

Social Data Science

Методы науки о данных в социальных исследованиях

Иван Смирнов, заведующий лабораторией Методов науки о данных в исследованиях образования, Институт образования ibsmirnov@hse.ru

Социальные науки стремительно компьютеризируются. Данные из социальных сетей и другие цифровые следы позволяют проводить исследования на гораздо больших выборках и с гораздо большей детализацией, чем это возможно с помощью традиционных методов. Новыми становятся не только данные, но и методы их анализа. Так, например, методы машинного обучения все чаще используются в социальных исследованиях. Использование новых данных требует от исследователей новой оптики, позволяющей видеть альтернативные источники данных, и новых навыков (извлечения, предобработки и анализа). Вскоре методы науки о данных окажутся незаменимы для подавляющего большинства исследователей также, как сегодня для них незаменимо знание математической статистики.

Цели курса

Цель	Чем проверяется
Познакомить обучающихся с наиболее актуальными работами по теме курса, сформировать у них представление о том, какие данные и из каких источников можно извлекать с помощью современных методов	Подготовка и презентация проекта исследования, использующего интернет-данные
Снабдить обучающихся минимально необходимыми навыками для извлечения новых типов данных (интернет-данных, других цифровых следов и др.)	Выполнение практических заданий
Обучить методологии извлечения данных и познакомить с методами их анализа	Итоговый индивидуальный проект по извлечению и анализу данных

Тематический план

1. Новые типы данных: интернет-данные, другие цифровые следы и возможности их применения в социальных науках
2. Знакомство с API ВКонтакте

3. Использование API ВКонтакте: интересы пользователей
4. Построение сетей дружбы
5. Презентация проектов исследований
6. Сопоставление интернет-данных с данными, полученными из других источников
7. Извлечение интернет-данных: пример ЦИАН
8. Извлечение интернет-данных: примере livejournal
9. Работа с текстовыми данными
10. Работа с геоданными
11. Презентация итогового проекта

Оценка

Требования	%, от оценки
Проект исследования	30%
Практические задания	30%
Итоговый проект	40%

Описание заданий

Самостоятельная работа слушателей включает три типа заданий:

Проект исследования

Описание: Обучаемым предлагается подготовить проект исследования, использующего интернет-данные, объемом 1-2 страницы и представить его. Проект должен включать в себя постановку исследовательского вопроса, описание интернет-данных, которые предполагается использовать, и методов их анализа.

Оценка: Оценивается то, насколько четко сформулирован и обоснован исследовательский вопрос, и то, насколько предлагаемые методы и данные действительно позволяют ответить на поставленный вопрос. Проект оценивается по десятибалльной шкале, вес в итоговой оценке: 30%.

Практические задания

Описание: После каждого занятия обучаемым предлагается выполнить практические упражнения. Упражнения базируются на том коде, который обсуждался на занятиях, однако требуют его модификации или дополнения и позволяют выяснить, насколько хорошо обучаемыми был усвоен материал.

Оценка: Каждое задание оценивается по трехбалльной шкале 0 — задание не выполнено, 1 — задание выполнено частично, 2 — задание полностью выполнено. Всего предполагается 9 заданий, из них учитывается 8 лучших. Вес в итоговой оценке: 30%.

Итоговый проект

Описание: Итоговый проект включает в себя сбор интернет-данных, их анализ и интерпретацию полученных результатов. Обучаемые готовят письменный отчет о выполнении проекта, включающий в себя написанный ими код, основные результаты и их интерпретацию. Помимо письменного отчета обучаемые готовят презентацию результатов и представляют ее на заключительном занятии.

Оценка: Учитывается техническое выполнение задания (корректность процедуры сбора данных и их анализа), а также то, насколько обоснованно были интерпретированы результаты. Итоговый проект оценивается по десятибалльной шкале, вес в итоговой оценке: 40%.

Список литературы

1. Salganik, M. J. (2017) *Bit By Bit: Social Research in the Digital Age*
2. Stephens-Davidowitz, S. (2017). *Everybody lies: Big data, new data, and what the internet can tell us about who we really are*. HarperCollins
3. Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15), 5802-5805.
4. Hobbs, W. R., Burke, M., Christakis, N. A., & Fowler, J. H. (2016). Online social integration is associated with reduced mortality risk. *Proceedings of the National Academy of Sciences*, 113(46), 12980-12984.
5. Palchykov, V., Kaski, K., Kertész, J., Barabási, A. L., & Dunbar, R. I. (2012). Sex differences in intimate relationships. *Scientific Reports*, 2, 370.
6. Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788-8790.
7. Smirnov, I., Sivak, E., & Kozmina, Y. (2016). *In Search of Lost Profiles*. Educational Studies Moscow No 4, 2016.
8. Smirnov, I., & Thurner, S. (2017). Formation of homophily in academic performance: Students change their friends rather than performance. *PloS one*, 12(8), e0183473.