Combining Information Sources Using Minimum Cross-Entropy Principle

Vladimíra Sečkárová[1,2]

Second Bayesian Young Statisticians Meeting (BAYS 2014)
Vienna, September 18–19, 2014

[1] Department of Adaptive Systems, Institute of Information Theory and Automation of the ASCE, Prague, Czech Republic
seckarov@utia.cas.cz

[2] Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University in Prague, Czech Republic

Abstract

Combining information sources, exploited in many areas of mathematics, has been an open problem for decades. It was approached from many point of views and different methods for assignment of the weights to the sources were developed. Some of them prefer subjective influence. Sometimes a reduction of the space of sources is used. We propose a method based on tools of information theory, i.e. the minimum cross-entropy principle and the Kullback-Leibler divergence. These help us eliminate the problem of subjectivity and allow us compute the weight for each considered source.

Keywords: Kullback-Leibler divergence; constrained nonlinear optimization;

1 Introduction

Combining information sources has received much attention in the past few decades in many fields of mathematics, e.g. statistical inference in sensor networks and opinion pooling in market scenarios. Towards the biggest issue, the assignment of weights to the sources (sensors, experts) many methods were developed. Here, we are particularly interested in the amount of subjectivity included in assigned weights. In order to eliminate the subjectivity we propose a combining procedure based on the tools of information theory applied before, e.g., in reduction of the space of sources [2]. Rather than reducing the space of information sources we explore the divergences between the sources, i.e. the Kullback-Leibler divergence [3], and principles ensuring we obtain one particular solution, i.e. the minimum cross-entropy principle [4].

2 Combining Information Sources

Let us have an unknown n-dimensional probability vector q and a finite number of information sources, $j = 1, \ldots, s$. Assume also that each of them provides a probability vector

p_j, an observation about an unknown probability vector q ($p_{ji}, q_i > 0$ and $\sum_{i=1}^n p_{ji} = 1$, $\sum_{i=1}^n q_i = 1$, $j = 1, \ldots, s$).

We search for the estimate $\hat{q}$ of q satisfying the following (see [1]):

$$\hat{q} = \arg\min_{\tilde{q}} \mathrm{E}_{\pi(q|p_1,\ldots,p_s)}(\mathrm{KLD}(q||\tilde{q})|p_1, \ldots, p_s)$$

where E(.|.) is the conditional expected value conditioned on $p_1, \ldots, p_s$, $\pi(.|.)$ is a posterior probability density function (pdf) conditioned on $p_1, \ldots, p_s$ and KLD(.||.) is the Kullback-Leibler divergence

$$\mathrm{KLD}(q||\tilde{q}) = \sum_{i=1}^{n} q_i \ln \frac{q_i}{\tilde{q}_i}. \tag{1}$$

The minimizing element is the conditional expectation of q conditioned on $p_1, \ldots, p_s$

$$\hat{q} = \mathrm{E}_{\pi(q|p_1,\ldots,p_s)}(q|p_1, \ldots, p_s) \tag{2}$$

with respect to yet unspecified posterior pdf $\pi(q|p_1, \ldots, p_s)$.

2.1 Posterior probability density determined by minimum cross-entropy principle

To obtain the posterior pdf $\pi(q|p_1, \ldots, p_s)$ we exploit the minimum cross-entropy principle. This principle states that from the set of all possible pdfs, satisfying additional constraints on expected values of q, we should choose the one minimizing the following function:

$$\int \pi(q|p_1, \ldots, p_s) \ln \frac{\pi(q|p_1, \ldots, p_s)}{\pi_0(q)} dq, \tag{3}$$

which is the Kullback-Leibler divergence of known prior pdf $\pi_0(q)$ from unknown posterior pdf $\pi(q|p_1, \ldots, p_s)$.

2.2 Constraints on Expected Values of Kullback-Leibler divergence

We expect that each source provides a reasonable observation about q in the sense that the conditional expectation of the Kullback-Leibler divergence (3) of q from source's probability vector is bounded. In our case the expected value is conditioned on $p_1, \ldots, p_s$ and taken with respect to the posterior pdf $\pi(q|p_1, \ldots, p_s)$. We also assume that these conditional expectations are equal for all sources, thus none of the sources is preferred. The constraints look as follows:

$$\mathrm{E}_{\pi(q|p_1,\ldots,p_s)}\mathrm{KLD}(p_j||q) = \mathrm{E}_{\pi(q|p_1,\ldots,p_s)}\mathrm{KLD}(p_s||q), \tag{4}$$

where $j = 1, \ldots, s-1$.

We see that (3) together with (4) leads to a constrained nonlinear optimization task. If the prior distribution $\pi_0(q)$ is Dirichlet distribution with prior parameters $(\nu_{01}, \ldots, \nu_{0n})$, the pdf minimizing (3) and satisfying (4) is the pdf of the Dirichlet distribution with updated parameters

$$\nu_i = \nu_{0i} + \sum_{j=1}^{s-1} \lambda_j (p_{ji} - p_{si}), \tag{5}$$

where λ_j are the Lagrange multipliers. Values of the multipliers are then simply computed by minimization of (3) with $\pi(q|p_1, \ldots, p_s)$ being the pdf of the Dirichlet distribution with parameters given in (5).

2.3 Final Combination of Information Sources

According to (2) and properties of Dirichlet distribution we obtain

$$\hat{q}_i = \frac{\nu_i}{\sum_{i=1}^{n} \nu_i} = \frac{\nu_{0i} + \sum_{j=1}^{s-1} \lambda_j (p_{ji} - p_{si})}{\sum_{i=1}^{n} \nu_{0i}}. \tag{6}$$

3 Illustrative Example

Let us assume q is a 3-dimensional probability vector and 3 information sources provided the following observations:

$$\begin{aligned} p_1 &= (0.35, \quad 0.25, \quad 0.40) \\ p_2 &= (0.40, \quad 0.50, \quad 0.10) \\ p_3 &= (0.65, \quad 0.15, \quad 0.20). \end{aligned}$$

Prior pdf $\pi_0(q)$ is the pdf of the Dirichlet distribution with parameters $(\nu_{01}, \nu_{02}, \nu_{03}) = (1, 1, 1)$. The optimization task (3) with constraints (4) was solved numerically in Matlab for 100 randomly chosen initial values of the Lagrange multipliers λ_1, λ_2. For all 100 initial values of the Lagrange multipliers the resulting values of λ_1 and λ_2 (and thus the value of $\hat{q}$ in (6)) were the same:

$$\begin{aligned} \lambda_1 &= -0.72 \\ \lambda_2 &= -0.17 \\ \hat{q} &= (0.55, \quad 0.26, \quad 0.19). \end{aligned}$$

4 Conclusion

This contribution brings insight into a procedure for combining information sources based on information theory. We successfully applied the minimum cross-entropy principle and the Kullback-Leibler divergence and obtained the weighted combination where the weights are not subjectively influenced and no reduction of the space of sources was needed. Experimental results suggest that the Lagrange multipliers λ_j, used in the final combination, converge to the same value for different starting points in the nonlinear optimization.

Acknowledgement: This work was supported by the grant GACR 13-13502S.

References

[1] Bernardo, J.M. (1979). "Expected information as expected utility." *Ann. Stat.*, **7**, 686-690.

[2] Fassinut-Mombot, B., Choquel, J.B. (2004). "A new probabilistic and entropy fusion approach for management of information sources." *Inform. Fusion*, **5**(1), 35–47.

[3] Kullback, S., Leibler, R.A. (1951). "On information and sufficiency." *Ann. Math. Stat.*, **22**, 79–86.

[4] Shore, J.E., Johnson, R.W. (1980). "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy." *IEEE Trans. Inf. Theory*, **26**, 26–37.