

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОРГАНИЗАЦИИ ДИСТАНЦИОННОГО ОБУЧЕНИЯ

Л.В. Серебряная¹, О.А. Шушина², Г.Э. Романюк³

^{1,2}БГУИР, кафедра программного обеспечения информационных технологий
220013, ул. П.Бровки, 6, г.Минск, Беларусь
+375 297739509, +375 336641379, e-mail: l_silver@mail.ru, o_shushina@gmail.com

³БНТУ, кафедра интеллектуальных систем
220000, пр. Независимости, 65, г.Минск, Беларусь
+375 172939325, e-mail: galarom@tut.by

Рассмотрены информационные технологии, применяемые в организации дистанционной формы обучения. Предложен метод обработки текстовой информации на основе искусственных нейронных сетей (ИНС). В качестве модели нейросети выбрана самоорганизующаяся карта Кохонена. Приведен алгоритм авторубрикации текстов с помощью сегои Кохонена. Рассмотрены вопросы контроля знаний студентов.

Ключевые слова: дистанционное обучение, искусственная нейронная сеть, карта Кохонена.

1 ОСОБЕННОСТИ ДИСТАНЦИОННОГО ОБУЧЕНИЯ

Дистанционное обучение базируется на традиционных и инновационных методах и средствах, использующих компьютерные и телекоммуникационные технологии, связывающие преподавателя, обучаемого и информационные ресурсы [1]. Постоянная переписка между участниками данного процесса предполагает активный документооборот. Электронный документооборот (ЭДО) – это единый механизм для работы с документами, представленными в электронном виде. Информационные технологии позволяют применять различные виды обработки электронных документов, сокращая время создания и пересылки сообщений.

Обмен электронной корреспонденцией на базе электронной почты, позволяет организовать надежную доставку сообщений, осуществлять контроль доступа к передаваемым данным, сохранять хронологию поступления сообщений. В работе [2] изложены особенности аппаратной реализации ЭДО, а также описано многофункциональное программное средство, предназначенное для организации документооборота в процессе дистанционного обучения.

Применение ИНС к обработке текстовой информации позволяет решить такие задачи, как: поиск, автореферирование, авторубрификация и др. В свою очередь, авторубрификация облегчает поиск необходимых данных. В дистанционном обучении это может быть использовано как основа для системы, структурирующей учебный материал.

2 КЛАССИФИКАЦИЯ ТЕКСТОВ НА ОСНОВЕ ИНС

В основе ИНС лежит формальный нейрон. ИНС состоит из произвольного количества слоев нейронов. Их количество в слоях также может быть произвольным. В каждом слое выполняется нелинейное преобразование линейной комбинации сигналов предыдущего слоя [3]. Многослойная сеть обычно формирует на выходе многомерную функцию при удачном выборе количества слоев, диапазона изменения сигналов и параметров нейронов.

Задача классификации текстовой информации решается с помощью самоорганизующейся сети (карты) Кохонена, поскольку она способна самостоятельно разбивать данные на классы и создавать новые рубрики в случае появления текстов, не относящихся ни к одной из уже имеющихся групп. Карта Кохонена представляет собой ИНС в виде двумерной решетки, в которой каждый нейрон имеет количество входов, равное размерности входного вектора. Чем больше размерность нейронной решетки, тем детальнее происходит анализ данных.

В ходе работы сети данные, относящиеся к одной рубрике, будут активизировать один и тот же нейрон. Поскольку речь идет об обучении без учителя, заранее не известно, какому конкретно классу соответствует выделенный нейрон, гарантируется лишь разделение различных наборов данных на классы.

Тематику текста можно определить, например, на основе ключевых слов. Часто они определяются с помощью алгоритмов вычислительной лингвистики, основанных на законах Zipфа. Затем слова приводят к основной словоформе, т.е. к виду, в котором они представлены в словаре. Первоначально словарь состоит из выделенных терминов, относящихся к определенным темам. Словарь представляет собой вектор, каждой позиции которого соответствует определенное ключевое слово. По аналогии со словарем документ тоже можно представить в виде вектора, поступающего на вход ИНС. В результате ее работы документ относится к той или иной рубрике.

Размер словаря равен количеству входов у нейронов сети, и каждому ключевому слову соответствует своя позиция в векторе. Если документ характеризуется определенным ключевым словом, то при построении вектора документа в соответствующую слову позицию заносится 1, иначе - заносится 0.

Рассмотрим алгоритм решения задачи рубрикации текстов с помощью сети Кохонена.

1. Весовым коэффициентам нейронов присваиваются случайные начальные значения, а затем данные подаются на вход сети.

2. Вычисляется расстояние между входным вектором и векторами весовых коэффициентов. «Победителем» назначается нейрон, обеспечивший минимальное расстояние.

3. Каждый весовой коэффициент «победившего» нейрона изменяется пропорционально его разнице с соответствующим входным значением и заданной скоростью обучения на текущем этапе обучения. Изменения коэффициентов осуществляются по следующей формуле $w_n = w_c + \alpha(x - w_c)$, где w_n – новое значение коэффициента; w_c – значение коэффициента на предыдущем шаге; x – значение входа нейрона; α – коэффициент скорости обучения. Первоначально его значение берется равным 0,7 и на каждом шаге обучения уменьшается на заданную величину.

Возможности ИНС были проверены на выборке из 14 текстов, относящихся к заранее определенным четырем классам. Словарь ключевых слов содержал 16 терминов. По вышеописанному правилу для каждого текста составлялись векторы входных данных. Они подбирались таким образом, что по некоторым позициям в векторах имелись совпадения. Первоначально сеть Кохонена задавалась решеткой размером 2x2, т.е. количество нейронов было равно предполагаемому количеству классов. В результате сеть не смогла разделить тексты, относящиеся к двум разным классам. Они были отнесены к одному нейрону и, следовательно, распознаны как объекты, принадлежащие одному и тому же классу. Некорректная работа сети могла быть вызвана несколькими причинами. Во-первых, большим сходством входных данных в текстах, относящихся к разным классам. Во-вторых, недостаточным размером сети, состоящей из четырех нейронов при классификации объектов на четыре класса. При увеличении решетки до девяти нейронов проблема была решена, и сеть корректно определяла классы для текстов уже после тридцати итераций обучения.

Тестирование ИНС на различных наборах исходных данных подтвердило ее работоспособность. Помимо разделения текстов на рубрики с помощью нейросети можно устанавливать логические связи между документами с целью организации их в единую информационную структуру, обработка которой значительно упрощается, а также увеличиваются возможности по расширению ее функционала.

Разные модификации алгоритма работы ИНС делают карту Кохонена универсальным средством для решения ряда задач: прогнозирования, моделирования процессов, классификации, сжатия и восстановления данных, усиления сигналов и др.

4 КОНТРОЛЬ ЗНАНИЙ

Образовательный процесс при дистанционной форме обучения в основном состоит из самостоятельной работы студентов. Важную роль при этом играют самоконтроль и анализ полученных результатов. Наиболее эффективным оказывается комплексный контроль, включающий в себя несколько взаимосвязанных этапов, интегрированных в процесс обучения. Такой подход позволяет получить полное представление о знаниях и навыках студентов. Методика комплексного контроля стимулирует постоянное повторение и закрепление изученного материала, поскольку предполагает тестирование по всем пройденным темам в рамках раздела или курса. По мере получения ответов пользователя формируется информация о том, насколько хорошо он освоил изучаемый материал.

Предложенная методика реализована в работе в форме электронной обучающей системы, где обработка текстов выполняется на базе ИНС. Подобная система может самостоятельно оценивать ситуацию и динамически формировать структуру изучаемого материала в виде семантически связанных текстов либо гипертекста. Вопросы формируются в соответствии с имеющимися в материале контрольными точками. По результатам анализа ответов пользователя определяется множество тем для повторения и с ними сопоставляется новый набор вопросов. Алгоритм, реализованный в системе, обеспечивает «равноправие» всех тем, то есть по каждой из них как минимум в один из этапов контроля включается вопрос. В результате тестирования обучаемый получает объективную информацию о своем уровне знаний и имеет возможность его улучшить.

Использование ИНС в организации дистанционного обучения связано с тем, что современные нейросети характеризуются такими свойствами, как: обучаемость, способность к обобщению и к абстрагированию. Поэтому процедура поиска решения задачи с помощью сети, прошедшей обучение, оказывается более гибкой, чем процесс программирования, поскольку ИНС может повышать точность результатов по мере накопления ею опыта и адаптироваться к происходящим изменениям.

ЛИТЕРАТУРА

- [1] Система дистанционного обучения в Беларуси: факторы и перспективы развития / И.А. Тавгень // Белорусский экономический журнал. – 2003. – № 2. – С. 27-32.
- [2] Бочкарева Л.В., Шушина О.А. Программно-инструментальная среда организации дистанционного обучения / Л.В.Бочкарева, О.А.Шушина // сборник научных трудов МНГК «Актуальные вопросы методики преподавания математики и информатики», Биробиджан, ДВГСИ, 16 апреля 2009, ч. 2, С. 57-61.
- [3] Тархов Д. А. Нейронные сети. Модели и алгоритмы. – М.: Радиотехника, 2005. – 256 с.