

ТЕХНОЛОГИИ СБОРА И АНАЛИЗА БОЛЬШИХ ДАННЫХ В КОНТЕКСТЕ СОЦМЕДИА

И. С. Хрипович

Белорусский государственный университет, г. Минск,

hripovich@bsu.by;

науч. рук. – А. Н. Курбацкий, д-р тех. наук, проф.

В рамках данного доклада предлагается обзор технологий и инструментов, которые могут быть использованы для решения поставленных задач.

Ключевые слова: большие данные; социальные сети; технологии сбора и анализа.

Активное развитие коммуникационных технологий, в первую очередь сети Интернет, привело к тому, что значительная часть социального взаимодействия происходит посредством социальных сетей, мессенджеров и других интернет-ресурсов. Ключевой особенностью такого взаимодействия является публичность и сохранность данных.

На сегодняшний день информационное пространство содержит огромное количество сведений о людях. Каждый человек на протяжении жизни оставляет свой след в информационном пространстве, при этом информация может быть представлена в явном виде это: анкетные данные, фотографии, видео, сообщения на различных ресурсах, профили, а может быть представлена в неявном виде это: упоминания об этом человеке другими людьми, история его посещений веб-страниц, перемещений в пространстве, использования отдельных сервисов и ресурсов. Данная информация содержится как в открытом доступе, так и в закрытых базах данных. Сбор и накопление персональных данных, а также любых данных, которые могут быть сопоставлены с некоторым человеком или группой лиц, осуществляется широким кругом организаций и структур. Собранные данные могут использоваться для рекламы, маркетинга, манипулирования общественным сознанием и информационных атак.

Прогнозирование поведения социальных групп традиционными методами сопряжено со следующими трудностями:

- отсутствие автоматизации процесса;
- недостаточность актуальных данных;
- проверка гипотез в ретроспективе.

Использование информационного пространства в качестве источника исходных данных позволяет получать изменяющиеся данные в реальном времени, а также позволяет получать ретроспективные выборки данных, что в свою очередь позволяет строить прогнозные модели по периоду в прошлом для проверки их адекватности на реальных данных.

Задача прогнозирования поведения социальных групп на основе больших данных разбивается на следующие подзадачи: сбор данных, агрегация данных и сопоставление их с определенной личностью, кластеризация личностей по социальным группам, исследование поведения социальных групп и построения моделей, непосредственно прогнозирование поведения.

Построение модели поведения социальных групп позволит прогнозировать реакцию социума на те или иные воздействия, что обеспечит: предсказуемость действий при наступлении некоторого события, планирование мероприятий для достижения максимального эффекта, возможность противодействия информационным атакам.

В рамках данного доклада будет дан обзор технологий и инструментов, которые могут быть использованы для решения поставленных задач.