

Codage Distribué dans des Réseaux de Capteurs avec Connaissance Incertaine des Corrélations

Elsa DUPRAZ¹, Aline ROUMY², Michel KIEFFER¹

¹L2S - CNRS - SUPELEC - Univ Paris-Sud, 3, rue Joliot Curie, 91192 Gif-sur-Yvette, France

²INRIA, Campus de Beaulieu, 35042 Rennes, France

elsa.dupraz@lss.supelec.fr, aline.roumy@inria.fr, michel.kieffer@lss.supelec.fr

Thème – 2.1 - Théorie de l'information, 2.7 - Coopération, Diversité, Réseaux de capteurs, 2.8 - Systèmes de communications

Problème traité – Cet article considère le problème de la compression de données sans perte dans un réseau de capteurs.

Originalité – Les schémas classiques de compression dans les réseaux de capteurs supposent que les corrélations entre les mesures prélevées par les capteurs sont parfaitement connues. Ici, nous prenons en compte une possible incertitude sur les niveaux de ces corrélations et modélisons cette incertitude par un paramètre θ inconnu à valeurs dans un ensemble connu.

Résultats – Nous comparons plusieurs stratégies de codage : 1) codage conjoint, où une coopération entre capteurs est autorisée 2) codage séparé, où la corrélation entre les sources est prise en compte, sans coopération entre capteurs, et en tenant compte des incertitudes sur les corrélations 3) sans coopération entre capteurs mais avec une séquence d'apprentissage pour que le récepteur estime la corrélation entre les capteurs. Les coûts de codage sont évalués analytiquement pour chaque stratégie en fonction des coûts de transmission entre les nœuds du réseau.

1 Introduction

Considérons un réseau de capteurs dans lequel les capteurs doivent transmettre régulièrement leurs mesures à un point de collecte dont le but est de reconstruire l'ensemble des informations collectées dans le réseau. Un certain nombre de solutions pratiques ont été proposées dans le cas où la distribution de probabilité jointe entre les sources est parfaitement connue [1, 5]. Dans ce cas, le codage séparé des sources est aussi efficace en terme de débit qu'un codage conjoint [4]. Pour le réseau de capteurs, on choisit donc en général d'effectuer un codage séparé, ce qui évite aux capteurs d'avoir à dépenser du débit pour communiquer entre eux. Malheureusement, la distribution jointe est souvent difficile à caractériser avec précision, puisqu'elle peut dépendre de grandeurs incertaines : position relative des capteurs, variations de l'environnement, etc.

Cet article considère un réseau à trois nœuds : deux capteurs \mathcal{X} , \mathcal{Y} et un récepteur \mathcal{S} . On suppose que la distribution jointe est paramétrée par un vecteur θ inconnu, qui peut varier d'une transmission à l'autre. Il s'agit d'un modèle non ergodique. On montre que pour ce modèle, le codage conjoint peut être plus efficace que le codage séparé et on explique comment choisir l'une ou l'autre des méthodes en fonction des paramètres du modèle et du réseau. De plus, nous proposons également l'analyse d'une solution mettant en œuvre des séquences d'apprentissage pour estimer θ .

2 Modèle de sources et de réseau

Les capteurs \mathcal{X} et \mathcal{Y} sont associés à des sources discrètes X et Y qui génèrent des suites de symboles $\{(X_n, Y_n)\}_{n=1}^{+\infty}$ indépendants et identiquement distribués. La loi jointe est paramétrée par un vecteur θ avec $P(X, Y|\theta) = P(X)P(Y|X, \theta)$. Le vecteur θ est inconnu mais fixé pour une suite de symboles $\{(X_n, Y_n)\}_{n=1}^{+\infty}$. Il appartient à un ensemble \mathcal{P}_θ connu. En revanche, aucune distribution de probabilité *a priori* n'est disponible pour θ . Le récepteur \mathcal{S} doit reconstruire $\{(X_n, Y_n)\}_{n=1}^{+\infty}$ sans perte.

\mathcal{X} , \mathcal{Y} et \mathcal{S} peuvent communiquer entre eux selon le schéma de communication représenté sur la Figure 1. Les arêtes représentent les liens de communication entre les nœuds et sont associées à des poids μ_1, μ_2, μ_3 . Le coût de transmission à un débit R sur le

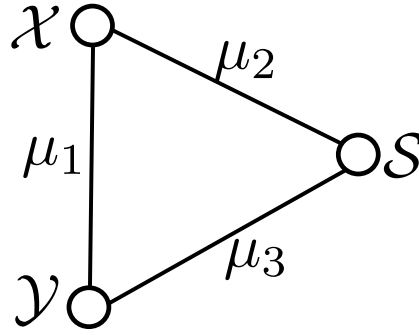


FIGURE 1 – Modèle de réseau

lien i est $\mu_i R$. La détermination des μ_i résulte des problèmes de codage canal associés au réseau, nous ne les analysons pas ici. En revanche, nous nous intéressons au problème de codage des sources et évaluons le coût des transmissions point à point pour des μ_i fixés.

3 Codage conjoint ou Codage séparé

Nous proposons ici une analyse asymptotique, c'est-à-dire que l'on considère le codage des suites de symboles $\{X_n\}_{n=1}^{+\infty}$ et $\{Y_n\}_{n=1}^{+\infty}$. Quelque soit le type de codage considéré, nous supposons que le débit moyen par symbole est donné par le débit moyen minimum atteignable au sens de la théorie de l'information. Même si ces débits théoriques ne sont pas toujours atteignables, ils nous permettent de comparer deux stratégies de codage conjoint et deux stratégies de codage séparé.

Avec la première stratégie de codage *conjoint*, \mathcal{X} transmet $\{X_n\}_{n=1}^{+\infty}$ à \mathcal{Y} et à \mathcal{S} à un débit par symbole $H(X)$, et \mathcal{Y} transmet $\{Y_n\}_{n=1}^{+\infty}$ à \mathcal{S} en utilisant $\{X_n\}_{n=1}^{+\infty}$ comme information adjacente disponible à la fois au codeur et au décodeur. Le codage de X avec l'information adjacente Y présente à la fois au codeur et au décodeur se fait à un débit $H(X|Y, \theta)$, où θ est le vrai paramètre. Avec la deuxième stratégie de codage conjoint, les rôles de X et Y sont simplement inversés. On note $m_c^{(X)}(\theta)$ et $m_c^{(Y)}(\theta)$ les coûts moyens par symbole correspondant aux deux stratégies pour un θ donné.

Dans la première stratégie de codage *séparé*, \mathcal{X} transmet $\{X_n\}_{n=1}^{+\infty}$ uniquement à \mathcal{S} , et \mathcal{Y} transmet $\{Y_n\}_{n=1}^{+\infty}$ à \mathcal{S} en utilisant $\{X_n\}_{n=1}^{+\infty}$ comme information adjacente disponible seulement au décodeur. D'après [2] le débit par symbole nécessaire à cette dernière opération est donné par $\sup_{\theta \in \mathcal{P}_\theta} H(X|Y, \theta)$. Dans la deuxième stratégie de codage séparé, les rôles de X et Y sont à nouveau inversés. D'après [1], si $\mu_3 \geq \mu_2$, la première stratégie de codage séparé donne toujours un coût plus faible que la deuxième. On considère donc ici le cas $\mu_3 \geq \mu_2$ et on note $m_s^{(X)}(\theta)$ le coût moyen par symbole associé à la première stratégie. L'analyse s'adapte directement si l'inégalité est inversée. On remarque également que ces deux stratégies correspondent à des sommets de la région des débits atteignables [4]. Ce sont en effet les seules qui peuvent être optimales pour ce modèle de réseau [1].

On a alors

$$m_c^{(X)}(\theta) = (\mu_1 + \mu_2)H(X) + \mu_3H(Y|X, \theta) \quad (1)$$

$$m_c^{(Y)}(\theta) = (\mu_1 + \mu_3)H(Y|\theta) + \mu_2H(X|Y, \theta) \quad (2)$$

$$m_s^{(X)}(\theta) = \mu_2H(X) + \mu_3 \sup_{\theta \in \mathcal{P}_\theta} H(Y|X, \theta). \quad (3)$$

Une stratégie est optimale si le sous-ensemble de paramètres θ pour lesquels le coût de transmission est minimal (parmi toutes les stratégies), est de mesure maximale. Pour déterminer la stratégie optimale, nous les comparons deux à deux. Pour comparer les deux stratégies de codage conjoint, considérons la différence de coût $\Delta_c(\theta) = m_c^{(X)}(\theta) - m_c^{(Y)}(\theta)$. Soit \mathcal{P}_θ^c l'ensemble des θ pour lesquels $\Delta_c(\theta) \geq 0$ et soit $\gamma_c = 1 - \int_{\mathcal{P}_\theta^c} d\theta / \int_{\mathcal{P}_\theta} d\theta$ la mesure associée à \mathcal{P}_θ^c . On conserve la stratégie où Y est information adjacente si $\gamma_c > 0.5$, on conserve l'autre stratégie sinon. Ensuite, on compare la stratégie conservée à la stratégie de codage séparé. Par exemple, si $\gamma_c > 0.5$, on calcule $\Delta_{cs}(\theta) = m_c^{(Y)}(\theta) - m_s^{(X)}(\theta)$, on définit \mathcal{P}_θ^{cs} comme l'ensemble des θ tels que $\Delta_{cs}(\theta) \geq 0$ et on note γ_{cs} la mesure associée. On choisit l'une ou l'autre des stratégies en conséquence. On observe par exemple que lorsque le rapport $\mu_1/\mu_3 \ll 1$ ou $\mu_1/\mu_2 \ll 1$, la meilleure stratégie est systématiquement le codage conjoint.

Remarquons que si l'on souhaite considérer le cas du broadcast, les expressions des coûts sont modifiées. Par exemple, la transmission de $\{Y_n\}_{n=1}^{+\infty}$ à \mathcal{X} et \mathcal{S} a un coût $\max(\mu_1, \mu_3)H(Y)$.

4 Estimation de paramètres

Nous avons montré dans [2] que la connaissance d'une estimée $\hat{\theta}$ de θ au décodeur pouvait réduire le temps de décodage et améliorer sa précision, à condition que l'estimation soit suffisamment précise. Nous étudions donc l'intérêt pour les sources de transmettre au préalable des séquences d'apprentissage, pour permettre au décodeur d'estimer θ . Il s'agit ici de caractériser l'augmentation de débit qui serait due au choix d'une stratégie avec séquence d'apprentissage. Dans cette partie, nous supposons que la stratégie optimale est la solution séparée, *i.e.*, μ_1 est grand devant μ_2 et μ_3 . De plus, nous considérons le codage de séquences $\{X_n\}_{n=1}^N$ et $\{Y_n\}_{n=1}^N$ de longueurs finies N . Sans perte de généralité, nous supposons que $\mu_3 \geq \mu_2$ c'est-à-dire que X est l'information adjacente. L'analyse est la même dans l'autre cas.

On étudie deux stratégies possibles. Soit $\{u_N\}_{N \in \mathbb{N}}$ une suite d'entiers tels que $\lim_{N \rightarrow \infty} u_N = +\infty$ et $\lim_{N \rightarrow \infty} \frac{u_N}{N} = 0$. Pour la stratégie *avec* séquence d'apprentissage, \mathcal{X} transmet $\{X_n\}_{n=1}^N$ à S à un débit $H(X)$, \mathcal{Y} transmet une séquence d'apprentissage $\{Y_n\}_{n=1}^{u_N}$ de longueur u_N à S à un débit $H(Y)$ puis transmet les $\{Y_n\}_{n=u_N+1}^N$ restants en utilisant $\{Y_n\}_{n=u_N+1}^N$ comme information adjacente présente uniquement au décodeur. La stratégie *sans* séquence d'apprentissage correspond à la stratégie de codage séparé décrite dans la section précédente. On note $m_1(\theta)$ et $m_{wl}(\theta)$ les coûts moyens par symbole correspondant à chaque stratégie pour un θ donné, et on a

$$m_1(\theta) = \mu_2 H(X) + \mu_3 \left(\frac{u_N}{N} H(Y|\theta) + \frac{N - u_N}{N} \sup_{\theta \in \mathcal{P}_\theta} H(Y|X, \theta) \right) \quad (4)$$

$$m_{wl}(\theta) = \mu_2 H(X) + \mu_3 \sup_{\theta \in \mathcal{P}_\theta} H(Y|X, \theta). \quad (5)$$

La solution avec séquence d'apprentissage implique un coût plus important à longueur N finie, mais les deux solutions sont équivalentes quand $N \rightarrow \infty$. A longueur finie, soit $\Delta(u_N, \theta)$ la perte en débit de la solution avec séquence d'apprentissage en fonction de u_N et de θ :

$$\Delta(u_N, \theta) = \frac{m_1 - m_{wl}}{\mu_3} = \frac{u_N}{N} \left(H(Y|\theta) - \sup_{\theta \in \mathcal{P}_\theta} H(Y|X, \theta) \right). \quad (6)$$

La perte ne dépend pas des paramètres μ_i . On suppose que l'estimée de θ est obtenue avec un estimateur au sens du maximum de vraisemblance pour lequel la variance de l'erreur d'estimation est inférieure à K/u_N , où K est le nombre de paramètres à estimer [3, Chap. 7]. Pour une contrainte de qualité d'estimation $K/u_N \leq \epsilon$ fixée, on a alors

$$\Delta(u_N, \theta) \geq \frac{\epsilon}{KN} \left(\sup_{\theta \in \mathcal{P}_\theta} H(Y|X, \theta) - H(Y|\theta) \right). \quad (7)$$

Pour finir, on peut expérimentalement déterminer la valeur de ϵ qui garantit un décodage correct, et obtenir u_n et $\Delta(u_N)$ à partir de cette valeur.

Références

- [1] R. Cristescu, B. Beferull-Lozano, and M. Vetterli. Networked Slepian-Wolf : theory, algorithms, and scaling laws. *IEEE Transactions on Information Theory*, 51(12) :4057–4073, Dec 2005.
- [2] E. Dupraz, A. Roumy, and M. Kieffer. Source coding with side information at the decoder : models with uncertainty, performance bounds, and practical coding schemes. In *International Symposium on Information Theory and its Applications, 2012*.
- [3] S.M. Kay. *Fundamentals of Statistical Signal Processing, Estimation theory*. Prentice Hall PTR, 1993.
- [4] D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19(4) :471–480, July 1973.
- [5] Z. Xiong, A.D. Liveris, and S. Cheng. Distributed source coding for sensor networks. *IEEE Signal Processing Magazine*, 21(5) :80–94, Sep 2004.