École Nationale Supérieure de Techniques Avancées examen du cours SOD333 "Filtrage bayésien et approximation particulaire"

vendredi 22 octobre 2021, 13:30 à 16:00

On rappelle l'équation récurrente

$$\mu_{k-1} \xrightarrow{\text{prédiction}} \eta_k = \mu_{k-1}\, Q_k \xrightarrow{\text{correction}} \mu_k = g_k \cdot \eta_k \ ,$$

vérifiée par le filtre bayésien, avec

$$\eta_k(dx') = \mu_{k-1}\, Q_k(dx') = \int_E \mu_{k-1}(dx)\, Q_k(x, dx') \ .$$

Si on remplace le filtre μ_{k-1} par son approximation particulaire

$$\mu_{k-1}^N = \sum_{i=1}^N w_{k-1}^i \ \delta_{\xi_{k-1}^i} \qquad \text{avec} \qquad \sum_{i=1}^N w_{k-1}^i = 1 \ ,$$

alors on obtient le mélange fini

$$\mu_{k-1}^N\, Q_k(dx') = \sum_{i=1}^N w_{k-1}^i \ Q_k(\xi_{k-1}^i, dx') \ ,$$

comme approximation du prédicteur η_k, et la distribution de probabilité

$$g_k(x') \ [\sum_{i=1}^N w_{k-1}^i \ Q_k(\xi_{k-1}^i, dx')\,] \ ,$$

définie à une constante de normalisation près, comme approximation du filtre μ_k.

Désormais, on considère la distribution de probabilité cible $\mu(dx')$ sur E définie par la factorisation, ou décomposition d'importance

$$\mu(dx') \propto \underbrace{g_k(x')}_{g(x')} \ \underbrace{\sum_{i=1}^N w_{k-1}^i \ Q_k(\xi_{k-1}^i, dx')}_{\eta(dx')} \ , \qquad (\star)$$

qui permet d'interpréter la distribution de probabilité cible $\mu(dx')$ comme une distribution de Gibbs–Boltzmann, et l'objectif de ce problème est d'étudier différentes approximations particulaires pour la distribution de probabilité cible $\mu(dx')$ et pour la constante de normalisation $\langle \eta, g \rangle$.

(i) **Décrire l'algorithme d'approximation par échantillonnage pondéré de la distribution de probabilité cible $\mu(dx')$, basé sur la décomposition d'importance $(\star)$.**

Comment les particules sont–elles générées ? Comment les poids sont–ils calculés ? Comment la constante de normalisation $\langle \eta, g \rangle$ est–elle approximée ?

(ii) **Quelle est la variance (non–asymptotique) de l'erreur d'estimation de la constante de normalisation $\langle \eta, g \rangle$ par son approximation particulaire ?**

On propose d'étudier de nouvelles approximations particulaires pour la distribution de probabilité cible $\mu(dx')$ et pour la constante de normalisation $\langle \eta, g \rangle$, en appliquant le paradigme suivant :

> Si on peut interpréter la distribution de probabilité cible comme la distribution de probabilité marginale sur E d'une distribution de probabilité jointe définie sur un espace produit, par exemple $\{1, \cdots, N\} \times E$ ou $\{1, \cdots, N\} \times E \times E$, alors on peut obtenir une approximation particulaire de la distribution de probabilité cible de la manière suivante : on construit par échantillonnage pondéré une approximation particulaire de la distribution de probabilité jointe, puis on marginalise sur E cette approximation particulaire.

Notations

De même qu'on utilise indifféremment la notation

$$\mu,\ \eta,\ \text{etc.} \qquad [\text{respectivement } g_k,\ g,\ \text{etc.}]$$

ou la notation

$$\mu(dx'),\ \eta(dx'),\ \text{etc.} \qquad [\text{respectivement } g_k(x'),\ g(x'),\ \text{etc.}]$$

pour désigner des distributions de probabilité [respectivement des fonctions] définies sur E, de même on utilisera indifféremment la notation

$$\mu_{\text{aux}},\ \eta_{\text{aux}},\ \text{etc.} \qquad [\text{respectivement } g_{\text{aux}},\ \text{etc.}]$$

ou la notation

$$\mu^a_{\text{aux}}(dx'),\ \eta^a_{\text{aux}}(dx'),\ \text{etc.} \qquad [\text{respectivement } g^a_{\text{aux}}(x'),\ \text{etc.}]$$

pour désigner des distributions de probabilité [respectivement des fonctions] définies sur l'espace produit $\{1, \cdots, N\} \times E$, et de même on utilisera indifféremment la notation

$$\mu_{\rm new},\ \eta_{\rm new},\ \text{etc.} \qquad [\text{respectivement } g_{\rm new},\ \text{etc.}]$$

ou la notation

$$\mu^a_{\rm new}(dz, dx'),\ \eta^a_{\rm new}(dz, dx'),\ \text{etc.} \qquad [\text{respectivement } g^a_{\rm new}(z, x'),\ \text{etc.}]$$

pour désigner des distributions de probabilité [respectivement des fonctions] définies sur l'espace produit $\{1, \cdots, N\} \times E \times E$.

Filtrage particulaire auxiliaire

Le point de départ du filtrage particulaire auxiliaire consiste à interpréter la distribution de probabilité cible $\mu(dx')$ comme la distribution de probabilité marginale sur E associée à la distribution de probabilité jointe

$$\mu^a_{\rm aux}(dx') \propto g_k(x')\ w^a_{k-1}\ Q_k(\xi^a_{k-1}, dx')\ ,$$

définie sur l'espace produit $\{1, \cdots, N\} \times E$.

(iii) **Vérifier que la distribution de probabilité marginale définie sur E à partir de la distribution de probabilité jointe $\mu^a_{\rm aux}(dx')$ définie sur l'espace produit $\{1, \cdots, N\} \times E$, coïncide avec la distribution de probabilité cible $\mu(dx')$.**

On introduit les poids auxiliaires $(\lambda^1_{\rm aux}, \cdots, \lambda^N_{\rm aux})$, qui forment un vecteur de probabilité de dimension N, ce qui permet de définir la factorisation, ou décomposition d'importance

$$\mu^a_{\rm aux}(dx') \propto \underbrace{g_k(x')\ \frac{w^a_{k-1}}{\lambda^a_{\rm aux}}}_{g^a_{\rm aux}(x')}\ \underbrace{\lambda^a_{\rm aux}\ Q_k(\xi^a_{k-1}, dx')}_{\eta^a_{\rm aux}(dx')}\ , \qquad (\star\star)$$

et d'interpréter la distribution de probabilité $\mu^a_{\rm aux}(dx')$ comme une distribution de Gibbs–Boltzmann.

(iv) **Montrer que**

$$\langle \eta_{\rm aux}, g_{\rm aux} \rangle = \langle \eta, g \rangle\ ,$$

c'est–à–dire que la constante de normalisation $\langle \eta_{\rm aux}, g_{\rm aux} \rangle$ associée à la décomposition d'importance $(\star\star)$ coïncide avec la constante de normalisation $\langle \eta, g \rangle$ associée à la décomposition d'importance $(\star)$.

(v) **Décrire l'algorithme d'approximation par échantillonnage pondéré de la distribution de probabilité $\mu_{\text{aux}}^a(dx')$, basé sur la décomposition d'importance $(\star\star)$.**

Comment les particules sont–elles générées ? Comment les poids sont–ils calculés ? Comment la constante de normalisation $\langle \eta_{\text{aux}}, g_{\text{aux}} \rangle$ est–elle approximée ? Comment la distribution de probabilité cible $\mu(dx')$ est–elle approximée, après marginalisation ?

Pour chaque vecteur de probabilité $(\lambda_{\text{aux}}^1, \cdots, \lambda_{\text{aux}}^N)$, on obtient une décomposition d'importance de la distribution de probabilité $\mu_{\text{aux}}^a(dx')$ et une approximation particulaire par échantillonnage pondéré.

(vi) **Montrer que la classe d'approximations particulaires décrite à la question (v) contient l'approximation particulaire décrite à la question (i) comme cas particulier.**

[Indice : Pour un choix particulier du vecteur de probabilité $(\lambda_{\text{aux}}^1, \cdots, \lambda_{\text{aux}}^N)$, l'approximation particulaire décrite à la question (v) coïncide avec l'approximation particulaire décrite à la question (i).]

(vii) **Montrer que la variance (non–asymptotique) de l'erreur d'estimation de la constante de normalisation $\langle \eta_{\text{aux}}, g_{\text{aux}} \rangle$ par son approximation particulaire est égale à**

$$\frac{1}{N} \, [\sum_{a=1}^N \frac{|w_{k-1}^a|^2}{\lambda_{\text{aux}}^a} \, |u_a|^2 - \langle \eta, g \rangle^2 \,] \ ,$$

avec

$$u_a = \{ \int_E |g_k(x')|^2 \; Q_k(\xi_{k-1}^a, dx') \}^{1/2} \ ,$$

pour tout $a = 1, \cdots, N$.

(viii) **Pour quel choix particulier du vecteur de probabilité $(\lambda_{\text{aux}}^1, \cdots, \lambda_{\text{aux}}^N)$, la variance de l'erreur d'estimation étudiée à la question (vii) est–elle minimale ? Monter que cette variance minimale est égale à**

$$\frac{1}{N} \, [\, [\sum_{a=1}^N w_{k-1}^a \; u_a \,]^2 - \langle \eta, g \rangle^2 \,] \ .$$

Comment cette variance se compare–t–elle à la variance obtenue à la question (ii) ?

Le point de départ de ce nouveau concept de filtrage particulaire auxiliaire consiste à introduire un noyau de probabilités de transition auxiliaire de l'ensemble fini $\{1, \cdots, N\}$ vers E, c'est–à–dire un vecteur $(\pi^1_{\rm new}(dz), \cdots, \pi^N_{\rm new}(dz))$ de distributions de probabilité sur E, et à interpréter la distribution de probabilité cible $\mu(dx')$ comme la distribution de probabilité marginale sur E associée à la distribution de probabilité jointe

$$\mu^a_{\rm new}(dz, dx') \propto g_k(x')\; w^a_{k-1}\; \pi^a_{\rm new}(dz)\; Q_k(\xi^a_{k-1}, dx') \;,$$

définie sur l'espace produit $\{1, \cdots, N\} \times E \times E$.

(ix) **Vérifier que la distribution de probabilité marginale définie sur E à partir de la distribution de probabilité jointe $\mu^a_{\rm new}(dz, dx')$ définie sur l'espace produit $\{1, \cdots, N\} \times E \times E$, coïncide avec la distribution de probabilité cible $\mu(dx')$.**

On suppose désormais que les noyaux de probabilités de transition $Q_k(x, dx')$ possèdent une densité, i.e. on suppose que

$$Q_k(x, dx') = q_k(x, x')\; dx \;,$$

ce qui permet de définir la factorisation, ou décomposition d'importance

$$\mu^a_{\rm new}(dz, dx') \propto \underbrace{g_k(x')\; \frac{q_k(\xi^a_{k-1}, x')}{q_k(z, x')}\; \frac{w^a_{k-1}}{\lambda^a_{\rm new}}}_{g^a_{\rm new}(z, x')}\; \underbrace{\lambda^a_{\rm new}\; \pi^a_{\rm new}(dz)\; Q_k(z, dx')}_{\eta^a_{\rm new}(dz, dx')} \;, \qquad (\star\star\star)$$

et d'interpréter la distribution de probabilité $\mu^a_{\rm new}(dz, dx')$ comme une distribution de Gibbs–Boltzmann.

(x) **Montrer que**

$$\langle \eta_{\rm new}, g_{\rm new} \rangle = \langle \eta, g \rangle \;,$$

c'est–à–dire que la constante de normalisation $\langle \eta_{\rm new}, g_{\rm new} \rangle$ associée à la décomposition d'importance $(\star\star\star)$ coïncide avec la constante de normalisation $\langle \eta, g \rangle$ associée à la décomposition d'importance $(\star)$.

(xi) **Décrire l'algorithme d'approximation par échantillonnage pondéré de la distribution de probabilité $\mu^a_{\rm new}(dz, dx')$, basé sur la décomposition d'importance $(\star\star\star)$.**

Comment les particules sont–elles générées ? Comment les poids sont–ils calculés ? Comment la constante de normalisation $\langle \eta_{\rm new}, g_{\rm new} \rangle$ est–elle approximée ? Comment la distribution de probabilité cible $\mu(dx')$ est–elle approximée, après marginalisation ?

Pour chaque vecteur $(\pi^1_{\text{new}}(dz), \cdots, \pi^N_{\text{new}}(dz))$ de distributions de probabilité sur E et pour chaque vecteur de probabilité $(\lambda^1_{\text{new}}, \cdots, \lambda^N_{\text{new}})$, on obtient une décomposition d'importance de la distribution de probabilité $\mu^a_{\text{new}}(dz, dx')$ et une approximation particulaire par échantillonnage pondéré.

(xii) **Montrer que la classe d'approximations particulaires décrite à la question (xi) contient la classe d'approximations particulaires décrite à la question (v), et a fortiori contient l'approximation particulaire décrite à la question (i) comme cas particulier.**

[Indice : Pour un choix particulier du vecteur $(\pi^1_{\text{new}}(dz), \cdots, \pi^N_{\text{new}}(dz))$ de distributions de probabilité sur E, l'approximation particulaire décrite à la question (xi) coïncide avec l'approximation particulaire décrite à la question (v).]

(xiii) **Montrer que la variance (non–asymptotique) de l'erreur d'estimation de la constante de normalisation $\langle \eta_{\text{new}}, g_{\text{new}} \rangle$ par son approximation particulaire est égale à**

$$\frac{1}{N} \, [\sum_{a=1}^N \frac{|w^a_{k-1}|^2}{\lambda^a_{\text{new}}} \int_E |u_a(z)|^2 \, \pi^a_{\text{new}}(dz) - \langle \eta, g \rangle^2 \,] \ ,$$

avec

$$u_a(z) = \{ \int_E |g_k(x')|^2 \, |\frac{q_k(\xi^a_{k-1}, x')}{q_k(z, x')}|^2 \, Q_k(z, dx') \}^{1/2} \ ,$$

pour tout $a = 1, \cdots, N$ **et pour tout** $z \in E$**.**

Dans le cas particulier où $z = \xi^a_{k-1}$**, montrer que** $u_a(\xi^a_{k-1}) = u_a$ **expression déjà vue à la question (vii), pour tout** $a = 1, \cdots, N$**.**

(xiv) **Pour quel choix particulier du vecteur de probabilité** $(\lambda^1_{\text{aux}}, \cdots, \lambda^N_{\text{aux}})$**, et du vecteur** $(\pi^1_{\text{new}}(dz), \cdots, \pi^N_{\text{new}}(dz))$ **de distributions de probabilité sur** E**, la variance de l'erreur d'estimation étudiée à la question (xiii) est–elle minimale ? Montrer que cette variance minimale est égale à**

$$\frac{1}{N} \, [\, [\sum_{a=1}^N w^a_{k-1} \, u^\bullet_a \,] - \langle \eta, g \rangle^2 \,] \ ,$$

où $u^\bullet_a = \min_{z \in E} u_a(z)$ **pour tout** $a = 1, \cdots, N$**.**

Comment cette variance se compare–t–elle à la variance obtenue à la question (vii) ?

[Indice : On supposera pour simplifier qu'il existe un unique minimiseur $z^\bullet_a$ pour lequel la valeur minimum $u^\bullet_a$ est atteinte.]