Большина Ангелина Сергеевна Методы разрешения лексической неоднозначности на основе автоматически размеченных семантических корпусов

ОГЛАВЛЕНИЕ ДИССЕРТАЦИИ

кандидат наук Большина Ангелина Сергеевна

Введение

Глава 1. Теоретические аспекты задачи автоматического разрешения неоднозначности

12

1.1. Представление значений слов

1.2. Лингвистические ресурсы

1.3. Представление контекста

1.4. Оценка методов

1.5. Методы автоматизации построения обучающей коллекции

1.5.1. Метод однозначных родственных слов

1.5.2. Использование параллельных корпусов

1.5.3. Методы, использующие базы знаний

1.5.4. Алгоритм распространения меток, бутстрэппинг и активное обучение

1.5.5. Подходы, направленные на увеличение покрытия значений и слов

1.5.6. Результаты, достигаемые на автоматически порождаемых наборах данных

1.6. История развития методов автоматического разрешения неоднозначности

1.6.1. Методы, основанные на знаниях

1.6.2. Методы машинного обучения с учителем

1.6.3. Методы машинного обучения без учителя

1.7. Исследования на материале русского языка

Выводы к главе

Глава 2. Автоматическое порождение корпуса с семантической разметкой на основе однозначных кандидатов

2.1. Описание метода

2.2. Данные

2.3. Подготовка обучающей коллекции с помощью однозначных родственных слов

68

Выводы к главе

Глава 3. Снятие лексической многозначности

3.1. Разрешение лексической неоднозначности на наборе данных RUSSE-Я^ога^

3.2. Разрешение лексической неоднозначности для всех частей речи

3.2.1. Количественные характеристики многозначных слов и их однозначных родственных слов

3.2.2. Полуавтоматическая разметка всех многозначных слов в тексте

3.2.3. Анализ ошибок

3.2.4. Итоги эксперимента по предсказанию значений для всех частей речи

3.3. Разрешение неоднозначности на основе псевдоаннотированной коллекции

3.3.1. Используемые данные и модели

3.3.2. Метод порождения псевдоразметки текстов на основе ансамбля моделей

3.3.3. Результаты и выводы

3.4. Визуализация контекстуализированных представлений примеров из обучающей коллекции

3.5. Задача Word-in-Context

Выводы к главе

Заключение

Список литературы

ПЕРЕЧЕНЬ ТАБЛИЦ

СПИСОК РИСУНКОВ

ПРИЛОЖЕНИЕ 1. Результаты оценки моделей, обученных на автоматически сгенерированных коллекциях

ПРИЛОЖЕНИЕ 2. Количество примеров для слов из набора данных RUSSE-RuWordNet в сбалансированной обучающей коллекции и Корпус-1000