**Булатов Виктор Геннадьевич Методы оценивания качества и многокритериальной оптимизации тематических моделей в библиотеке TopicNet**

ОГЛАВЛЕНИЕ ДИССЕРТАЦИИ

кандидат наук Булатов Виктор Геннадьевич

Введение

Глава 1. Вероятностное тематическое моделирование

1.1 Задача тематического моделирования

1.2 Классические тематические модели

1.3 Аддитивная регуляризация тематических моделей

1.3.1 Тематическое моделирование в рамках подхода ЛКГМ

1.3.2 Мультимодальное тематическое моделирование

1.4 Приложения тематического моделирования

1.4.1 Анализ программных продуктов

1.4.2 Разведочный информационный поиск

1.5 Интерпретируемость тематических моделей

1.6 Проблемы тематического моделирования

1.6.1 Неинтерпретируемые темы

1.6.2 Вводящие в заблуждение темы

1.6.3 Неустойчивость

1.6.4 Доступность

1.6.5 Подбор гиперпараметров

Глава 2. Критерии качества тематических моделей

2.1 Экспертная оценка тематических моделей

2.2 Устойчивость

2.3 Различность тем

2.3.1 Оценки попарной различности между темами

2.3.2 Расстояние до известных распределений

2.4 Разреженность

2.5 Значимость

2.6 Теоретико-информационные критерии

2.7 Энтропия

2.8 Величины, полезные для диагностики

2.9 Анализ верхних токенов

2.9.1 Похожесть списков верхних токенов

2.9.2 Когерентность

2.9.3 Качество кластеризации

2.9.4 Критерий Ьо§Ый

2.9.5 Критерий ВЫ-эсоге

Глава 3. Внутритекстовая когерентность

3.1 Каркас вычисления когерентности

3.2 Критика метрик, основанных на верхних токенах

3.2.1 Учёт специфики коллекции

3.2.2 Когерентность и стоп-слова

3.3 Эксперимент: никаких оценок без представительства

3.4 Предлагаемая мера: внутритекстовая когрентность

3.5 Постановка задачи

3.6 Вычислительный эксперимент

3.6.1 Результаты

Глава 4. Повышение интерпретируемости тематических

моделей при помощи регуляризации

4.1 Относительные коэффициенты регуляризации

4.1.1 Регуляризация Ф и важные частные случаи

4.1.2 Регуляризация в и важные частные случаи

4.1.3 Использование на практике

4.2 Аддитивная регуляризация тематических моделей с быстрой векторизацией текста

4.2.1 ЕМ-алгоритм с быстрой векторизацией документов

4.2.2 Эксперименты

4.2.3 Метрики

4.2.4 Результаты

4.2.5 Интуитивное объяснение особенностей ТЛЯТМ

4.3 Заключение

Глава 5. Библиотека TopicNet и её применение в конкретных

задачах

5.1 Сравнимые проекты

5.2 Технология в основе

5.2.1 Достоинства BigARTM

5.2.2 Недостатки BigARTM

5.3 Мотивация и видение

5.4 Архитектура

5.5 Механизм отбора моделей в TopicNet

5.6 Сравнение с конкурентами

5.6.1 Использование ресурсов

5.6.2 Качество построенных моделей

5.6.3 Использование псевдорегуляризатора быстрой векторизации

5.7 Адаптивная траектория регуляризации

5.7.1 Ухудшение модели

5.7.2 Архитектура Regularization Controller Cube

5.8 Заключение

Глава 6. Применение регуляризации для кластеризации

интентов

6.1 Предобработка

6.1.1 Объединение и фильтрация реплик

6.1.2 Автоматическое выделение п-грам

6.1.3 Распознавание именованных сущностей

6.1.4 Коррекция ошибок

6.2 Построение модели

6.2.1 Построение уровней иерархии

6.2.2 Обработка результатов моделирования

6.3 Эксперименты

6.3.1 Оценка качества

6.3.2 Базовые модели

6.3.3 Качество предложенной модели

6.4 Выводы

Заключение

Список сокращений и условных обозначений

Список литературы

Список рисунков

Список таблиц

Приложение А. Свидетельства о государственной регистрации

программ