Amplification de la Distance Contextuelle dans les Langages de programmation probabilistes d'ordre supérieur

Présenté par:
Houssein MANSOUR

Encadrante:
Raphaëlle CRUBILLÉ

Stage de M2 IMD
LIS, Équipe LSC

Juin 2024

Introduction

Le programme peut faire des choix probabilistes à tout moment pendant son exécution.

Exemple : **(Marche Aléatoire)**

$$n-1 \xleftarrow{(1-p)} n \xrightarrow{p} n+1$$

Algorithmes plus rapides

Par exemple, les algorithmes de tri randomisés :

- **Tri rapide** :
  - Complexité temporelle : $O(n^2)$.
- **Tri rapide randomisé** :
  - Complexité en moyenne : $O(n \log n)$.

Langages de Programmation Probabilistes O.S

Les fonctions peuvent être :

- passées en argument ;
- retournées comme résultat.

Exemples : Langages d'ordre supérieur :

- Haskell, ML, Java, Python.

Exemple de langage de programmation Probabiliste (O.S)

PCF$_\oplus$: un lambda calcul typé avec un opérateur de point fixe.

Types dans PCF$_\oplus$: $\sigma, \tau \in \mathcal{T}_\oplus ::= \mathrm{Nat} \mid \sigma \to \tau$.

Termes de PCF$_\oplus$:

$$M ::= x \mid n \mid \mathrm{succ}(M) \mid \mathrm{if}(M, N, L) \mid \mathrm{fix}\, M$$
$$\mid \lambda x.M \mid MN \mid M \oplus N, \quad \text{où} \;\; x \in V, \;\; n \in \mathbb{N}.$$

- $\lambda y.\lambda x_1.\lambda x_2.if(x_1 \oplus x_2, y\,\underline{0}, y\,\underline{0}) : (\mathrm{Nat} \to \tau) \to \mathrm{Nat} \to \mathrm{Nat} \to \tau$.

Équivalence Contextuelle

Équivalence Contextuelle[1]

- **Contextes** : programmes avec trous dans le même langage.
- **Observable** : ce qu'on peut observer sur l'exécution.
- **Équivalence contextuelle** : $M \sim N$ si $\forall C$ (contexte), $\mathrm{Obs}(C[M]) = \mathrm{Obs}(C[N])$.

Cas déterministe PCF :

- Type d'observables : Nat.
- Observer si le résultat de l'exécution est 0.

Cas probabiliste PCF$_\oplus$:

- Type d'observables : Nat.
- $\mathrm{Obs}(C[M]) = \mathrm{proba}(C[M]$ renvoie $0)$.

1. J. H. Morris Jr., Lambda-calculus models of programming languages, PhD thesis, Massachusetts Institute of Technology, 1969.

Distance Contextuelle

Exemple de motivation

La classe BPP : problèmes de décision résolus par une machine de Turing probabiliste en temps polynomial, avec une proba d'erreur $< \frac{1}{3}$.

La distance entre le programme et sa spécification $< \frac{1}{3}$.

Distance Contextuelle[2]

$d^{ctx}(M, N) = \sup_{C \text{ contexte}} |\mathrm{Obs}(C[M]) - \mathrm{Obs}(C[N])|$.

Est-ce que cette définition est raisonnable ?

2. R. Crubillé and U. Dal Lago. Metric reasoning about lambda-terms : The affine case. In Proc. of LICS, pages 633–644, 2015.

Trivialisation de la distance

Théorème (Crubillé et Dal Lago)[2]

Soit $\mathcal{L}$ un langage probabiliste où tous les programmes terminent, et où les contextes possèdent au moins les capacités suivantes :

- copier leur argument,
- calculer les fonctions construites à l'aide de la récursion primitive,

Alors $\forall (M, N), d^{\text{ctx}}(M, N) \in \{0, 1\}$.

Contribution

Il existe une profonde connexion entre ce résultat de trivialisation et la loi des grands nombres, qui stipule que la moyenne empirique d'une distribution de probabilité convérge vers son espérance.

$\mathbb{T}_{\oplus}$: Variante Probabiliste du Système $\mathbb{T}$.

Pourquoi $\mathbb{T}_\oplus$?

- Tout programme M termine avec probabilité 1. [3]
- Le système $\mathbb{T}_\oplus$ n'est pas Turing-complet, mais il est raisonnablement expressif. [3]

Syntaxe de $\mathbb{T}_\oplus$

Types dans $\mathbb{T}_\oplus$:

$$\sigma, \tau \in \mathcal{T}_\oplus ::= \mathsf{Nat} \mid \sigma \to \tau \mid \sigma \times \tau.$$

Terme de $\mathbb{T}_\oplus$:

$$M :: x \mid \underline{0} \mid succ(M) \mid \langle M, N \rangle \mid rec \mid \pi_1 \mid \pi_2 \mid M \oplus N \mid \lambda x.M \mid MN \mid \cancel{fix\ M}$$

3. F. Breuvart, U. D. Lago, and A. Herrou. On higher-order probabilistic subrecursion. 2017.

Exécution	$\mathbb{T}$	$\mathbb{T}_\oplus$
Une étape	$M \to N$	$M \to \alpha \cdot \{N_1\}^1 + \beta \cdot \{N_2\}^1 + \ldots$
$\to^*$	cloture ref. tran. $(\to)$	?
Sémantique	$\{V \text{ t.q. } M \to^* V\}$	?

Réduction en un pas (CBV)

Programme dans $\mathbb{T}_\oplus(P_{\mathbb{T}_\oplus})$: termes fermés.

Relation de réduction en un pas pour $\mathbb{T}_\oplus$, $\rightarrow \subseteq P_{\mathbb{T}_\oplus} \times \mathcal{D}(P_{\mathbb{T}_\oplus})$. [3]

Exemple :

- $(\lambda x.x)((\lambda x.succ(x))\,\underline{0}) \rightarrow \{(\lambda x.x)(succ(\underline{0}))\}^1$.
- $(\lambda x.x)(succ(\underline{0}) \oplus \underline{0}) \rightarrow \frac{1}{2}\{(\lambda x.x)(succ(\underline{0}))\}^1 + \frac{1}{2}\{(\lambda x.x)(\underline{0})\}^1$.

Sémantique Opérationnelle

Une relation pour $\mathbb{T}_\oplus$ par rapport à $\rightarrow$: $\rightarrow_{\mathcal{D}(P_{\mathbb{T}_\oplus})} \subseteq \mathcal{D}(P_{\mathbb{T}_\oplus}) \times \mathcal{D}(P_{\mathbb{T}_\oplus})$. [3]

Exemple :

- $\{(\lambda x.x)(succ(\underline{0}))\}^1 \rightarrow \{succ(\underline{0})\}^1$.
- $\frac{1}{2}\{(\lambda x.x)(succ(\underline{0}))\}^1 + \frac{1}{2}\{(\lambda x.x)(\underline{0})\}^1 \rightarrow \frac{1}{2}\{succ(\underline{0})\}^1 + \frac{1}{2}\{\underline{0}\}^1$.

Sémantique Opérationnelle

$[\![M]\!] = \left\{\text{l'unique } D, \text{ distribution sur } \mathcal{V}_{\mathbb{T}\oplus}, \text{ t.q. } \{M\}^1 \rightarrow^* D\right\}$.

Distance Contextuelle

- **Contextes de $\mathbb{T}_\oplus$**
  $C \in C_{\mathbb{T}_\oplus} ::= [\cdot] \mid \lambda x.C \mid C\,M \mid M\,C \mid \langle C, M\rangle \mid \langle M, C\rangle \mid C \oplus M \mid M \oplus C.$
- **Équivalence contextuelle :** $M_1 \equiv_\alpha M_2$
  $\forall C$ t.q. $[\cdot] : \alpha \vdash \mathcal{C} : \mathrm{Nat}$, $[\![\mathcal{C}[M_1]]\!](\underline{0}) = [\![\mathcal{C}[M_2]]\!](\underline{0})$.
- **Distance contextuelle entre M_1 et M_2**

$$(\delta^{\mathrm{ctx}}_{\mathbb{T}_\oplus})_\alpha(M_1, M_2) = \sup_{\mathcal{C} \text{ t.q. } [\cdot]:\alpha\vdash\mathcal{C}:\mathrm{Nat}} \{|[\![\mathcal{C}[M_1]]\!](\underline{0}) - [\![\mathcal{C}[M_2]]\!](\underline{0})|\}\,.$$

Théorème [2]

La distance contextuelle $\delta^{\mathrm{ctx}}_{\mathbb{T}_\oplus}$ trivialise.

Preuve (Crubillé et Dal Lago) [2]

Définition : famille de contextes d'amplification (f.c.a)

C_n est une f.c.a. pour $\alpha, \beta \in [0,1]$ avec $\alpha \neq \beta$ si :

- $[\cdot] : \mathrm{Nat} \vdash C_n : \mathrm{Nat}$.
- $\forall M_1, M_2$ avec $\vdash M_1, M_2 :$ Nat avec $[\![M_1]\!] =$ Bernoulli $(\alpha) := \alpha \cdot \{\underline{0}\}^1 + (1-\alpha) \cdot \{\underline{1}\}^1$ et $[\![M_2]\!] =$ Bernoulli (β),

$$\lim_{n\to\infty} |\mathrm{Obs}(C_n[M_1]) - \mathrm{Obs}(C_n[M_2])| = 1,$$

Trivialisation par Contextes d'Amplification

Proposition

L'existence d'une f.c.a. pour tout $\alpha, \beta \in [0,1]$ avec $\alpha \neq \beta$ implique que $\delta^{\text{ctx}}_{\mathbb{T}_\oplus}$ est trivial.

Les étapes de la preuve :

Pour $M_1 \not\equiv M_2$ avec $\vdash M_1, M_2 : \tau$,

- $\exists \mathcal{D}$ s.t. $[.] : \sigma \vdash \mathcal{D} : \text{Nat}$ tel que

$$|[\![\mathcal{D}[M_1]]\!](\underline{0}) - [\![\mathcal{D}[M_2]]\!](\underline{0})| > 0.$$

- $\mathcal{C}[\cdot] := \mathit{if}\ \mathcal{D}[\cdot]\ \underline{0}\ \underline{1}$.
- $\lim_{n\to\infty} |\text{Obs}(C_n[\mathcal{C}[M_1]]) - \text{Obs}(C_n[\mathcal{C}[M_2]])| = 1.$

Objectif

(C. et D.L)^2 : une f.c.a. construite par des techniques d'analyse réelle.

Notre **objectif** est de comprendre la trivialisation avec la théorie des proabilités .

Variables Aléatoires

Variables Aléatoires

- Une **variable aléatoire** est une fonction mesurable $X : \Omega \to E$ définie sur un espace probabilisé (Ω, μ) et prenant ses valeurs dans un espace mesurable E.
- **Loi d'une v.a.** $X : (\Omega, \mu) \to E$:

$$\mathrm{loi}(X)(B) = \mu(X^{-1}(B)) \quad \text{pour tout } B \in \mathcal{E}.$$

- **V.a. Indépendantes** :
  Soit X_1, X_2 deux v.a., $X_i : (\Omega, \mu) \to E$. Ces deux v.a. sont indépendantes si :

$$\mathrm{loi}(X_1, X_2) = \mathrm{loi}(X_1) \times \mathrm{loi}(X_2) \in \mathcal{D}(\mathcal{E}^2).$$

Loi faible de grands nombres

- **Espérance d'une v.a. discrète** Soit $X : (\Omega, \text{ Proba }) \to E$ une variable aléatoire, l'espérance mathématique de X est définie par :

$$\mathbb{E}(X) = \sum_{x \in E} x \cdot Proba(X = x),$$

loi faible de grands nombres

Soit (X_k) une suite de v. a. réelles indépendantes et de même loi.
$\forall \epsilon > 0$, on a

$$\lim_{k \to \infty} \text{Proba}(|\frac{1}{k} \sum_{i=1}^{k} X_i - \mathbb{E}(X_i)| \leq \epsilon) = 1.$$

Trivialisation en $\mathbb{T}_\oplus$

Contextes qui calculent la moyenne empirique

Proposition

$\forall e, \epsilon \in \mathbb{Q}_2$, $\exists$ une f.c. $C_k^{e,\epsilon}$ t.q. $\forall M$, avec $[\![M]\!] = Bernoulli(\alpha)$,

$$[\![\mathcal{C}_k^{e,\epsilon}[M]]\!]\ (\underline{0}) = \text{Proba}\left(\frac{X_1 + \cdots + X_k}{k} \in [e-\epsilon, e+\epsilon]\right).$$

où X_i sont des v.a.i.i. suivant la loi de Bernoulli avec paramètre $\mathbb{E}([\![M]\!])$.

L'idée de la preuve :

$$C_k^{e,\epsilon}[\cdot] = \lambda x.((\lambda y_1 \ldots \lambda y_k.\underline{f_k^{e,\epsilon}}(y_1, \ldots, y_k))\underbrace{(x\,\underline{0})\ldots(x\,\underline{0})}_{kfois})\lambda x.[\cdot].$$

$$f_k^{e,\epsilon}(n_1, n_2, \ldots, n_k) = \begin{cases} 0 & \text{si } \left|\frac{\sum_{i=1}^k n_i}{k} - e\right| \leq \epsilon, \\ 1 & \text{sinon} \end{cases}$$

$$\underline{f_k^{e,\epsilon}}(\underline{n_1}, \ldots, \underline{n_k}) \to^* \left\{\underline{f_k^{e,\epsilon}(n_1, \ldots, n_k)}\right\}^1.$$

Contextes d'amplification et loi faible

Lemme

Si $[\![M]\!] = \mathit{Bernoulli}(\alpha)$, alors

- $\lim_{k\to\infty} |\mathrm{Obs}(C_k^{e,\epsilon}[M]) = 1$ si $\alpha \in]e-\epsilon, e+\epsilon[$.
- $\lim_{k\to\infty} |\mathrm{Obs}(C_k^{e,\epsilon}[M]) = 0$ si $\alpha \notin [e-\epsilon, e+\epsilon]$.

L'idée de la preuve :

- $\mathrm{Obs}(C_k^{e,\epsilon}[M]) = \mathrm{Proba}\left(\frac{X_1+\cdots+X_k}{k} \in [e-\epsilon, e+\epsilon]\right)$.
- Loi faible de grands nombres.

Théorème

Pour $M_1 \not\equiv M_2$ deux programmes avec $[\![M_1]\!] = \mathit{Bernoulli}(e_1)$ et $[\![M_2]\!] = \mathit{Bernoulli}(e_2)$,

$$C_k^{e,\epsilon}[\cdot] = C_k^{e_1, \frac{|e_1 - e_2|}{3}}[\cdot]$$

forme une f.c.a. c.à.d.

$$\lim_{k \to \infty} |\mathrm{Obs}(C_k^{e_1,\epsilon}[M_1]) - \mathrm{Obs}(C_k^{e_1,\epsilon}[M_2])| = 1.$$

Conclusion et Perspectives

- **Conclusion** : Nouvelle preuve
  - Plus simple.
  - Purement probabiliste.
- **Perspectives** : Taux de convergence de la distance contextuelle
  - Résultats non bloquants pour $\delta^1(M, N)$, $\delta^2(M, N)$, $\ldots$
  - Taux de convergence de la loi des grands nombres.

Déroulement de stage :

- Étude bibliographique : lambda calcul probabiliste, distance contextuelle...
- Contribution originale : nouvelle preuve du trivialisation.
- Soumission les résultats à HOPE'2024 (Higher order programming with effects).
- Assister au sémnaires et écoles des recherches.
  1. Séminaires de l'equipe LDP-LSC.
  2. Rencontres "CHoCoLa" Curry-Howard : Calcul et Logique.
  3. Workshop DiλLL2024.
  4. École des Jeunes Chercheurs et Chercheuses en Informatique Mathématique, (GRIM).

Merci pour votre attention !