

## ОТЗЫВ

научного руководителя на диссертацию Дударина Павла Владимировича «Исследование и разработка моделей и методов нечеткой кластеризации коротких текстов», представленной на соискание ученой степени кандидата технических наук по специальности 05.13.01 – «Системный анализ, управление и обработка информации (информационные технологии и промышленность)»

### Актуальность исследования

В настоящее время поддержка принятия решений в системах управления предполагает обработку больших массивов информации, представленной короткими текстовыми сообщениями. Обработка информации предполагает кластеризацию коротких текстов, выполняемую на основе методов машинного обучения. Процесс обработки в контуре управления чаще всего осуществляется на основе как экспертной информации, так и использования словаря проблемной области. Эффективность кластеризации определяется словарем, который должен быть расширен за счет использования современных языковых моделей, фиксирующих семантическую близость терминов проблемной области. Машинное обучение, как научное направление, располагает эффективными структурами нейронных сетей глубокого обучения для обработки текстов на естественном языке. Но для использования данных методов при кластеризации коротких текстов необходима их адаптация и разработка новой архитектуры и алгоритма обучения нейронной сети. В связи с этим работа соискателя Дударина П.В., целевым назначением которой является повышение эффективности нечеткой кластеризации коротких текстов путем разработки модели, метода и алгоритма в системе поддержки принятия решений для кластеризации коротких текстов на русском языке с учетом экспертной информации, обеспечивающих сокращение времени и трудоемкости работы, выполняемой экспертом при достаточной точности кластеризации, является, несомненно, актуальной.

## Научная новизна результатов

Проведенный автором анализ особенностей кластеризации коротких текстов при построении систем принятия решений определил цель и задачи работы. Результаты исследования, полученные автором на основе теории вероятностей, методов машинного обучения, теории нечетких систем, нейронных сетей обладают научной новизной. Автором проведены вычислительные эксперименты, подтверждающие достоверность выдвигаемых положений.

Несомненной научной новизной обладает полученные автором результаты:

1. Предложенная архитектура искусственной нейронной сети, отличающаяся от известных тем, что позволяет решать задачу кластеризации на базе скрытого пространства признаков языковой модели;
2. Предложенный метод обработки текстов для расширения словаря языковой модели на базе нейронной сети с использованием нечеткого иерархического классификатора, отличающийся от известных тем, что позволяет учитывать семантическую близость слов;
3. Предложенный метод обработки обратной связи от эксперта, отличающийся от известных тем, что позволяет корректировать весовые коэффициенты нейронной сети и проводить интерактивную кластеризацию наборов коротких текстов;
4. Разработанный алгоритм в системе поддержки принятия решений, автоматизирующий применение предложенных модели и методов для выполнения нечеткой интерактивной кластеризации наборов коротких текстов.

## Самостоятельность получения результатов

Все перечисленные ниже научные результаты получены соискателем самостоятельно, в частности:

1. Проведено исследование моделей и методов машинного обучения для обработки текстов, и разработана архитектура искусственной нейронной сети, реализующей кластеризацию на базе пространства признаков языковой модели русского языка.
2. Разработан метод обработки текстов для расширения словаря языковой модели на базе нейронной сети с использованием нечеткого иерархического классификатора, который позволяет повысить точность кластеризации в среднем на 10%.
3. Разработан метод для обработки обратной связи от эксперта, используемый для корректировки весовых коэффициентов нейронной сети, что позволяет проводить интерактивную кластеризацию наборов коротких текстов.
4. Сформулирован перечень этапов программы проведения испытаний метода нечеткой интерактивной кластеризации коротких текстов. Проведенные исследования по данной программе позволили установить границы применения предлагаемого метода. Метод наиболее эффективен для текстов с количеством слов от 10 до 100. В ходе проведенных исследований была достигнута средняя точность кластеризации 80%, при более низком числе дополнительных ограничений по сравнению с аналогичными методами. На разработанный программный модуль, использованный для проведения численных экспериментов получено свидетельство о государственной регистрации программы для ЭВМ № 2021615642.
5. Составлен алгоритм автоматизации работ по нечеткой интерактивной кластеризации коротких текстов в системе поддержки принятия решений СППР “Федеральная Система Стратегического Планирования РФ (ФИС СП)”.
6. Проведена апробация разработанных модели, методов и алгоритма нечеткой интерактивной кластеризации коротких текстов в качестве элементов СППР “ФИС СП”. Внедрение алгоритма позволило решить задачу составления автоматического классификатора ключевых показателей эффективности документов стратегического планирования, при этом экономия трудозатрат

эксперта оценивается более чем в 9 человеко-месяцев. Наличие автоматического классификатора позволяет снизить временные затраты экспертов при проведении процедуры проверки корректности классификации входящих (новых) документов.

7. Разработанный метод интерактивной кластеризации является универсальным и может быть применен для различных наборов коротких текстов. Предложенный метод может быть доработан для совместного использования с различными языковыми моделями и обобщен на случай совместной работы ряда экспертов

### **Практическая ценность результатов**

Практическая ценность работы состоит, прежде всего, в том, что реализован программный модуль интерактивной кластеризации для системы Планета.Аналитика 4.0 (входит в реестр отечественного ПО) компании ООО “ИБС “Экспертиза”.

Практическая значимость работы подтверждается использованием ее результатов и рекомендаций в рамках НИР в интересах Министерства экономического развития РФ по темам: “Разработка рекомендаций по совершенствованию информационного обеспечения участников стратегического планирования в части осуществления мониторинга и контроля реализации документов стратегического планирования с использованием Федеральной информационной системы стратегического планирования (ФИС СП)” и “Разработка методического обеспечения интеллектуальной системы проверки уведомления об утверждении (одобрении) документа стратегического планирования или внесении в него изменений при ведении федерального государственного реестра документов стратегического планирования Федеральной информационной системы стратегического планирования (ФИС СП)”. Результаты НИР внедрены в системе ГАС “Управление”, что

подтверждено соответствующими актами об использовании результатов диссертации.

### Направления развития исследований

Проведенные исследования ставят новые исследовательские задачи, которые могут быть положены в развитие настоящей работы, в частности целесообразно продолжить исследования и разработку методов и средств построения систем принятия решений, использующих нечеткую интерактивную кластеризацию текстовой информацию на основе машинного обучения и современных языковых моделей.

### Выводы

Содержание результатов исследований позволяет сделать вывод, что диссертационная работа представляет научный интерес и несомненную практическую ценность и полностью удовлетворяет требованиям ВАК РФ, предъявляемым к кандидатским диссертациям, а соискатель Дударин П.В. заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.01 – «Системный анализ, управление и обработка информации (информационные технологии и промышленность)»

Д.т.н., профессор каф. «Информационные системы»,  
ректор Ульяновского государственного  
технического университета,  
научный руководитель Дударина П.В.

Ярушкина Н.Г.



01.06.2021