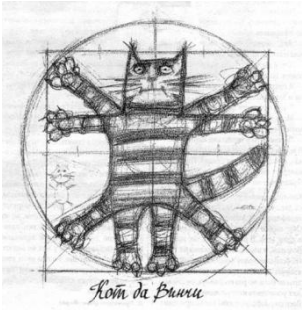


ПРИМЕНЕНИЕ АЛГОРИТМА БАЙЕСОВСКОЙ ОПТИМИЗАЦИИ ДЛЯ ОБУЧЕНИЯ ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ

Ю. Цой, yurytsoy@gmail.com

Томский политехнический университет

Томский государственный университет систем управления и радиоэлектроники

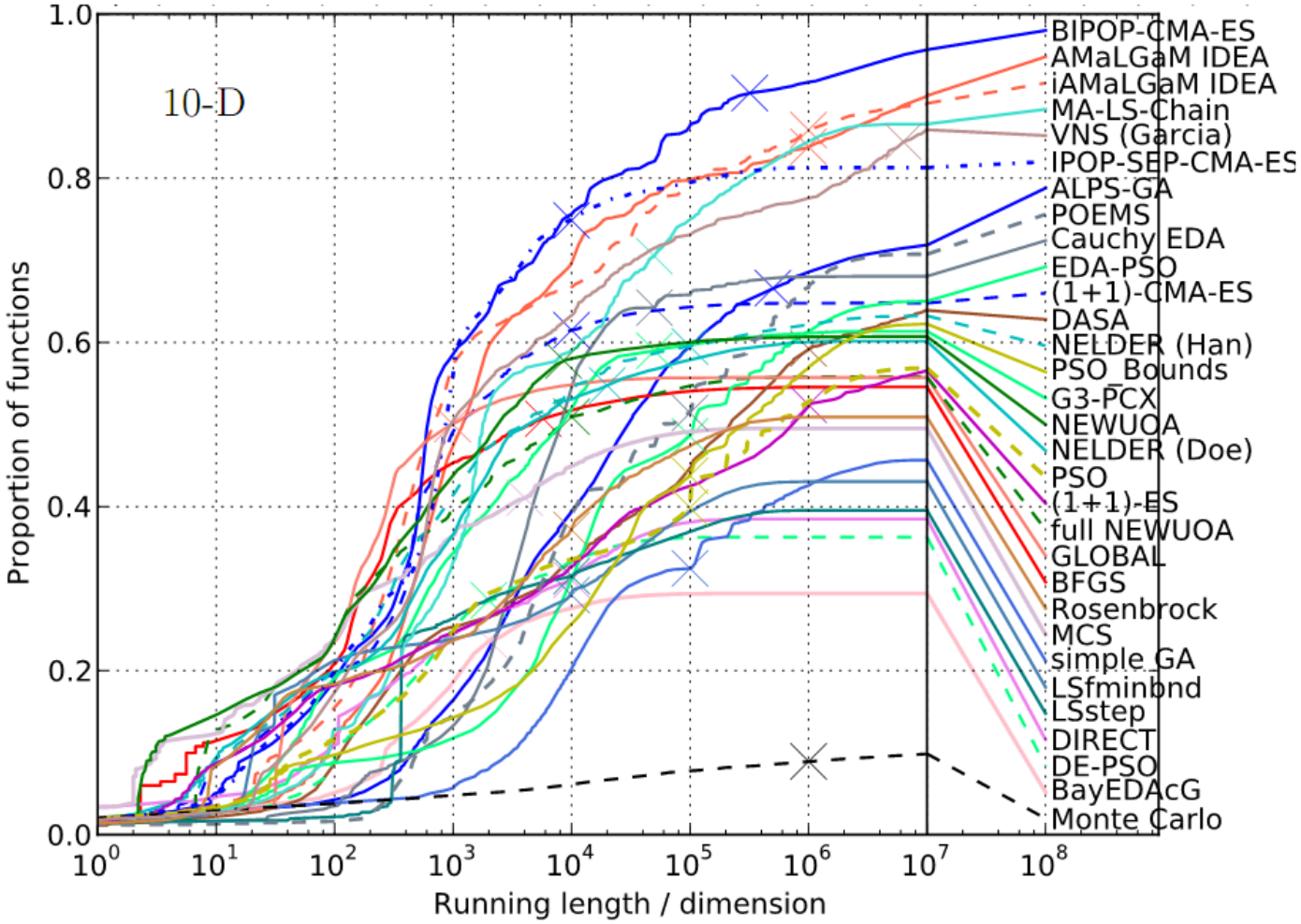
г. Томск

V Всероссийская конференция "Проблемы оптимизации и экономические приложения", Омск, 2-6 июля 2012г.

Содержание

- Алгоритмы оценки распределений и байесовской оптимизации.
- AMaLGaM IDEA.
- Декомпозиция весов нейронной сети.
- Коэволюционный алгоритм.
- Результаты экспериментов.
- Заключение.

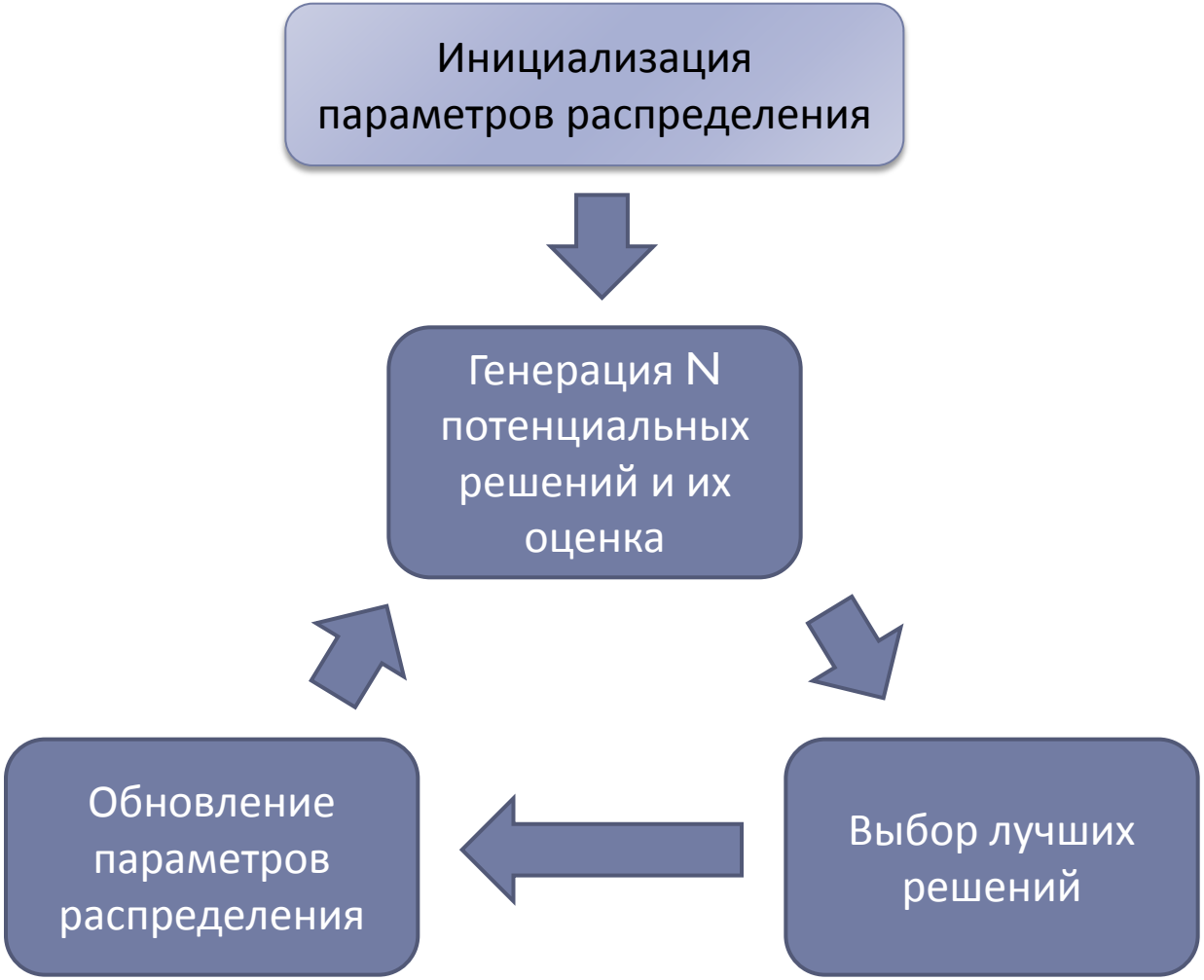
Результаты Black Box Optimization Benchmark-2009



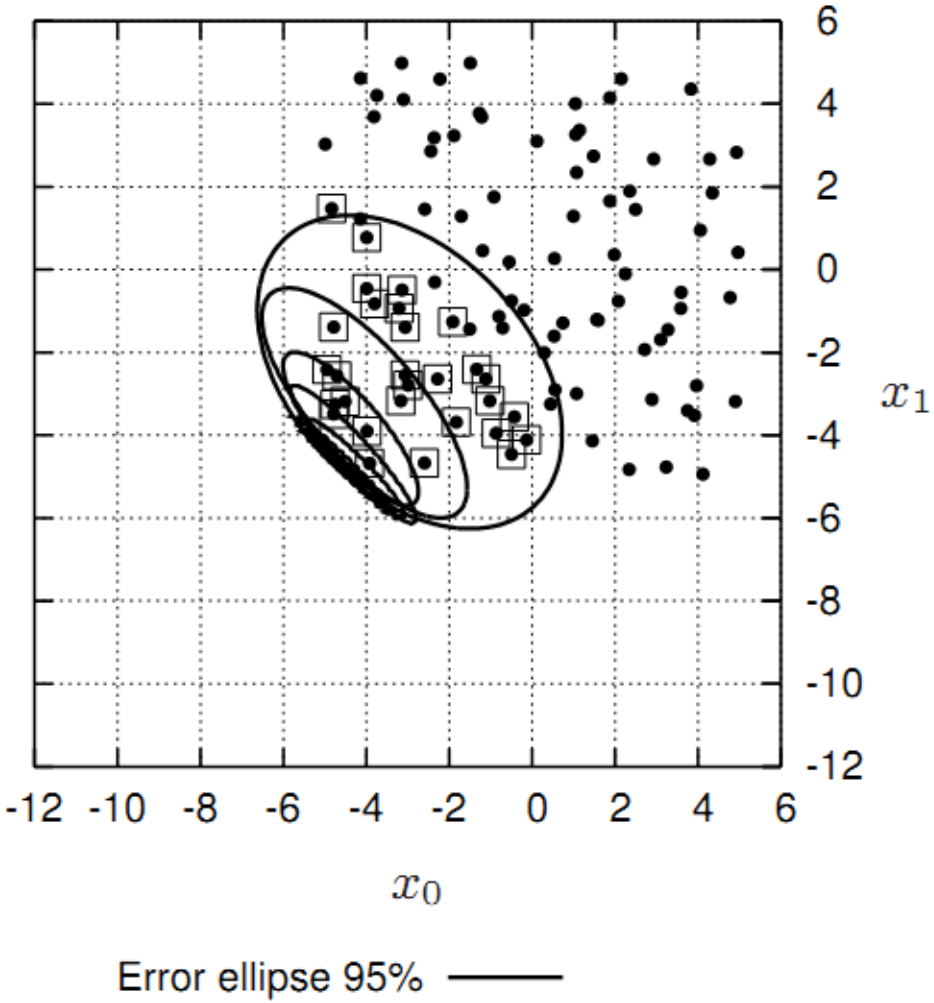
Алгоритмы с оценкой распределения

- M. Pelikan, F. Muhlenbein. ~1996-1997.
- Семейство алгоритмов оптимизации.
- «Подкласс» эволюционных алгоритмов.
- Поиск параметров распределения, описывающего области с высокой приспособленностью в пространстве переменных.

Алгоритмы с оценкой распределения



Алгоритмы с оценкой распределения



- Bosman P.A.N. , Grahl J., Thierens D. Adapted maximum-likelihood Gaussian models for numerical optimization with continuous EDAs. Tech report no. SEN-E0704. Dec. 2007.

Алгоритмы с оценкой распределения

- Одномерные

$$p(X_1, X_2, \ldots, X_n) = p(X_1)p(X_2)\ldots p(X_n)$$

- Двумерные

$$p(X) = p(X_{i_1} \mid X_{i_2})p(X_{i_2} \mid X_{i_3})\ldots p(X_{i_{n-1}} \mid X_{i_n})p(X_{i_n})$$

- Многомерные

$$p(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} p(X_i \mid \Pi_i)$$

Алгоритмы байесовской оптимизации

- Pelikan M., Goldberg D.E., Cantu-Paz E., 2000.
- Разновидность АОР с построением байесовской модели зависимостей между переменными.
- Модель часто строится с коррекцией на каждой итерации.

AMaLGaM IDEA (Bosman, 2007)

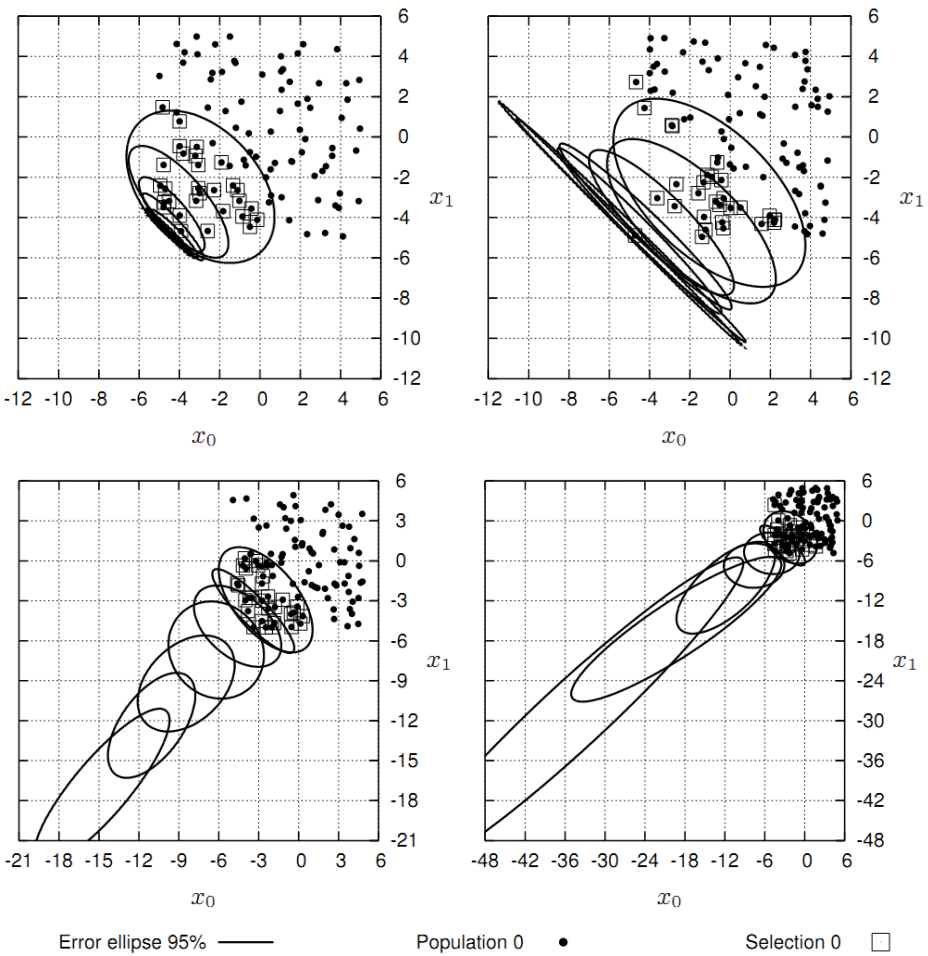
- Adaptive Variance Scaling (AVS)

$$\hat{\Sigma} \leftarrow c^{\text{Multiplier}} \hat{\Sigma}$$

- Anticipated Mean Shift (AMS)

$$\boldsymbol{x} \leftarrow \boldsymbol{x} + \delta \hat{\boldsymbol{\mu}}^{\text{Shift}}(t)$$

$$\hat{\boldsymbol{\mu}}^{\text{Shift}}(t) = \hat{\boldsymbol{\mu}}(t) - \hat{\boldsymbol{\mu}}(t-1)$$

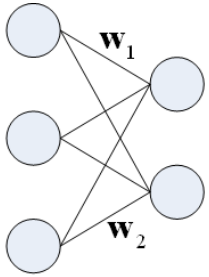


- Bosman P.A.N., Grahl J., Thierens D. Adapted maximum-likelihood Gaussian models for numerical optimization with continuous EDAs. Tech report no. SEN-E0704. Dec. 2007.

Проблемы с АОР

- Сложно использовать для задач большой размерности: необходимо решать полную проблему собственных значений для текущей оценки матрицы ковариации.
- Пример: ИНС, 9 входов, 18 нейронов в скрытом слое, 6 выходов: 294 параметра (включая смещения).
- А бывают ИНС с 20, 30, 50, … входами и б*о*льшим количеством нейронов в скрытом слое…
- Байесовская декомпозиция «в лоб» достаточно трудоемка.

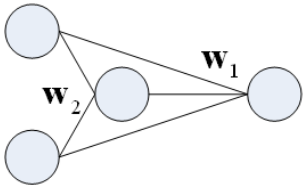
Декомпозиция весов искусственной нейронной сети



$$\mathbf{W} = \{\mathbf{w}_1, \mathbf{w}_2\}$$
$$= \{w_{11}, w_{12}, w_{13}, \mathbf{w}_2\}$$

$$P(\mathbf{W}) = P(\mathbf{w}_1)P(\mathbf{w}_2)$$
$$= P(w_{11}, w_{12}, w_{13})P(\mathbf{w}_2)$$

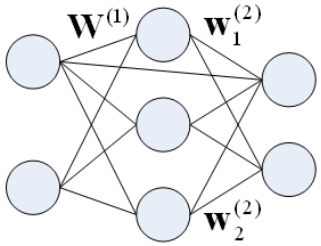
а)



$$\mathbf{W} = \{\mathbf{w}_1, \mathbf{w}_2\}$$

$$P(\mathbf{W}) = P(\mathbf{w}_1 \mid \mathbf{w}_2)P(\mathbf{w}_2)$$

б)



$$\mathbf{W} = \{\mathbf{W}^{(1)}, \mathbf{W}^{(2)}\}$$
$$= \{\mathbf{W}^{(1)}, \mathbf{w}_1^{(2)}, \mathbf{w}_2^{(2)}\}$$
$$= \{\mathbf{w}_1^{(1)}, \mathbf{w}_2^{(1)}, \mathbf{w}_3^{(1)}, \mathbf{w}_1^{(2)}, \mathbf{w}_2^{(2)}\}$$

$$P(\mathbf{W}) = P(\mathbf{W}^{(2)} \mid \mathbf{W}^{(1)})P(\mathbf{W}^{(1)})$$
$$= P(\mathbf{w}_1^{(2)} \mid \mathbf{W}^{(1)})P(\mathbf{w}_2^{(2)} \mid \mathbf{W}^{(1)})P(\mathbf{W}^{(1)})$$
$$= P(\mathbf{w}_1^{(2)} \mid \mathbf{w}_1^{(1)}, \mathbf{w}_2^{(1)}, \mathbf{w}_3^{(1)})P(\mathbf{w}_2^{(2)} \mid \mathbf{w}_1^{(1)}, \mathbf{w}_2^{(1)}, \mathbf{w}_3^{(1)})P(\mathbf{w}_1^{(1)}, \mathbf{w}_2^{(1)}, \mathbf{w}_3^{(1)})$$

в)

Будем рассматривать простейший вариант: независимые группы переменных

Декомпозиция весов искусственной нейронной сети

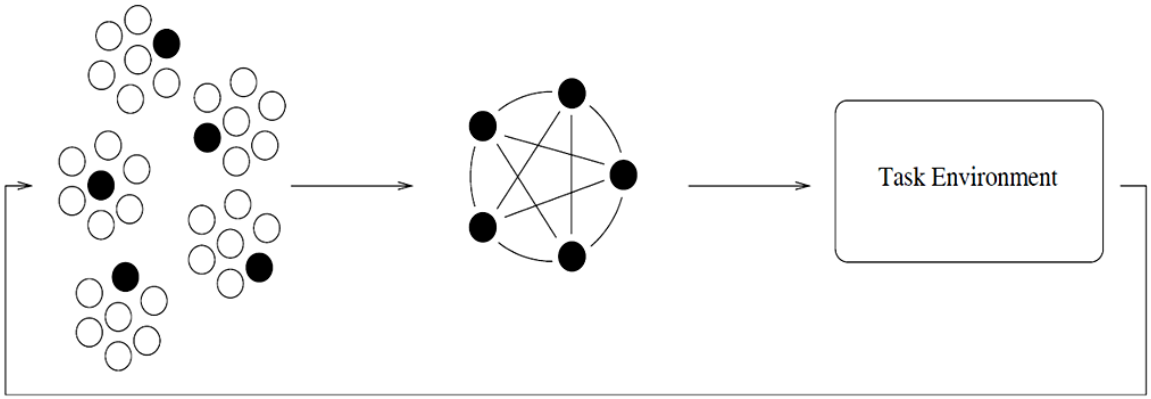
Варианты декомпозиции

- Без декомпозиции
- По слоям
- По нейронам
  - SANE (D. Moriarty, 1997)
  - ESP (F. Gomez, 2003)
- Смешанная: по слоям, последний слой по нейронам.

Получаем **коэволюционный алгоритм**.

Коэволюционный алгоритм

1. Определение параметров структуры ИНС.
2. Декомпозиция весов связей ИНС на *K* групп в зависимости от структуры.
3. Коэволюционная оптимизация с использованием *K* алгоритмов AMaLGaM IDEA.
4. Синтез лучшего решения (объединение лучших решений каждого из *K* алгоритмов).



Faustino Gomez. Robust Non-Linear Control through Neuroevolution. PhD thesis, Department of Computer Sciences, 2003.

Особенности

- Из-за быстрой сходимости алгоритма «страдает» определитель матрицы ковариации, сокращается окрестность поиска.

  После 40-50 итераций коэффициенты порядка 10^{-7}.
- Что если попробовать заменить матрицу ковариации на матрицу корреляции?

$$\mathbf{\Sigma} = \mathbf{X}_{\Sigma}\mathbf{\Lambda}_{\Sigma}\mathbf{X}_{\Sigma}^{T},$$

$$\mathbf{C} = \mathbf{X}_{C}\mathbf{\Lambda}_{C}\mathbf{X}_{C}^{T},$$

$$\mathbf{C} = \mathbf{S}^{-1}\mathbf{\Sigma}\mathbf{S}^{-1}$$

$$\Rightarrow$$

$$\mathbf{\Lambda}_{C} = \mathbf{K}^{1/2}\mathbf{\Lambda}_{\Sigma}\mathbf{K}^{1/2},$$

$$\mathbf{X}_{C} = \mathbf{S}^{-1}\mathbf{X}_{\Sigma}\mathbf{K}^{-1/2}$$

где $\mathbf{S} = diag(\sigma_1, .. \sigma_n)$

$$\mathbf{K} = diag\left(\lambda_{C,1} / \lambda_{\Sigma,1}, \ldots, \lambda_{C,n} / \lambda_{\Sigma,n}\right)$$

Задачи

- Proben1: cancer1, diabetes1, glass1.
- Критерий (ЦФ): Среднеквадратичная ошибка.
- Варианты декомпозиции:
  - без декомпозиции (no decomposition);
  - по слоям (layers);
  - по нейронам (nodes);
  - смешанная (mixed).
- AMaLGaM IDEA.
- MentalAlchemy, Alglib, Math.Net.

Задачи

- **cancer1**

Диагностика рака груди.

9 признаков, 2 класса. 350/175/174 примеров.

- **diabetes l**

Диагностика сахарного диабета.

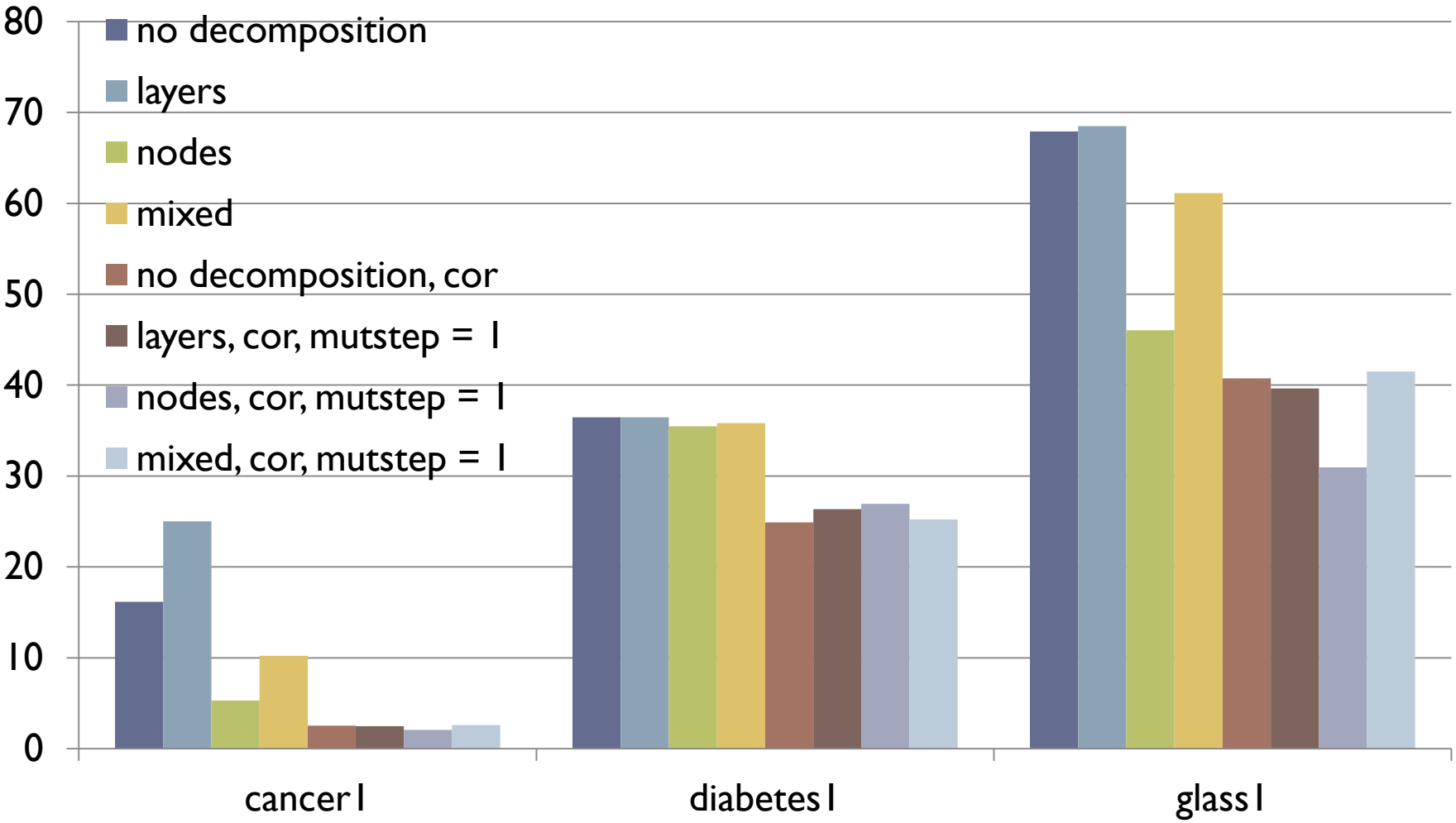
8 признаков, 2 класса. 384/192/192 примера.

- **glass l**

Классификация типов стекла.

9 переменных, 6 классов. 107/54/53 примера.

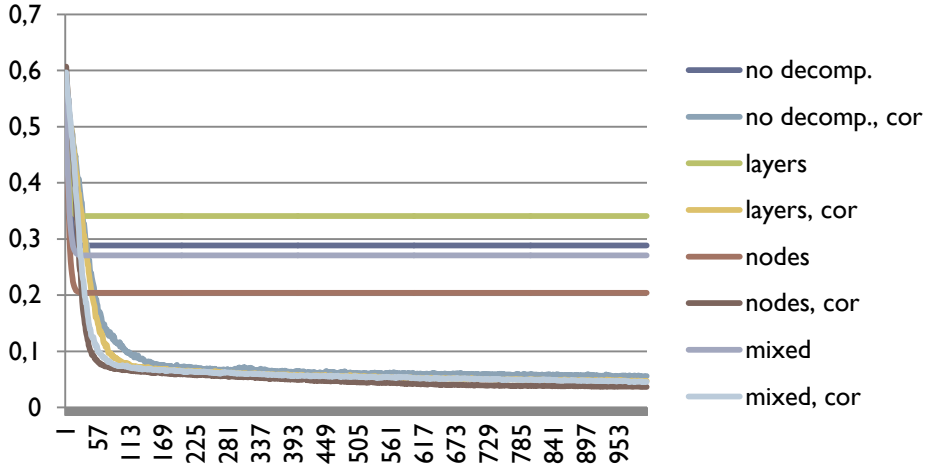
Результаты экспериментов



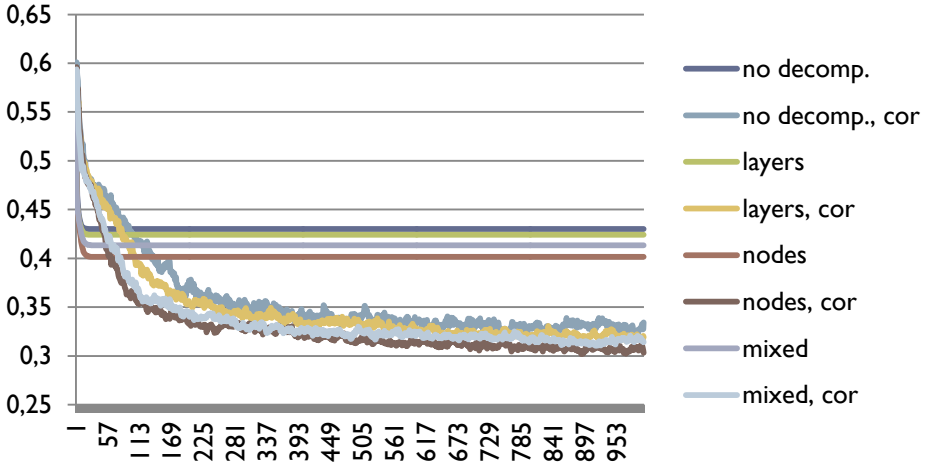
Ошибка классификации на тестовом множестве

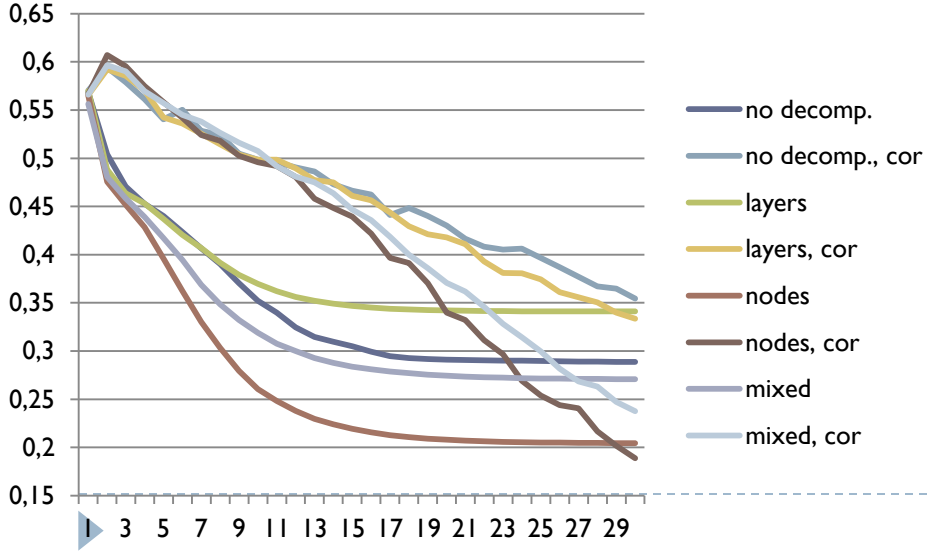
Результаты экспериментов

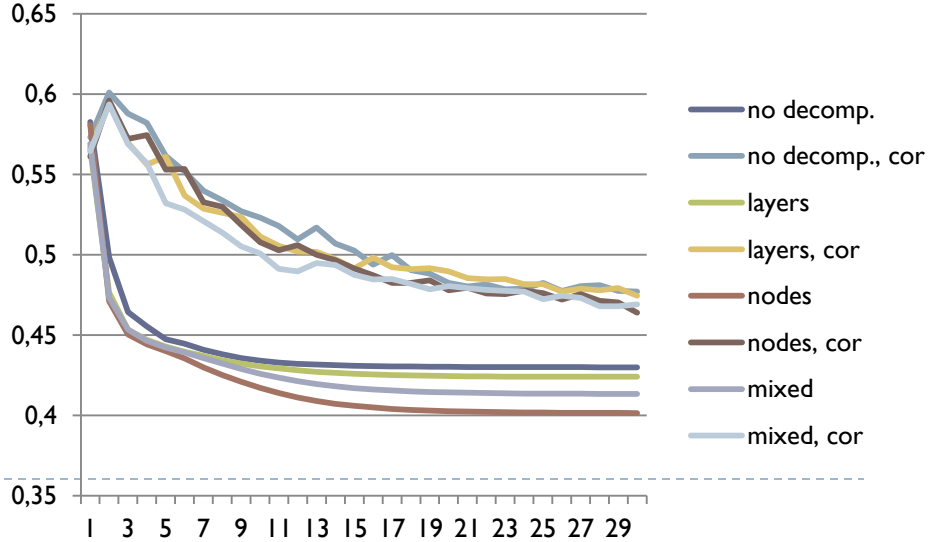
cancer1



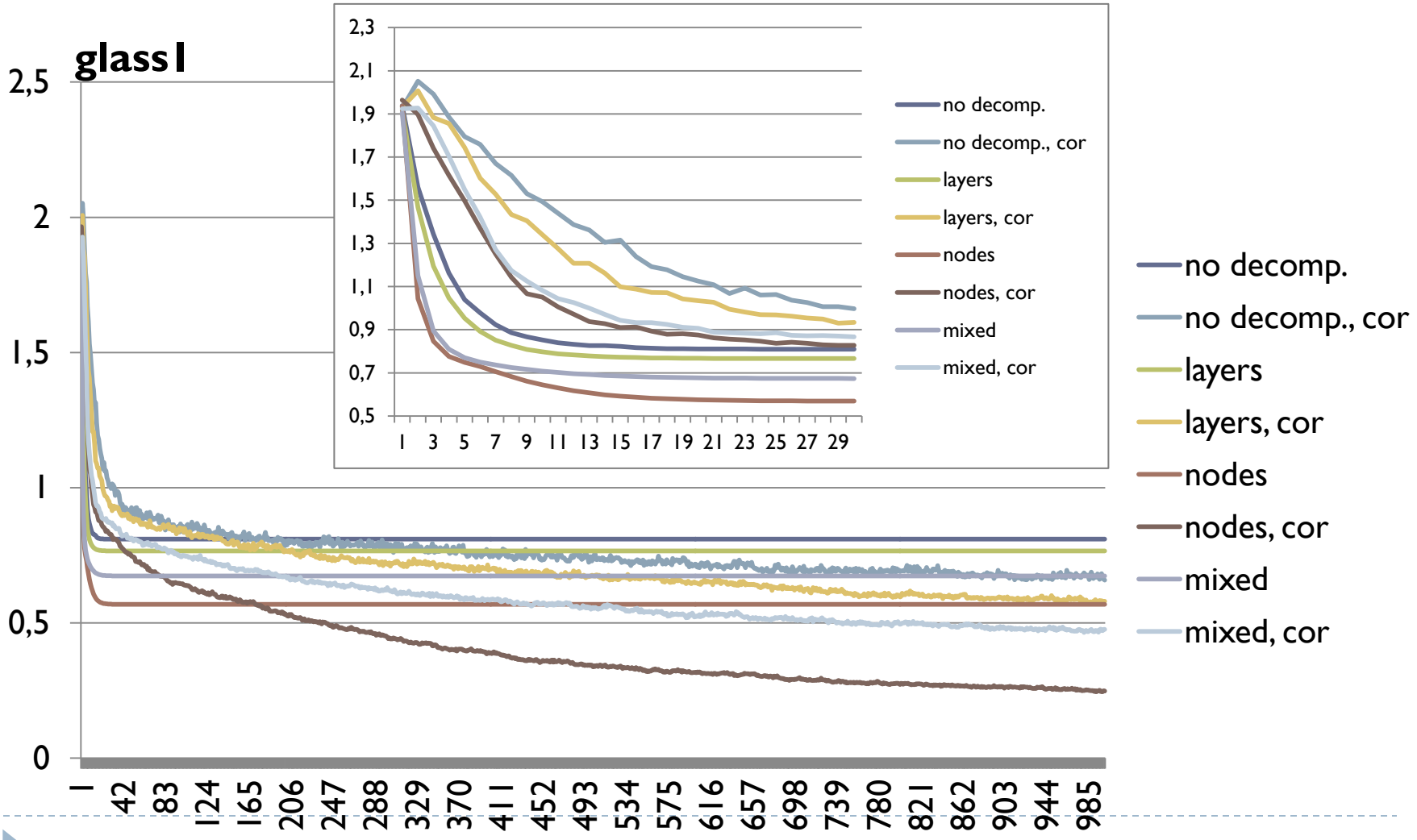
diabetes1







Результаты экспериментов



Результаты экспериментов

Среднее время на одно поколение (сек.)

	no decomposition	layers	nodes	mixed
cancer l	0,097	0,179	0,543	0,286
diabetes l	0,082	0,156	0,463	0,238
glass l	0,881	0,459	2,808	1,026

Среднее время на одно вычисление целевой функции (сек.)

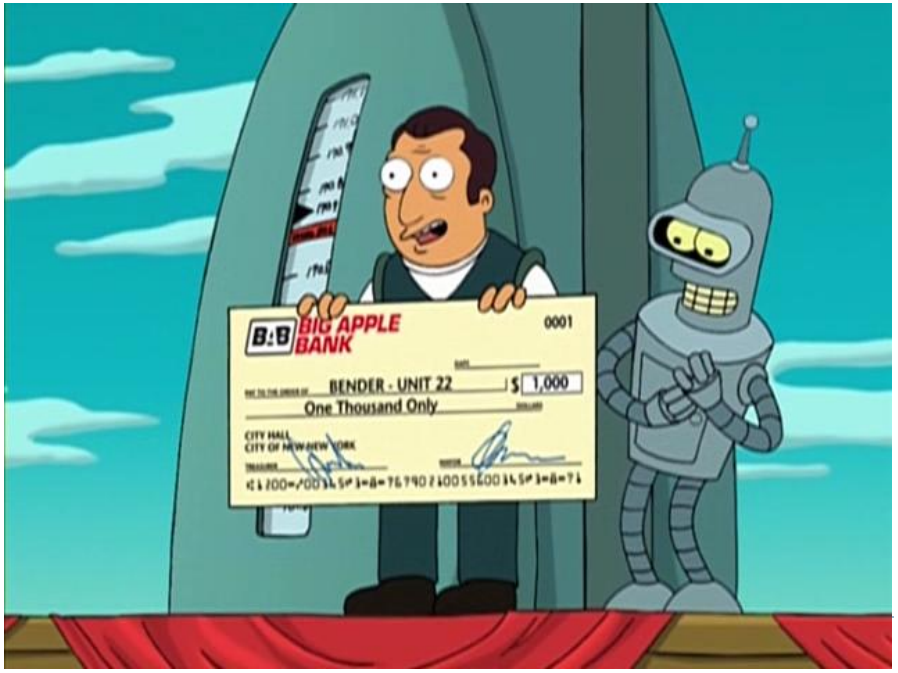
	no decomposition	layers	nodes	mixed
cancer l	0,00194	0,00179	0,00181	0,001907
diabetes l	0,00164	0,00156	0,001543	0,001587
glass l	0,01762	0,00459	0,00234	0,002931

Заключение

- Применили AMaLGaM IDEA к оптимизации весов ИНС.
- Использование матрицы корреляции вместо матрицы ковариации может быть оправдано.
- Можно декомпозировать веса ИНС, исходя из структуры.

Благодарности

- Исследование выполнено при поддержке грантов РФФИ, проекты № 11-07-00027-а и 12-08-00296-а.



- Брисбен, Австралия. 10-15 июня 2012.
- International Joint Conference on Neural Networks (IJCNN 2012).
- IEEE Congress on Evolutionary Computations (IEEE CEC).
- IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2012).
- **Есть труды конференции в электронном виде**. (труды, сборник пленарных докладов, лекции)

ПРИМЕНЕНИЕ АЛГОРИТМА БАЙЕСОВСКОЙ ОПТИМИЗАЦИИ ДЛЯ ОБУЧЕНИЯ ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ

Ю. Цой
yurytsoy@gmail.com