# پردازش تکاملی

برنامه ریزی تکاملی

**دانشگاه صنعتی مالک اشتر**

مجتمع دانشگاهی فن آوری اطلاعات و امنیت

زمستان ۱۳۹۲

# مروری کوتاه بر برنامه ریزی تکاملی

- توسعه اولیه در ایالات متحده آمریکا و در سال ۱۹۶۰
- اولین افراد توسعه دهنده این حوزهD. Fogel
- نوعا به منظور کاربردهای ذیل استفاده میگردید:
  - EP اولیه : کاربردهای یادگیری ماشین به وسیله ماشین های متناهی
  - EP معاصر: بهینه سازی عددی
- ویژگی ها:
  - چارچوب بسیار باز: هر گونه تغییر و تحول و عملیات های کامپیوتری در آن امکان پذیر میباشد.
  - پیوند خورده با ES میباشد.( در EP معاصر)
  - نهایتا: به سختی میتوان گفت چه چیزی به عنوان استاندارد EP می باشد.
- خصوصیات
  - عدم ترکیب
  - خود انطباقی درپارامترهای استاندارد(EP معاصر)

# متد اصلی

- فرم اصلی (L. J. Fogel)
  - جهش های تصادفی واحد
  - الفبای گسسته
  - انتخاب $(\mu+\mu)$
- برنامه ریزی تکاملی توسعه یافته(D. B. Fogel)
  - بهینه سازی پارامترهای پیوسته
  - شباهت هایی با ES
    - دگرگونی توزیع شده معمول
    - خود انطباقی در پارامترهای دگرگونی

# انواع EP

- EP استاندارد ( مبتنی بر ارائه)
- EP استاندارد متوالی
  - EP مشابه GA در حالت دائمی ( پایدار)
- متا EP ( state-of-the-art)
  - یکپارچه نمودن واریانس ها به منظور خود انطباقی
- متوالی( پیوسته)meta-EP
- Rmeta-EP
  - یکپارچه نمودن کو واریانس به منظور خود انطباقی

# خلاصه ای فنی از EP

| | |
|---|---|
| بردارهای مقدار دهی شده واقعی | معرفی |
| None | ترکیب مجدد |
| اختلالات گوسی | دگرگونی |
| قطعی | انتخاب والدین |
| احتمالی(μ+μ) | انتخاب بازماندگان |
| خود انطباقی در اندازه گامهای جهش (در متا EP) | اختصاصی |

# دورنمایی از EP

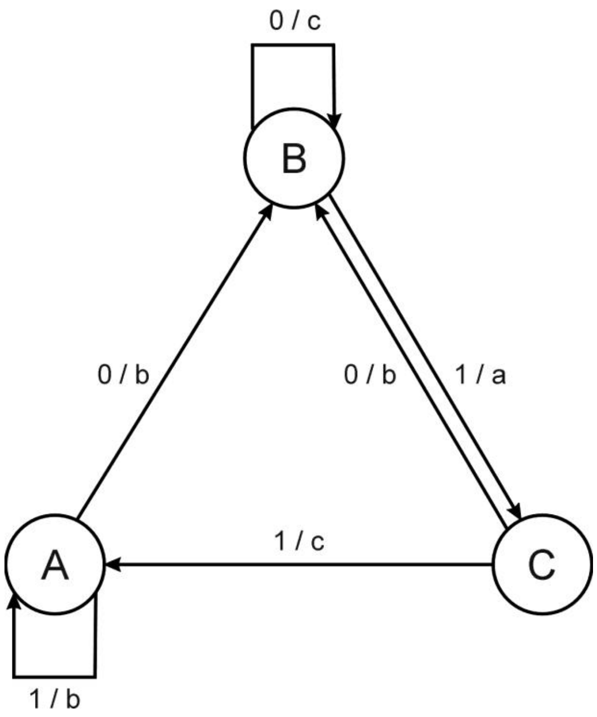
- EP در راستای هوشمند سازی روی کار آمد.
- هوشمندی در قالب رفتارهای قابل تطبیق دیده میشد.
- پیش بینی محیط به عنوان یک پیشنیاز در راستای رفتار قابل تطبیق در نظر گرفته میشد.
- بدین ترتیب: توانایی انجام پیشگویی یک کلید به منظور هوشمندی قلمداد میشود.

# پیشگویی به کمک ماشین های حالت متناهی

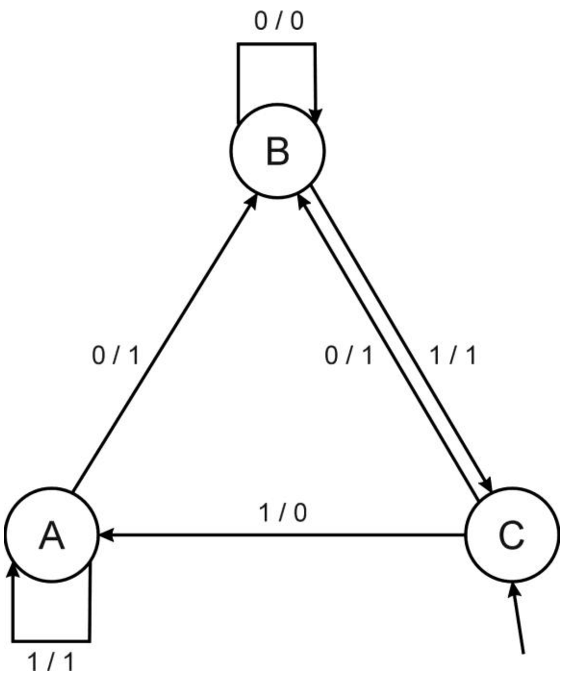
- ماشین حالت متناهی(FSM):
  - States S
  - Inputs I
  - Outputs O
  - تابع انتقال $\delta : S \times I \rightarrow S \times O$
  - انتقال رشته ورودی به رشته خروجی
  - میتواند به منظور پیشگویی مورد استفاده قرار گیرد. به صورتی که نماد ورودی بعدی در یک توالی را پیشگویی کند.

# مثال FSM

- یک FSM را در نظر بگیرد :
  - S = {A, B, C}
  - I = {0, 1}
  - O = {a, b, c}
  - δ given by a diagram



# FSM به عنوان پیشگو




- FSM زیر را در نظر بگیرید:
- وظیفه: پیشگویی ورودی بعدی
- Quality: % of $in_{(i+1)} = out_i$
- Given initial state C
- Input sequence 011101
- Leads to output 110111
- Quality: 3 out of 5

## مثال مقدماتی:
## به کارگیری FSM به منظور پیشگویی primeها

- $P(n) = 1$ if $n$ is prime, 0 otherwise
- $I = \boldsymbol{N} = \{1,2,3,\ldots, n, \ldots\}$
- $O = \{0,1\}$
- تصحیح پیش گویی: $out_i = P(in_{(i+1)})$
- تابع سازگاری:
  - ۱ امتیاز برای هر پیشگویی صحیح ورودی بعدی
  - ۰ امتیاز برای پیش گویی غیر صحیح
  - غرامت برای مرحل "بسیار زیاد"

# مثال مقدماتی:
# به کارگیری FSM به منظور پیشگویی primeها

- انتخاب والدین: هر FSM یکبار جهش یافته است.
- Parent selection: each FSM is mutated once
- عملگرهای جهش : ( یک انتخاب تصادفی)
  - تغییر یک نماد خروجی
  - تغییر یک مرحله انتقال (i.e. redirect edge)
  - اضافه نمودن یک مرحله
  - پاک کردن یک مرحله
  - تغییر مرحله آغازین
- انتخاب باقی مانده: (μ+μ)
- نتایج: در حال مناسب پس از ۲۰۲ ورودی بهترین FSM یک حالت دارد و دو خروجی آن ۰ می باشد. همیشه "not prime" را پیشگویی نموده است.

# جهش

- حالت عمومی (std. EP)

$$x_i' = x_i + \sqrt{\beta_i \cdot \Phi(\vec{x}) + \gamma_i} \cdot N_i(0,1)$$

  - پارمترهای خارجی $\beta_i, \gamma_i$ میبایست به منظور وظیفه ای خاص تطبیق داده شوند.
  - معمولا

$$x_i' = x_i + \sqrt{\Phi(\vec{x})} \cdot N_i(0,1)$$

  - مسئله
    - چنانچه حداقل های عمومی سازگار مقادیر صفر نباشد دستیابی به اهداف مورد نظر امکان پذیر نخواهد بود.
    - چنانچه مقادیر سازگاری بسیار بزرگ باشد جستجو معمولا به صورت گامهای تصادفی می باشد.
    - چنانچه کاربر اطلاع نسبی از محل مینیمم سراسری نداشته باشد میزان سازی پارامترها امکان پذیر نخواهد بود.

# مدرن EP

- ارائه از پیش تعریف شده ای در حالت عمومی وجود ندارد.
- بدین ترتیب: جهش از پیش تعریف شده ای ندارد. (must match representation)
- اغلب خود انطباقی را در پارامترهای جهت به کار میگیرد.
- در ادامه آن یک متغیر EP توسط ما ارائه خواهد شد( not the canonical EP)

# جهش

- In meta-EP $$x_i' = x_i + \sqrt{v_i} \cdot N_i(0,1)$$ $$v_i' = v_i + \sqrt{\zeta v_i} \cdot N_i(0,1)$$
  – for guaranteeing positive variances
  $$v_i' \leq 0 \Rightarrow v_i' := \varepsilon_\sigma > 0$$
  – difference with ES
    - delayed effect of strategy parameter changes
    - ability of zero selection pressure and no guarantee positive variance values

# ارائه

- برای بهینه سازی پارامترهای پیوسته
- کروموزوم ها از دو بخش تشکیل شده اند:
  - متغیرهای شیء: $x_1,\ldots,x_n$
  - اندازه گامهای جهش: $\sigma_1,\ldots,\sigma_n$
- اندازه کامل: $\langle x_1,\ldots,x_n, \sigma_1,\ldots,\sigma_n \rangle$

# جهش

- کروموزوم ها: $\langle x_1,\ldots,x_n, \sigma_1,\ldots,\sigma_n \rangle$
- $\sigma_i' = \sigma_i \cdot (1 + \alpha \cdot N(0,1))$
- $x'_i = x_i + \sigma_i' \cdot N_i(0,1)$
- $\alpha \approx 0.2$
- محدوده رول : $\sigma' < \varepsilon_0 \Rightarrow \sigma' = \varepsilon_0$
- دیگر گزینه های پیشنهاد شده و مورد آزمایش قرار گرفته:
  - رویه Lognormalهمانند ES
  - استفاده از واریانس به جای انحراف از معیار
  - جهش σ-last
  - دیگر توزیع ها ( Cauchy به جای گوسی)

# ترکیب مجدد

- None
- منطق: یک نقطه در فضای جستجو بر انواع مختلفی استوار میباشد، نه به صورت اختصاصی.
- not for an individual and there can be no crossover between species
- بحث و مناظره میان “mutation”و“crossover”
- اینگونه به نظر میرسد که هدف برنامه ریزی شده امروزه شایع تر شده است.

# ترکیب مجدد

- عدم ترکیب
  - جهش گوسی روش مناسب تری میباشد (Fogel and Atmar)
    - نه در همه موقعیت ها
  - زیست شناسی تکاملی
  - نقش متقاطع اغلب بیشتر مورد تایید بوده است.

- mutation-enhancing evolutionary optimization
- crossover-segregating defects
- the main point of view from researchers in the field of Genetic Algorithms

# انتخاب والدین

- هر کدام به صورت جداگانه یک فرزند به وسیله جهش ایجاد میکند.
- بدین ترتیب:
  - قطعی
  - به وسیله fitness جهت دار نشده است.

# انتخاب بازماندگان

- $P(t)$: μ parents, $P'(t)$: μ offspring
- جفت ها در قالب round-robin در رقابت میباشند.
  - هر راه حل X از $P(t) \cup P'(t)$ در مقابل q راه حل های انتخابی تصادفی دیگر مورد ارزیابی قرار میگیرد.
  - در هر مقایسه یک "win" تخصیص خواهد یافت چنانچه X از رقیب آن بهتر باشد.
  - راه حل های μ با بزرگترین تعداد winها ضمانت کننده والدین در نسل بعدی میباشند.
  - پارامتر q اجازه میزان سازی فشار انتخاب را میدهد به عنوان نمونه q=10

# انتخاب

- مسابقه انتخاب Q

$$s_{\{q\}} : I^{2\mu} \rightarrow I^{\mu},$$
$$s_{\{q\}}(P(t) \cup P'(t)) = P(t+1), q \in N$$

$$w_i = \sum_{j=1}^{q} \begin{cases} 1, \text{if } \Phi(\vec{a}_i) \leq \Phi(\vec{a}_{\chi_j}) \\ 0, otherwise \end{cases}$$

$$\chi_j \in \{1, \ldots, 2\mu\} \text{ a uniform random integer}$$

# پارامترهای استاندارد

- پارامترهای استاندارد
  - ارائه
    - std. EP

$$EP(\beta_i, \gamma_i, q, \mu)$$

    - meta-EP

$$mEP(\zeta, c, q, \mu)$$

  - default

$$mEP(6,25,10,200)$$

# تنظیمات پیش فرض پارامترهای خارجی

| Parameter | Occurs in: EP | Occurs in: meta-EP | Default |
|---|---|---|---|
| Range bounds $u_i, v_i$ | × | × | $u_i = -50$<br>$v_i = 50$ |
| Upper bound $c$ of $\sigma_i$ | | × | $c = 25$ |
| Proportionality constants $\beta_i$ | × | | $\beta_i = 1$ |
| Offset constants $\gamma_i$ | × | | $\gamma_i = 0$ |
| Meta-parameter $\zeta$ for self-adaptation | | × | $\zeta = 6$ |
| Tournament size $q$ | × | × | $q = 10$ |
| Population size $\mu$ | × | × | $\mu = 200$ |

**Table 2.2.** Default settings of exogenous parameters of standard- and meta-Evolutionary Programming.

# مثال کاربردی :
# Ackley(Bäck et al ’93) تابع

- تابع Ackley دراینجا با n=30

$$f(x) = -20 \cdot \exp\left(-0.2\sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)\right) + 20 + e$$

- ارائه:
  - $-30 < x_i < 30$ (coincidence of 30’s!)
  - 30 variances as step sizes
- ابتدا جهش به وسیله تغییر متغیرهای شیء
- اندازه جمعیت $\mu = 200$, انتخاب با $q = 10$
- خاتمه : پس از ۲۰۰۰۰۰ ارزیابی شایستگی
- نتایج: بهترین مقدار میانگین $1.4 \cdot 10^{-2}$

# تئوری

- تحلیل std. EP(Fogel)
  - aims at giving a proof of convergence with prob. 1 for resulting algorithm
  - جهش :

$$EP(1,0,q,\mu) \quad x_i' = x_i + \sqrt{\Phi(\vec{x})} \cdot N_i(0,1)$$

  - تحلیل یک مورد خاص EP(1,0,q,1)
    - identical to a (1+1)-ES having

$$\sigma = \sqrt{f(\vec{x})}$$

  - تابع هدف
    - ساده سازی حوزه مدل

$$\tilde{f}_2(x_i) = \sum_{i=1}^{n} x_i^2 = r^2$$

– ترکیب با SD بهینه

$$\sqrt{\tilde{f}_2(x_i)} = r = \frac{\sigma_2^* n}{1.224} \qquad \sigma_2^* = 1.224 \cdot r / n$$

- زمانی که ابعاد افزایش می یابد عملکرد نسبت به الگوریتیم که قادر است .opt SD را باز گرداند بدتر میشود.
- از Eq. (2.30)

$$\varphi_2 = \frac{\sigma_2}{\sqrt{2\pi}} \exp\left( -\frac{1}{8}\left(\frac{\sigma_2 n}{r}\right)^2 \right) - \frac{\sigma_2^2 n}{4r}\left(1 - \mathrm{erf}\left(\frac{\sigma_2 n}{r\sqrt{8}}\right)\right)$$

- نرخ کواریانس یک EP-(1+1)بوسیله

$$\tilde{\varphi}_2 = \varphi_2(\sigma_2 = r)$$

$$\tilde{\varphi}_2 = \left( \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{n^2}{8} \right) - \frac{n}{4}\left(1 - \mathrm{erf}\left(\frac{n}{\sqrt{8}}\right)\right)\right) \cdot r$$

# نسبت نرخ کوواریانس

- نسبت نرخ کوواریانس

$$\phi_2(n) := \frac{\tilde{\varphi}_2}{\varphi_2^*}$$

$$= \frac{n}{0.2025}\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{n^2}{8}\right)-\frac{n}{4}\left(1-\operatorname{erf}\left(\frac{n}{\sqrt{8}}\right)\right)\right)$$

  – با دقت بیشتر n*=1.48

$$\left.\frac{d\phi_2(n)}{dn}\right|_{n^*} = 0$$

# نسبت نرخ کوواریانس

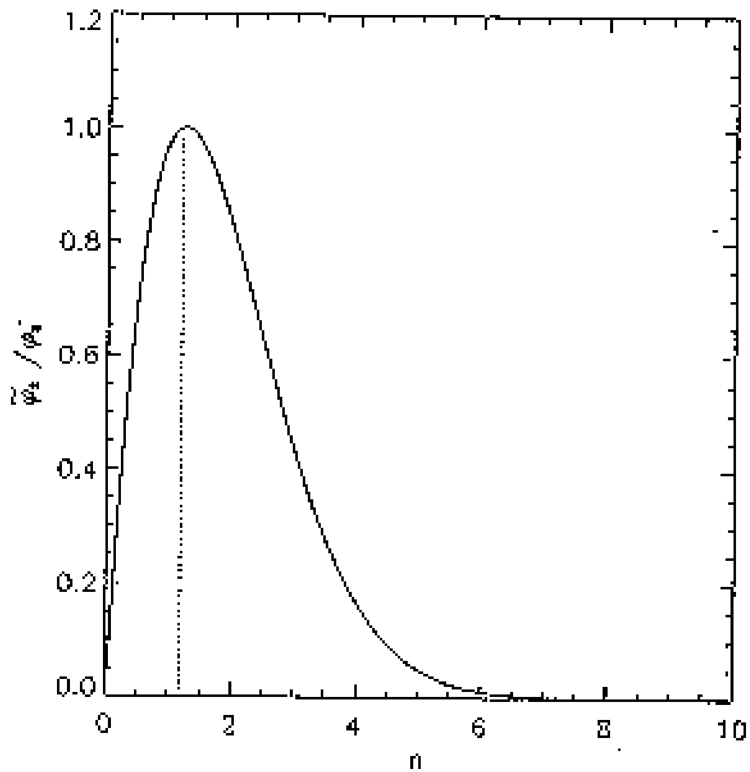



Fig. 2.7: Ratio of convergence rate for standard Evolutionary Programming ($\varphi_2$) and optimal convergence rate ($\varphi_2^*$) over the dimension ($n$) of the objective function sphere model.

- یک فاکتور مقیاس گذاری زمانی که 2<n

$$\beta_i = n^{-2}$$

– مورد دیگر

$$\sqrt{\frac{\tilde{f}_2(\vec{x})}{n^2}} = \frac{r}{n} = \frac{\sigma_2^*}{1.224}$$

- تفاوت کمتر میان ESوEP

# فرآیند مارکوف

- تعریف(Stochastic process, Markov process)

$$P\{X(t_k) \le x_k \mid X(t_{k-1}) = x_{k-1}, \ldots, X(t_1) = x_1\} =$$
$$\{X(t_k) \le x_k \mid X(t_{k-1}) = x_{k-1}\}$$

  – چنانچه تعداد مراحل $x_i$ متناهی است یا قابل شمارش نامتناهی است X(t) یک زنجیره مارکوف نامیده میشود.
  – احتمالات انتقال

$$p_{ij}(k) = P\{X_{k+1} = j \mid X_k = i\}$$

– زنجیره مارکوف همگن

- چانچه $p_{ij}(k)$ مستقل از K باشد
- ماتریس انتقال

$$\mathbf{P} = \left(p_{ij}\right)$$

– احتمالات انتقال l-step

$$p_{ij}^{(l)} = P\left\{X_{k+l} = j \mid X_k = i\right\}$$

- به وسیله معادلات چپمن کولموگروف

$$p_{ij}^{(m+l)} = \sum_k p_{ik}^{(m)} p_{kj}^{(l)}$$

– expressed by matrix multiplication

$$\mathbf{P}^{(m+l)} = \left(p_{ij}^{(m+l)}\right) = \mathbf{P}^m \mathbf{P}^l$$

– جذب مرحله ۱

$$p_{ij} = 1$$

- از زمانی که فرآیندها وارد یک مرحله جذب میشوند هرگز نمیتوانند آن فضا را ترک نمایند

– استدلال Fogel’s

– تنها مرحله جذب یک زنجیره مارکوف به عنوان یک واحد از تمامی مراحل است که شامل نقطه بهینه عمومی میباشد.

- تئوری
  - احتمال اینکه p1(k) وارد مرحله جذب در زمان K شوند به وسیله رابطه زیر بدست می آید:

$$p_1(k) = \left(\vec{P}(0)\mathbf{P}^k\right)_1 = \vec{P}(0)\begin{pmatrix} 1 \\ \mathbf{N}_k\mathbf{R} \end{pmatrix}$$

  - then

$$\lim_{k\to\infty} p_1(k) = \vec{P}(0)\begin{pmatrix} 1 \\ \cdot \\ \cdot \\ 1 \end{pmatrix} = 1$$

- Result based on elitist property of selection
  - guarantee
    - monotone behavior of evolution
    - existence of an absorbing state
    - discretization of the search space
    - ability to get anywhere in the search space either in one step or many
  - no guarantee
    - when

$$f : R^n \to R$$