

ЭВОЛЮЦИОННОЕ ПРОЕКТИРОВАНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ КЛАССИФИКАЦИИ.

Коромылова А. А.

научный руководитель **Семенкина М. Е.**

Сибирский Государственный Аэрокосмический Университет им. М.Ф. Решетнёва

В настоящее время интеллектуальные информационные технологии (ИИТ) используются во многих отраслях науки. Это обусловлено увеличивающимися вычислительными мощностями современных компьютеров, которые применяются для решения практических задач, и способностью систем, использующих ИИТ, эффективно решать широкий круг задач.

Однако, как правило, реализация ИИТ – это трудоемкий процесс, требующий больших временных и материальных затрат. Например, при решении конкретной задачи с помощью искусственных нейронных сетей (ИНС) существует проблема выбора их эффективных структур и настройки весовых коэффициентов, а при использовании систем на нечёткой логике (НЛ) необходимо проектировать базы правил и настраивать лингвистические переменные.

В данной работе, был реализован генетический алгоритм (ГА) для настройки весовых коэффициентов полносвязных нейронных сетей прямой проводимости (GA-ANNW) с произвольным числом скрытых слоев и нейронов на них, так как ГА позволяет эффективно решать задачи глобальной оптимизации. В качестве активационной функции для нейронов на выходном и скрытых слоях используется сигмоида. Весовые коэффициенты записываются в хромосому последовательно в виде двоичного кода. Количество используемых бит зависит от точности настройки и разброса возможных значений весов, и отличаются для каждой конкретной задачи.

Эффективность GA-ANNW проверялась на 14 тестовых задачах аппроксимации, для решения которых использовались нейронные сети с тремя скрытыми слоями по три нейрона на каждом. Результаты работы реализованного генетического алгоритма для настройки весовых коэффициентов представлены в Таблице 1.

Таблица 1.

Результаты работы GA-ANNW на некоторых задачах

№ функции	GA-ANNW		Обратное распространение ошибки
	Средняя	Лучшая	
1	4,74%	3,75%	36,99%
2	6,73%	5,63%	37,16%
5	10,34%	8,28%	38,26%
10	16,93%	16,14%	39,57%

Тестирование проводилось при помощи разработанной программной системы по следующим правилам:

- при каждой настройке генетического алгоритма выполнялось по 100 запусков;
- количество поколений 1000;
- количество индивидов 500;
- размер выборки 500.
- объем обучающей выборки 70% от общего числа точек, а тестовой – 30%;

- результаты представлены для лучших настроек генетического алгоритма;
- ошибка по всем запускам усреднялась (в таблице – «Средняя»);
- в каждом запуске считалась средняя ошибка аппроксимации по формуле

$$\text{error} = \frac{100\%}{n \cdot |Y_{\max} - Y_{\min}|} \sum_{i=0}^n |\tilde{Y}_i - Y_i|; \quad (1)$$

- в таблице с результатами также указана минимальная из найденных ошибок («лучшая»).

Эффективность работы ИНС так же сильно зависит от выбранной структуры. Оптимальную структуру могут определить эксперты, но это потребует больших затрат временных и материальных ресурсов, поэтому предлагается использовать генетический алгоритм для автоматического проектирования структуры ИНС (GA-ANNS), который так же использует бинарные хромосомы. Для каждого нейрона сначала, случайным образом с определённой вероятностью равной 1/3, решается, будет ли он использоваться в данной сети. Если нейрона в сети нет, то на его место в хромосому записываются нули. В противном случае для него, случайным образом выбирается одна из пятнадцати функций активации, номер которой записывается в двоичном коде.

Далее для каждой выбранной структуры ИНС запускается генетический алгоритм GA-ANNW для выбора весовых коэффициентов. Тестирование, результаты которого представлены в Таблице 2, проводилось по тем же правилам, что и для GA-ANNW.

Таблица 2.

Результаты работы GA-ANNS

№ функции	GA-ANNS			
	Среднее		Лучшее	
	ошибка	количество нейронов	ошибка	количество нейронов
1	0,27%	3,2	0,13%	3
2	1,87%	4,5	1,6%	3
5	7,18%	6,3	7,03%	5
10	7,54%	7,8	6,86%	7

GA-ANNS решает задачу на приемлемом уровне, значительно лучше, чем GA-ANNW. Однако, точность заметно падает при увеличении сложности решаемой задачи, в частности при росте размерности. Поэтому необходимо производить выбор наиболее информативных признаков, чтобы оставить точность на приемлемом уровне, используя меньше ресурсов.

Особенность генетического алгоритма с выбором наиболее информативных признаков при автоматической настройке ИНС (GA-ANNinput) заключается в способе кодирования нейронной сети в бинарную хромосому: сначала записываются все входные переменные, по одному биту на признак, а затем нейроны со скрытых слоев, кодируемые аналогично GA-ANNS. При инициализации признак считался неинформативным с вероятностью 0,5.

Методика тестирования алгоритма была описана выше. Результаты представлены в Таблице 3.

Сравнивая результаты Таблиц 2 и 3, видно, что модификация не приносит существенных улучшений, так как размерность тестовых задач мала, т.е. все входы являются информативными. Поэтому для более полной проверки необходимо использовать значительно более сложные задачи.

Генетический алгоритм, настраивающий ИНС, решает задачу на приемлемом уровне, однако является «черным ящиком». Поэтому предлагается использовать системы на нечёткой логике.

Таблица 3.

Результаты тестирования GA-ANNinput

№ функции	GA-ANNinput					
	Среднее			Лучшее		
	Ошибка	количество нейронов	количество входов	ошибка	количество нейронов	количество входов
1	0,27%	3,4	1	0,15%	3	1
2	1,87%	4,1	1	1,4%	4	1
5	7,65%	5	1,9	7,03%	4	2
10	9,12%	8,7	1,6	6,86%	6	1

При работе с системами на НЛ, возникают две проблемы:

- определение способа кодирования лингвистических переменных;
- выбор оптимальной базы правил.

Лучший способ решения этих проблем – работа с экспертом. Однако это требует материальных и временных затрат, а иногда и вообще не представляется возможным. Поэтому предлагается использовать генетический алгоритм для автоматической настройки систем, основанных на нечёткой логике (GA-FL).

Для кодирования лингвистических переменных генерировалась бинарная строка, к которой далее применялись операторы генетического алгоритма. Все нечеткие переменные описываются набором (N,X,A), где N – это название переменной, X – универсальное множество (область рассуждений), A – нечеткое множество на X.

В данной работе одновременно использовались три вида функции принадлежности: треугольная, трапециевидальная и гауссова. Для формирования базы правил так же использовалась бинарная хромосома. Количество битов на одно правило зависело от размерности задачи. Результаты работы GA-FL представлены в Таблице 4. Его тестирование проводилось по правилам описанным выше.

Таблица 4.

Результаты работы генетического алгоритма при настройке систем на нечёткой логике

№ функции	GA-FL			
	Средняя		Лучшая	
	Ошибка	количество правил	ошибка	количество правил
1	0,41%	3,7	0,11%	4
2	2,37%	5,2	1,45%	5
5	7,03%	5,9	6,89%	7
10	7,69%	6,3	7,21%	7

GA-FL решает задачу с приемлемой точностью. Однако точность заметно падает при росте размерности. Чтобы оставить точность на приемлемом уровне, используя меньше ресурсов, предлагается производить выбор наиболее информативных признаков.

Генетический алгоритм для автоматической настройки систем на нечёткой логике с выбором наиболее информативных признаков (GA-FLinput) основан на том же принципе выбора информативных признаков, что и GA-ANNinput.

Результаты тестирования разработанного алгоритма представлены в Таблице 5. Так же, как при использовании GA-ANNinput, GA-FLinput, при решении тестовых задач,

не сильно влияет на эффективность. Это обосновано малой размерностью вектора входных переменных.

Таблица 5.

Результаты тестирования GA-FLinput

№ функции	GA-FLinput					
	Среднее			Лучшее		
	Ошибка	количество входов	количество правил	ошибка	количество входов	количество правил
1	0,36%	1	3,9	0,1%	1	5
2	1,85%	1	5	1,51%	1	6
5	7,43%	1,9	5,7	7,14%	2	6
10	7,67%	1,3	6,8	6,93%	1	7

Апробация реализованных алгоритмов проводилась на 3 задачах анализа данных: классификация ирисов, австралийская и немецкая задачи о кредитах. Данные для анализа были взяты из репозитория автоматического обучения. Результаты представлены в Таблице 6. Тестирование проводилось при лучших настройках генетического алгоритма (пропорциональная селекция, двухточечное скрещивание и средняя мутация).

Таблица 6.

Результаты работы алгоритмов при решении практических задач

Название метода	Ирис		Австралийская задача		Немецкая задача	
	Среднее	Лучшее	Среднее	Лучшее	Среднее	Лучшее
GA-ANNW	0,098	0,049	0,132	0,106	0,204	0,127
GA-ANNS	0,043	0,021	0,099	0,087	0,137	0,089
GA-ANN input	0,021	0,017	0,096	0,091	0,121	0,113
GA-FL	0,039	0,03	0,124	0,097	0,195	0,129
GA-FL input	0,024	0,023	0,113	0,098	0,212	0,144

Тестирование модифицированных алгоритмов на задачах аппроксимации не показало существенного улучшения по сравнению с генетическими алгоритмами, которые настраивали структуру нейронной сети или обычных систем на нечеткой логике. Это связано с тем, что у тестовых задач вектор входных переменных небольшой размерности, а следовательно исключение каких либо признаков либо невозможно, либо ведет к уменьшению точности решения задачи.

Однако при решении задач большой размерности, как видно из Таблицы 6, модифицированные алгоритмы показали более высокую точность, при этом для нахождения решения понадобилось меньшее количество вычислительных ресурсов.

Так же была выполнена проверка эффективности реализованных алгоритмов в сравнении с другими методами, которая показала, что GA-ANNinput занимает 2-ое место для австралийской задачи, то есть уступает только специально разработанному для таких задач методу, и 6 место для немецкой задачи. А GA-FLinput занимает 4 и 3 места соответственно. Что является хорошим результатом для неточного под задачу метода. Генетические алгоритмы без выбора наиболее информативных признаков показали эффективность, уступающую как модифицированным алгоритмам, так и некоторым из стандартных.