Correct Aggregation Operations with Algorithms

Construction of correct algorithms from other correct algorithms using aggregation operations

Z.M. Shibzukhov

Mathematical Methods of Pattern Recognition, 2015

What is about

Consider models of algorithms and learning methods which make possible to build *collections* of *correct algorithms*.

Correct algorithm produces correct output for all learning samples and we call them *basic* correct algorithms

Question

How to build new correct algorithms from other correct algorithms that could extend capabilities of basic correct algorithms?

Some kinds of *aggregation operations* can be used for building new correct algorithms from other correct algorithms

Initial definitions

$\mathbf{X}$ – space of input informations about objects

$\mathbf{Y}$ – set of possible answers to given question about objects

$y\colon \mathbf{X} \to \mathbf{Y}$ – mapping from input informations to answers

$\mathfrak{a}\colon \mathbf{X} \to \mathbf{Y}$ – algorithm ($\mathfrak{a} \in \mathbb{A}$), which *approximates* y

$\mathbb{A}$ – space of *basic* algorithms

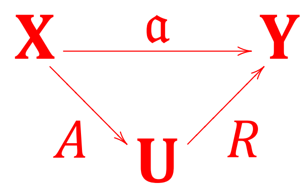
$\widetilde{\mathbf{X}}$ – finite subset of $\mathbf{X}$ (*samples*)

$\mathfrak{a}$ is *correct* on $\widetilde{\mathbf{X}}$ if it evaluates *valid answers* on $\widetilde{\mathbf{X}}$

F is *correct operation* with algorithms from $\mathbb{A}$ on $\widetilde{\mathbf{X}}$ if for any tuple $\{\mathfrak{a}_1, \dots, \mathfrak{a}_m\}$ of correct algorithms on $\widetilde{\mathbf{X}}$ the new algorithm $\mathfrak{a} = \mathrm{F}\{\mathfrak{a}_1, \dots, \mathfrak{a}_m\}$ is correct on $\widetilde{\mathbf{X}}$

Quality functions of answers

Consider algorithms that represented in the form $\mathfrak{a} = R \circ A$:



A – *estimating operator*, R – *decision rule*

$\mathbf{U} \subseteq \mathbb{R}^q$ – *space of estimations*

Quality function of answer $Q(\mathbf{u}, \tilde{y})$ evaluates quality of answer $y = \mathfrak{a}(\tilde{\mathbf{x}})$ on the base of estimation $\mathbf{u} = A(\tilde{\mathbf{x}})$ and correct answer $\tilde{y}$

Subset $\mathbf{Q}_{\tilde{y}} \subset \mathrm{image}\, Q$ associates with answers $y = \mathfrak{a}(\tilde{\mathbf{x}})$ that accepted as correct as $\tilde{y}$

Answer $y = \mathfrak{a}(\tilde{\mathbf{x}})$ is *correct* if $Q(\mathbf{u} \mid \tilde{y}) \in \mathbf{Q}_{\tilde{y}}$.

Monotone/unimodal functions

$f\colon \mathbf{U} \to \mathbb{R}$ – function of many arguments

f is *monotone* if $\forall y \in \mathbb{R}$: $f^{-1}(y)$ is connected set

Denote $D_f(y) = \{\mathbf{u} : f(\mathbf{u}) \leqslant y\}$

f is *monotone increasing* if $\forall y \in \mathbb{R}$:

$$y_1 \leqslant y_2 \Longrightarrow D_f(y_1) \subseteq D_f(y_2)$$

f is *monotone decreasing* if $\forall y \in \mathbb{R}$:

$$y_1 \geqslant y_2 \Longrightarrow D_f(y_1) \subseteq D_f(y_2)$$

f is *unimodal* if the set $\{\mathbf{u} : \mathbf{u}$ is local min/max of $f\}$ is connected

Loss functions

Loss function is mapping $Q: \mathbf{U} \times \mathbf{Y} \to \mathbb{R}_+$ if
for any $\tilde{y}$: $Q_{\tilde{y}}(\mathbf{u}) = Q(\mathbf{u}, \tilde{y})$ is:

- *monotone*,
- *increasing*,
- *unimodal* (min).

- $Q_{\tilde{y}} = \min_{\mathbf{u} \in \mathbf{U}} Q(\mathbf{u}, \tilde{y})$ – minimal losses for correct answers
- $\mathbf{U}_{\tilde{y}} = \{\mathbf{u}: Q(\mathbf{u}, \tilde{y}) = Q_{\tilde{y}}\}$ – set of estimates that associates with correct answer $\tilde{y}$

Margin functions

Margin function is mapping $Q\colon \mathbf{U} \times \mathbf{Y} \to \bar{\mathbb{R}}$ if for any $\tilde{y}$: $Q_{\tilde{y}}(\mathbf{u}) = Q(\mathbf{u}, \tilde{y})$ is:

- *monotone*,
- *unimodal* (max).

- $\mathbf{U}_{\tilde{y}} = \{\mathbf{u}\colon Q(\mathbf{u}, \tilde{y}) > 0\}$ – set of estimates that associates with correct answer
- $\bar{\mathbf{U}}_{\tilde{y}} = \{\mathbf{u}\colon Q(\mathbf{u}, \tilde{y}) < 0\}$ – set of estimates that associates with incorrect answer
- $\delta\mathbf{U}_{\tilde{y}} = \{\mathbf{u}\colon Q(\mathbf{u}, \tilde{y}) = 0\}$ – set of estimates when definite answer isn't possible

Regression problems ($\mathbf{Y} \subset \mathbb{R}$)

$Q(\mathfrak{a} \mid \mathbf{x}) = \ell(\mathfrak{a}(\tilde{\mathbf{x}}), \tilde{y})$ – *loss function*, $\mathbf{V}_Q = [0, \varepsilon]$

Definition

Function $\ell(y, \tilde{y})$ is a loss function if:

1. $\ell(\tilde{y}, \tilde{y}) = \inf_y \ell(y, \tilde{y})$
2. $y_1 \leqslant y_2 \leqslant \tilde{y} \Rightarrow \ell(y_1, \tilde{y}) \geqslant \ell(y_2, \tilde{y}) \geqslant \ell(\tilde{y}, \tilde{y})$
3. $\tilde{y} \leqslant y_1 \leqslant y_2 \Rightarrow \ell(\tilde{y}, \tilde{y}) \leqslant \ell(y_1, \tilde{y}) \leqslant \ell(y_2, \tilde{y})$

Definitions (correct answer)

- answer $y = \mathfrak{a}(\tilde{\mathbf{x}})$ is as correct as $\tilde{y}$ if $\ell(y, \tilde{y}) \in \mathbf{V}_Q$.
- answer $y = \mathfrak{a}(\tilde{\mathbf{x}})$ is correct if $y \in \mathbf{U}_{\tilde{y}} = \{y \in \mathbf{Y} \colon \ell(y, \tilde{y}) \in \mathbf{V}_Q\}$.

Examples: loss functions

$$\langle y-\tilde{y}\rangle_\varepsilon = \left(\frac{|y-\tilde{y}|}{1+|\tilde{y}|}-\varepsilon\right)_+, E_+ = [E \geqslant 0]\cdot E, \varepsilon > 0$$

Example (symmetrical)

$$\ell(y,\tilde{y}) = \langle y-\tilde{y}\rangle_\varepsilon^\lambda, \mathbf{V}_Q = [0,\varepsilon)$$

Example (nonsymmetrical)

$$\ell(y,\tilde{y}) = \begin{cases} \alpha\langle y-\tilde{y}\rangle_\varepsilon^\lambda & \text{if } y > \tilde{y}+\varepsilon \\ 0, & \text{if } |y-\tilde{y}| \leqslant \varepsilon, \mathbf{V}_Q = [0,\varepsilon), \alpha,\beta > 0 \\ \beta\langle y-\tilde{y}\rangle_\varepsilon^\lambda & \text{if } y < \tilde{y}-\varepsilon \end{cases}$$

Classification problems ($\mathbf{Y}$ is discrete)

Algorithms for classification are compositions:
$\mathfrak{a}(\mathbf{x}) = R \circ A(\mathbf{x})$.

Quality functions:

- $Q(\mathfrak{a} \mid \mathbf{x}) = \mu(\mathbf{u}, \tilde{y})$, $\mu\colon \mathbf{U} \times \mathbf{Y} \to \bar{\mathbb{R}}$ is *margin function* and $\mathbf{V}_Q = [\delta, \infty]$
- $Q(\mathfrak{a} \mid \mathbf{x}) = \ell(\mathbf{u}, \tilde{y})$, $\ell\colon \mathbf{U} \times \mathbf{Y} \to \mathbb{R}_+$ is *loss function* and $\mathbf{V}_Q = [0, \varepsilon]$

Definitions (correct estimate)

- estimation $\mathbf{u} = A(\mathbf{x})$ is correct if $\mu(\mathbf{u}, \tilde{y}) \in \mathbf{V}_Q$
- estimation $\mathbf{u} = A(\mathbf{x})$ is correct if $\mathbf{u} \in \mathbf{U}_{\tilde{y}} = \{\mathbf{u}\colon \mu(\mathbf{u}, \tilde{y}) \in \mathbf{V}_Q\}$

Examples: classification

Example (2-class classification)

$\mathbf{Y} = \{-1, 0, +1\}$, $R(u) = \operatorname{sign} u$, $\mu(u, \tilde{y}) = uy$

$\mathbf{V}_Q = [\delta, \infty]$, $\mathbf{U}_{\tilde{y}} = \{u\colon u\tilde{y} > \delta\}$.

Example (q-class classification)

$\mathbf{Y} = \{0, 1, \dots, q\}$, $\mathbf{A} = (A_1, \dots, A_m)$, $u_j = A_j(\mathbf{x})$

$$R(\mathbf{u}) = \begin{cases} y^\star = \arg\max_{y \in \{1,\dots,m\}} \{u_y\}, & \text{if } \mu(\mathbf{u}, \tilde{y}) \geqslant \varepsilon \\ \text{fail}, & \text{if } \mu(\mathbf{u}, \tilde{y}) \in (-\varepsilon, \varepsilon) \\ 0, & \text{if } \mu(\mathbf{u}, \tilde{y}) \leqslant -\varepsilon \end{cases}$$

$\ell(\mathbf{u}, \tilde{y}) = e^{-\mu(\mathbf{u}, \tilde{y})}$, $\qquad \mu(\mathbf{u}, \tilde{y}) = u_{\tilde{y}} - \max_{j = \{1,\dots,q\} \setminus \{\tilde{y}\}} \{u_j\}$

$\mathbf{V}_Q = (\varepsilon, \infty)$, $\mathbf{U}_{\tilde{y}} = \{\mathbf{u}\colon u_{\tilde{y}} > u_y \text{ for all } y \neq \tilde{y}\}$

Aggregation and mean functions

Definition (aggregation function)

M – *aggregation function* on $\mathbf{Y} \subseteq \mathbb{R}$, i.e.

- for any $m \in \mathbb{N}$ and any tuple $\{y_1, \dots, y_m\}$: $\mathrm{M}\{y_1, \dots, y_m\} \in \mathbf{Y}$
- for any $m \in \mathbb{N}$ and any pair of tuples $\{y_1', \dots, y_m'\} \leqslant \{y_1'', \dots, y_m''\}$: $\mathrm{M}\{y_1', \dots, y_m'\} \leqslant \mathrm{M}\{y_1'', \dots, y_m''\}$

Definition (mean function)

M – *mean function* if

$$\min\{y_1, \dots, y_m\} \leqslant \mathrm{M}\{y_1, \dots, y_m\} \leqslant \max\{y_1, \dots, y_m\}$$

Definition (idempotent function)

M – *idempotent function* if

$$\mathrm{M}\{y, \dots, y\} = y$$

Point-wise correct operations

Definition (point-wise correct algorithm)

$\mathfrak{a}$ is *correct algorithm* on $\tilde{\mathbf{X}}$ if $\mathfrak{a}(\tilde{\mathbf{x}})$ is correct for all $\tilde{\mathbf{x}} \in \tilde{\mathbf{X}}$

Condition (point-wise correct operation)

M is *correct operation* if for any tuple $\{y_1, \ldots, y_m\} \subset \mathbf{U}_{\tilde{y}}$:

$$\mathrm{M}\{y_1, \ldots, y_m\} \in \mathbf{U}_{\tilde{y}}$$

Criterion (mean function – correct point-wise operation)

If M is idempotent mean function, then M is correct operation.

Generalized Kolmogorov's mean

Definition (generalized Kolmogorov's mean)

$$\mathrm{M}_g\{u_1,\dots,u_m\} = g^{-1}\big(w_1 g(u_1) + \cdots + w_m g(u_m)\big),$$

$w_1,\dots,w_m \geqslant 0$ and $w_1 + \cdots + w_m = 1$

$g : \mathbf{U} \to \mathbb{R}$ or $g : \mathbf{U} \to \mathbb{R}_+$ or $g : \mathbf{U} \to [0,1]$ – inversible function

Examples (weighted Kolmogorov's mean)

- $\mathrm{M}_g\{u_1,\dots,u_m\} = (\sum w_j u_j^{\langle p\rangle})^{\langle 1/p\rangle}$ $\qquad g(s) = u^{\langle p\rangle}$
- $\mathrm{M}_g\{u_1,\dots,u_m\} = \frac{1}{p}\ln(\sum w_j e^{p u_j})$ $\qquad g(u) = e^{pu}$
- $\mathrm{M}_g\{u_1,\dots,u_m\} = \prod u_j^{w_j}$ $\qquad g(u) = \ln u$

Mean aggregation operators

Let $\preccurlyeq$ – partial order on $\mathbf{U}$

Definition

$\mathsf{M} = (\mathsf{M}_1, \dots, \mathsf{M}_m)$ is *aggregation operator* on $\mathbf{U}$:

- for each m and any tuple $\{\mathbf{u}_1, \dots, \mathbf{u}_m\} \subset \mathbf{U}$:
  $\mathsf{M}\{\mathbf{u}_1, \dots, \mathbf{u}_m\} \in \mathbf{U}$;
- for each m and any pairs of tuples
  $\{\mathbf{u}'_1, \dots, \mathbf{u}'_m\} \preccurlyeq \{\mathbf{u}''_1, \dots, \mathbf{u}''_m\}$:
  $\mathsf{M}\{\mathbf{u}'_1, \dots, \mathbf{u}'_m\} \preccurlyeq \mathsf{M}\{\mathbf{u}''_1, \dots, \mathbf{u}''_m\}$

Definition

M is *mean aggregation operator* if

$$\inf\{\mathbf{u}_1, \dots, \mathbf{u}_m\} \preccurlyeq \mathsf{M}\{\mathbf{u}_1, \dots, \mathbf{u}_m\} \preccurlyeq \sup\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$$

M is point-wise correct operation if $\mathsf{M}\{\mathbf{U}_{\tilde{y}}, \dots, \mathbf{U}_{\tilde{y}}\} \subseteq \mathbf{U}_{\tilde{y}}$.

Multivariate Kolmogorov's mean

If $\mathbf{M}$ is idempotent operation, i.e. $\mathbf{M}\{\mathbf{u},\ldots,\mathbf{u}\}=\mathbf{u}$, then $\mathbf{M}$ is correct point-wise operation.

Multivariate weighted Kolmogorov's mean:

$$\mathbf{M}_{\boldsymbol{g}}\{\mathbf{u}_1,\ldots,\mathbf{u}_m\}=\boldsymbol{g}^{-1}\Big(\sum \mathbf{w}_j\boldsymbol{g}(\mathbf{u}_j)\Big),$$

where $\mathbf{w}_1+\cdots+\mathbf{w}_m=1$, $\boldsymbol{g}\colon \mathbf{U}\to\mathbb{R}^q$ is inversible continuous mapping.

Fact

If $\boldsymbol{g}(\mathbf{U}_{\tilde{y}})$ is a convex set then $\mathbf{M}_{\boldsymbol{g}}\{\mathbf{U}_{\tilde{y}},\ldots,\mathbf{U}_{\tilde{y}}\}\subseteq\mathbf{U}_{\tilde{y}}$, i.e. multivariate weighted Kolmogorov's mean $\mathbf{M}_{\boldsymbol{g}}$ is correct point-wise operation.

Aggregationally correct algorithm

M – aggregation function on image $\mathcal{Q}$.

Aggregation quality functional

$$\mathcal{Q}(\mathfrak{a} \mid \widetilde{\mathbf{X}}) = \mathrm{M}\{\mathcal{Q}(\mathfrak{a} \mid \tilde{\mathbf{x}}) : \tilde{\mathbf{x}} \in \widetilde{\mathbf{X}}\}$$

evaluates the quality of algorithm on whole set $\widetilde{\mathbf{X}}$ on the base of quality of all answers on $\widetilde{\mathbf{X}}$.

Let $\mathbf{V}_{\mathcal{Q}}$ is subset of values of $\mathcal{Q}$, which corresponds correct algorithms on $\widetilde{\mathbf{X}}$.

Definition (Aggregationally correct algorithm)

Algorithm $\mathfrak{a}$ is *aggregationally correct* on $\widetilde{\mathbf{X}}$ if $\mathcal{Q}(\mathfrak{a} \mid \widetilde{\mathbf{X}}) \in \mathbf{V}_{\mathcal{Q}}$.

Linear operations

M – arithmetic weigted mean:

$$\mathrm{M}\{z_1, \ldots, z_N\} = \sum_{k=1}^{N} w_k z_k$$

F – linear operation:

$$\mathrm{F}\{u_1, \ldots, u_m\} = \alpha_1 u_1 + \cdots + \alpha_m u_m,$$

$\alpha_1, \ldots, \alpha_m \geqslant 0, \alpha_1 + \cdots + \alpha_m = 1.$

Let

1. $Q_{\tilde{y}}(\mathbf{u})$ – *convex for any* $\tilde{y} \in \mathbf{Y}$ *;*
2. $\mathbf{V}_Q$ *is convex;*
3. *if* $Q' \in \mathbf{V}_Q$ *u* $Q'' \leqslant Q'$, *mo* $Q'' \in \mathbf{V}_Q$.

Then linear operation is aggreagtionally correct operation with algorithms with respect to Q.

F/G-convex

$f\colon \mathbf{U} \to \mathbf{R}$

F – idempotent aggregation function on $\mathbf{U} \subseteq \mathbb{R}$

G – idempotent aggregation function on $\mathbf{R} \subseteq \mathbb{R}$

Definition

$f(\mathbf{u})$ is F/G-convex, if

$$f(\mathrm{F}\{\mathbf{u}_1, \dots, \mathbf{u}_m\}) \leqslant \mathrm{G}\{f(\mathbf{u}_1), \dots, f(\mathbf{u}_m)\}.$$

M – aggregation function on $\mathbf{R}$

Example

$$f(\alpha_1\mathbf{u}_1 + \cdots + \alpha_m\mathbf{u}_m) \leqslant \alpha_1 f(\mathbf{u}_1) + \cdots + \alpha_m f(\mathbf{u}_m)$$

$$f(\alpha_1\mathbf{u}_1 + \cdots + \alpha_m\mathbf{u}_m) \leqslant \max\{f(\mathbf{u}_1), \dots, f(\mathbf{u}_m)\}$$

Domination of aggregation functions

Definition

G is dominated over M if

$$\mathrm{M}\{\mathrm{G}\{u_{11},\dots,u_{1m}\},\dots,\mathrm{G}\{u_{N1},\dots,u_{Nm}\}\} \leqslant \mathrm{G}\{\mathrm{M}\{u_{11},\dots,u_{N1}\},\dots,\mathrm{M}\{u_{1m},\dots,u_{Nm}\}\}$$

If $\forall N \in \mathbb{N}$ *function* $H(u_1,\dots,u_N) = \mathrm{M}\{u_1,\dots,u_N\}$ *is convex, then*

$$\mathrm{M}\Big\{\sum_j \alpha_j u_{1j},\dots,\sum_j \alpha_j u_{Nj}\Big\} \leqslant \sum_j \alpha_j \mathrm{M}\{u_{1j},\dots,u_{Nj}\},$$

i.e. linear weighted mean dominated over M.

Aggregationally correct operations

Theorem

Let

1. $Q_{\tilde{y}}(\mathbf{u})$ – F/G-convex for any $\tilde{y} \in \mathbf{Y}$;
2. G is dominated over M;
3. $\mathbf{V}_Q$ is closed within G;
4. if $Q' \in \mathbf{V}_Q$ и $Q'' \leqslant Q'$, то $Q'' \in \mathbf{V}_Q$.

Then F – aggreagtionally correct operation with algorithms with respect to Q.

From domination to convexity

G – aggregation operation on $\mathbb{R}$

F – aggregation operation on $\mathbb{R}^m$:

$$\mathrm{F}\{\mathbf{u}_1, \dots, \mathbf{u}_N\} = (\mathrm{G}\{u_{11}, \dots, u_{1m}\}, \dots, \mathrm{G}\{u_{N1}, \dots, u_{Nm}\})$$

H – aggregation operation on image of M

Definition

M – G/H-convex if

$$\mathrm{M}\{\mathrm{G}\{u_{11}, \dots, u_{1m}\}, \dots, \mathrm{G}\{u_{N1}, \dots, u_{Nm}\}\} \leqslant \mathrm{H}\{\mathrm{M}\{u_{11}, \dots, u_{N1}\}, \dots, \mathrm{M}\{u_{1m}, \dots, u_{Nm}\}\}.$$

Aggregationally correct operations

Theorem

Let

1. $Q_{\tilde{y}}(\mathbf{u})$ – F/G-convex;
2. M – G/H-convex;
3. $\mathbf{V}_Q$ is closed within H;
4. if $Q' \in \mathbf{V}_Q$ and $Q'' \leqslant Q'$ then $Q'' \in \mathbf{V}_Q$.

Then F – aggreagtionally correct operation with algorithms with respect to Q.

Thank you!