Multi-source Transfer Learning for Signal Detection over a Fading Channel with Co-channel Interference

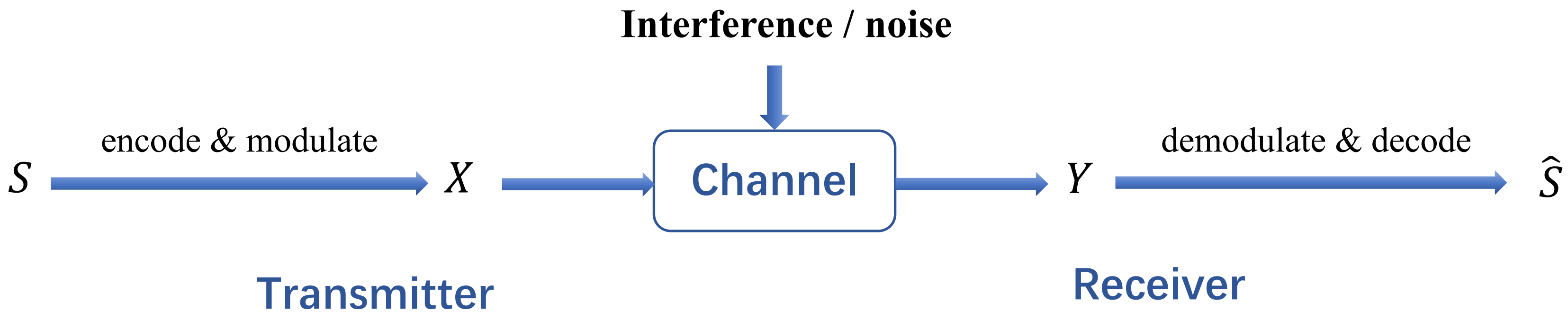
Ziyan Zheng*, Xinyi Tong*, Xinchun Yu*, Xiangxiang Xu†‡ and Shao-Lun Huang*§
* Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, China
† Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, USA

Presenter: Ziyan Zheng

ICC 2022

1 Background

Background



◆ Signal detection task -- detect X from Y

Standard model-based approaches may be inapplicable if:

- Precise channel model is unavailable
- Optimal receiver is complicated or unknown

e.g. → Existence of co-channel intereference

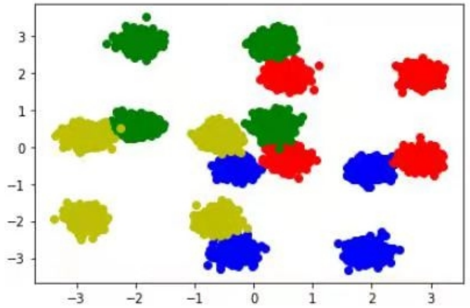


Fig 1. A constellation of transmitted QPSK-modulated signals with QPSK-modulated interference

General ways to study co-channel interference / non-Gaussian noise

- Ignore
- Treat as Gaussian noise

● Ziyan Zheng, et al. "Multi-source Transfer Learning for Signal Detection over a Fading Channel with Co-channel Interference." *IEEE International Conference on Communications* (2022).

Background

◆ Deep Learning

Strength -- Feature extraction from complex-distributed data
Difficulty -- Lack of pilots for training in time-varying channels

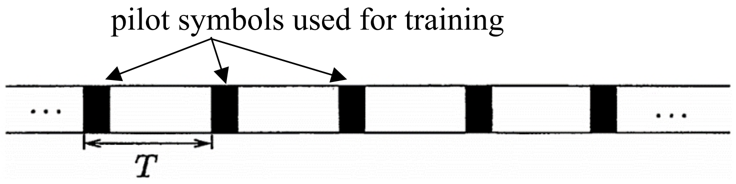


Fig 2. A data stream with pilots

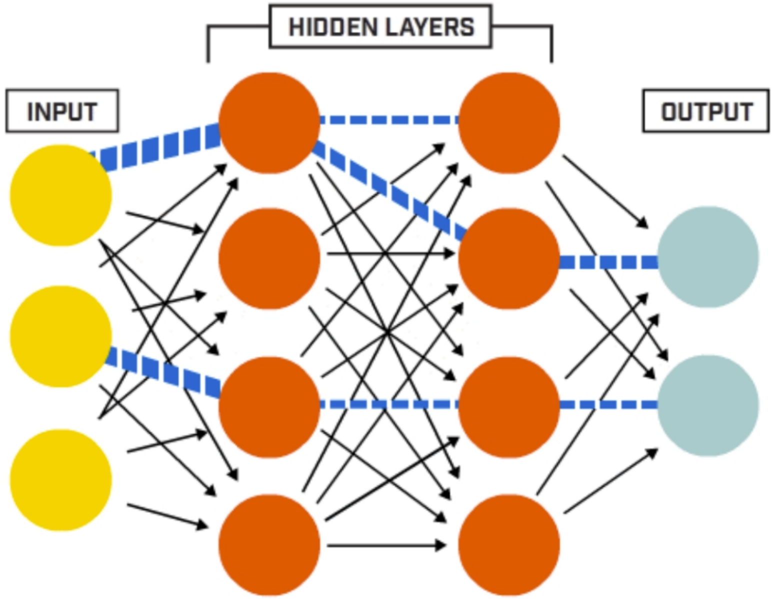


Fig 3. A simple deep neural network

◆ Transfer Learning

In a detector for time-varying channel,
can historical data be used in training? -- **Yes**

- Ziyan Zheng, et al. "Multi-source Transfer Learning for Signal Detection over a Fading Channel with Co-channel Interference." *IEEE International Conference on Communications* (2022).

2 Model and Theory

Model

received signal   discrete transmitted symbol

$$Y_t = H_t X_t + I_t + n_t$$

time-varying channel state matrix   co-channel interference   AWGN

◆ Notations and Assumptions

- $t \in \mathbb{Z}$: time period / packet index
- Each period t contains many symbols to transmit, the first several ones are pilots
- $H_t,\ X_t,\ I_t,\ n_t$ are pairwise independent

◆ Goal

- Detect X_T from Y_T at the latest period $T \in \mathbb{Z}$

◆ Available training samples for the detector

- Source samples: $\left\{\left(x_t^{(i)}, y_t^{(i)}\right)\right\}_{i=1}^{N_t}$ for $t = 1,\ 2,\ \ldots,\ T-1$
- Target samples: $\left\{\left(x_T^{(i)}, y_T^{(i)}\right)\right\}_{i=1}^{N_T}$, note that $N_T \ll N_t\ (\forall t < T)$

MSTL Framework

◆ Intuition

What kind of sources would provide more useful knowledge to target task?

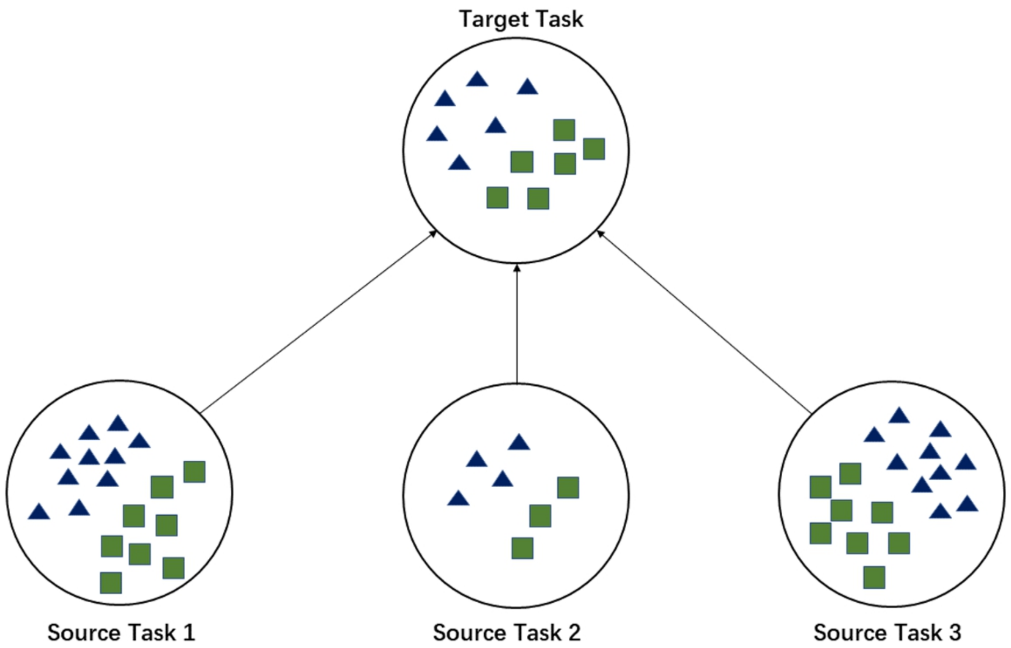


Fig 4. Intuition for transfer learning

- Source task 1: quite good
- Source task 2: less samples → Sample size is important
- Source task 3: not like target task → Similarity is important

- Ziyan Zheng, et al. "Multi-source Transfer Learning for Signal Detection over a Fading Channel with Co-channel Interference." *IEEE International Conference on Communications* (2022).

MSTL Framework

First – Assume received symbols Y_t are discrete

- Define empirical distributions (learned model) as $\hat{P}_{X_t Y_t}(x, y) \triangleq \frac{1}{N_t} \sum_{i=1}^{N_t} \mathbb{1}\left\{x_t^{(i)} = x, y_t^{(i)} = y\right\}$, $1 \leq t \leq T$
- Consider the linear combination of sources and target

$$Q_{X_T Y_T}^{(\boldsymbol{w})} \triangleq \sum_{t=1}^{T} w_t \hat{P}_{X_t Y_t}$$

with combining weights

$$\boldsymbol{w} \in \{(w_1, w_2, \ldots, w_T) : \sum_{t=1}^{T} w_t = 1, w_t \geq 0\}$$

Characterize the knowledge transferred from each source to the target

- Testing loss

$$L_{\text{test}}^{(\boldsymbol{w})} \triangleq \mathbb{E}\left[\chi^2\left(P_{X_T Y_T}, Q_{X_T Y_T}^{(\boldsymbol{w})}\right)\right]$$

the corresponding optimal weights

$$\boldsymbol{w}^* \triangleq \arg\min_{\boldsymbol{w}} \ L_{\text{test}}^{(\boldsymbol{w})}$$

Definition 1. *For random variable X and Y, given a reference distribution R_{XY} for any distribution P_{XY} and Q_{XY}, the referenced χ^2-distance between them is defined as*

$$\chi_R^2(P_{XY}, Q_{XY}) \triangleq \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \frac{(P_{XY}(x, y) - Q_{XY}(x, y))^2}{R_{XY}(x, y)}. \tag{7}$$

In particular, $\chi_R^2(P_{XY}, Q_{XY})$ becomes Pearson χ^2-distance denoted by $\chi^2(P_{XY}, Q_{XY})$ when $R_{XY} = P_{XY}$.

- Ziyan Zheng, et al. "Multi-source Transfer Learning for Signal Detection over a Fading Channel with Co-channel Interference." *IEEE International Conference on Communications* (2022).

MSTL Framework

Proposition 1 ([15], Theorem 3). *The testing loss (5) for the target task is*

$$L_{test}^{(\boldsymbol{w})} = \chi^2 \left(P_{X_T Y_T}, \sum_{t=1}^{T} w_t P_{X_t Y_t} \right) + \sum_{t=1}^{T} \frac{w_t^2}{N_t} V_t, \quad (8)$$

where V_t is defined as

$$V_t \triangleq \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \frac{P_{X_t Y_t}(x, y)\left(1 - P_{X_t Y_t}(x, y)\right)}{P_{X_T Y_T}(x, y)}. \quad (9)$$

Key factors to determine usefulness of historical data

- **Channel similarity** measured by $\chi^2 \left(P_{X_T Y_T}, \sum_{t=1}^{T} w_t P_{X_t Y_t} \right)$
- **Number of pilots and previous signals** for training, i.e. N_t for $t = 1, 2, ..., T$
- **Complexity of the model** characterized by the domain needed to capture, i.e. $|\mathcal{X}||\mathcal{Y}| - 1$ in V_T

- Ziyan Zheng, et al. "Multi-source Transfer Learning for Signal Detection over a Fading Channel with Co-channel Interference." *IEEE International Conference on Communications* (2022).

[15] X. Tong, X. Xu, S.-L. Huang, and L. Zheng, "A mathematical framework for quantifying transferability in multi-source transfer learning," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

MSTL Framework

When Y_t are continuous:
How the DNN models the distribution?
How to avoid the high dimensionality $|\mathcal{X}||\mathcal{Y}|$?

- Adopt discriminative model

$$\tilde{P}^{(\boldsymbol{f},\boldsymbol{g})}_{X_T|Y_T}(x|y) \triangleq P_{X_T}(x)\left(1 + \boldsymbol{f}^T(x)\,\boldsymbol{g}(y)\right)$$

- Weight $\hat{\boldsymbol{f}}_t$ is defined as

$$\hat{\boldsymbol{f}}_t \triangleq \arg\min_{\boldsymbol{f}} \; \chi^2_{R_{XY}}\left(\hat{P}_{X_tY_t}, P_{Y_T}\tilde{P}^{(\boldsymbol{f},\boldsymbol{g})}_{X_T|Y_T}\right)$$

- The convex combination becomes

$$Q^{(\boldsymbol{w})}_{X_T|Y_T} \triangleq \sum_{t=0}^{T} w_t \tilde{P}^{(\hat{\boldsymbol{f}}_t,\boldsymbol{g})}_{X_T|Y_T} = \tilde{P}^{(\hat{\boldsymbol{f}},\boldsymbol{g})}_{X_T|Y_T} \quad \text{with } \hat{\boldsymbol{f}} \triangleq \textstyle\sum_{t=1}^{T} w_t \hat{\boldsymbol{f}}_t$$

- Testing loss

$$L^{(\boldsymbol{w})}_{\text{test}} \triangleq \mathbb{E}\left[\chi^2_{R_{XY}}\left(P_{X_TY_T}, P_{Y_T}Q^{(\boldsymbol{w})}_{X_T|Y_T}\right)\right]$$

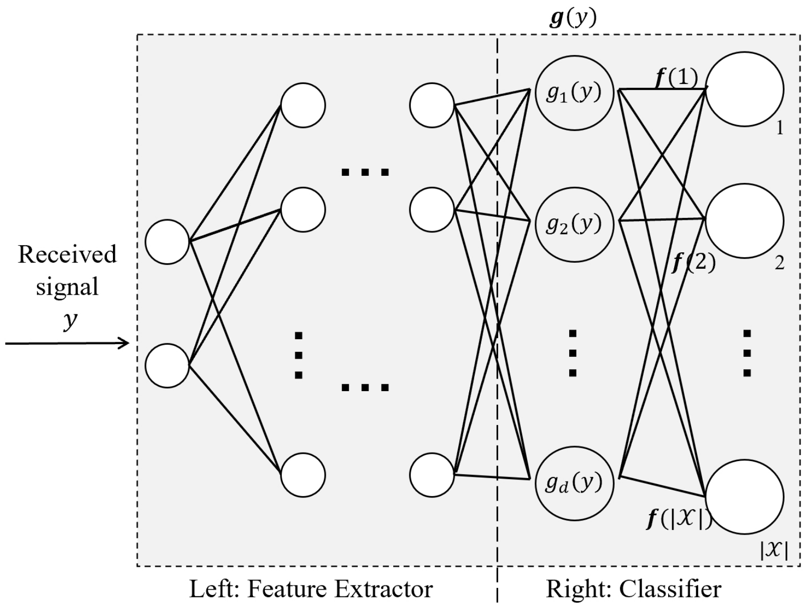


Fig 5. A DNN for classification

Proposition 2 ([15], Theorem 4)**.** *The testing loss (13) associated with the model (12) is*

$$\begin{aligned} L^{(\boldsymbol{w})}_{\textit{test}} &= \chi^2_{R_{XY}}\left(P_{Y_T}\tilde{P}^{(\boldsymbol{f}_T,\boldsymbol{g})}_{X_T|Y_T}, \sum_{t=1}^{T} w_t \tilde{P}^{(\boldsymbol{f}_t,\boldsymbol{g})}_{X_T|Y_T}\right) \\ &\quad + \sum_{t=1}^{T}\frac{w_t^2}{N_t}\tilde{V}_t + \chi^2_{R_{XY}}\left(P_{X_TY_T}, P_{Y_T}\tilde{P}^{(\boldsymbol{f}_T,\boldsymbol{g})}_{X_T|Y_T}\right), \end{aligned} \tag{15}$$

where $\boldsymbol{f}_t \triangleq \arg\min_{\boldsymbol{f}} \chi^2_{R_{XY}}\left(P_{X_tY_t}, P_{Y_T}\tilde{P}^{(\boldsymbol{f},\boldsymbol{g})}_{X_T|Y_T}\right)$ *and* $\tilde{V}_t$ *is a constant independent of* $\boldsymbol{w}$*.*

- Ziyan Zheng, et al. "Multi-source Transfer Learning for Signal Detection over a Fading Channel with Co-channel Interference." *IEEE International Conference on Communications* (2022).

[15] X. Tong, X. Xu, S.-L. Huang, and L. Zheng, "A mathematical framework for quantifying transferability in multi-source transfer learning," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

3 Algorithm

Algorithm for MSTL Detector

Algorithm 1 Algorithm for training the MSTL detector

Input:

Historical samples $\left\{\left(x_t^{(i)} y_t^{(i)}\right)\right\}_{i=1}^{N_t}$ $(t = 1, 2, ..., T-1)$

Pilots $\left\{\left(x_T^{(i)} y_T^{(i)}\right)\right\}_{i=1}^{N_T}$

Randomly initialize $\boldsymbol{w}^*$ and DNN parameters

Output:

1: **repeat**

2: $(\boldsymbol{f}^*, \boldsymbol{g}^*) \leftarrow \arg\min_{\boldsymbol{f},\boldsymbol{g}} L^{(\boldsymbol{w}^*,\boldsymbol{f},\boldsymbol{g})}$

3: $\boldsymbol{w}^* \leftarrow \arg\min_{\boldsymbol{w}} L_{\text{test}}^{(\boldsymbol{w})}$

4: **until** $\boldsymbol{w}^*$ converges

5: $(\boldsymbol{f}^*, \boldsymbol{g}^*) \leftarrow \arg\min_{\boldsymbol{f},\boldsymbol{g}} L^{(\boldsymbol{w}^*,\boldsymbol{f},\boldsymbol{g})}$

6: **return** $\boldsymbol{f}^*, \boldsymbol{g}^*$;

Train neural networks

Solve a non-negative quadratic programming problem

Determine $\hat{x}$ by the MAP decision rule

$$\hat{x}(y) = \arg\max_{x\in\mathcal{X}} \tilde{P}_{X_T|Y_T}^{(\boldsymbol{f}^*,\boldsymbol{g}^*)}(x|y) = \arg\max_{x\in\mathcal{X}} P_{Y_T}(y)\left(1 + \boldsymbol{f}^{*T}(x)\,\boldsymbol{g}^*(y)\right)$$

- Loss function optimizing $(\boldsymbol{f}, \boldsymbol{g})$

$$L^{(\boldsymbol{w},\boldsymbol{f},\boldsymbol{g})} \triangleq \sum_{t=1}^{T} w_t \chi^2_{R_{XY}}\left(\hat{P}_{X_tY_t}, P_{Y_T}\tilde{P}_{X_T|Y_T}^{(\boldsymbol{f},\boldsymbol{g})}\right)$$

- Loss function optimizing $\boldsymbol{w}$

$$L_{test}^{(\boldsymbol{w})} = \chi^2_{R_{XY}}\left(P_{Y_T}\tilde{P}_{X_T|Y_T}^{(\boldsymbol{f}_T,\boldsymbol{g})}, \sum_{t=1}^{T} w_t \tilde{P}_{X_T|Y_T}^{(\boldsymbol{f}_t,\boldsymbol{g})}\right) + \sum_{t=1}^{T}\frac{w_t^2}{N_t}\tilde{V}_t + \chi^2_{R_{XY}}\left(P_{X_TY_T}, P_{Y_T}\tilde{P}_{X_T|Y_T}^{(\boldsymbol{f}_T,\boldsymbol{g})}\right)$$

- Ziyan Zheng, et al. "Multi-source Transfer Learning for Signal Detection over a Fading Channel with Co-channel Interference." *IEEE International Conference on Communications* (2022).

4 Simulation

Setting

$$Y_t = H_t X_t + I_t + n_t$$

◆ Time-varying flat Rayleigh fading channel with co-channel interference

- H_t -- Gauss-Markov process $H_t = aH_{t-1} + u_t, \quad u_t \sim \mathcal{CN}\left(0, \left(1-a^2\right)\sigma_H^2\right)$
- X_t -- QPSK-modulated signals $X_t = K_t + jQ_t, (K_t, Q_t) \in \{(1,1),(-1,1),(-1,-1),(1,-1)\}$
- I_t -- QPSK-modulated interference $I_t \in \{q+jq, -q+jq, -q-jq, q-jq\}$ for some constant q
- n_t -- AWGN with variance σ_n^2

◆ Update process of the MSTL detector

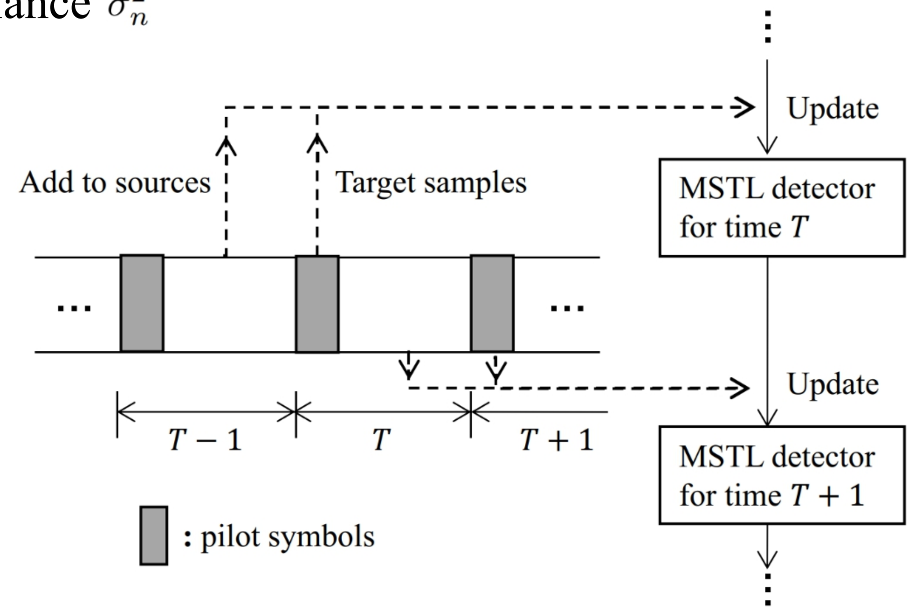


Fig 6. The update process of the detector

- Ziyan Zheng, et al. "Multi-source Transfer Learning for Signal Detection over a Fading Channel with Co-channel Interference." *IEEE International Conference on Communications* (2022).

Simulation

◆ Comparisons

- FCDNN-1 Conventional fully-connected DNN based on L_2 loss, using N_T pilots for training
- FCDNN-2 Using N_T pilots and $\sum_{t=T-9}^{T-1} N_t$ reserved samples for training
- LMMSE Estimate CSI by N_T pilots

◆ Implementation

- Measure of interference level

$$\mathrm{SINR} \triangleq 10\lg \frac{\sigma_H^2 |X_t|^2}{\sigma_n^2}$$

  Set $q \propto \sigma_n$ for the consistency of interference and noise

- For each task time T, $N_t = 200$ for $t = T-9,\ T-8,\ \dots,\ T-1$, discard samples before $t = T-10$

- We vary: (i) Fading coefficient a
  (ii) Number of pilots N_T
  (iii) SINR

● Ziyan Zheng, et al. "Multi-source Transfer Learning for Signal Detection over a Fading Channel with Co-channel Interference." *IEEE International Conference on Communications* (2022).

Simulation

◆ Results

- Largest gain occurs in the range $a \in [0.9875,\ 0.9975]$

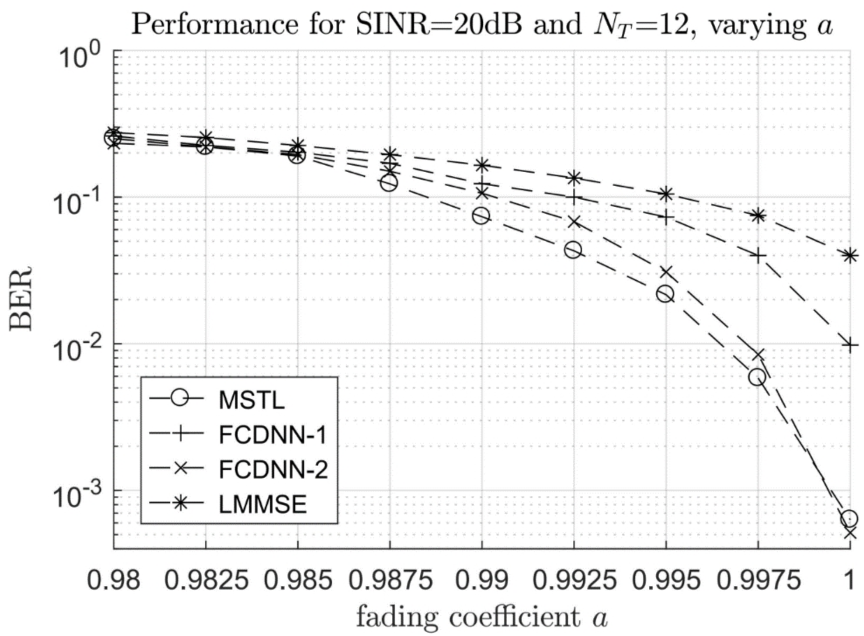


Fig 7

- Low BER for high SINR

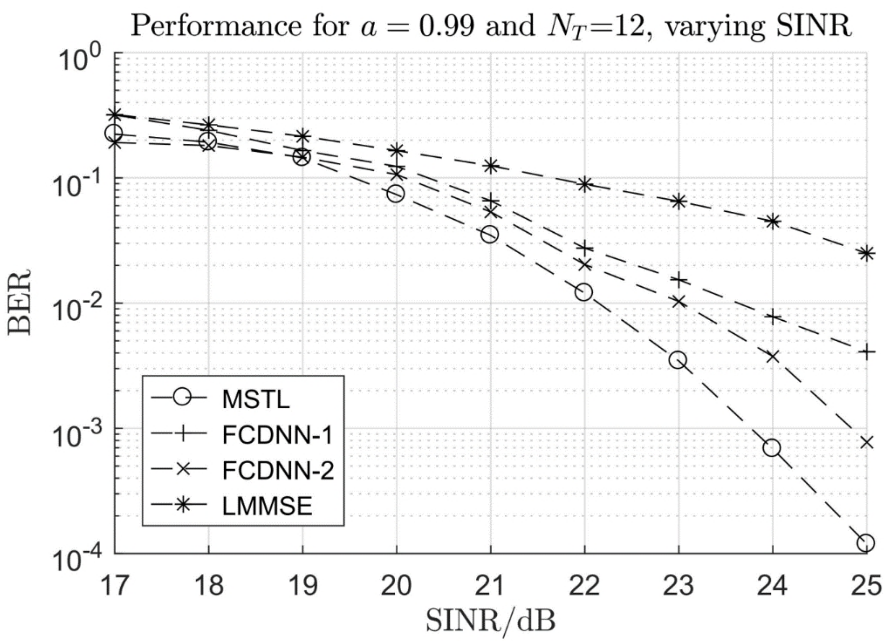


Fig 8

- Ziyan Zheng, et al. "Multi-source Transfer Learning for Signal Detection over a Fading Channel with Co-channel Interference." *IEEE International Conference on Communications* (2022).

Simulation

◆ Results

- Outperforms other algorithms in a large range of pilots length

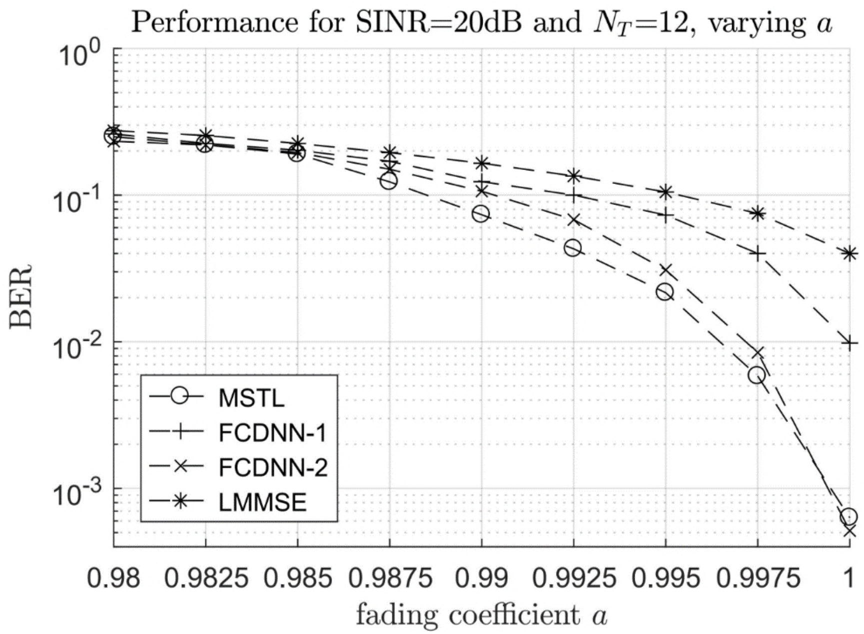


Fig 9

- Monotonous similarity of the sources

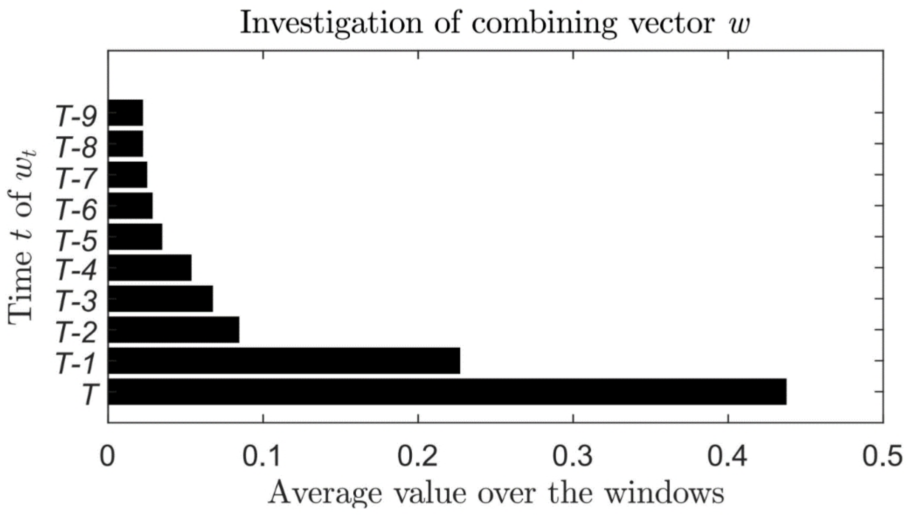


Fig 10

- Ziyan Zheng, et al. "Multi-source Transfer Learning for Signal Detection over a Fading Channel with Co-channel Interference." *IEEE International Conference on Communications* (2022).

5 Conclusion

Main contributions

- We adopt a mathematical framework of MSTL and develop a detection algorithm for time-varying channels. Our MSTL-based detector involves pilots and much previous data in the training scheme to mitigate the effect of interference, without the direct estimation for CSI.
- Unlike most of the previous works that either ignored interference or treated interference as Gaussian noise, non-Gaussian-distributed co-channel interference is taken into consideration in the channel model.
- The proposed MSTL detector is implemented in the signals stream over a simulated time-varying fading channel in the presence of co-channel interference, in which we show that our algorithm outperforms both the traditional method and conventional neural networks without the design of knowledge transfer in a variety of scenarios.

- Ziyan Zheng, et al. "Multi-source Transfer Learning for Signal Detection over a Fading Channel with Co-channel Interference." *IEEE International Conference on Communications* (2022).

References

[1] Z. Guo and P. Nilsson, "Algorithm and implementation of the k-best sphere decoding for mimo detection," *IEEE Journal on selected areas in communications*, vol. 24, no. 3, pp. 491–503, 2006.

[2] S. Wu, L. Kuang, Z. Ni, J. Lu, D. Huang, and Q. Guo, "Low-complexity iterative detection for large-scale multiuser mimo-ofdm systems using approximate message passing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 902–915, 2014.

[3] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Model-driven deep learning for mimo detection," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1702–1715, 2020.

[4] S. Park, H. Jang, O. Simeone, and J. Kang, "Learning to demodulate from few pilots via offline and online meta-learning," *IEEE Transactions on Signal Processing*, vol. 69, pp. 226–239, 2020.

[5] O. Simeone, "A very brief introduction to machine learning with applications to communication systems," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, no. 4, pp. 648–664, 2018.

[6] S. Bouchired, D. Roviras, and F. Castanié, "Equalisation of satellite mobile channels with neural network techniques," *Space Communications*, vol. 15, no. 4, pp. 209–220, 1998.

[7] C. Liu, Y. Chen, and S.-H. Yang, "Signal detection with co-channel interference using deep learning," *Physical Communication*, vol. 47, p. 101343, 2021.

[8] S. Catreux, P. F. Driessen, and L. J. Greenstein, "Attainable throughput of an interference-limited multiple-input multiple-output (mimo) cellular system," *IEEE transactions on communications*, vol. 49, no. 8, pp. 1307–1311, 2001.

[9] V. S. Annapureddy and V. V. Veeravalli, "Gaussian interference networks: Sum capacity in the low-interference regime and new outer bounds on the capacity region," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3032–3050, 2009.

[10] A. Carleial, "A case where interference does not reduce capacity (corresp.)," *IEEE Transactions on Information Theory*, vol. 21, no. 5, pp. 569–570, 1975.

[11] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in ofdm systems," *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 114–117, 2017.

[12] N. Farsad and A. Goldsmith, "Neural network detection of data sequences in communication systems," *IEEE Transactions on Signal Processing*, vol. 66, no. 21, pp. 5663–5678, 2018.

[13] S. Thrun, L. K. Saul, and B. Schölkopf, *Advances in Neural Information Processing Systems 16: Proceedings of the 2003 Conference*, vol. 16. MIT press, 2004.

[14] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *International conference on artificial neural networks*, pp. 270–279, Springer, 2018.

[15] X. Tong, X. Xu, S.-L. Huang, and L. Zheng, "A mathematical framework for quantifying transferability in multi-source transfer learning," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

[16] S. Rendle, "Factorization machines," in *2010 IEEE International conference on data mining*, pp. 995–1000, IEEE, 2010.

[17] O. Levy and Y. Goldberg, "Neural word embedding as implicit matrix factorization," *Advances in neural information processing systems*, vol. 27, pp. 2177–2185, 2014.

[18] M. Dong, L. Tong, and B. M. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 5, pp. 1403–1418, 2004.

[19] G. A. Arredondo, W. H. Chriss, and E. H. Walker, "A multipath fading simulator for mobile radio," *IEEE Transactions on Vehicular Technology*, vol. 22, no. 4, pp. 241–244, 1973.

[20] R. A. Iltis, "Joint estimation of pn code delay and multipath using the extended kalman filter," *IEEE Transactions on communications*, vol. 38, no. 10, pp. 1677–1685, 1990.

[21] M. Stojanovic, J. G. Proakis, and J. A. Catipovic, "Analysis of the impact of channel estimation errors on the performance of a decision-feedback equalizer in fading multipath channels," *IEEE Transactions on Communications*, vol. 43, no. 2/3/4, pp. 877–886, 1995.

[22] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Transactions on Information theory*, vol. 46, no. 3, pp. 933–946, 2000.

[23] D. Qiu, *Representation and transfer learning using information-theoretic approximations*. PhD thesis, Massachusetts Institute of Technology, 2020.

[24] H. Kim, Y. Jiang, R. Rana, S. Kannan, S. Oh, and P. Viswanath, "Communication algorithms via deep learning," *arXiv preprint arXiv:1805.09317*, 2018.

[25] S. Verdu *et al.*, *Multiuser detection*. Cambridge university press, 1998.

Thanks !