On Linear Probabilistic Opinion Pooling Based on Kullback-Leibler Divergence

Vladimíra Sečkárová
Department of Adaptive Systems
Institute of Information Theory and Automation of the CAS
Pod Vodárenskou věží 4, Prague 8, CZ 182 08
e-mail: seckarov@utia.cas.cz

Abstract

In this contribution we focus on the finite collection of sources, providing their opinions about a hidden (stochastic) phenomenon, that is not directly observable. The assumption on obtaining opinions yields a decision making process commonly referred to as opinion pooling. Due to the complexity of the space of possible decisions we consider the probability distributions over this set rather than single values, exploited before, e.g., in [2]. The final decision (result of pooling) is then a combination of probability distributions provided by sources. Here, we in particular exploit the combination introduced in [4], assuming each source is cooperating and willing to share its opinion with others, but *selfishly* requires the combination to be close to its opinion. The summary of basic steps is given below.

Kullback-Leibler divergence based combination of sources' opinions

Let us have $s < \infty$ sources providing discrete probability distributions represented by probability mass functions (pmf) $\mathbf{p}_1, \ldots, \mathbf{p}_s$:

$$\mathbf{p}_j = (p_{j1}, \ldots, p_{jn}) : \quad p_{ji} > 0, \ \sum_{i=1}^{n} p_{ji} = 1, \ n < \infty, \ j = 1, \ldots, s. \tag{1}$$

By exploiting theory of the Bayesian decision making [3] we search for their combination as the estimator $\hat{\mathbf{q}}$ of an unknown pmf $\mathbf{q}$ minimizing the expected Kullback-Leibler divergence [1]:

$$\mathrm{E}_{\pi(\mathbf{q}|\mathbf{p}_1,\ldots,\mathbf{p}_s)} \mathrm{KLD}(\mathbf{q}||\hat{\mathbf{q}}). \tag{2}$$

The minimizer of (2) is

$$\hat{\mathbf{q}} = \mathrm{E}_{\pi(\mathbf{q}|\mathbf{p}_1,\ldots,\mathbf{p}_s)}[\mathbf{q}|\mathbf{p}_1, \ldots, \mathbf{p}_s]. \tag{3}$$

To obtain the conditional expectation in (3) the conditional probability density function (pdf) $\pi(\mathbf{q}|\mathbf{p}_1, \ldots, \mathbf{p}_s)$ has to be specified. We formalize the notion of *selfishness* among sources by considering the following equality constraints:

$$\mathrm{E}_{\pi(\mathbf{q}|\mathbf{p}_1,\ldots,\mathbf{p}_s)}[\mathrm{KLD}(\mathbf{p}_j||\mathbf{q})|\mathbf{p}_1,\ldots,\mathbf{p}_s] = \mathrm{E}_{\pi(\mathbf{q}|\mathbf{p}_1,\ldots,\mathbf{p}_s)}[\mathrm{KLD}(\mathbf{p}_s||\mathbf{q})|\mathbf{p}_1,\ldots,\mathbf{p}_s], \tag{4}$$

$j = 1,\ldots,s-1$. Let $\mathcal{S}$ denote the set of all pdfs $\pi(\mathbf{q}|\mathbf{p}_1,\ldots,\mathbf{p}_s)$ satisfying (4). We now exploit the minimum cross-entropy principle [5] and choose the conditional pdf $\pi(\mathbf{q}|\mathbf{p}_1,\ldots,\mathbf{p}_s) \in \mathcal{S}$ that solves:

$$\min_{\pi(\mathbf{q}|\mathbf{p}_1,\ldots,\mathbf{p}_s)\in\mathcal{S}} \mathrm{KLD}(\pi(\mathbf{q}|\mathbf{p}_1,\ldots,\mathbf{p}_s)||\pi_0(\mathbf{q})), \tag{5}$$

where $\pi_0(\mathbf{q})$ denotes the prior guess on the conditional pdf $\pi(\mathbf{q}|\mathbf{p}_1,\ldots,\mathbf{p}_s)$.

We choose the pdf of the Dirichlet distribution with parameters $\nu_{01},\ldots,\nu_{0n}$ as the prior guess $\pi_0(\mathbf{q})$ for its computationally advantageous properties. Then, the conditional pdf $\hat{\pi}(\mathbf{q}|\mathbf{p}_1,\ldots,\mathbf{p}_s)$ minimizing (5) is also the pdf of the Dirichlet distribution $Dir(\hat{\nu}_1,\ldots,\hat{\nu}_n)$. The values of its parameters $\hat{\nu}_1,\ldots,\hat{\nu}_n$ are expressed by the following formula:

$$\hat{\nu}_i = \nu_{0i} + \sum_{j=1}^{s} \lambda_j (p_{ji} - p_{si}), \quad i = 1,\ldots,n, \tag{6}$$

where λ_j are the Lagrange multipliers resulting from minimization of (5) with respect to $(s-1)$ equations in (4), and the combination (3) is

$$\hat{q}_i = \frac{\nu_{0i}}{\sum_{k=1}^{n}\nu_{0k}} + \sum_{j=1}^{s} \frac{\lambda_j}{\sum_{k=1}^{n}\nu_{0k}} (p_{ji} - p_{si}), \quad i = 1,\ldots,n. \tag{7}$$

Although the combination has been introduced earlier in [4], its properties have not received much attention. We next discuss the choice of prior parameters $\nu_{01},\ldots,\nu_{0n}$ and the changes in the value of the combination when we deal with duplicate opinions.

Properties of the combination

It is somewhat surprising that the equation (6) combines simultaneously both, the parameters of the Dirichlet distribution and pmfs $\mathbf{p}_1,\ldots,\mathbf{p}_s$. Pmfs provided by sources can be viewed as individual guess for $\nu_1,\ldots,\nu_n$ when $\sum_{k=1}^{n}\nu_k = \sum_{k=1}^{n}\nu_{0k} = 1$. By plugging this relation into (7) we obtain

$$\hat{q}_i = p_{0i} + \sum_{j=1}^{s-1} \lambda_j p_{ji} + \left(-\sum_{j=1}^{s-1}\lambda_j\right) p_{si}, \tag{8}$$

where prior pmf $(p_{01},\ldots,p_{0n})$, generally $p_{0i} = \frac{\nu_{0i}}{\sum_{k=1}^{n}\nu_{0k}}$, coincides with $(\nu_{01},\ldots,\nu_{0n})$, a part of $\hat{\mathbf{q}}$ induced by prior pdf prior pdf $\pi_0(\mathbf{q})$.

Remind that we focus on combining sources' (experts') opinions, where the prior information about the studied problem may not be available. For the prior guess on $(p_{01},\ldots,p_{0n})$ one should then exploit provided pmfs $\mathbf{p}_1,\ldots,\mathbf{p}_s$. Based on the additive nature of the derived optimal estimator $\hat{q}$ and the considered relation between $(\nu_{01},\ldots,\nu_{0n})$ and $(p_{01},\ldots,p_{0n})$ in (8), we focus on the weighted linear combination of $\mathbf{p}_1,\ldots,\mathbf{p}_s$, e.g., arithmetic mean. Preferences can be assigned by delegated person or depend on other available information, e.g., sources' prior information about parameters of the Dirichlet distribution. The constraints (equality of the expected KL-divergences) should then be modified accordingly.

We next study how the value of the combination (7) changes with the duplicate data. Let us now have $s+1$ pmfs $\mathbf{p}_1, \ldots, \mathbf{p}_s, \mathbf{p}_{s+1}$ and for simplicity assume that $p_{s+1,i} = p_{s,i}$, $i = 1, \ldots, n$. Let $\lambda_1, \ldots, \lambda_s$ be the Lagrange multipliers related to s equality constraints in (4). Then, for a fixed prior pmf $\mathbf{p}_0$, the combination of $\mathbf{p}_1, \ldots, \mathbf{p}_s, \mathbf{p}_{s+1}$ coincides with combination evaluated with omission of $\mathbf{p}_{s+1}$ and unchanged $\mathbf{p}_0$:

$$\hat{q}_i = p_{0i} + \sum_{j=1}^{s-1} \lambda_j (p_{ji} - p_{si}) + \lambda_s (p_{si} - p_{si}). \tag{9}$$

The additivity property of combination (7) implies that if other s_1 sources gave the same pmf $\mathbf{p}_k$, then the coefficient of each source equals $\frac{\lambda_k}{s_1}$.

It may seem strange that repeated sources' opinion are not taken more "seriously", with a higher weight. This is consequence of the fact that individual sources are not qualified by a weight reflecting their reliability. When such a weighting will be introduced, the coincidence of opinions can be taken into account and distinguished from cheating by repetitions of the same opinion.

Conclusion and future work

In this contribution we focused on approach to combining sources' opinions described in [4]. This combination is of a conservative type and qualifies all repetitions as "cheating", prevents overweighing of such source. The analogue of (7), where the prior guess $\mathbf{p}_0$ as well as the constraints (4) are influenced by preferences among sources, is of interest in the future.

Keywords: combining discrete probability distributions, linear opinion pooling, minimum cross-entropy principle

Acknowledgement: This research has been supported by GAČR 13-13502S.

References:

[1] Bernardo, J. M.: Expected information as expected utility. *Ann. Stat.*, vol. 7 (1979)

[2] Kárný, M., Guy, T. V., Bodini, A., Ruggeri, F.: Cooperation via sharing of probabilistic information. *Int. J. of Computational Intelligence Studies*, vol. 1 (2009)

[3] Savage, L. J.: The Foundations of Statistics. Dover Books on Mathematics Series (1972)

[4] Sečkárová, V.: Dynamic Parameter Estimation Based on Minimum Cross-Entropy Method For Combining Information Sources. *Pliska Studia Mathematica Bulgarica*, vol. 24 (2015)

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