Semantic Communication

Research Presentation

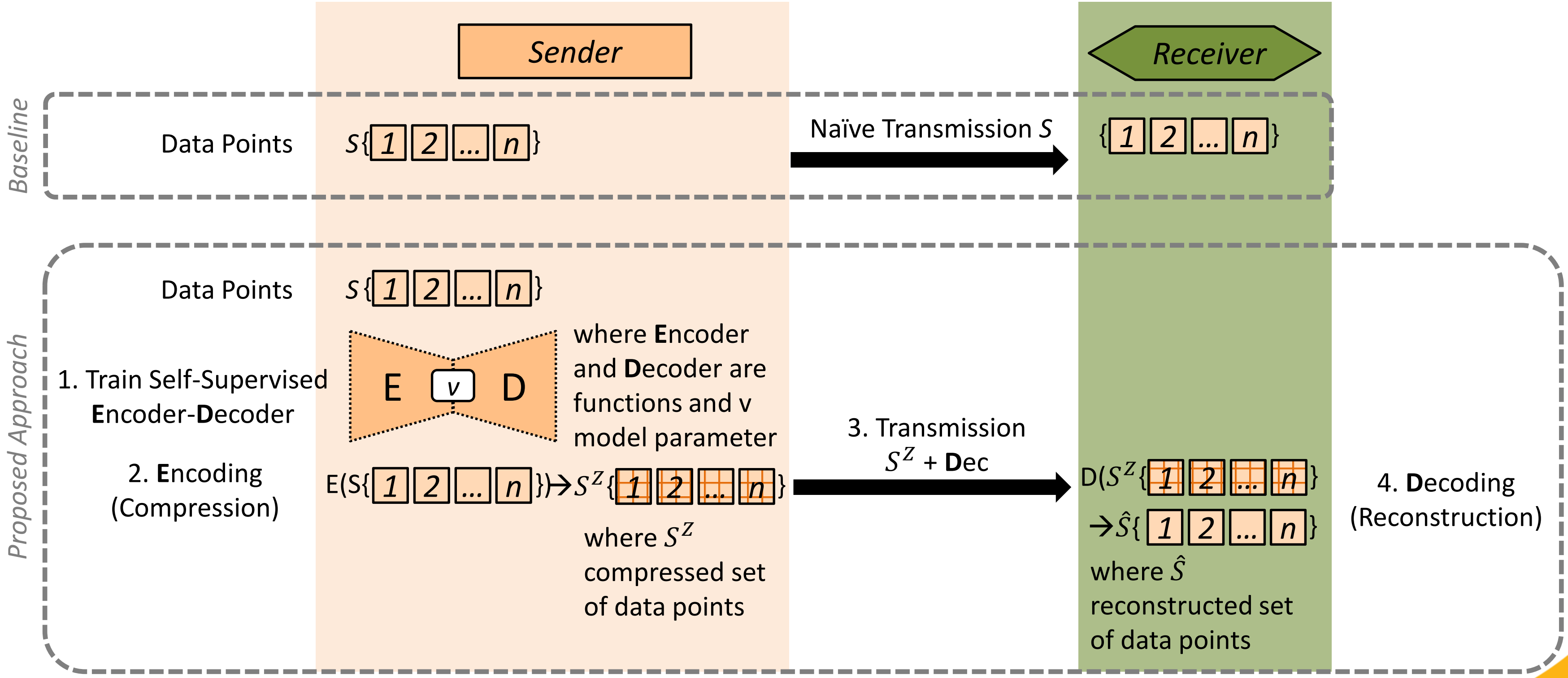
Leonard Traeger
Department of Information Systems,
University of Maryland Baltimore County

Prof. Dr. George Karabatis

Athanasios Ritas
Department of Electronics Engineering
Hellenic Mediterranean University

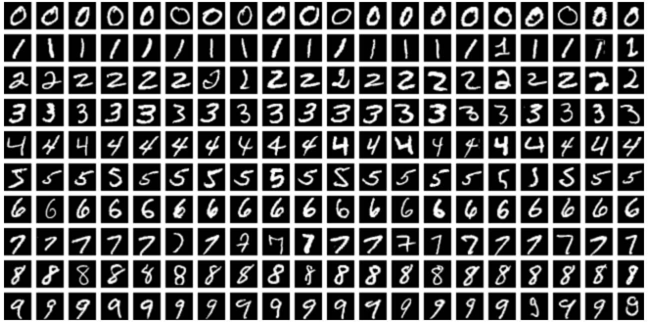
Prof. Dr. Ioannis Barbounakis

Communication



MNIST dataset

The MNIST database of handwritten digits with 784 features (28x28 pixels)



https://www.openml.org/search?type=data&sort=nr_of_likes&status=active&id=554

Communication Example

Sender

1. Train Self-Supervised Encoder-Decoder

```
mean, pca = method_pca_variance(X=S, v=0.8)
```

✓ 7.6s

2. Encoding (Compression)

```
S_z = pca_encode(S, mean, pca)
```

✓ 0.1s

```
#compressed second digit
S_z[1]
```

✓ 0.0s

```
array([1010.494   , -289.9636  , -576.1207  , -485.08362 ,
        841.47833 , -145.46716 ,  -48.333355 ,  102.231064 ,
        -34.786186 , -151.26501 , -282.70004 , -110.107765 ,
        329.9303   ,  -97.00281 , -226.4773  , -156.05844 ,
        -49.416885 ,  -37.92279 ,   63.192932 ,  113.47844 ,
        202.56096 ,  -94.196815 ,  -12.712128 ,  -29.628128 ,
        233.85992 ,  196.91718 ,   18.939308 ,  -28.379498 ,
        363.63416 ,  -84.49492 ,   -7.011917 , -100.57101 ,
        192.7258  ,   86.99288 , -137.27632 ,  -79.502464 ,
         61.11911 ,    4.0909624,  -43.707153 ,  -83.74587 ,
        219.2553  ,  -64.80328 ,   68.30545 ], dtype=float32)
```

3. Transmission

S^Z + **Dec**

3. Transmission

```
print("Original size of a digit:")
print(len(S[0]))
print("Compressed size of a digit:")
print(len(S_z[0]))
print("Reduction:")
print(f"{1 - (len(S_z[0])/len(S[0])):.2%}")

print("-------")
print("Original size of a digit times dataset size:")
print(str(len(S[0])) + " x " + str(len(S)) + " = " + str(len(S[0])*len(S)))
print("Compressed size of a digit times dataset size:")
print(str(len(S_z[0])) + " x " + str(len(S)) + " = " + str(len(S_z[0])*len(S)))
```

✓ 0.0s

```
Original size of a digit:
784
Compressed size of a digit:
43
Reduction:
94.52%
-------
Original size of a digit times dataset size:
784 x 70000 = 54880000
Compressed size of a digit times dataset size:
43 x 70000 = 3010000
```

Receiver

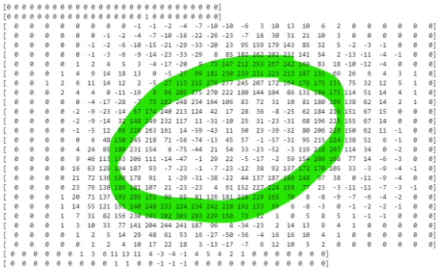
4. Decoding (Reconstruction)

```
S_hat = pca_decode(S_z, mean, pca)
```

✓ 0.2s

```
print_digit_array(np.array([round(pixel) for pixel in S_hat[1]]))
```

✓ 0.0s



Remarks and Questions

R1: The size of the compressed set of data points S^Z and the decoder function D depends on the parameter v (generalizability of the model). The higher v, the larger the size of S^Z and the decoder function D and the more accurate the reconstruction.

R2: There must be a trade-off in transmitting the data points from Sender to Receiver naively S versus with the proposed approach S^Z+ D. Performance metrics:

- Volume of transmission: S > S^Z+ D
- Data Accuracy:
  - Sender S → Receiver S: 100%
  - Sender E(S) = S^Z → Receiver D(S^Z)= $\hat{S}$: Reconstruction Error(S, $\hat{S}$)

Q1: What baseline compression methods exist?

Q2: When is the volume of the transmission indeed lower than the baseline while maintaining accurate data interpretation at receiver's end?

Back-Up

Irrelevant Lines

Alg.1: 14,15

Alg.2: 7-12

Algorithms

Algorithm 1 Local Self-Supervised Models

Input: S_k local schema, T^a, T^t global textual serialization, E global language model encoder, $v \in (1..0)$ global variance

Output: Local model: $M_k = \{\mu_k$ local signature mean, PC_k local principal components, l_k local linkability range$\}$

1: $S_k^t \leftarrow (e_{k_j}^t \leftarrow T^a(a_{k_j}) | a_{k_j} \in t_{k_i} \in S_k) \cup$
$(e_{k_i}^t \leftarrow T^t(t_{k_i}) | t_{k_i} \in S_k)$ `//Local serialization.`

2: $S_k^{\vec{v}} \leftarrow (e_{k_i}^{\vec{v}} \leftarrow E(e_{k_i}^t) | e_{k_i}^t \in S_k^t))$ `//Local signatures.`

3: $\mu_k = \text{mean}(S_k^{\vec{v}})$ `//Compute local signatures mean.`

4: $X_{origin} = S_k^{\vec{v}} - \mu_k$ `//Project signatures onto origin.`

5: $SV = \{sv_1, sv_2, \ldots\}, PC = \{pc_1, pc_2, \ldots\} = SVD(X_{origin})$ `//Compute full Singular Value Decomposition and return Singular Values and Principal Components.`

6: $EV^{sum} = \sum_{j=1}^{SV} sv_j^2$ `//Compute the sum of the squared SV for Explained Variance.`

7: $EV \leftarrow (ev_i = \frac{sv_i^2}{EV^{sum}} | \forall sv_i \in SV)$ `//Compute EV per PC.`

8: $CEV = (ev_1, ev_1 + ev_2, \ldots) \leftarrow \text{CumulativeSum}(EV)$ `//Cumulate EV for each added PC.`

9: $n_{comp} \leftarrow \text{GetIndex}(CEV, v) + 1$ `//Find PC number needed to locally explain the global variance so that` $> v$.

10: $PC_k \leftarrow \{pc_1, pc_2, \ldots\}$ with $pc_l \in PC \wedge l < n$ `//Reduce set of all PC to the top-`n.

11: $X^Z = X_{origin} \cdot PC_k^T$ `//Encode projected signatures.`

12: $\hat{X}_{origin} = X^Z \cdot PC_k$ `//Decode signatures.`

13: $\hat{X} = \hat{X}_{origin} + \mu_k$ `//Reverse projection onto origin.`

14: $S_k^s \leftarrow \{s_{k_i} = MSE(e_{k_i}^{\vec{v}}, \hat{x}_i) | \forall (e_{k_i}^{\vec{v}}, \hat{x}_i) \in (S_k^{\vec{v}}, \hat{X})\}$ `//Compute mean reconstruction error score of original and encoded-decoded signatures.`

15: $l_k \leftarrow max(s_{k_i} \in S_k^s)$ `//Select maximum outlier score as local linkability range.`

16: **return** $M_k = \{\mu_k, PC_k, l_k\}$ `//Local model components.`

Corresponding paper in review

Algorithm 2 Local Linkability Assessment

Input: $S_k^{\vec{v}}$ local schema signatures, $M = \{M_1, M_2, \ldots, M_m\} \setminus \{M_k\}$ models of all other local schemas where $M_m = \{\mu_m, PC_m, l_m\}$

Output: Streamlined schema: $S_k' = \{e_{k_1}, e_{k_2}, \ldots, e_{k_i}\}$

1: **for all** $M_m \in M$ **do**

2: $\quad X_{origin} = S_k^{\vec{v}} - \mu_m$

3: $\quad X^Z = X_{origin} \cdot PC_m^T$ `//Encode projected signatures.`

4: $\quad \hat{X}_{origin} = X^Z \cdot PC_m$ `//Decode signatures.`

5: $\quad \hat{X} = \hat{X}_{origin} + \mu_m$ `//Reverse projection onto origin.`

6: $\quad S_k^s \leftarrow \{s_{k_i} = MSE(e_{k_i}^{\vec{v}}, \hat{x}_i) | \forall (e_{k_i}^{\vec{v}}, \hat{x}_i) \in (S_k^{\vec{v}}, \hat{X})\}$ `//Compute mean reconstruction error score of original and encoded-decoded signatures.`

7: $\quad$ **for all** $s_{k_i} \in S_k^s$ **do**

8: $\quad\quad$ **if** $s_{k_i} \leq l_m$ **then**

9: $\quad\quad\quad S_k' \leftarrow Append(S_k', e_{k_i})$ `//Append linkable table or attribute signature` $e_{k_i} \in S_k$ `to streamlined schema` S_k'.

10: $\quad\quad$ **end if**

11: $\quad$ **end for**

12: **end for**

13: **return** S_k'

Corresponding paper in review