Энтропийная рандомизация в задачах машинного обучения

Ю.С. Попков[1,2] Ю.А. Дубнов[1,3] А.Ю. Попков[1]

[1]Федеральный исследовательский центр «Информатика и управление» РАН

[2]Институт проблем управления РАН

[3]НИУ «Высшая школа экономики»

Семинар «Теория автоматического управления»
Институт проблем управления РАН Москва, 10 ноября 2020 г.

Содержание

1. Рандомизация — имитация неопределенности

Источники неопределенности

$$\left.\begin{aligned} X &= \{x^{(1)},\dots,x^{(m)}\} \text{ – вход} \\ Z &= \{z^{(1)},\dots,z^{(r)}\} \text{ – выход} \end{aligned}\right\} \text{данные}$$

$$\hat{\mathbf{z}}^{(j)} = \Omega(X,\mathbf{a}) \text{ – модель}$$

$$v^{(j)} = \hat{\mathbf{z}}^{(j)} + \xi^{(j)}, \quad \xi^{(j)} \text{ – шум}$$

Искусственные ансамбли

$$\left.\begin{aligned} \mathcal{Z}(\mathbf{a}) &= \{z^{(1)},\dots,z^{(r)}\} \\ \mathcal{V}(\mathbf{a}) &= \{v^{(1)},\dots,v^{(r)}\} \end{aligned}\right\} \begin{matrix} P(\mathbf{a}) \\ Q(\xi) \end{matrix}$$

Оптимизация ансамблей

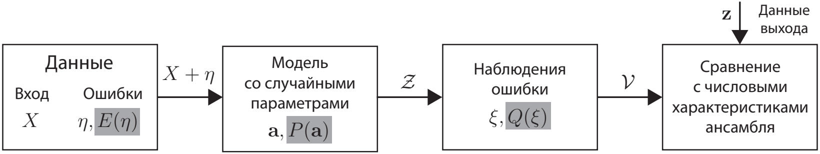
$P(\mathbf{a})$ — оптимальная
$Q(\xi)$ — оптимальная

Рандомизированная имитация

Генерация $\mathbf{a}$, ξ с оптимальными ПРВ $P(\mathbf{a})$, $Q(\xi)$

2. Рандомизированное машинное обучение (РМО)

2.1 Классы РМО



Жесткое РМО

$$\mathbf{z} = \mathcal{M}(\mathcal{V})$$

Мягкое РМО

$$\|\mathbf{z} - \mathcal{M}(\mathcal{V})\| \rightarrow \begin{cases} \min \\ \leq \delta \end{cases}$$

1. *Y.S. Popkov, Y.A. Dubnov, A.Y. Popkov*. Randomized Machine Learning: Statements, Solutions, Applications. 2016 IEEE 8th International Conference on Intelligent Systems (IS-2016), doi:10.1109/IS.2016.7737456.
2. *Y.S. Popkov, Y.A. Dubnov, A.Y. Popkov*. Introduction to the Theory of Randomized Machine Learning, 2018. Studies in Computational Intelligence 756, pp. 199-220, doi:10.1007/978-3-319-75181-8-10.

2.2 Ошибки в данных

- Ошибки в выходных данных

$$\xi, Q(\xi)$$

- Ошибки во входных данных

$$\eta, G(\eta, \mathbf{a})$$

- Ошибки во входных и выходных данных

$$\eta, \xi, G(\eta, \mathbf{a}), Q(\xi)$$

Линейная модель

$$G(\eta,\mathbf{a}) = E(\eta)P(\mathbf{a})$$

- Наихудшие (при максимальной неопределенности) ПРВ ошибок

$$Q^*(\xi)$$

$$G^*(\eta, \mathbf{a})$$

$$G^*(\eta,\mathbf{a}), Q^*(\xi)$$

1. *Y.S. Popkov*. Soft Randomized Machine Learning. Doklady Mathematics, 2018, V. 98, No. 3, pp. 1-2.

3. Восстановление пропущенных данных

- Пропуски в выходных данных

$$z[0],\ldots,z[k^- -1], z[k^+ +1],\ldots,z[N]$$

ошибки

$$\xi[k],\ k \in \overline{0,N},\ Q(\xi)$$

- Вспомогательная рандомизированная модель

$$\hat{z}[k] = \Omega(\mathbf{a},X),\ P(\mathbf{a})$$

входные данные

$$X,\ k = \overline{0,N}$$

- Обучение модели

$$P^*(\mathbf{a}),\ Q^*(\xi)$$

- Генерация ансамблей на интервале пропусков $[k^-, k^+]$

- Вычисление характеристик ансамбля (средняя, медиана, дисперсионная трубка, ...)

1. *Дубнов Ю.А., Сокол Е.С., Полищук В.М., Попков Ю.С., Полищук Ю.М., Мельников А.В.*. Метод энтропийно-рандомизированного восстановления пропущенных данных // АиТ, в печати.

4.1. Прямое и обратное проектирование

- Матрица данных «объекты m, признаки s»

$$U_{m\times s}$$

- Прямое проектирование

$$Y_{m\times r} = U_{m\times s}\boxed{Q_{s\times r}},\ \boxed{r < s}$$

- Обратное проектирование $X_{m\times s} = Y_{m\times r}\boxed{T_{r\times s}}$

- Оптимизация матриц проекторов Q, T $\Rightarrow$

- минимизация «расстояния» $\widehat{U_{m\times s}, X_{m\times s}}$

- минимизация по Q максимального по T «расстояния» $\widehat{U_{m\times s}, X_{m\times s}}$

1. *Попков Ю.С., Попков А.Ю., Дубнов Ю.А.* Кросс-энтропийная редукция матрицы данных с ограничением информационной емкости матриц-проекторов и их норм // Математическое моделирование, 2020, Т. 32, № 9, С. 35-52.

4.2. Рандомизированное проектирование

- Отображение матрицы данных в точки признакового пространства

$$U_{m\times s} = \begin{pmatrix} \mathbf{u}^{(1)} \\ \vdots \\ \mathbf{u}^{(m)} \end{pmatrix}, \ \mathbf{u}^{(j)} \in R^s$$

- Индикатор компактности

$$\rho_U = \frac{2}{m(m-1)} \sum_{\substack{(\alpha,\beta)=1 \\ \alpha<\beta}}^{m} \rho(\mathbf{u}^{(\alpha)}, \mathbf{u}^{(\beta)})$$

$\rho(\mathbf{u}^{(\alpha)}, \mathbf{u}^{(\beta)})$ — расстояние

- Случайные матрицы–проекторы

$$Q_{s\times r}$$
$$Q_{s\times r} \in \mathcal{Q} = [Q^-, Q^+]$$
$$P(Q)$$

$$T_{n\times m}$$
$$T_{n\times m} \in \mathcal{T} = [T^-, T^+]$$
$$W(T)$$

4.2. Рандомизированное проектирование

Классификация преобразований

- Проектирование на $R^r_{(m)}$

$$Y_{m\times r} = \begin{pmatrix} \mathbf{y}^{(1)} \\ \vdots \\ \mathbf{y}^{(r)} \end{pmatrix} = U_{m\times s} Q_{s\times r}$$

$$\rho_Y(Q) = \frac{2}{m(m-1)} \sum_{\alpha<\beta}^{m} \rho(\mathbf{y}^{(\alpha)}, \mathbf{y}^{(\beta)} | Q)$$

- Проектирование на $R^s_{(n)}$

$$B_{n\times s} = \begin{pmatrix} \mathbf{b}^{(1)} \\ \vdots \\ \mathbf{b}^{(n)} \end{pmatrix} = T_{n\times m} U_{m\times s}$$

$$\rho_B(T) = \frac{2}{m(m-1)} \sum_{\alpha<\beta}^{m} \rho(\mathbf{b}^{(\alpha)}, \mathbf{b}^{(\beta)} | T)$$

- Проектирование на $R^r_{(n)}$

$$Z_{n\times r} = \begin{pmatrix} \mathbf{z}^{(1)} \\ \vdots \\ \mathbf{z}^{(n)} \end{pmatrix} = T_{n\times m} U_{m\times s} Q_{s\times r}; \ \rho_Z(T,Q) = \frac{2}{m(m-1)} \sum_{\alpha<\beta}^{n} \rho(\mathbf{z}^{(\alpha)}, \mathbf{z}^{(\beta)} | T,Q)$$

- Оптимизация

$$P^*(Q),\ W^*(T),\ F^*(T,Q)$$

1. *Попков Ю.С., Дубнов Ю.А., Попков А.Ю.* Энтропийно-рандомизированное проектирование // АиТ, 2021, в печати.

- Обученная модель — генератор ансамблей прогнозных траекторий

$$\hat{\mathbf{z}}[k] = \Omega(\mathbf{a}, P^*(\mathbf{a}), X_{pred}), \quad X_{pred} \text{ — прогнозируемый вход}$$

- Обученная вспомогательная модель — генератор X_{pred}

$$X_{pred} = \mathcal{B}(\mathbf{b}, \zeta, W^*(\mathbf{b}, \zeta)), \quad \zeta \text{ — случайная последовательность}$$

- Генерация ансамбля

$$\mathcal{L} = \{\hat{z}[k^*], \dots, \hat{z}[M]\}, \quad [z^*, M] \text{ — интервал проноза}$$

1. *Y.S. Popkov, A.Y. Popkov, Y.A. Dubnov, D. Solomatine*. Entropy-Randomized Forecasting of Stochastic Dynamic Regression Models // Mathematics, 2020, V. 8, p. 1119, doi:10.3390/math8071119.

$$\mathcal{H}[P(\mathbf{a}, Q(\xi))] = -\int\limits_{\mathcal{A}} P(\mathbf{a}) \ln P(\mathbf{a}) d\mathbf{a} - \sum_{j=1}^{N} \int\limits_{\Xi_j} Q_j(\xi[j]) \ln Q_j(\xi[j]) d\xi[j] \Rightarrow \max$$

- нормировка

$$\int\limits_{\mathcal{A}} P(\mathbf{a}) d\mathbf{a} = 1; \quad \int\limits_{\div_{|}} Q_j(\xi[j]) d\xi[j] = 1; \quad j = \overline{1,N}$$

- эмпирические балансы

$$\int\limits_{\mathcal{A}} P(\mathbf{a}) \Omega(\mathbf{a}, X) d\mathbf{a} + \int\limits_{\div_{|}} Q_j(\xi[j]) \xi[j] d\xi[j] = \mathbf{z}[j]; \quad j = \overline{1,N}$$

Решение

$$P^*(\mathbf{a}) = \frac{\exp(-\langle\theta, \Omega(\mathbf{a}, X)\rangle)}{\mathcal{P}(\theta)}, \quad Q_j(\xi[j]) = \frac{\exp(-\theta_j, \xi[j])}{\mathcal{Q}_j(\theta_j)}, \quad j = \overline{1,N}$$

$$\mathcal{P}(\theta) = \int\limits_{\mathcal{A}} \exp(-\langle\theta, \Omega(\mathbf{a}, X)\rangle) d\mathbf{a}, \quad \mathcal{Q}_j(\theta_j) = \int\limits_{\Xi_j} \exp(-\theta_j, \xi[j]) d\xi[j]$$

Множители Лагранжа

$$\mathcal{P}^{-1}(\theta) \int\limits_{\mathcal{A}} \Omega(\mathbf{a}, X) \exp(-\langle\theta, \Omega(\mathbf{a}, X)\rangle) d\mathbf{a} +$$
$$+\mathcal{Q}_j^{-1}(\theta_j) \int\limits_{\Xi_j} \xi[j] \exp(-\theta_j, \xi[j]) d\xi[j] = \mathbf{z}[j], \quad j = \overline{1,N}$$

6. Математическая задача и ее свойства

Оценки максимальной энтропии

$$P^*(\mathbf{a}) = P^*(\mathbf{a},\theta(\mathbf{z}_{(N)})), \quad Q_j^*(\xi[j]) = Q_j^*(\xi[j],\theta_j(\mathbf{z}_{(N)}))$$

$$\mathbf{z}_{(N)} = \{z_1, \ldots, z_N\} \text{ — данные}$$

Ассимптотическая эффективность

$$\lim_{N\to\infty} P^*(\mathbf{a},\theta(\mathbf{z}_{(N)})) = P^*(\mathbf{a},\theta^*)$$

$$\lim_{N\to\infty} Q_j^*(\xi[j],\theta_j(\mathbf{z}_{(N)})) = Q_j^*(\xi[j],\theta_j^*)$$

$$\theta^* = \lim_{N\to\infty} \theta(\mathbf{z}_{(N)})$$

1. *Попков Ю.С.* Ассимптотическая эффективность оценок максимальной энтропии // ДАН, 2020, Т. 493. С. 104-107.

7. Приложения (рандомизированное прогнозирование)

- Динамика мирового населения

Y.S. Popkov, Y.A. Dubnov, A.Y. Popkov. New Method of Randomized Forecasting Using Entropy-Robustation to the World Population Prediction. Mathematics, 2016, V. 4, doi:10.3390/math4010016.

- Суточная электрическая нагрузка энергетической системы

Попков Ю.С., Попков А.Ю., Дубнов Ю.А. Элементы рандомизированного прогнозирования и его применение для предсказания суточной электрической нагрузки энергетической системы // АиТ, 2020, № 7, С. 148-172.

- Прогнозирование эволюции площади термокарстовых озер

Дубнов Ю.А., Полищук В.Ю., Попков А.Ю., Сокол Е.С., Мельников А.В., Полищук Ю.М., Попков Ю.С. Энтропийно-рандомизированное моделирование и прогнозирование эволюции площади термокарстовых озер в зонах вечной мерзлоты. Планируется в Environment Modelling and Software, 2021, в печати.

СПАСИБО
ЗА
ВНИМАНИЕ