Lossless Coding with Generalized Criteria

T. Charalambous, C. D. Charalambous and F. Rezaei

Charalambos D. Charalambous

Department of Electrical and Computer Engineering
University of Cyprus

University of Cyprus

ISIT 2011
SAINT PETERSBURG

Outline

1. Introduction
   - Objectives - Problems Considered
   - Related Problems
2. Optimal Weights and Merging Rule
   - Optimal weigths
   - Algorithm for optimal weights
   - A schematic representation
3. Optimal Codeword Lengths
4. Illustrative Examples
5. Conclusions and Future Directions

Outline

Coding problems considered I

Two lossless coding problems are considered, in which the pay-offs are the following:

Problem1: A convex combination of the maximum codeword length and the average codeword length

Find codeword length vector $\ell^{S,\dagger} \in \mathcal{L}(\mathbb{R}^n)$ which solves

$$\mathbb{L}_{\alpha}^{MO}(\mathbf{l}, \mathbf{p}) \triangleq \left\{ \alpha \max_{x \in \mathcal{X}} l(x) + (1-\alpha) \sum_{x \in \mathcal{X}} l(x) p(x) \right\}$$

where $\alpha \in [0,1]$ is a weighting parameter.

The extreme cases:

- $\alpha = 0$ corresponds to the average codeword length, and
- $\alpha = 1$ corresponds to the maximum codeword length.

Coding problems considered II

Problem 2: A convex combination of the average of an exponential function of the codeword length and the average codeword length

Find codeword length vector $\ell^{S,\dagger} \in \mathcal{L}(\mathbb{R}^n)$ which solves

$$\mathbb{L}_{t,\alpha}^{MO}(\mathbf{l},\mathbf{p}) \quad \stackrel{\triangle}{=} \frac{\alpha}{t}\log\Big(\sum_{x\in\mathcal{X}} p(x) D^{t(x)}\Big) + (1-\alpha)\sum_{x\in\mathcal{X}} l(x)p(x)$$

where $\alpha \in [0,1]$ is a weighting parameter and $t \in (-\infty,\infty)$.

Although, Problem 2 is also defined for $t \leq 0$, its solution will be discussed for $t \geq 0$.

Related Problems

The problems considered are multiobjective pay-offs whose solution bridges together an anthology of source coding problems with different pay-offs:

- the average redundancy of the codeword length [6, 2],
- the average of an exponential function of the codeword length [4, 7, 1],
- the average of an exponential function of the redundancy of the codeword length [3, 2, 1].

Related problems

- For $\alpha = 0$ or $\alpha = 1$ the resulting special cases of Problem 2 are found in [6, 4, 7, 3, 2, 1].
- The special case $\mathbb{L}_{t,1}^{MO}(\mathbf{l}$ is the dual problem of universal coding problems formulated as a minimax (see [8]).

Outline

1. Introduction
2. Optimal Weights and Merging Rule
3. Optimal Codeword Lengths
4. Illustrative Examples
5. Conclusions and Future Directions

Restructuring Problem 1

Weights

- Define

$$\ell_u = ||\ell||_\infty \triangleq \max_i(\ell_i), \qquad \mathcal{U} \triangleq \{i \in \mathbb{Z}_+^n : \ell_i = \ell_u\}$$

- Pay-off $\mathbb{L}_\alpha^{MO}(\mathbf{l}, \mathbf{p})$

$$\begin{aligned} \alpha\ell_u + (1-\alpha)\sum_i p_i\ell_i &= \left[\alpha + (1-\alpha)\sum_{i\notin\mathcal{U}} p_i\right]\ell_u + \sum_{i\in\mathcal{U}}(1-\alpha)p_i\ell_i \\ &= \sum_i w_i(\alpha)\ell_i, \end{aligned}$$

where $0 \le w_i(\alpha) \le 1,\ \forall\alpha \in [0,1)$ and $\sum_i w_i(\alpha) = 1$

- The code is optimal if $L(\alpha) = \sum_i w_i(\alpha)\ell_i$ is minimal [5].

Optimal weights

Theorem 1

Given the probability vector $p = (p_1 \ p_2 \ \dots \ p_n)$ and $\alpha \in [\alpha_k, \alpha_{k+1})$, $k \in \{0, 1, \dots, n-1\}$, the optimal weights $w^\dagger = (w_1^\dagger \ w_2^\dagger \ \dots \ w_n^\dagger)$ are

$$w_i^\dagger(\alpha) = \begin{cases} (1-\alpha)p_i = w_i(\alpha_k) - (\alpha - \alpha_k)p_i \ , & \forall \ i \notin \mathcal{U}_k \\ w_i^*(\alpha_k) + (\alpha - \alpha_k)\dfrac{\sum_{i \notin \mathcal{U}_k} p_i}{|\mathcal{U}_k|} \ , & \forall \ i \in \mathcal{U}_k \end{cases}$$

where $w_i^*(\alpha) \triangleq \min_{j \in \mathbb{N}_+^n} w_j(\alpha)$,

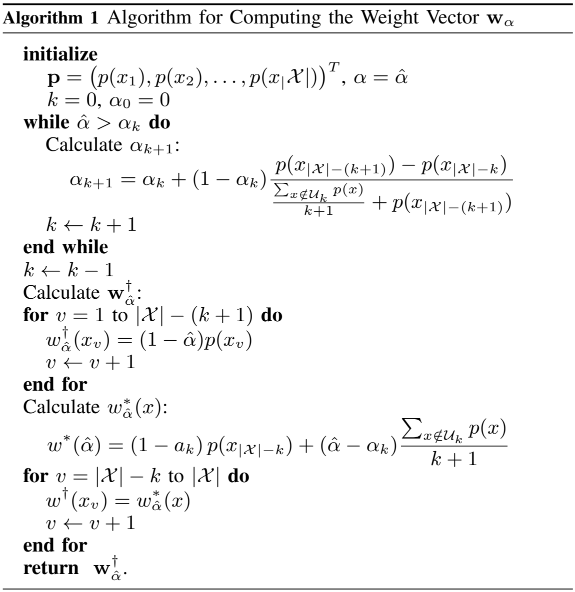
$$\alpha_k \triangleq \min\left\{\alpha \in [0,1) : w_{n-(k-1)}(\alpha) = w_{n-k}(\alpha), \ k \in \{1, \dots, n-1\}\right\}$$

$$\mathcal{U}_k \triangleq \{i \in \{n-k, \dots, n\} : w_i(\alpha_k) = w_i^*(\alpha_k), \ k \in \{0, \dots, n-1\}\}$$

and

$$\alpha_{k+1} = \alpha_k + (1 - \alpha_k)\frac{(p_{n-(k+1)} - p_{n-k})}{\frac{\sum_{i \notin \mathcal{U}_k} p_i}{|\mathcal{U}_k|} + p_{n-(k+1)}}$$

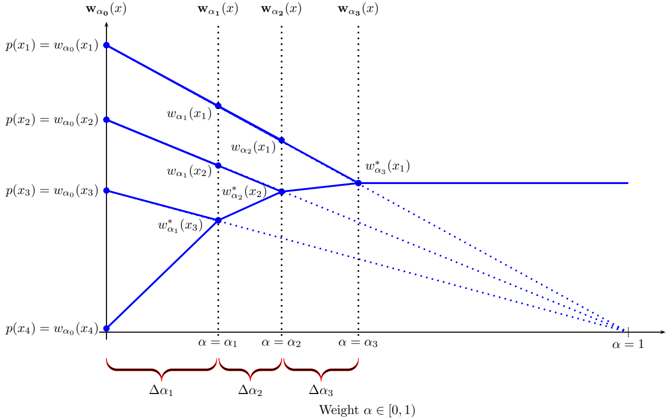
Algorithm for optimal weights



Algorithm 1 Algorithm for Computing the Weight Vector $\mathbf{w}_\alpha$

```
initialize
    $\mathbf{p} = \big(p(x_1), p(x_2), \dots, p(x_{|}\mathcal{X}|)\big)^T$, $\alpha = \hat{\alpha}$
    $k = 0$, $\alpha_0 = 0$
while $\hat{\alpha} > \alpha_k$ do
    Calculate $\alpha_{k+1}$:
        $\alpha_{k+1} = \alpha_k + (1 - \alpha_k) \frac{p(x_{|\mathcal{X}|-(k+1)}) - p(x_{|\mathcal{X}|-k})}{\frac{\sum_{x \notin \mathcal{U}_k} p(x)}{k+1} + p(x_{|\mathcal{X}|-(k+1)})}$
    $k \leftarrow k + 1$
end while
$k \leftarrow k - 1$
Calculate $\mathbf{w}^{\dagger}_{\hat{\alpha}}$:
for $v = 1$ to $|\mathcal{X}| - (k+1)$ do
    $w^{\dagger}_{\hat{\alpha}}(x_v) = (1 - \hat{\alpha}) p(x_v)$
    $v \leftarrow v + 1$
end for
Calculate $w^{*}_{\hat{\alpha}}(x)$:
    $w^{*}(\hat{\alpha}) = (1 - a_k)\, p(x_{|\mathcal{X}|-k}) + (\hat{\alpha} - \alpha_k) \frac{\sum_{x \notin \mathcal{U}_k} p(x)}{k+1}$
for $v = |\mathcal{X}| - k$ to $|\mathcal{X}|$ do
    $w^{\dagger}(x_v) = w^{*}_{\hat{\alpha}}(x)$
    $v \leftarrow v + 1$
end for
return $\mathbf{w}^{\dagger}_{\hat{\alpha}}$.
```

A schematic representation



A schematic representation of the weights for different values of α.

Outline

1. Introduction
2. Optimal Weights and Merging Rule
3. Optimal Codeword Lengths
4. Illustrative Examples
5. Conclusions and Future Directions

Optimal Codeword Lengths I

The optimal real-valued codeword length vectors $\mathbf{l} \in \mathcal{L}(\mathbb{R}_+^{|\mathcal{X}|})$ of the multiobjective pay-off of Problem 1 are shown to be

$$l_i(\alpha) = -\log_2 w_i(\alpha).$$

Theorem 2

Consider Problem 1. For any probability distribution $\mathbf{p} \in \mathbb{P}(\mathcal{X})$ and $\alpha \in [0, 1]$ the optimal prefix real-valued code $\mathbf{l} \in \mathbb{R}_+^{|\mathcal{X}|}$ minimizing the pay-off $\mathbb{L}_\alpha^{MO}(\mathbf{l}, \mathbf{p})$ is given by

$$l_\alpha^\dagger(x) = \begin{cases} -\log\Big((1-\alpha)p(x)\Big) & \text{for } x \notin \mathcal{U}_k \\ -\log\left(\frac{\alpha + (1-\alpha)\sum_{x \in \mathcal{U}_k} p(x)}{|\mathcal{U}_k|}\right) & \text{for } x \in \mathcal{U}_k \end{cases}$$

where $\alpha \in [\alpha_k, \alpha_{k+1}) \subset [0, 1], \ \forall k \in \{1, \ldots, |\mathcal{X}| - 1\}$.

Optimal Codeword Lengths II

Theorem 3

Consider Problem 2. For any probability distribution $\mathbf{p} \in \mathbb{P}(\mathcal{X})$ and $\alpha \in [0,1]$ the optimal prefix real-valued code $\mathbf{l} \in \mathbb{R}_+^{|\mathcal{X}|}$ minimizing the pay-off $\mathbb{L}_{t,\alpha}^{MO}(\mathbf{l},\mathbf{p})$ is given by

$$l_{t,\alpha}^{\dagger}(x) = -\log\Big(\alpha \nu_{t,\alpha}(x) + (1-\alpha)p(x)\Big), \quad x \in \mathcal{X} \tag{1}$$

where $\{\nu_{t,\alpha}(x) : x \in \mathcal{X}\}$ is defined via the tilted probability distribution

$$\nu_{t,\alpha}(x) \triangleq \frac{D^{t\, l_{t,\alpha}^{\dagger}(x)} p(x)}{\sum_{x \in \mathcal{X}} p(x) D^{t\, l_{t,\alpha}^{\dagger}(x)}}, \ x \in \mathcal{X} \tag{2}$$

Optimal Codeword Lengths III

Remarks

- *The Limiting Case as $t \to \infty$*:
  The minimization of the multiobjective pay-off $\mathbb{L}_{\alpha}^{MO}(\mathbf{l}, \mathbf{p})$ obtained in Theorem 2 is indeed obtained from the minimization of the two parameter multiobjective pay-off $\mathbb{L}_{t,\alpha}^{MO}(\mathbf{l}, \mathbf{p})$ in the limit, as $t \to \infty$.

$$\lim_{t\to\infty} \mathbb{L}_{t,\alpha}^{MO}(\mathbf{l}, \mathbf{p}) = \mathbb{L}_{\alpha}^{MO}(\mathbf{l}, \mathbf{p}), \ \forall\ \mathbf{l} \text{ and hence at } \mathbf{l} = \mathbf{l}^{\dagger}.$$

- *Coding Theorems*:
  Coding theorems for Problem 1 and Problem 2 can be easily obtained either from the closed form solutions or by following [4].

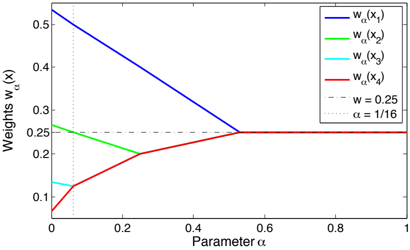
Outline

1. Introduction
2. Optimal Weights and Merging Rule
3. Optimal Codeword Lengths
4. Illustrative Examples
5. Conclusions and Future Directions

Examples I

Consider binary codewords and a source with $|\mathcal{X}| = 4$ and probability distribution

$$\mathbf{p} = \left(\begin{array}{cccc} \frac{8}{15} & \frac{4}{15} & \frac{2}{15} & \frac{1}{15} \end{array} \right).$$

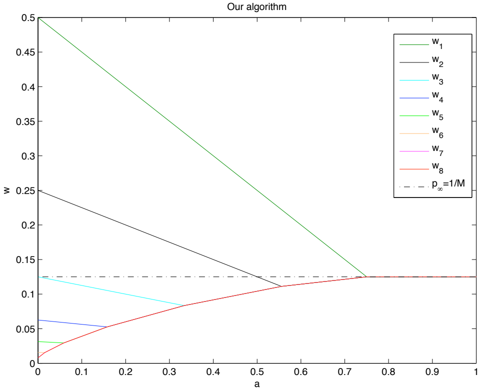


- $\alpha = \alpha_1 = 1/16$:
  Optimal weights are:
  $w_3^\dagger(\alpha) = w_4^\dagger(\alpha) = 1/8$,
  $w_2^\dagger(\alpha) = (1-\alpha)p_2 = 1/4$
  and
  $w_1^\dagger(\alpha) = (1-\alpha)p_1 = 1/2$.
- The resulting codeword lengths correspond to the optimal Huffman code.

Examples II

Consider binary codewords and a source with $|\mathcal{X}| = 8$ and probability distribution

$$\mathbf{p} = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} & \frac{1}{32} & \frac{1}{64} & \frac{1}{128} & \frac{1}{128} \end{pmatrix}.$$



Outline

Conclusions and Future Directions

Conclusions

- Two lossless coding problems with generalised pay-offs investigated.
- Real-valued codeword length solutions presented.
- Relations to problems discussed in the literature obtained.

Future Directions

Huffman algorithms which solve this problem:

Find codeword length vector $\ell^{H,\dagger} \in \mathcal{L}(\mathbb{Z}_+^n)$ which solves

$$\mathcal{R}_{cc}^{H}(\ell^{H,\dagger}) \triangleq \min_{\ell^{H,\dagger} \in \mathcal{L}(\mathbb{Z}_+^n)} \Big\{ \alpha ||\ell||_\infty + (1-\alpha) \sum_{j \in \mathcal{X}} \ell_j p_j \Big\}$$

are part of ongoing research.

References I

M.B. Baer.

Optimal Prefix Codes for Infinite Alphabets With Nonlinear Costs.

IEEE Transactions on Information Theory, 54(3):1273 –1286, march 2008.

M.B. Baer.

Tight bounds on minimum maximum pointwise redundancy.

In *IEEE International Symposium on Information Theory*, pages 1944 –1948, july 2008.

Michael Baer.

A general framework for codes involving redundancy minimization.

IEEE Trans. of Information Theory, 52:344–349, 2006.

L.L. Campbell.

A coding theorem and Rényi's entropy.

Information and Control, 8(4):423–429, August 1965.

References II

Thomas M. Cover and Joy A. Thomas.
Elements of Information Theory.
Wiley-Interscience, second edition, 2006.

Michael Drmota and Wojciech Szpankowski.
Precise minimax redundancy and regret.
IEEE Transactions of Information Theory, 50:2686–2707, 2004.

P. Humblet.
Generalization of huffman coding to minimize the probability of buffer overflow.
IEEE Transactions on Information Theory, 27(2):230–232, 1981.

F. Rezaei and C.D. Charalambous.
Robust coding for uncertain sources: a minimax approach.
In *Proceedings. International Symposium on Information Theory, ISIT*, pages 1539–1543, 2005.