### 

**НЕЙРОСЕТЕВЫЕ ПОДХОДЫ К КЛАССИФИКАЦИИ ТЕКСТОВ НА ОСНОВЕ МОРФОЛОГИЧЕСКОГО РАЗБОРА**

**Ле Мань Ха**

*Московский физико-технический институт (МФТИ)*

Морфология — это раздел грамматики, изучающий грамматические свойства слов. В центре морфологии находятся понятия слова и морфемы, и грамматическую семантику, изучающую свойства грамматических морфологических значений и категорий, которые позволяют изучать смысловое значение и структуру текстов.  
  
В работе предложен метод латентно-семантической анализа на основе морфологического разбора с использованием нейронной сети. Для классификации текстов был применен анализ взаимосвязи между коллекцией документов и терминами в них встречающимися.  
  
Также предложен рекурсивный автоэнкодер с двумя частями - одна объединяет двух векторов-слов, а вторая объединяет двух векторов-морфологий. Морфорлогическая часть рекурсивного автоэнкодера позволяет повышать точности выбора векторов-слов в процедуре оформления векторного представления текста. Векторное представление текста получается повторением процесса объединения двух слов - векторов с использованием рекурсивным автоэнкодом. На каждом этапе выбор пары слов - векторов для объединения происходит с помощью данного рекурсивного автоэнкодера.  
  
Предложены подходы к оптимизации алгоритма KNN с использованием морфологического разбора и двоичной кучи.   
  
Шифрование текстов с помощью морфологического разбора. Для каждого слова, анализируем его морфологический разбор и решаем морфологическую многозначность, получаем представление текста в виде массива индексов слов в словаре и их морфологическими разборами.  
  
Уменьшение количества потенциальных ближайших соседей. Для уменьшения количества вычислительных операций, будем сравнивать только тексты, имеющие общие термины, другими словами, нет смысла сравнивать тексты, которые не имеют никаких связей.  
  
Использование двоичной кучи при поиске K ближайщих соседей для повышения быстродействия алгоритма.