

Кучер В.О. — рецензент *Киселёв Г.Д.*

УНК “Институт прикладного системного анализа” НТУУ “КПИ”, Киев, Украина

Нейросетевая модель выбора эвристики предоставления ресурсов на уровне потока заданий в grid-системе

Моделирование процесса планирования выполнения заданий в grid-системе представлено в работе автора [1], в которой рассматривалась grid-система с многоуровневой иерархической структурой, на каждом из уровней которой существуют свои сервисы, взаимодействующие посредством определенных протоколов.

В указанной работе ставилась цель: разработать алгоритм планирования выполнения заданий, обеспечивающий минимизацию времени выполнения задания и сбалансированность загрузки поставщиков ресурсов.

Использование нечетко-множественного подхода, принципа прямоугольной упаковки в полосу заданной ширины, выбор агрегативной системы в качестве математической модели, позволили разработать модель процесса планирования, обеспечивающую выполнение поставленной цели.

Разработанная модель позволяет оценивать эффективность предложенного алгоритма, сравнивать оценки вариантов при проектировании, при выборе режимов функционирования с точки зрения выбранного критерия эффективности, оценивать динамику изменения системы, например, при изменении потока задач, конфигурации кластеров, сетевых соединений и т.п.

Анализ свойств заданий потока позволяет выявить, какие из факторов, характеризующих задания: ресурсные запросы, структура, статистика загрузки процессоров, потребности в данных, наиболее существенны при выборе стратегии управления. Возникла задача об интеграции методов планирования на уровне потока заданий с целью эффективного использования вычислительных ресурсов распределенных сред.

Постановка задачи: Построить модель по определению категории задания потока. Для каждой из категорий стратегия управления предполагает наличие определенных правил предоставления и потребления ресурсов, что позволяет повысить эффективность планирования и распределения ресурсов на уровне потоков заданий.

Для решения поставленной задачи выбрана многослойная нейронная сеть прямого распространения (персептронного типа) [2], которая, согласно теореме Колмогорова, может эффективно применяться для решения задач классификации. Персептрон переводит входной образ, определяющий степени возбуждения нейронов самого нижнего уровня иерархии, в выходной образ, определяемый нейронами самого верхнего уровня. Состояние возбужденного нейрона на верхнем уровне говорит о принадлежности входного образа к той или иной категории.

При построении классификатора определены параметры, влияющие на принятие решения о том, к какой категории принадлежит задание. Задание представляется вектором, в качестве компонент которого используются данные о приоритете задания, о запрашиваемых вычислительных ресурсах (о выполнении задания в пределах одного кластера или на ресурсах нескольких кластеров), о предполагаемой длительности его выполнения, о необходимости обработки и анализа больших объемов данных и др. Исходные данные могут иметь произвольный диапазон, поэтому проводится нормировка входных данных для подачи на вход сети. Для этого используется способ линейного преобразования, при котором операция нормировки выполняется с помощью деления каждой компоненты входного вектора на длину вектора. Таким образом входной вектор превращается в вектор единичной длины в n -мерном пространстве.

Представление выходных данных осуществляется с помощью вектора, компоненты которого соответствуют различным номерам категорий: i -я компонента вектора соответствует i -ой категории. При подаче на вход сети вектора-задания, на выходе имеем вероятности соответствия этого задания каждой из введенных категорий. При интерпретации результата считаем, что номер категории задания определяется номером выхода сети, на котором появилось макси-

мальное значение. При таком способе кодирования можно ввести также оценку уверенности сети в том, что задание относится к этой категории, например, определять разность между максимальным значением выхода и значением выхода, которое является ближайшим к максимальному. Чем выше уверенность, тем больше вероятность того, что сеть дала правильный ответ.

В качестве активационной функции выбранной двухслойной нейронной сети прямого распространения информации с одним скрытым слоем, обучаемой по методу обратного распространения ошибки, используется сигмоидная функция:

$$OUT = \frac{1}{1 + e^{-NET}}.$$

При обучении сети для каждой обучающей пары из обучающего множества подается входной вектор на вход сети и вычисляется выход сети. Разность между выходом сети и требуемым выходом (целевым вектором обучающей пары *Target*) используется для корректировки весов сети. Подстройка весов осуществляется с использованием дельта правила:

$$\delta = OUT(1 - OUT)(Target - OUT),$$

где $OUT(1 - OUT)$ – производная активационной функции.

Для каждого веса от нейрона скрытого слоя к нейрону в выходном слое выполняются следующие процедуры:

$$\begin{aligned} \Delta w_{pg} &= \eta \delta_g OUT_p, \\ w_{pg}(n+1) &= w_{pg}(n) + \Delta w_{pg}, \end{aligned} \quad (1)$$

где $w_{pg}(n)$ – величина веса от нейрона p в скрытом слое к нейрону g в выходном слое на шаге n (до коррекции); $w_{pg}(n+1)$ – величина веса на шаге $n+1$ (после коррекции); δ_g – величина δ для нейрона g ; OUT_p – величина OUT для нейрона p в скрытом слое; η – коэффициент скорости обучения.

Подстройка весов скрытого слоя осуществляется следующим образом. При проходе вперед каждый нейрон скрытого слоя передает свой выходной сигнал нейронам в выходном слое через соединяющие их веса. Во время обучения эти веса функционируют в обратном порядке, пропуская величину δ от выходного слоя назад к скрытому слою. Каждый из этих весов умножается на величину δ нейрона, к которому он присоединен в выходном слое. Величина δ , необходимая для нейрона скрытого слоя, определяется соотношением:

$$\delta_p = OUT_p(1 - OUT_p) \left[\sum_g \delta_g w_{pg} \right].$$

Имея значение δ , веса, питающие скрытый уровень, корректируются с помощью уравнений (1).

При выборе степени сложности сети используется механизм контрольной кросс-проверки: резервируется часть обучающих наблюдений, которая не используется в обучении по алгоритму обратного распространения. По мере работы алгоритма они используются для независимого контроля результата. Оптимальной размерностью скрытого слоя нейронной сети считается такая размерность, при которой средний процент правильной классификации категорий заданий для тестовой выборки достигает приемлемого значения и дальнейшее увеличение количества нейронов в скрытом слое не повышает процент правильной классификации.

В выходном слое количество нейронов равно количеству категорий заданий потока.

Литература. 1. Кучер В.О. Моделирование процессу планування виконання завдань в grid-системах з використанням нечітко-множинного підходу // Економіко-математичне моделювання соціально-економічних систем. Збірник наукових праць. Випуск 16. – Київ: МННЦ ІТС НАНУ. – 2011. – С. 256–278. 2. Яхьяева Г.Э. Нечеткие множества и нейронные сети. Учебное пособие // Яхьяева Г.Э. – М.: БИНОМ, 2006. – 317 с.