Uncertainties in Identification Systems

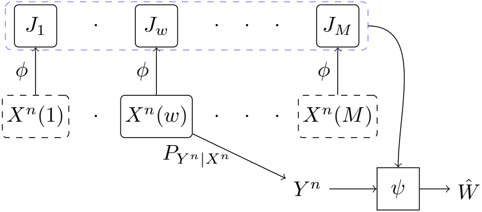
Minh Thanh Vu[1], Tobias Oechtering[1],
Mikael Skoglund[1] and Holger Boche[2]

[1]Dept. ISE, KTH Royal Institute of Technology

[2]LTI, Technische Universität München

ISIT June 22, 2018

Identification Systems Recap



- *Enrollment Phase:* Data from each user is compressed and stored via

$$\phi_n\colon \mathcal{X}^n \to \mathcal{M}_1.$$

- *Identification Phase:*
  - Observation sequence y^n is related to one randomly chosen user in the system.
  - The processing center searches for the true user

$$\psi_n\colon \mathcal{Y}^n \times \mathcal{M}_1^M \to \mathcal{W} \cup \{e\}.$$

- The pair (ϕ_n, ψ_n) is called an ID-code.

Relevant previous results for given P_X and $P_{Y|X}$

- Requirements for an arbitrary $\delta > 0$

$$\frac{1}{n}\log|\mathcal{M}_1| \leq R_c + \delta,\ \frac{1}{n}\log M \geq R_i - \delta,$$
$$\Pr\{W \neq \hat{W}\} \leq \delta,\ \forall n \geq n_0(\delta).$$

- Willems[1] *et.al* gave the characterization of identification capacity

$$C = I(X;Y).$$

- Tuncel[2] studied the identification-compression trade-off $\mathcal{R}_{\mathrm{ID}}$

$$R_c \geq I(X;U),\ R_i \leq I(Y;U)$$
$$Y - X - U,\ |\mathcal{U}| \leq |\mathcal{X}| + 1.$$

---

[1] F. Willems, T. Kalker, and J.-P. Linnartz, "On the capacity of a biometrical identification system," in *ISIT 2003*

[2] E. Tuncel, "Capacity/storage tradeoff in high-dimensional identification systems," *T-IT*, vol. 55, no. 5.

Related works

Compression of data

- E. Tuncel and D. Gündüz, "Identification and lossy reconstruction in noisy databases," *IEEE Trans. Inf. Theory*, vol. 60
- M. B. Westover and J. A. O'Sullivan, "Achievable rates for pattern recognition," *IEEE Trans. Inf. Theory*, vol. 54

Arbitrary/Compound settings

- D. Blackwell, L. Breiman, and A. Thomasian, "The capacity of a class of channels," *The Annals of Mathematical Statistics*, 1959.
- —"The capacities of certain channel classes under random coding," *The Annals of Mathematical Statistics*, 1960.
- R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Probability Theory and Related Fields*, vol. 44.

Uncertainties

Motivation:

- Knowing the exact users' data distribution and the observation channel is restrictive.
- We relax this assumption by assume that the data distribution comes from the set

$$\mathcal{P}_s = \{P_{X|S=s},\ s \in \mathcal{S}\}$$

  and the channel is selected by nature from

$$\mathcal{P}_c = \{P_{Y|X,\tau},\ \tau \in \mathcal{T}\}.$$

- $\mathcal{S}$ can be the set of different locations, while $\mathcal{T}$ could be the set of different collecting methods.

Compound source-channel

- We assume that all users' data are generated from the same but unknown distribution, i.e., $X^n(i) \sim P_{X|S=s}^{\otimes n}$ for all $i \in [1:M]$.
- The channel is given by $P_{Y|X,\tau}^{\otimes n}$ for an unknown τ.
- A pair (R_c, R_i) is *achievable* if

$$\frac{1}{n}\log|\mathcal{M}_1| < R_c + \delta, \frac{1}{n}\log M > R_i - \delta,$$
$$\sup_{\tau,s} \mathsf{Pr}\{W \neq \hat{W}|S = s, T = \tau\} < \delta,$$

for all sufficiently large n. The set of all achievable rate pairs is denoted by $\mathcal{R}_{sc}$.

Compound source-channel

Theorem

In case both $\mathcal{X}$ and $\mathcal{Y}$ are finite, $\mathcal{R}_{sc}$ is given by set of all (R_c, R_i) such that

$$R_c \geq \max_s I(X;U|S=s)$$
$$R_i \leq \min_{s,\tau} I(Y;U|S=s,T=\tau),$$

where $P_{XYU|T=\tau,S=s} = P_{Y|X,\tau} \times P_{X|S=s} \times P_{U|X,S=s}$ and $|\mathcal{U}| \leq |\mathcal{X}| + |\mathcal{T}|$.

Compound source-channel: Coding scheme

- **Enrollment:**
  - State estimation for each user $\hat{s}_i = T_X(x^n(i))$. It can be shown that

$$\Pr\{\hat{S}_i = s | S = s\} \to 1, \forall s \in \mathcal{S}.$$

  - Compress $x^n(i)$ into $u^n(m_{\hat{s}_i,i})$ using random coding. Store $(\hat{s}_i, m_{\hat{s}_i,i})$ in the database. We need $R_c > \max_s I(X;U|S=s)$.
- **Identification:**
  - Estimate the underlying observation state $\hat{\kappa} = T_Y(y^n)$ and deduce the channel state $\hat{\tau}$ with high probability.
  - Since all $x^n(i)$ are generated from $P_{X|S=s}^{\otimes n}$, set $\hat{s} = s_1$.
  - Search for a unique $\hat{w}$ such that

$$(y^n, u^n(m_{\hat{s}_{\hat{w}},\hat{w}})) \in \mathcal{T}_\epsilon^n(P_{YU|T=\hat{\tau},S=\hat{s}}).$$

We need $R_i < \min_{s,\tau} I(Y;U|S=s,T=\tau)$.

Individual state varying

- Each user has its own state s_i and the data are generated according to $X^n(i) \sim P_{X|S=s_i}^{\otimes n}$ for all $i \in [1:M]$.
- We assume that the channel $P_{Y|X}$ is fixed.
- A pair (R_c, R_i) is *achievable* if

$$\frac{1}{n}\log|\mathcal{M}_1| < R_c + \delta,\ \frac{1}{n}\log M > R_i - \delta,$$
$$\sup_{(s_i)_{i=1}^M \in \mathcal{S}^M} \Pr\{W \neq \hat{W} | (S_i)_{i=1}^M = (s_i)_{i=1}^M\} < \delta,$$

  for all sufficiently large n. We denote the set of all achievable rate pairs by $\mathcal{R}_{iis}$.
- The number of constraints is exponential.

Individual state varying

Theorem

For finite alphabets $\mathcal{R}_{iis}$ *is the collection of* (R_c, R_i) *such that*

$$R_c \geq \max_s I(X;U|S=s)$$
$$R_i \leq \min_s I(Y;U|S=s),$$

where $P_{XYU|S=s} = P_{Y|X} \times P_{X|S=s} \times P_{U|X,S=s}$ *and* $|\mathcal{U}| \leq |\mathcal{X}| + 1$.

- We observe that when $|\mathcal{T}| = 1$, $\mathcal{R}_{sc} = \mathcal{R}_{iis}$.
- $\mathcal{R}_{iis}$ is a convex set.
- *Proof Sketch:* We observe that $\mathcal{R}_{iis} \subseteq \mathcal{R}_{sc}$ due to the independent of users. The achievability follows from the same coding arguments as in the proof of $\mathcal{R}_{sc}$.

A connection to WAK-network

- **Motivation:** The achievability proofs of $\mathcal{R}_{sc}$ and $\mathcal{R}_{iis}$ when $|\mathcal{T}| = 1$ are shown by similar random coding arguments. Perhaps we can show them for the large class.
- Recap WAK-network: Assume that $(X^n, Y^n) \sim P_{XY}^{\otimes n}$. A WAK-code consists of
  - Encoding and decoding mapping

$$\phi_{1n} \colon \mathcal{X}^n \to \mathcal{M}_1,\, \phi_{2n} \colon \mathcal{Y}^n \to \mathcal{M}_2$$
$$\psi_n \colon \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{Y}^n$$

  - Requirement

$$\Pr\{Y^n \neq \hat{Y}^n\} \to 0, \text{ as } n \to \infty.$$

- Denote the trade-off region by $\mathcal{R}_{\mathrm{WAK}}$.

A connection to WAK-network

WAK-region	ID-region
$R_1 \geq I(X;U)$, $R_2 \geq H(Y\|U)$	$R_c \geq I(X;U)$, $R_i \leq I(Y;U)$
$Y - X - U$	$Y - X - U$

- It can be seen that

$$(R_c, R_i) \in \mathcal{R}_{\text{ID}} \Leftrightarrow (R_c, H(Y) - R_i) \in \mathcal{R}_{\text{WAK}}.$$

- Our observation:

Proposition

From a WAK-code $(\phi_{1n}, \phi_{2n}, \psi_n)$ *we can construct a corresponding* (ϕ_{1n}, ψ_n') *ID-code such that*

$$\begin{aligned}\Pr\{\hat{W} \neq w | W = w\} \leq P_{\text{WAK}}\{\text{error}\} + \Pr\{Y^n \notin \mathcal{A}_\gamma^n\} \\ + M|\mathcal{M}_2|e^{-n(H(Y)-\gamma)},\end{aligned}$$

where $\gamma > 0$ *and* $\mathcal{A}_\gamma^n$ *is the weak typical set.*

A connection to WAK-network

- In case $|\mathcal{T}| = 1$ the code constructed from an arbitrary collection $\{\phi_{1n,s}, \phi_{2n,s}, \psi_{n,s}\}_{s \in \mathcal{S}}$ performs similarly for both settings.
- **The keys**
  - Users' data are independent conditioning on the underlying states
  - vanishing estimation error probability.
- We also show that: Given an ID-code (ϕ_n, ψ_n) there exists a WAK-code $(\phi_n, \phi'_{2n}, \psi'_n)$ such that

$$\Pr\{\hat{Y}^n \neq Y^n\} \leq P_{\mathrm{ID}}(\mathrm{error}) + e^{-n\gamma} + \Pr\{Y^n \notin \mathcal{A}_\gamma^n\} + e^{-nR_2} e^{n(H(Y)+3\gamma-R_i)}.$$

- Combining both directions we have

Theorem

Suppose $(R_a, R_b) \in \mathbb{R}_+^2$ *with* $R_b \leq H(Y)$ *then*

$$(R_a, R_b) \in \mathcal{R}_{\mathrm{WAK},\epsilon} \Leftrightarrow (R_a, H(Y) - R_b) \in \mathcal{R}_{\mathrm{ID},\epsilon},$$

Summary

- We studied models for the uncertainties in identification systems.
- We showed a connection between the identification problem and the WAK problem.

Extensions:

- We also study several mixture models with countable state space.
- An argument for the strong converse of the ID problem which can be transferred to the Gaussian case is investigated.

Thank you for your attention!