affine)

structure preservation component



THE UNIVERSITY OF UTAH

VISA Research

Experiments

Synthetic datasets

Real-world datasets

Applications

THE UNIVERSITY OF UTAH

VISA Research

Experiments

Real datasets

We use three real-world datasets and choose MUSE and UMWE as our baselines.

THE UNIVERSITY OF UTAH

VISA Research

Experiments

Real datasets

- airport embedding dataset
- multi-language embeddings dataset
- merchant embedding dataset

THE UNIVERSITY OF UTAH

VISA Research

Experiments

Multi-language embeddings dataset

The unsupervised word vectors were trained using FastText.

The languages focused on in our experiments are English (en), Spanish (es), French (fr), German (de), Russian (ru) and Italian (it).

We use the standard K-nearest neighbor (NN) and Cross-Lingual Similarity Scaling (CSLS) as the evaluation approaches.

We measure how many times one of the correct translations of a source word are retrieved, and report the precisions for K = 1, 5, 10.

Experiments

Multi-language embeddings dataset

Our goal is to remove the "language'" feature, and retain the meanings of the words from different languages.

		en-es	es-en	en-fr	fr-en	en-de	de-en	en-ru	ru-en	en-it	it-en
MUSE	K@1	**69.07**	64.40	53.96	61.6	**61.02**	51.40	24.20	32.00	56.45	59.33
	K@5	**82.93**	78.40	65.08	75.93	71.19	66.60	44.20	50.00	66.13	73.73
	K@10	**86.87**	81.8	68.25	80.60	72.88	71.93	51.40	56.20	69.35	81.13
UWME	K@1	64.13	61.73	62.47	61.27	52.40	51.73	**27.47**	39.13	56.73	57.27
	K@5	76.20	75.80	75.73	75.67	74.27	67.33	**52.33**	57.67	71.67	71.27
	K@10	79.47	79.60	79.60	79.47	79.13	71.87	**60.87**	63.53	76.33	75.33
UCAN	K@1	68.93	**72.00**	**69.60**	**69.00**	60.27	**61.80**	25.73	**46.60**	**61.33**	**63.07**
	K@5	81.26	**84.60**	**82.93**	**83.13**	**79.33**	**76.26**	51.33	**65.73**	**77.93**	**79.07**
	K@10	85.00	**88.40**	**85.80**	**87.20**	**83.27**	**80.80**	59.80	**70.26**	**82.87**	**82.53**

Experiments

Multi-language embeddings dataset

We also demonstrate the results on languages which are less similar to English, including Greek (el), Vietnamese (vi), Arabic (ar), Czech (cs) and Dutch (nl).

		en-el	el-en	en-vi	vi-en	en-ar	ar-en	en-cs	cs-en	en-nl	nl-en
MUSE	K@1	13.87	0.00	0.00	0.07	12.47	0.00	24.73	41.67	49.07	45.53
	K@5	31.13	0.00	0.13	0.07	26.33	0.07	42.00	58.33	67.80	61.87
	K@10	38.33	0.00	0.33	0.07	32.27	0.07	49.67	64.40	73.33	67.00
UWME	K@1	18.73	26.27	3.27	1.87	14.60	22.49	26.87	31.13	43.07	33.67
	K@5	36.47	44.80	8.40	5.60	32.00	38.35	46.53	48.00	60.20	48.93
	K@10	43.27	50.53	10.73	8.53	38.87	44.51	53.20	53.07	66.07	54.13
UCAN	K@1	**28.20**	**41.47**	**12.13**	**31.93**	**19.97**	**34.80**	**28.33**	**49.73**	**60.00**	**61.67**
	K@5	**46.60**	**59.93**	**23.27**	**45.33**	**41.45**	**51.74**	**50.27**	**67.73**	**76.40**	**75.60**
	K@10	**52.53**	**65.27**	**27.67**	**49.60**	**48.67**	**57.50**	**58.93**	**71.60**	**80.60**	**79.27**

THE UNIVERSITY OF UTAH

VISA Research

Experiments

Real datasets

- airport embedding dataset
- multi-language embeddings dataset
- merchant embedding dataset

THE UNIVERSITY OF UTAH

VISA Research

Experiments

Merchant embedding dataset

The merchant embedding is generated from a real-world transaction dataset involving 70 million merchants and 26-million customers.

We focus on two features: the location and the merchant category code (MCC).

Experiments

Merchant embedding dataset

Our goal is to retain the merchant category code feature (F1) and remove the location feature (F2).

	ORG			UCAN			MUSE			UWME		
	O_F1	O_F2	O_F1/O_F2	C_F1	C_F2	C_F1/C_F2	M_F1	M_F2	M_F1/M_F2	U_F1	U_F2	U_F1/U_F2
LA→SF	0.64	0.88	0.73	0.62	0.59	**1.05**	0.62	0.63	0.98	0.62	0.77	0.81
SF→LA	0.64	0.88	0.73	0.62	0.61	**1.02**	0.61	0.78	0.78	0.62	0.75	0.83
LA→CHI	0.58	0.95	0.61	0.57	0.57	**1.00**	0.58	0.71	0.82	0.57	0.79	0.72
CHI→LA	0.58	0.95	0.61	0.56	0.57	**0.98**	0.57	0.63	0.90	0.57	0.75	0.76
SF→CHI	0.65	0.93	0.70	0.63	0.59	**1.07**	0.61	0.79	0.77	0.64	0.82	0.78
CHI→SF	0.65	0.93	0.70	0.63	0.56	**1.13**	0.63	0.66	0.95	0.64	0.76	0.84
MAN→LA	0.61	0.90	0.68	0.59	0.54	**1.09**	0.59	0.82	0.72	0.60	0.75	0.80
LA→MAN	0.61	0.90	0.68	0.59	0.54	**1.09**	0.59	0.60	0.98	0.59	0.73	0.81
MAN→SF	0.66	0.88	0.75	0.64	0.53	**1.21**	0.64	0.62	1.03	0.64	0.71	0.90
SF→MAN	0.66	0.88	0.75	0.64	0.53	**1.21**	0.64	0.62	1.03	0.65	0.68	0.96
MAN→CHI	0.62	0.93	0.67	0.60	0.52	**1.15**	0.59	0.80	0.74	0.61	0.80	0.76
CHI→MAN	0.62	0.93	0.67	0.60	0.52	**1.15**	0.61	0.61	1.00	0.62	0.75	0.83

THE UNIVERSITY OF UTAH

VISA Research

Experiments

Synthetic datasets

Real-world datasets

Applications

THE UNIVERSITY OF UTAH

VISA Research

Experiments

Application: on the merchant embedding dataset

- False Merchant Identity Detection
- Cross-City Restaurant Recommendations

THE UNIVERSITY OF UTAH

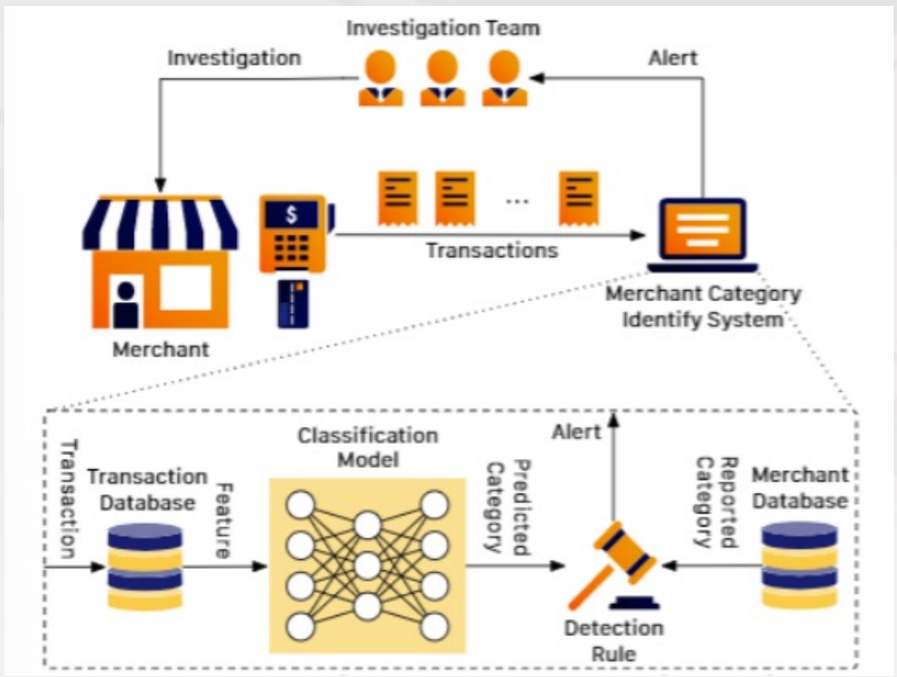
VISA Research

Experiments

False Merchant Identity Detection

Check whether a merchant's self-reported merchant category is within the top-kth most likely merchant. If It is not, the merchant will be reported as a suspicious merchant.

We trained our model with Los Angeles's merchants and tested the model on San Francisco's merchants.

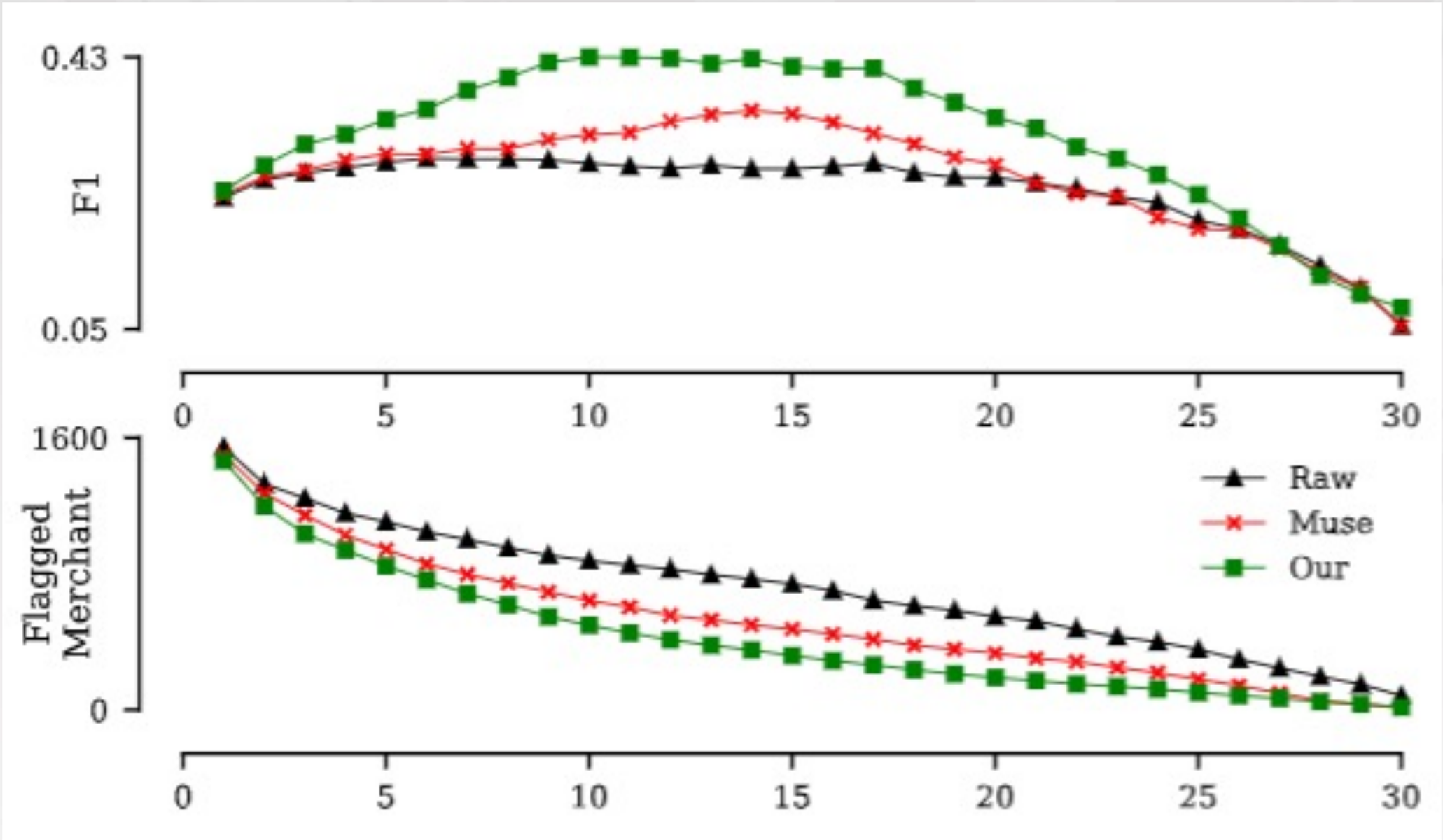


THE UNIVERSITY OF UTAH

VISA Research

Experiments

False Merchant Identity Detection



THE UNIVERSITY OF UTAH

VISA Research

Experiments

Application

- False Merchant Identity Detection
- Cross-City Restaurant Recommendations

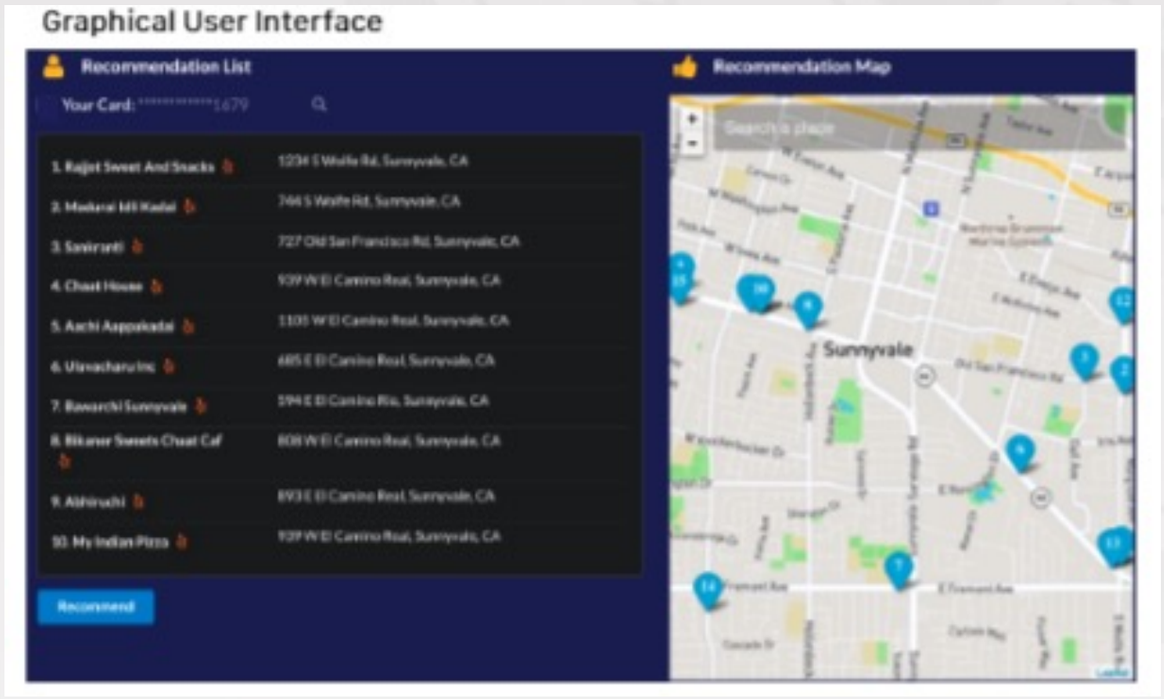
THE UNIVERSITY OF UTAH

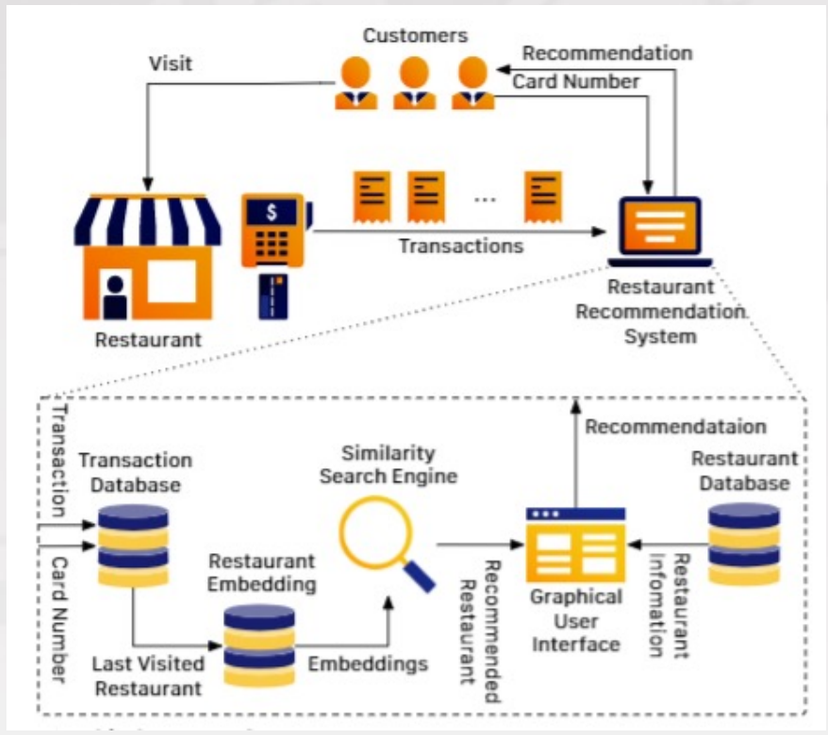
VISA Research

Experiments

Cross-City Restaurant Recommendations

Our goal is to recommend Los Angeles (LA) restaurants to the customers who usually have restaurant transactions in San Francisco (SF).





THE UNIVERSITY OF UTAH

VISA Research

Experiments

Cross-City Restaurant Recommendations

Our original Word2Vec embeddings give a score of 60.36%. The score increases to 63.60% after UCAN removes the location feature from the embeddings, while MUSE does not improve the score.

THE UNIVERSITY OF UTAH

VISA Research

Thank you

ywang@cs.utah.edu

https://arxiv.org/pdf/1910.05862.pdf

THE UNIVERSITY OF UTAH

[1] D. R. Figueiredo, L. F. Ribeiro, and P. H. Saverese, “struc2vec: Learning node representations from structural identity,” in KDD 2018, 2017, pp. 13–17

Video Format

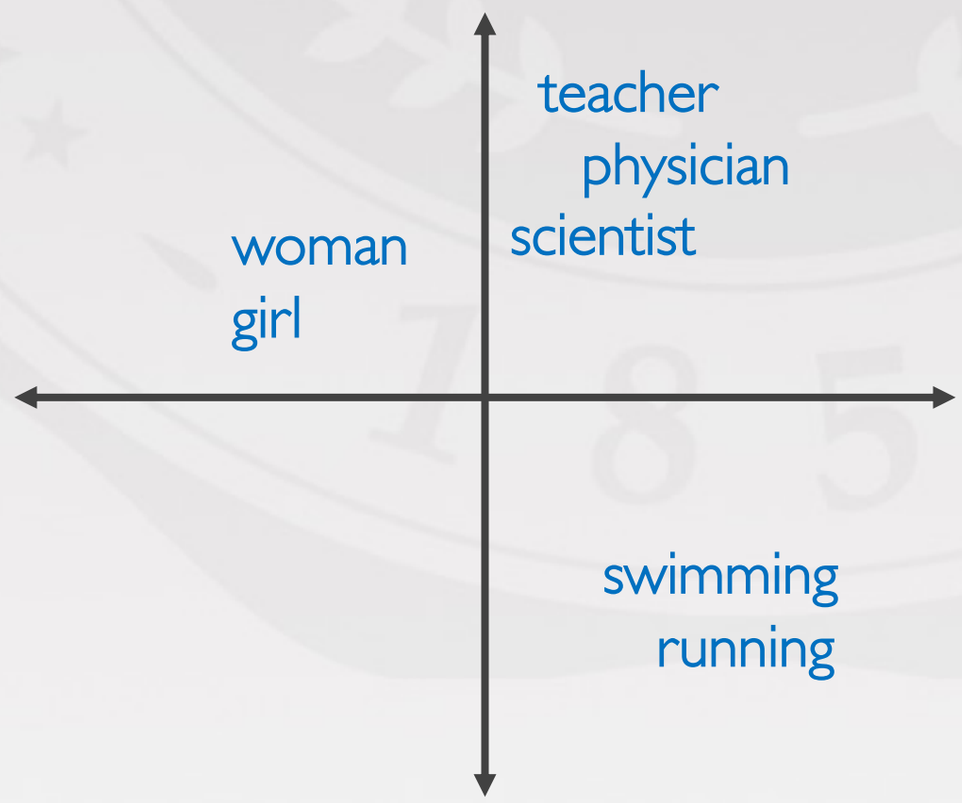
- Video resolution: 1280p Frame Width x 720p Frame Height
- Frame Rate: 25 Frames / Second
- File Type: mp4
- We suggest to authors to use for the creation of their video the ZOOM recording feature that by default support the aforementioned specs

- Video durations are specified in the following table:
- 7+5QA

Word Embeddings

One hot vectors or bag of words embeddings

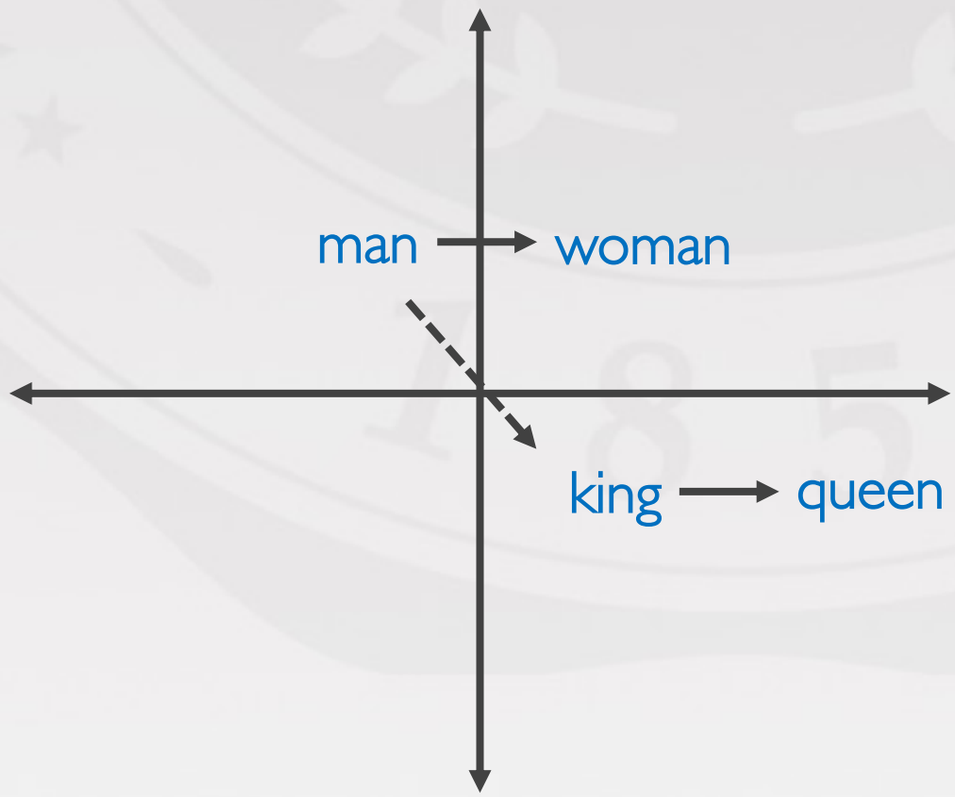
- 100K or more dimensional vectors
- sparse but inefficient
- including useful semantic and syntactic information



Word Embeddings

Word2Vec, GloVe, FastText

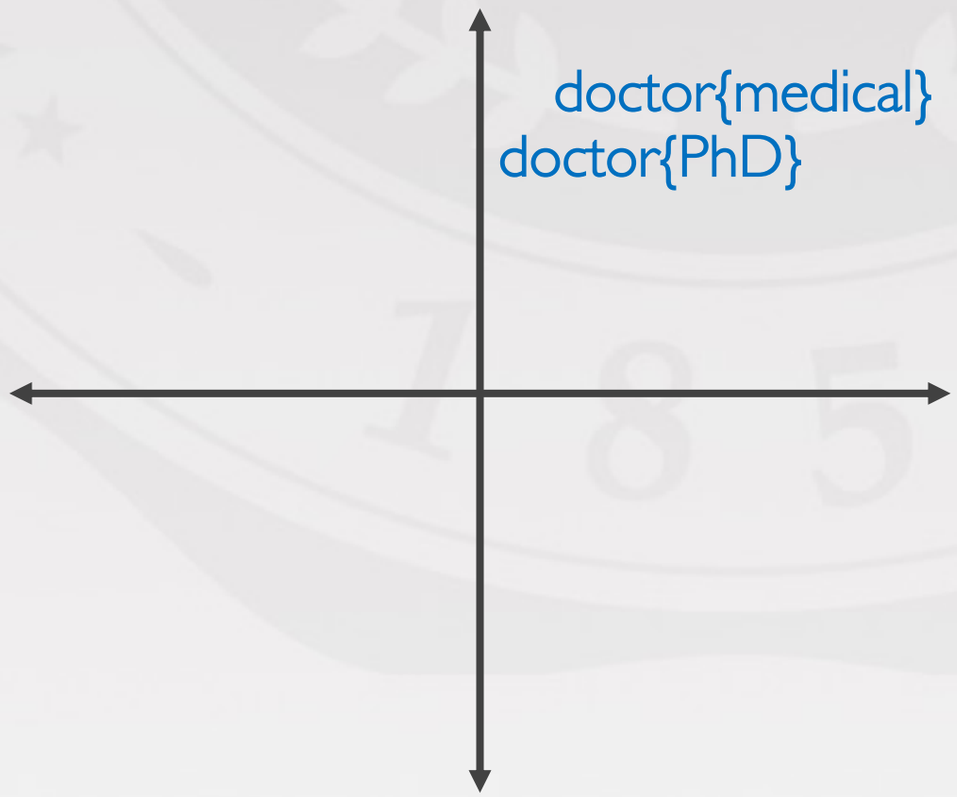
- distributed embeddings
- about ~300 dimensions
- additional useful linear relationships



Word Embeddings

ELMo, BERT

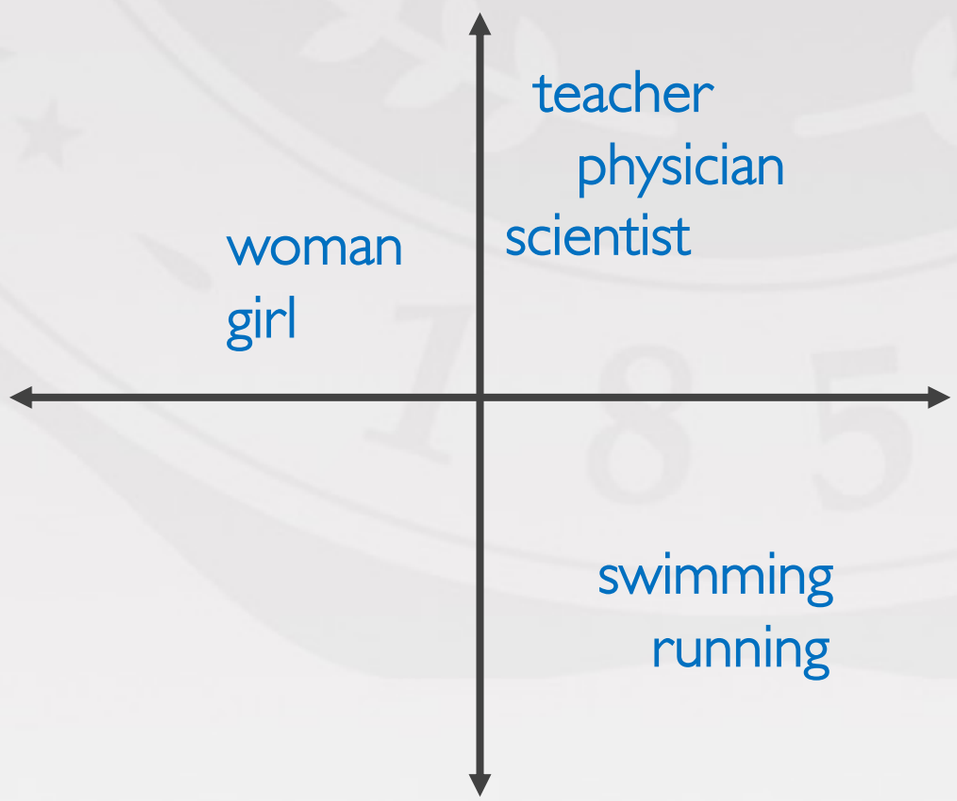
- contextual embeddings
- about 3000 (3*1024) dimensions
- distinguishable word senses



Embeddings

What is embeddings?

An embedding is a relatively low-dimensional space into which you can translate high-dimensional vectors.



Experiments

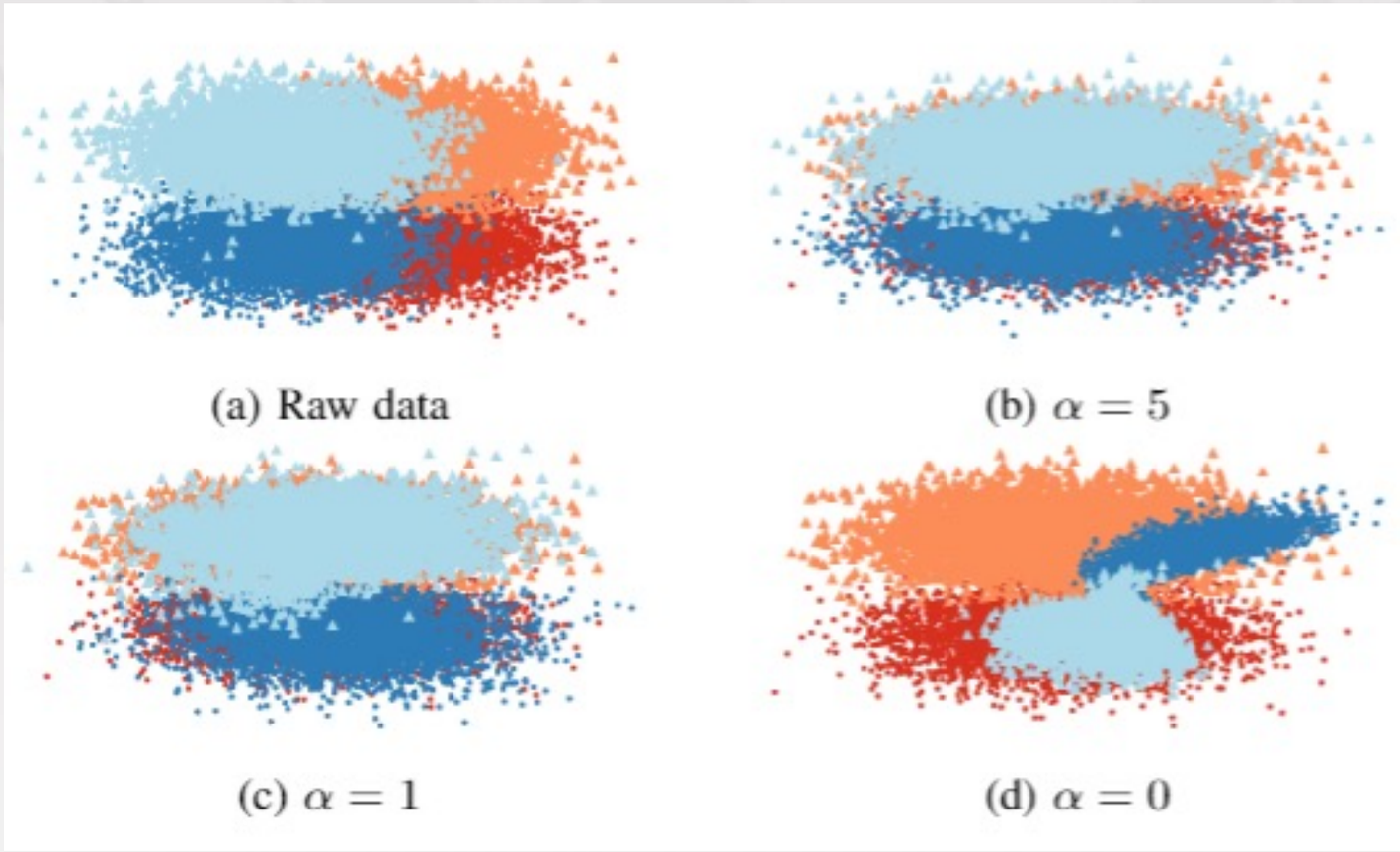
Synthetic datasets

Datasets have two dimensions – the color feature (the x-axis) and the lightness feature (the y-axis) and each corresponds to one type of binary feature. In this experiment, we aim to remove the color feature (F2) and retain the lightness feature (F1).

Experiments

Synthetic datasets

- balanced dataset
- unbalanced dataset



(a) Raw data

(b) $\alpha = 5$

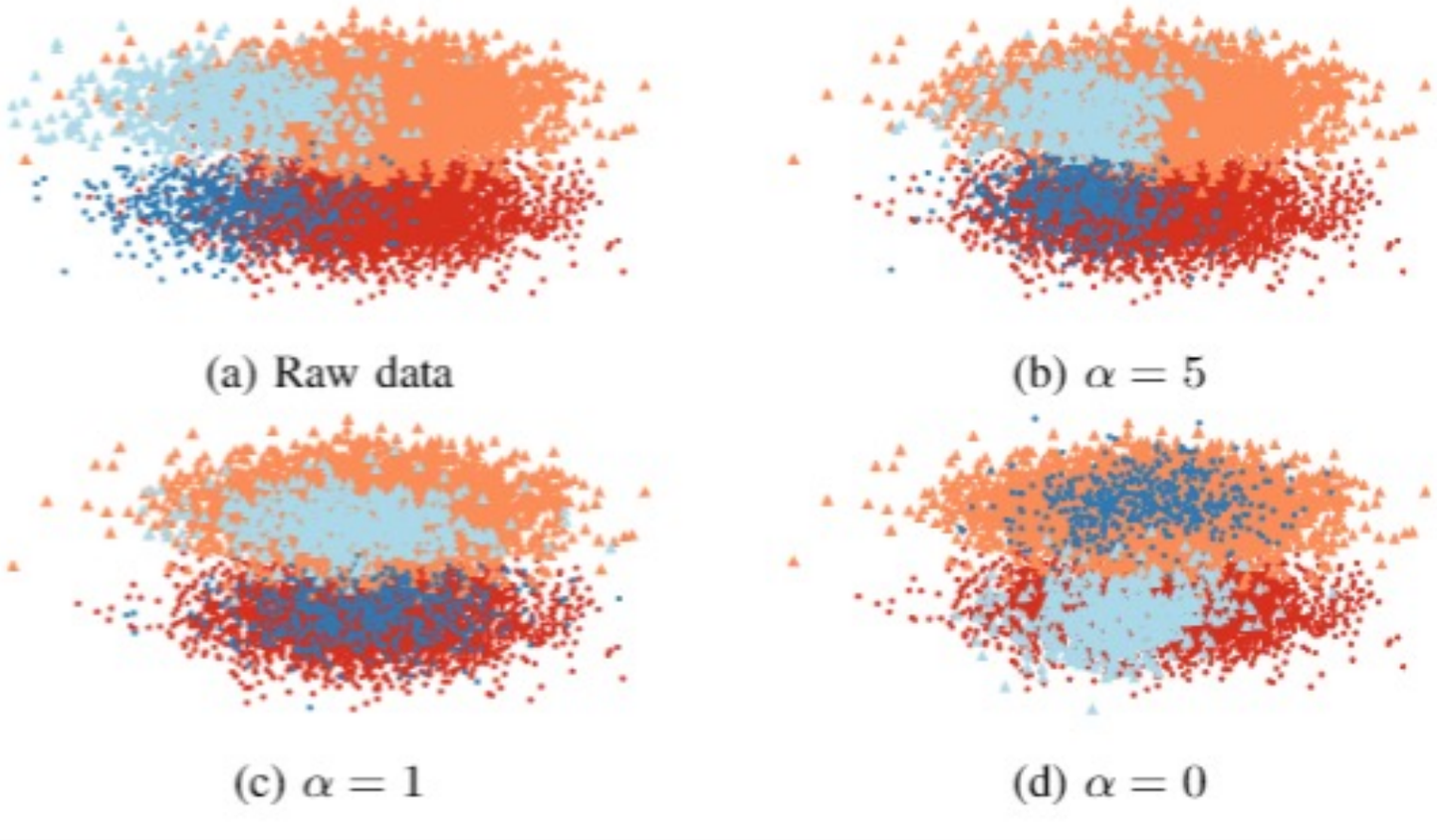
(c) $\alpha = 1$

(d) $\alpha = 0$

Experiments

Synthetic datasets

- balanced dataset
- unbalanced dataset



(a) Raw data

(b) $\alpha = 5$

(c) $\alpha = 1$

(d) $\alpha = 0$

Experiments

Real dataset

- airport embedding dataset
- multi-language embeddings dataset
- merchant embedding dataset

Experiments

Airport embedding dataset

The air-traffic networks dataset includes three air-traffic networks [3] from three countries.

We consider two features: the level of activity (F1) with 4 classes, and country location (F2) with 3 classes.

Experiments

Airport embedding dataset

We aim to retain the level of activity (F1) and remove country location (F2).

	O_F1	O_F2	O_F1/O_F2	C_F1	C_F2	C_F1/C_F2	M_F1	M_F2	M_F1/M_F2	U_F1	U_F2	U_F1/U_F2
Brazil→Europe	0.80	1	0.80	0.77	0.84	**0.92**	0.76	0.97	0.78	0.57	0.71	NA
Europe→Brazil	0.80	1	0.80	0.79	0.86	**0.92**	0.82	0.95	0.86	0.57	0.80	NA
Brazil→USA	0.85	1	0.85	0.85	0.61	**1.39**	0.83	0.98	0.85	0.5	1	NA
USA→Brazil	0.85	1	0.85	0.88	0.91	**0.97**	0.87	0.93	0.94	0.52	0.51	NA
USA→Europe	0.84	1	0.84	0.85	0.90	**0.94**	0.83	0.92	0.90	0.54	0.53	NA
Europe→USA	0.84	1	0.84	0.83	0.62	**1.34**	0.80	0.99	0.81	0.49	0.97	NA

THE UNIVERSITY OF UTAH

VISA Research

Feature Measurement

Imbalanced dataset and AUC

- "Accuracy Paradox"
- We use average one vs. all AUC as the metric to measure if a feature is embedded for both balanced and imbalanced data.