

Университет ИТМО

Определение хронотипа человека по его активности в социальной сети

Студент: Баев Владимир Алексеевич, гр. 4538

Научный руководитель: Фильченков А.А., к. ф.-м. н.,
ведущий инженер каф. ИС

Рецензент: Азаров А.А., к.т.н., с.н.с. лаб. ТИМПИ СПИИРАН

Санкт-Петербург 2015

Введение в предметную область

Хронотип человека – это характеристика, отражающая значения некоторых физических показателей человека, в зависимости от времени суток.

- «жаворонки» – наиболее активны в первой половине дня
- «совы» – наиболее активны во второй половине дня



Цель работы. Актуальность

Определять хромотип, анализируя активность в социальной сети (Twitter)

- таргетированный маркетинг
- оптимальный рабочий график
- выбор профессии
- диагностика заболеваний

Обзор аналогов

- Harbinger: an analyzing and predicting system for online social network users' behavior (Rui Guo, Hongzhi Wang, Lucheng Zhong, Jianzhong Li, and Hong Gao, DASFAA 2014, Bali)

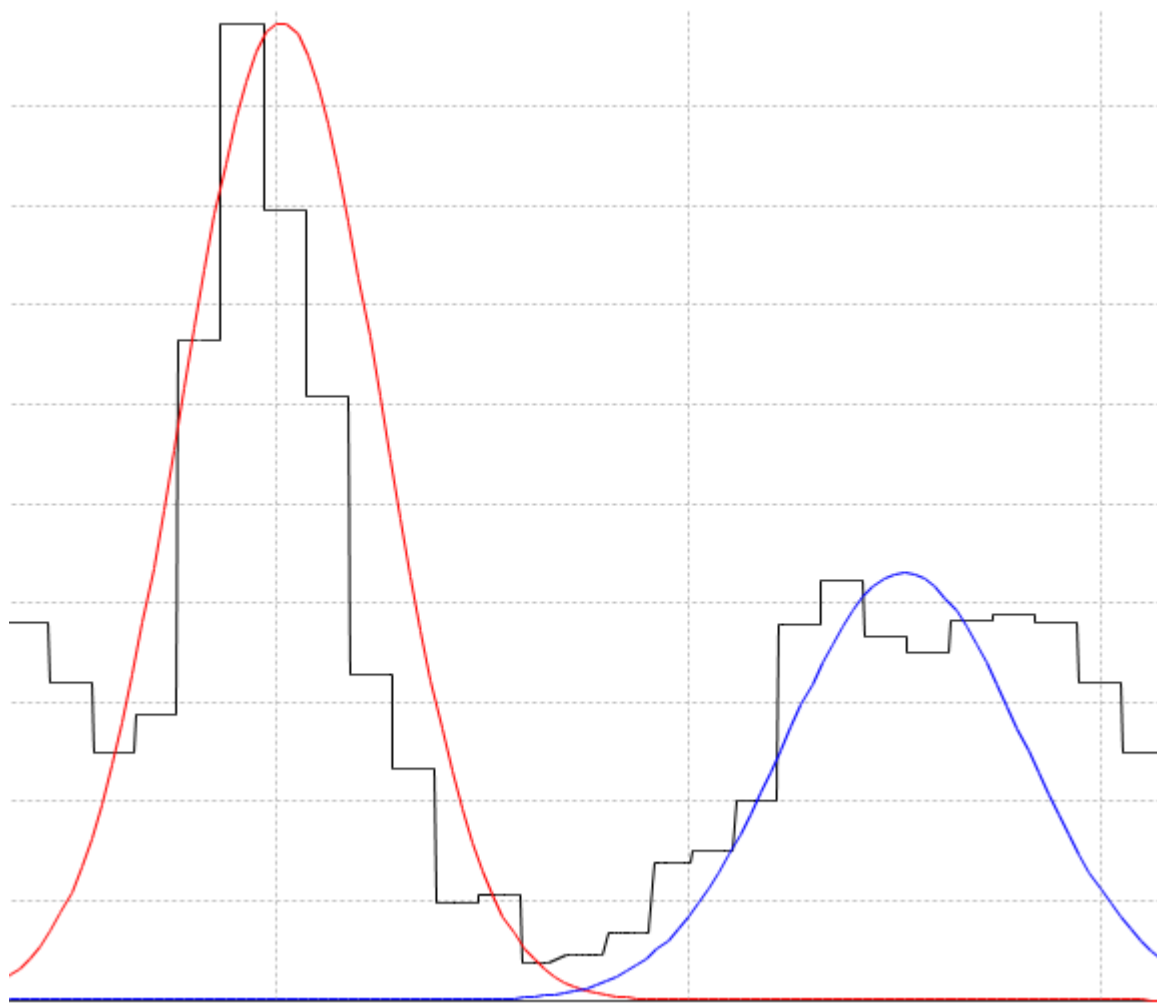
Предсказание дневной активности пользователя Twitter с помощью модели смеси гауссиан



Задачи

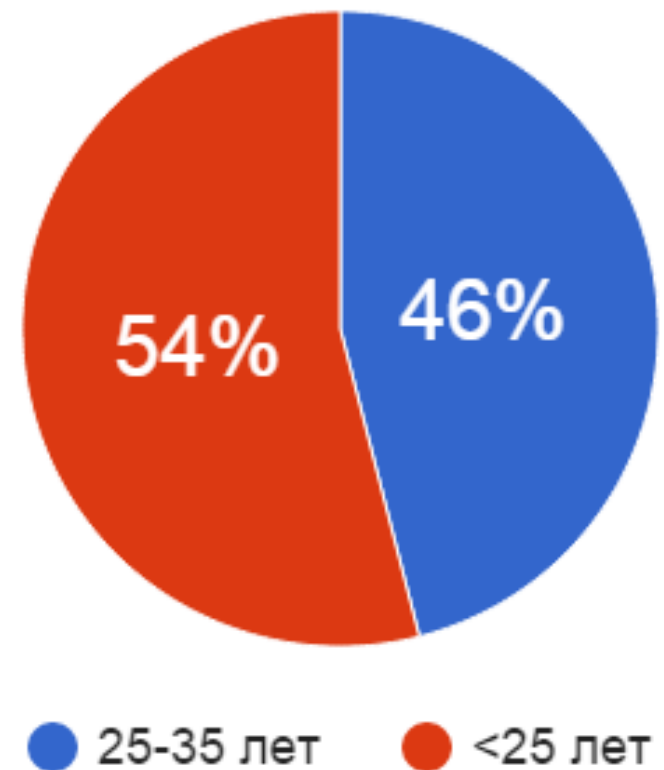
- сбор данных о хронотипах пользователей
- проверка гипотезы о связи хронотипа и суточной активности в сети
- автоматическая разметка данных
- выделение признаков на основе анализа текста сообщений
- построение классификатора

Суточная активность в сети



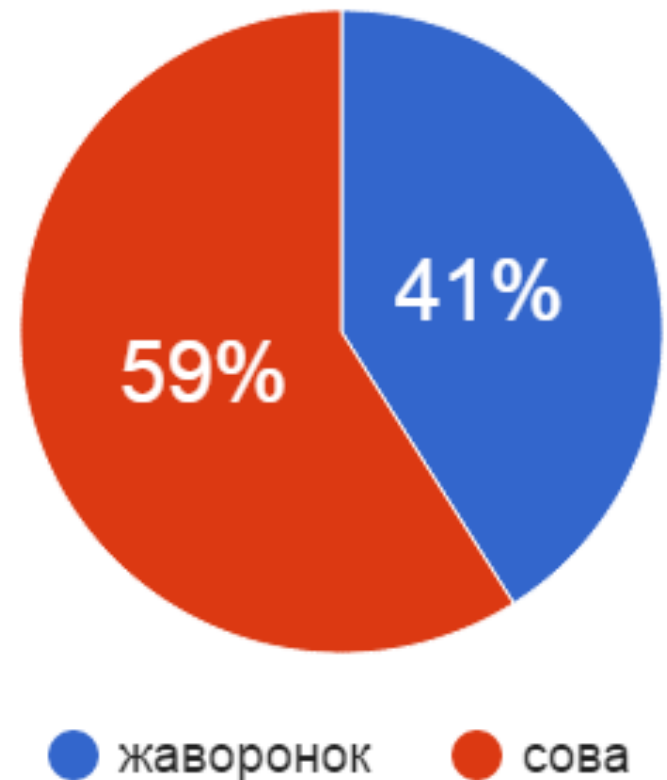
Связь хронотипа и временной активности

- 320 человек опрошено
- по статистике, 85% пользователей Twitter младше 40 лет



Связь хронотипа и временной активности

- 230 человек после фильтрации
- корреляция между реальным хронотипом и метками равна 0,69



Данные

- 230 пользователей Twitter с известным хронотипом (90 тыс. твитов)
- 1000 пользователей Twitter с «грязными» метками хронотипа (400 тыс. твитов)

Связь хронотипа и прочих показателей

- эмоциональная окраска
- склонность к депрессии
- склонность к употреблению кофеина
- предрасположенность к излишнему весу
- проблемы со сном

Признаки для анализа сообщений

- эмоциональная окраска
смайлы, оценка слов по шкале позитивности
- склонность к депрессии
депрессивная лексика, антидепрессанты
- склонность к употреблению кофеина
«кофе», «эспрессо», «американо», «глясе»
- предрасположенность к излишнему весу
названия блюд
- проблемы со сном
«бессонница», «усталость»

Связь признаков и хронотипа

Признак	Корреляция
Эмоциональная окраска	0,61
Склонность к депрессии	0,59
Склонность к употреблению кофе	0,3
Предрасположенность к излишнему весу	0,21
Проблемы со сном	0,32

Результаты. F_1 -мера

Алгоритм	Автоматическая разметка
KNN	0,63
NaiveBayes	0,62
Backpropagation	0,73
AdaBoost	0,74
RandomForest	0,70

Архитектура

Сбор сообщений пользователя
(Twitter API)



Разметка данных с помощью
модели смеси гауссиан



Выделение признаков для
классификатора



Построение классификатора

Результаты. F_1 -мера

Алгоритм	Ручная разметка
KNN	0,56
NaiveBayes	0,58
Backpropagation	0,65
AdaBoost	0,67
RandomForest	0,70

Результаты

- выявлена связь между хронотипом и суточной активностью в сети
- выделены признаки для построения классификатора
- построен классификатор для определения хронотипа

Университет ИТМО

Определение хронотипа человека по его активности в социальной сети

Студент: Баев Владимир Алексеевич, гр. 4538

Научный руководитель: Фильченков А.А., к. ф.-м. н.,
ведущий инженер каф. ИС

Рецензент: Азаров А.А., к.т.н., с.н.с. лаб. ТИМПИ СПИИРАН

Санкт-Петербург 2015