

# プライバシーを保護する特許検索

中川研究室 修士 2 年 胡 瀚林

指導教員： 中川 裕志 教授

2016 年 7 月 1 日

概要

## 1 INTRODUCTION

### 1.1 Patent Search

特許文章の特徴

特許検索の目的と方法

――新規性調査 (Novelty Search)

### 1.2 Patent Versus Non-patent Literature

特許文章と普通の文章の区別

## 2 PRIVATE INFORMATION RETRIEVAL

PIR の背景紹介

### 2.1 Private Information Retrieval

### 2.2 Obfuscation-Based Private Search

既存手法とその手法が特許検索に適用できない理由

## 3 LATENT SEMANTIC MODELS

### 3.1 tf-idf

### 3.2 Latent Semantic Indexing

長所: 計算簡単

短所: トピックベクトルが直交である

### 3.3 Probabilistic Latent Semantic Indexing

長所:確率的モデル

短所:トレーニングセットに含まれていない文章 (質問) の分析が困難である

### 3.4 Latent Dirichlet Allocation

長所:確率的モデルトレーニングセットに含まれていない文章 (質問) の分析が簡単

短所:学習するときは単語数  $\times$  トピック数の行列を用いて反復するので学習するには時間がかかる (30 トピック、1000 反復は 3 日かかる)

## 4 privacy-protecting patent search

提案手法

評価 (攻撃) 方法

## 5 EXPERIMENT

実験

1 質問者:tfidf 攻撃者 LSA

2 質問者:LSA 攻撃者 LSA

3 質問者:LDA 攻撃者 LSA

4 質問者:LSA 攻撃者 LDA

## 6 CONCLUSIONS

## 7 FUTURE WORKS