

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра дискретной математики и алгоритмики

Аннотация к магистерской диссертации

«Система поиска ответов в документе»

Свидунович Денис Владимирович

Научный руководитель – кандидат физико-математических наук,
доцент Соболевская Е.П.

Минск, 2020

Реферат

Магистерская диссертация, 56 страниц, 17 рисунков, 5 таблиц, 34 формулы, 22 источника.

МАШИННОЕ ОБУЧЕНИЕ, ПОНИМАНИЕ ТЕКСТА, ПОИСК ТОЧНОГО ОТВЕТА НА ВОПРОС, СТАТИЧЕСКИЕ И ДИНАМИЧЕСКИЕ ПРЕДСТАВЛЕНИЯ, ДИСТИЛЛЯЦИЯ ЗНАНИЙ.

Объект исследования – понимание ЭВМ текстов на естественных языках, задача поиска точного ответа на вопрос по тексту.

Цель работы – исследование методов представления текстов на естественном языке в нейронных сетях, разработка системы поиска ответов на вопросы по тексту на основании лучших методов представления текстов.

Методы проведения работы – изучение научных статей по теме работы, анализ, вычислительный эксперимент, тестирование, сравнение.

Результат – изучены основные статические и динамические методы представления текстов в нейронных сетях. По каждому типу представлений текстов разработаны нейронные сети для решения задачи поиска точного ответа на вопрос по тексту. Экспериментально доказано, что динамические методы представления текстов гораздо более эффективнее статических. На основе подхода дистилляции знаний разработана целевая система, обладающая преимуществами обоих методов: скорость применения и необходимые для этого вычислительные ресурсы такие же, как и при использовании статического представления текстов, а качество всей системы повышено в сторону динамических моделей.

Область применения – решение задач обработки естественных языков, информационный поиск.

Abstract

Master thesis, 56 pages, 17 figures, 5 tables, 34 formulas, 22 references.

MACHINE LEARNING, MACHINE READING COMPREHENSION, EXACT ANSWER RETRIEVAL, STATIC AND DYNAMIC EMBEDDINGS, KNOWLEDGE DISTILLATION.

Object of research – natural language machine reading comprehension, the task of exact answer retrieval to a question about a text.

Objective – research of methods for embedding texts in natural language in neural networks, development of a system for exact answer retrieval to a question about text based on the best methods of presenting texts.

Methods – study of scientific articles on the topic of work, analysis, computational experiment, testing, comparison.

Result – the main static and dynamic methods of text embedding in neural networks are studied. Neural networks have been developed for each type of text embedding to solve the problem of an exact answer retrieval to a question based on the text. It is experimentally proved that dynamic methods of text embedding are much more effective than static ones. Based on the knowledge distillation approach, a target system has been developed that has the advantages of both methods: the speed of application and the necessary computing resources are the same as when using static text representation, and the quality of the entire system is improved towards dynamic models.

Application area – solving problems of natural language processing, information retrieval.