УДК 519.68

ГЕНЕТИЧЕСКИЙ АЛГОРИТМ СИНТЕЗА НЕЙРОСЕТЕВОЙ МОДЕЛИ КАК НАБОРА САМООРГАНИЗУЮЩИХСЯ КАРТ КОХОНЕНА

Малухин Д. В.
Научный руководитель – доцент Тынченко В. В.
Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева

Самоорганизующиеся карты Кохонена (СКК) – это особый подкласс нейронных сетей, предназначенных, главным образом, для визуализации и первоначального анализа многомерных данных, описывающих состояние некоторой системы, посредством их проецирования в пространство меньшей размерности (обычно двумерное) и последующего окрашивания в различные цвета точек-проекций в соответствии со значениями анализируемых параметров. Сформированный набор СКК позволяет наглядно отобразить наличие или отсутствие кластерной структуры во входных данных и, если таковая имеется, описать ее характер, а также выявить имеющиеся в этих данных зависимости.

Для того, чтобы судить о степени адекватности построенной нейросетевой модели, необходимо проводить оценку корректности отображения набором СКК зависимостей между параметрами системы. С этой целью в данной работе предлагается способ расчета коэффициента связи C_{ij} между картами *i* и *j* определённого набора. Коэффициент C_{ij} представляет собой вектор, состоящий из двух величин, описывающих силу положительной и отрицательной связей: $C_{ij} = (S_+; S_-)$:

$$S_+ = \sum_{i \neq j} \cos\left(\angle\left(\nabla P_{XY,i}; \nabla P_{XY,j}\right)\right),$$ где отрицательный косинус равен нулю.

$$S_- = \sum_{i \neq j} \left|\cos\left(\angle\left(\nabla P_{XY,i}; \nabla P_{XY,j}\right)\right)\right|,$$ где положительный косинус равен нулю.

Здесь $\nabla P_{XY,i}$, $\nabla P_{XY,j}$ – проекции векторов градиента $\nabla P = (dx, dy, dz)$ (направление максимального возрастания «рельефа» карты) в точках с координатами XY карт *i* и *j* соответственно. Положительная связь максимальна в точке, если векторы ∇P_{XY} обеих карт сонаправлены. Отрицательная связь максимальна в точке, если векторы ∇P_{XY} обеих карт противоположно направлены.

При организации карт используется функция соседства h_i:

$$h_i = \alpha(t) \cdot \boldsymbol{exp}\left(-\frac{r_i^2}{2\sigma^2(t)}\right)$$

Здесь r_i - евклидово расстояние от ближайшего вектора до текущего вектора *i*.

После одной организации карт (итерации) можно посчитать коэффициент связи C_{ij} для каждой пары карт. Получим вектор характеристик: $(C_{12}, \ldots, C_{(N-1)N})$, где $N = \frac{m!}{2(m-2)!} = \frac{(m-1)m}{2}$, m - количество параметров системы.

После нескольких итераций мы получим несколько векторов $(C_{12}, \ldots, C_{(N-1)N})$. Таким образом, представляется возможным найти вектор средних значений корреляций каждой пары $\left(\dot{C}_{12},\ldots,\dot{C}_{(N-1)N}\right)$, который описывает связи между параметрами системы на порядок точнее, чем вектор, полученный от одной случайной итерации. Оптимальным решением будет набор карт, вектор связи $\left(C_{12},\ldots,C_{(N-1)N}\right)$ которого наиболее близок к вектору средних значений $\left(\dot{C}_{12},\ldots,\dot{C}_{(N-1)N}\right)$.

Характер убывания параметров α и σ опишем при помощи следующей формулы:

$$\chi(t)=\frac{1-\exp\left(-(1-t)^{n}\cdot k\right)}{1-\exp(-k)}\cdot(1-l)+l$$

$$\alpha=\alpha_0\cdot\chi_1(t)\quad \sigma=\sigma_0\cdot\chi_2(t)$$

Здесь n, k, l – коэффициенты, задающие характер убывания функции $\chi(t)$, t - время, которое при проведении итерации принимает значение в диапазоне $[0\ldots1]$. В начале итерации $t=0$, в конце итерации $t=1$.

Перебрав различные сочетания значений коэффициентов, задающих характер убывания параметров α и σ, можно найти такую комбинацию карт, для которой вектор $\left(C_{12},\ldots,C_{(N-1)N}\right)$ будет наиболее близок к вектору $\left(\dot{C}_{12},\ldots,\dot{C}_{(N-1)N}\right)$. Однако использование полного перебора для поиска оптимального решения в данном случае неприемлемо, так как он займет примерно 60 лет при проведении вычислений на одном компьютере средней мощности.

При решении подобного рода сложных многопараметрических оптимизационных задач хорошо себя зарекомендовали генетические алгоритмы, которые не требуют информации о свойствах оптимизируемой функции и позволяют вести глобальный поиск в пространстве решений, избегая локальных оптимумов.

Выполним адаптацию генетического алгоритма к решаемой задаче. В нашем случае индивиды (поисковые точки в многомерном пространстве допустимых решений) представляют собой совокупности описанных выше параметров, а именно: α_0, n_α, k_α, l_α, σ_0, n_σ, k_σ, l_σ. Индивиды кодируются в виде двоичных последовательностей – хромосом. Генетический алгоритм применяется к популяции хромосом путем реализации типового цикла эволюции – популяция случайно инициализируется и охватывает лучшие регионы поискового пространства посредством выполнения процессов селекции, мутации и рекомбинации.

Значение функции пригодности в нашем случае рассчитывается следующим образом:

1. Производится декодирование хромосомы, определение характера убывания параметров α и σ.
2. Проводится N итераций с данным характером убывания параметров α и σ, каждая итерация вносит свой вклад в вектор $\left(\dot{C}_{12},\ldots,\dot{C}_{(N-1)N}\right)$. На каждой итерации получаем очередной вектор $\left(C_{12},\ldots,C_{(N-1)N}\right)$.
3. Значением функции пригодности является минимальное евклидово расстояние между одним из векторов $\left(C_{12},\ldots,C_{(N-1)N}\right)$, полученных в результате выполнения шага 2, и вектором $\left(\dot{C}_{12},\ldots,\dot{C}_{(N-1)N}\right)$. Набор карт, соответствующий минимальному расстоянию, сохраняется.

Применение предложенного генетического алгоритма позволяет автоматизировать формирование такого набора самоорганизующихся карт Кохонена, который с достаточной степенью достоверности будет отображать существующие зависимости между параметрами системы, что, в свою очередь, даст возможность принимать более эффективные решения в сфере управления в условиях нарастающей неопределённости от избыточного потока информации.