

**Перспективы и возможности использования «Больших данных» в
политическом анализе**

Научный руководитель – Танцура Марина Сергеевна

Гриценко Руслан Артемович

Студент (бакалавр)

Дальневосточный федеральный университет, Восточный институт - Школа
региональных и международных исследований, Владивосток, Россия

E-mail: st.ruslan96@mail.ru

Для начала необходимо определиться с тем, что же такое Большие данные. В отчете McKinsey Institute дается следующее определение: «набор данных, размер которых превосходит возможности типичных баз данных по занесению, хранению, управлению и анализу информации» [1]. Другими словами, Большие данные - это весь тот массив информации, который мы оставляем после посещения сети интернет. То, какие сайты мы посещали, в какие разделы заходили, что искали в поисковых системах, какие страницы в социальных сетях посещали, на какие записи ставили «лайки», что комментировали и т.д.

Как можно понять из определения, Большие данные сегодня уже используются для анализа информации. Но чаще всего этот анализ не имеет никакого отношения к политике, или имеет, но только косвенное.

Сначала мы попробуем разобраться, для чего нужны Большие данные, а затем поговорим о перспективах использования Больших данных в политическом анализе.

В 2009 компания Google спровоцировала ажиотаж в медицинских кругах. Компания провела исследование и доказала возможность отслеживания сезонных вспышек гриппа. Была сопоставлена частота запросов в поисковой системе Google и данные об эпидемиях гриппа по регионам. Исследователи не претендовали на точное знание того, какие именно поисковые запросы являются наилучшими и наиболее точными индикаторами, поэтому просто пропускали все запросы через специальный алгоритм, который их ранжировал. Проанализировав почти полмиллиарда запросов, исследователи получили 45 терминов («боль в горле», «из носа течет»), которые сильно коррелировали со вспышками гриппа. Все эти термины так или иначе связаны с гриппом, но невооружённым глазом трудно из полмиллиарда запросов увидеть те, которые будут пригодны для проверки и дадут результат [2]. Данные исследования Google доказали, что по поведению людей в интернете можно предсказать некоторые события в том случае, если эти события уже происходили в прошлом и поведение людей в сети перед этим событием было примерно таким же, как и в данный момент.

Теперь в случае получения данных об увеличении частоты запросов «боль в горле», «из носа течет» в каком-либо регионе, можно с высокой вероятностью прогнозировать вспышку гриппа в этом регионе, и превентивно предпринять необходимые меры, чтобы пресечь эту эпидемию.

Анализ Больших данных не объясняет то, как связаны частота запросов и вспышки гриппа, даже не имеет подобной цели. Он просто показывает то, что эта связь есть.

Работа с Большими данными заключается в поиске корреляций, а не причинно-следственных связей. Анализ Больших данных базируется на совпадении различных явлений в прошлом и вероятности их совпадения в будущем. Мы всё ещё не знаем, почему происходят некоторые события и явления, но теперь мы можем предсказать, что эти события произойдут.

Рассмотрим, какие корреляции между определенными действиями людей в интернете и последующими событиями можно использовать в политическом анализе.

Приведем пример того, как это можно использовать. Недавно прошли митинги 26 марта 2017 года «Он вам не Димон», на которые вышло большое количество людей. Анализируя Большие данные можно предсказать вероятность того, когда повторится протестная волна такой же силы. Для этого необходимо проанализировать уровень активности в сообществах оппозиционной направленности, количество подобных запросов в поисковых системах «Расследование Навального», «Навальный», «Он вам не Димон» и т.д. количество подобных ключевых слов в постах, которые предшествовали митингам 26 марта за какой-то промежуток времени. В дальнейшем, при достижении вышеперечисленными показателями тех же частот, что и перед митингами 26 марта, можно прогнозировать повторение этой волны.

Если коротко: есть какой-то прецедент, есть данные, при которых произошел этот прецедент и в случае повторения этих данных, можно прогнозировать высокую вероятность повторения этого прецедента.

Прецедентом, или контрольной точкой, в данном случае может выступать не только какое-то событие, но так же и ситуация в обществе в целом, состояние этого общества и т.д. Например, можно измерить отношение жителей Приморского края к В. В. Миклушевскому. Для этого необходимо проанализировать частоту таких запросов жителей Приморского края в поисковых системах «губернатор», «Миклушевский» и т.д., что даст нам понимание того, насколько фигурой губернатора интересуются жители края. Затем нужно проанализировать, что это были за запросы, каких было больше, негативных (Миклушевский украл, Миклушевский и вертолет) или позитивных (Миклушевский сделал, Миклушевский проверил), чтобы понять, какое отношение преобладает. По тому же принципу необходимо проанализировать посты и комментарии жителей Приморского края в социальных сетях, чтобы выявить соотношение позитивных и негативных оценок. Проведя вышеперечисленные операции, мы получим данные об отношении к фигуре губернатора жителями края. Для получения более полной картины, можно сравнить эти данные с данными за предыдущие года, либо с данными отношения населения других регионов к своим главам.

Иными словами, можно не только прогнозировать, но и измерять. Можно прогнозировать и измерять уровень гражданской активности населения, повышение или падение рейтинга какой-либо политической силы, уровень поддержки власти, одобрение или неодобрение какого-либо политического решения.

Опосредованный метод анализа Больших данных используется в том случае, когда перед анализом Больших данных анализируются «малые данные».

Проводя опрос, мы выясняем возраст, пол, место жительства, место работы, интересы, политические предпочтения, ценности и убеждения, а так же характер и темперамент респондентов. Затем мы применяем психометрические методики, а именно метод OCEAN, суть которого заключается в том, что на основании пяти критериев можно разбить всех людей на несколько психотипов. Вот эти критерии: открытость (насколько вы готовы к новому), добросовестность (насколько вы перфекционист), экстраверсия (как вы относитесь к социуму), доброжелательность (насколько вы дружелюбны и готовы к сотрудничеству) и нейротизм (насколько легко вас вывести из себя) На основе этих измерений можно точно понимать, с каким человеком имеешь дело, в чем его желания и страхи, наконец, как он себя может вести [3]. После прохождения респондентами опроса, мы можем всех их поделить на психотипы по данным методике, плюс разделить их внутри этих психотипов по политическим взглядам, интересам, месту жительства, месту работы и т.д. Иными словами, знаем о респонденте практически всё. Это первый этап.

На втором этапе начинается работа с Большими данными. Мы находим наших респондентов в социальных сетях и анализируем их поведение: уровень их активности в сообществах каких-либо тематик. Затем, среди членов нашей генеральной совокупности, мы ищем пользователей, у которых схожий уровень активности в сообществах с теми же или схожими тематиками. И на основании этих данных распределяем всех членов генеральной совокупности по этим психотипам. В результате, мы получаем всю генеральную совокупность распределённую по интересам, темпераменту и политическим убеждениям с адресами и аккаунтами в социальных сетях.

Этот метод строится на утверждении: если люди посещают одни и те же сайты, «лайкают» похожие вещи, имеют схожие запросы в поисковых системах и ключевые слова в своих постах, то они так же имеют и схожие ценности, убеждения, интересы и темперамент.

Данный метод позволит сегментировать всю целевую аудиторию по ценностям, интересам и темпераменту, а не по демографическим и локальным признакам. Такой подход намного эффективнее, ведь зная ценности и интересы целевой аудитории гораздо проще на неё повлиять. Показать нужную рекламу, выбрать ЛОМов, к которым она прислушается и т.д.

Таким образом, использование Больших данных в политическом анализе позволяет качественно улучшить результат анализа, а так же дает более оперативные данные, ведь мы можем использовать большой массив данных тогда, когда нам это необходимо и без больших затрат. Анализ Больших данных экономичен и мобилен. Однако Большие данные могут быть применимы только в том случае, когда объектом являются группы людей, их мнение, поведение и т.д. Если задачей исследования является прогнозирование поведения политических лидеров, оценка какого-либо политического решения и т.д. Большие данные не представляют ценности для исследования.

Анализ Больших данные - это новый социологический метод, который будет становиться только актуальнее, потому что с каждым годом уровень интернализации населения увеличивается, и следовательно, увеличивается количество граждан, о которых можно получить Большие данные а так же увеличивается объем, степень проработки и глубина этих данных.

Источники и литература

- 1) Ковлягина, Т. Большие данные закрепились в России [Электронный ресурс] / CleverData. – Режим доступа: <http://cleverdata.ru/bolshie-dannye-zakrepilis-v-rosii/> (Дата обращения: 25.04.2017)
- 2) Кукикер, К. Майер-Шёнбергер, В. Большие данные [Электронный ресурс] / Россия в глобальной политике. – Режим доступа: <https://dlib.eastview.com/browse/doc/35311570?searchLink=%2Fsearch%2Fsimple> (Дата обращения: 25.04.2017)
- 3) The Insider. Расследование Das Magazin: как Big Data и пара ученых обеспечили победу Трампу и Brexit [Электронный ресурс] / The Insider. – Режим доступа: <http://theins.ru/politika/38490> (Дата обращения: 25.04.2017)