

ИСПОЛЬЗОВАНИЕ НЕЙРО-НЕЧЕТКИХ ТЕХНОЛОГИЙ ДЛЯ ЗАДАЧ КЛАССИФИКАЦИИ ПОЛИТЕМАТИЧЕСКИХ ТЕКСТОВЫХ ДОКУМЕНТОВ

Бодянский Евгений Владимирович, **Рябова** Наталия Владимировна,
Золотухин Олег Викторович

Харьковский национальный университет радиоэлектроники
кафедра искусственного интеллекта
zolotukhin.ov@gmail.com

Стремительное развитие Интернет-технологий и Web-пространства привели к значительному увеличению объема информации. Учитывая постоянно возрастающие объемы доступной информации в текстовом виде и связанную с этим проблему смыслового поиска, актуальность разработки методов автоматической классификации текстовых документов различного типа чрезвычайно высока.

Под классификацией понимают отнесение объектов (наблюдений, событий) к одному из заранее известных классов. Классификация относится к стратегии обучения с учителем (supervised learning), которое также называют контролируемым или управляемым обучением. Как правило, документы, которые необходимо классифицировать, характеризуются разнородностью информации (политематичностью), т.е. затрагивают сразу несколько тем, которые могут быть как близкие по смыслу, так и весьма различны. Классификация такого рода текстовых документов не является тривиальной задачей, поскольку в небольшом фрагменте текста может содержаться весьма ценная информация, и отнесение к соответствующему классу нельзя игнорировать, а близко расположенные классы могут пересекаться и/или сливаться. Поэтому желательно учесть принадлежность анализируемого документа к каждому из потенциально интересующих пользователя классов. В то же время большинство известных методов классификации относят текстовый документ к одному из четко различимых классов. Исходя из этого, задачу классификации можно рассматривать как задачу классификации политематических текстовых документов. В связи с этим актуальной является необходимость разработки нечетких методов классификации текстовых документов.

Предлагается использовать архитектуру и on-line алгоритм обучения адаптивной нечеткой нейронной сети обучаемого векторного квантования, которая предназначена для последовательной обработки текстовых документов в условиях перекрывающихся классов. Архитектура предлагаемой нейро-фаззи системы адаптивного обучаемого векторного квантования содержит два слоя обработки информации, при этом нейроны первого скрытого слоя связаны между собой латеральными связями, с помощью которых реализуются процессы конкуренции, а нейроны выходного слоя вычисляют уровни принадлежности предъявленного образа к каждому классу.