

Подходы к анализу данных социальных сетей и сообщений от населения в геоинформационных системах мониторинга наркоситуации в регионе

П.А. Митягин, С.Н. Болбин, С.А. Митягин

Санкт-Петербургский национальный исследовательский университет
информационных технологий, механики и оптики,
Санкт-Петербургский информационно-аналитический центр
petr_mityagin@mail.ru, sergey.bolbin@gmail.com, mityagin@iac.spb.ru

Аннотация

В работе рассматриваются подходы к анализу информации о социальных явлениях, скрытых от непосредственного наблюдения на примере наркомании. В качестве доступного информационного ресурса для анализа рассматриваются сообщения населения о фактах наркомании в районе проживания и данные исследований социальных сетей на предмет потребления наркотических веществ.

Введение

Создание и внедрение государственной системы мониторинга наркоситуации является одним из основных направлений реализации государственной антинаркотической политики Российской Федерации до 2020 года, утвержденной указом Президента РФ от 9 июня 2010 года № 690 [1].

Согласно стратегии антинаркотической политики, методике и порядку осуществления мониторинга, а также критериям оценки развития наркоситуации, утвержденных п. 2.1 протокола заседания Государственного антинаркотического комитета от 21 декабря 2011 г. № 14, необходима разработка государственной информационной системы мониторинга наркоситуации.

Сбор и систематизация данных об уровне распространения наркомании в регионах способствует повышению эффективности всей антинаркотической деятельности. При этом, особое внимание уделяется осуществлению социологических исследований по данной тематике. Следует отметить, что проведение социологических исследований традиционным, опросным путем, часто вызывает трудности как при самой организации опроса и при финансировании данных исследований, так и при анализе полученных результатов.

Альтернативой, ставшим в последнее время традиционными опросным исследованиям, может яв-

ляться исследование социальных сетей. Данное направление работ является весьма перспективным, поскольку независимо от специфики региона, за исключением уровня развития ИКТ.

Исследование социальных сетей позволяет получить еще один важный срез знаний о проблеме наркотизации населения помимо общепринятых на данный момент социологических исследований методами опроса населения и анализа данных официальной государственной статистики государственной власти, органов здравоохранения и органов наркоконтроля.

Факт того, что наркомания передается при личном контакте лиц, подверженных наркомании, и его учет позволяет более точно и эффективно противодействовать данной проблеме [7]. Однако, большинство традиционных подходов рассматривают анализ макроуровневых показателей. Таким образом, требуется разработка методик, позволяющих свести воедино традиционный статистический подход на высоком уровне и алгоритмы микро-социального анализа.

В Российской Федерации мониторинг наркоситуации осуществляется на федеральном уровне и уровне субъектов Российской Федерации с использованием единого банка данных по вопросам, касающимся оборота наркотических средств, психотропных веществ и их прекурсоров, а также противодействия их незаконному обороту, который формируется на указанных уровнях.

За счет привязки данных о распространении наркомании к фиксированным точкам территории региона достигается эффективное выявление областей напряженности, требующих особого внимания при планировании и организации деятельности по противодействию незаконному обороту наркотиков. Данные методы имеют широкое применение в задаче организации и планирования работы с населением, в частности, в задаче проведения профилактических мероприятий на территории.

Совместный анализ данных социометрических исследований социальных сетей интернета и картографических данных позволяет осуществлять анализ и прогнозирование распространения наркомании на основе новой логики: используя знания об уровне

пропаганды наркомании, вовлеченности населения в различные субкультурные сообщества, в которых поощряется потребление наркотиков осуществляется обобщенная оценка наркоиммунитета территории как готовности общества противостоять наркомании; сообщения граждан о местах распространения наркотиков и появления наркоманов позволяют выявить наиболее кризисные области на территории, а также места скопления представителей сообществ; общее же состояние территории и в частности кризисных областей описывается макроэкономическими показателями развития территории на основе государственной ведомственной статистики [11].

Данная логика повторяет логику мер противодействия от планирования целевых индикаторов на уровне основных направлений деятельности правительства и составления целевых программ, до планирования и реализации отдельных мероприятий.

1. Исследование социальных сетей в интернете на предмет пропаганды приема наркотиков

Отсутствие данных о распространении наркотиков от индивида к индивиду, заставляет нас использовать специальные методы установления связей между индивидами, например, анализ социальных сетей в интернете. Социальные сети, являющиеся своего рода слепком общества, доступным для наблюдения, и позволяют произвести оценивание распространённости наркокультуры по ряду формальных признаков. Определение численных характеристик сегмента социальной сети позволяет построить комплексную сетевую модель наркотизации населения на данных социальной сети. Оценка наркотизации населения на основе комплексной сетевой модели, полученной на основе мониторинга социальных сетей, может являться смещенной, в силу сложности получения репрезентативной выборки, но достоверно описывает психологическое состояние общества.

В настоящее время весьма актуальной задачей является разработка автоматизированного исследовательского инструментария, позволяющего проводить социометрические исследования в социальных сетях. На решение данных задач ориентирован проект, реализуемый в Санкт-Петербургском национальном исследовательском университете информационных технологий, механики и оптики (НИУ ИТМО), в рамках которого создан веб-ориентированный производственно-исследовательский центр для решения прикладных задач в области социодинамики [6, 3].

Задача исследования наркотизации населения на основе данных социальных сетей включает две подзадачи: мониторинга социальных сетей с восстановлением топологии графа социальных связей и классификации узлов сети по критерию вовлеченности в наркоманию и моделирование наркотизации населения.

Вследствие того, что объектом наблюдения является текстовый массив, содержащийся в узлах сети (на страницах пользователей) следует говорить не столько о вовлеченности в наркоманию, сколько о вовлеченности узла в наркокультуру.

В данной работе осуществлено исследование фрагмента социальной сети LiveJournal на основе технологии краулинга, заключающейся в обходе социальной сети по ссылкам между страницами пользователей. Данный подход позволяет не только осуществлять оценивание агентов социальной сети по публикуемой информации, но также восстанавливать структуру отношений агентов.

Алгоритм краулинга социальной сети можно описать следующим образом:

1. Определение стартовых страниц пользователей, мнения и взгляды которых соответствуют заданной тематике – потребления наркотиков.

2. Переход по ссылкам на страницы связанных пользователей.

3. Оценивание связанных пользователей на основе алгоритмов интеллектуального исследования текстов на предмет выявления искомых мнений и взглядов.

3.1. В случае отсутствия искомых мнений обход прекращается;

3.2. В противном случае алгоритм повторяется с шага №2.

Говоря об анализе информации узлов социальной сети, следует отметить, что сленг индивидуумов, потребляющих наркотики может в различной степени отличаться от общепринятых норм. С другой стороны сообщества потребителей наркотиков в большой степени закрыты в силу их криминального характера. По этой причине с целью выявления пользователей предположительно имеющих опыт потребления наркотиков предлагается использовать базу знаний сигнальных слов. Среди сигнальных слов рассматриваются: сленг потребителей и распространителей наркотиков, специальные термины наркологии, сопутствующие слова и термины, имеющие двойной смысл применительно к тематике наркомании.

На основе описанного алгоритма было осуществлено исследование фрагмента социальной сети LiveJournal. Обобщенные характеристики исследования приведены в табл.1

Таблица 1. Обобщенные характеристики исследования социальной сети LiveJournal

Свойство	Результат
Размер словаря	450 (слов), 1500 (фраз)
Размер фрагмента сети	785107 (узлов)
Количество пользователей, в текстах которых выявлена тематика наркомании	32899 (узлов)
Количество различных тематик и интересов пользователей фрагмента сети	192709

Полученный массив узлов (страниц) социальной сети в большей степени содержит узлы, тексты которых содержат эпизодические упоминания о наркотиках и наркомании.

Анализ сигнальных слов, выявленных в текстах, показал наличие характерных групп слов, определяющих вовлеченность лиц в наркокультуру. Полученное соотношение слов по критерию принадлежности жаргону свидетельствует о наличии центральной группы узлов, в текстах которых в равной мере представлены слова, описывающие приготовление, способ приема и производимый эффект.

2. Исследование картографической информации о наркоситуации на территории

Рассматривается задача мониторинга и анализа социальных процессов с использованием картографической информации. За счет привязки данных об исследуемом явлении к фиксированным точкам территории региона достигается эффективное выявление областей напряженности, требующих особого внимания при планировании и организации мер улучшения жизни населения. Данные методы имеют широкое применение в задаче организации и планирования работы с населением, в частности, в задаче проведения профилактических мероприятий на территории. Данная задача рассматривается в рамках разработки экспертной системы анализа развития региона.

На данный момент существует определенных разрыв в организации деятельности между органами наркоконтроля, органами здравоохранения и организациями профилактической направленности, в том числе деятельности антинаркотических комиссий. В преодолении данного разрыва в организации деятельности, а также анализе данных мониторинга и выработке управленческих решений, заключается цель создания и функционирования единого банка данных и автоматизированной информационно-аналитической системы, как на региональном, так и федеральном уровнях.

Перспективным направлением развития систем мониторинга наркоситуации в области организации межведомственного взаимодействия и информационной поддержки принятия управленческих решений является применение геоинформационных систем (ГИС) [8].

По определению картографические данные – это пространственная информация, обычно явно привязанная к некоторой части земной поверхности и используемая для управления ситуации [6]. В большей степени нас интересует информация, характеризующая уровень развития наркомании на территории. Как правило, такая информация включает в себя данные о преступности в сфере незаконного оборота наркотиков, данные о заболеваемости и болезненности наркоманий, а также данные о состоянии населения и общественном мнении по данной проблеме.

Достоинством данного подхода является его наглядность и простота восприятия, а также жесткая привязка к реальным объектам, что упрощает интерпретацию результатов и выработку управленческих решений. Вследствие описанных особенностей рассматриваемые методы применимы в задаче оценки эффективности работы с населением, в частности, в задаче организации мероприятий первичной профилактики. Организация профилактической работы подразумевает определение места проведения и целевую аудиторию ряда мероприятий, что позволяет осуществлять ее анализ посредством ГИС [8].

В качестве примера рассматривается модель криминального поведения, в частности незаконное распространение наркотических веществ [12]. На данный момент существует определенный разрыв в организации деятельности между органами наркоконтроля, органами здравоохранения и организациями профилактической направленности, в том числе деятельности антинаркотических комиссий.

Рассматриваются следующие множества картографических объектов:

- Сигналы о преступлениях, связанных с незаконным оборотом наркотиков, поступающие от граждан. Данная информация поступает в рамках инновационной программы «Наш город» [9]. Так каждый сигнал (заявление) несет информацию о месте (адресе) правонарушения и ему соответствует точка на карте;

- Факты выявления правонарушений, связанных с незаконным оборотом наркотиков по месту совершения правонарушения;

- Факты об отравлениях наркотическими веществами, рассматриваемые органами здравоохранения. В рамках построения множества больных наркоманией определяются адреса проживания зарегистрированных наркозависимых.

Целью обработки картографических данных является определение наиболее кризисных областей на рассматриваемой территории. Каждое мероприятие профилактики характеризуется местом проведения и целевой аудиторией. Выделение групп точек позволяет определить области большей и меньшей напряженности на территории. На основе полученных кластеров можно выделить кризисные и благоприятные зоны по обращениям граждан, см. рис. 1а.

Полученный результат отличается от результатов группировки точек на основе административного деления территории, но подобные данные являются непосредственной рекомендацией к организации работы по противодействию незаконному обороту наркотиков.

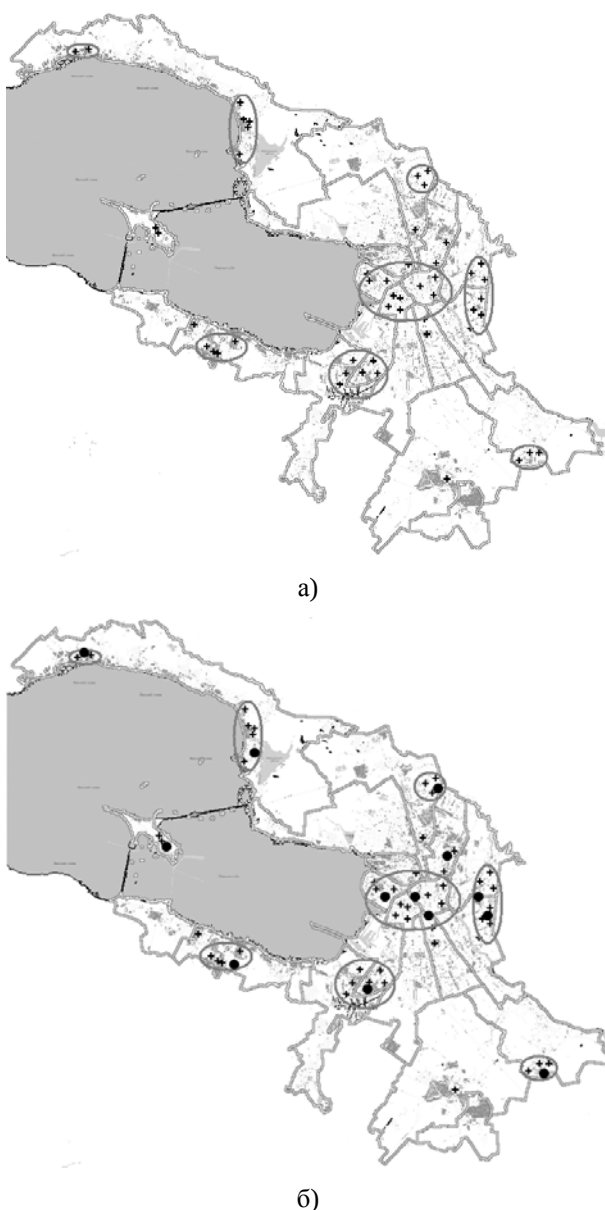


Рис 1. а) Распределение точек по кластерам; б) Расположение точек, минимизирующее критерии наложения

В качестве критериев эффективности противодействия незаконному обороту наркотиков имеет смысл рассматривать величину соответствия множеств, где мерой соответствия между множествами является близость точек или групп точек. С использованием методов, описанных в данной работе, задача определения перспективного места проведения текущих мероприятий сводится к задаче двумерного поиска. Пример такого расположения приведен на рис. 1б.

Таким образом, достигается возможность определить непосредственно места, школы, учебные центры в которых существует наибольшая необходимость в проведении профилактических мероприятий, либо выявить неэффективность подобных мероприятий на территории. Тем самым достигаются цели информационной поддержки принятия управленческих решений. Предложенные методы позво-

ляют связать данные мониторинга наркоситуации и проведение профилактических мероприятий, чем в некоторой степени ликвидируется разрыв между этими областями деятельности по противодействию незаконному обороту наркотиков.

3. Подходы к моделированию наркотизации населения

В качестве примера рассматривается модель наркотизации населения, основанная на графовом представлении группы риска по наркомании. С точки зрения имитационного моделирования процесс наркотизации рассматривается как цепная реакция процесса передачи наркотиков от источника наркомании (действующий наркоман, либо распространитель наркотиков) к восприимчивым лицам из группы риска, что позволяет представить данный процесс взвешенным неориентированным графом.

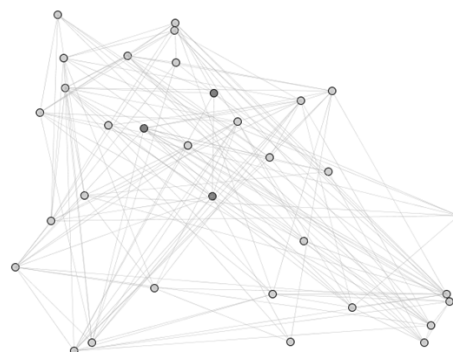


Рис 2. Представление графа социальных связей

Оценивание структуры социальных связей на данный момент является открытым вопросом, поскольку не поддерживается реальными данными. Однако исследование социометрических характеристик пользователей социальных сетей позволяет оценить обобщенных характеристики связей лиц, подверженных наркомании и вовлеченных в различные субкультуры, так и лиц с иммунитетом к наркотикам. Например, рассматриваются связи между узлами, среди которых: 10% составляют связи с узлами старших возрастных групп; 10% составляют связи с узлами младших возрастных групп; 80% составляют связи в своей возрастной группе с разницей 1-3 года.

В данном примере осуществлялась апробация площадной модели наркотизации населения на примере города Санкт-Петербурга. Рассматривается детализация данных структуры населения до домов Санкт-Петербурга. Таким образом, в модели учтены более 127 тыс. домов (адресов) Санкт-Петербурга, представленных в виде агентов, каждый из которых определяет состояние дома (адреса) как наличие или отсутствие наркозависимых среди жильцов.

Внешними данными представленной модели являются информация о преступности на уровне административных районов Санкт-Петербурга, точки (адреса), где была зафиксирована криминальная активность.



Рис. 3 Размещение агентов на карте Санкт-Петербурга и моделирование распространения наркотизма в обществе

Таким образом, данный подход позволяет оценить вовлеченность населения в криминальное поведение и осуществить прогнозирование развития данного явления в социуме на основе социальной структуры данного явления. Под социальной структурой понимается выявление группы риска по данному явлению, половозрастной структуры группы риска, факторов, сопутствующих исследуемому процессу.

Заключение.

Представленный подход позволяет оценить вовлеченность населения в криминальное поведение и осуществить прогнозирование развития данного явления в социуме на основе социальной структуры данного явления. Под социальной структурой понимается выявление группы риска по данному явлению, половозрастной структуры группы риска, факторов, сопутствующих исследуемому процессу. Тем самым решается поставленная задача.

Разработанная модель способствовала выделению факторов, воздействие которых на удовлетворенность жизнью наиболее значимо для всех стилей поведения, оценке их вклада в удовлетворенность жизнью и построению общей и частной дескриптивных линейных моделей удовлетворенности жизнью. На следующем этапе были созданы более сложные модели, учитывающие и нелинейные функциональные стохастические зависимости.

Все это позволило подойти к разработке модельного компьютерного эксперимента для исследования адаптации людей с различными стилями поведения и с разным уровнем субъективного качества жизни. При этом динамика удовлетворенности жизнью рассматривается нами как макропроцесс, который моделируется как результат микродинамического процесса циклических переходов из состояния удовлетворенностью жизнью в состояние неудовле-

творенностью жизнью. В качестве математического аппарата в данном случае может применяться теория агентов и коалиционные автоматы. Собственно циклический характер психических явлений был подробно исследован Д.В. Сочивко в рамках предложенного им психодинамического подхода [10].

Рассмотренный подход интеграции социального моделирования криминальных процессов позволяет осуществить интеграцию анализа статистических данных с ГИС и существенно упростить процесс планирования мероприятий противодействия негативным явлениям за счет большей детализации исходных данных, что позволяет его рассматривать как основу создания общей экспертной системы поддержки принятия управленческих решений. Данный метод имеет ряд недостатков, связанных с высокими требованиями к вычислительным средствам, необходимостью собирать, хранить и обрабатывать большие объемы данных.

Работа выполнена в рамках реализации постановлений № 220 и № 218 Правительства Российской Федерации при поддержке ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007-2012 годы» (проект № 07.517.11.4042 от 29 сентября 2011 г., шифр 2011-1.4-514-048-066).

На данном этапе особый исследовательский интерес заключается в выявлении взаимосвязи процессов, происходящих в он-лайн и офф-лайн режимах, а также развитии прикладных инструментов для решения этих задач. Эта проблема с октября 2012 г. решается исследовательской группой в рамках проекта «Исследование специфики распространения социально-значимой информации в социальных сетях на примере Живого Журнала с помощью автоматизированных программных средств» (грант Минобрнауки РФ № 2012.-1.2.2-12-000-3003-056).

Литература

- [1] Указ Президента РФ от 09.06.2010 № 690 «Об утверждении Стратегии государственной антинаркотической политики Российской Федерации до 2020 года».
- [2] Постановление Правительства РФ от 20.06.2011 № 485 «Об утверждении Положения о государственной системе мониторинга наркоситуации в Российской Федерации».
- [3] Бершадская Л.А. и др. Социометрические исследования в социальных сетях как инструмент социологии и политологии / Бершадская Л.А., Биккулов А.С., Болгова Е.В., Чугунов А.В., Якушев А.В. // Современные проблемы науки и образования. 2012. № 4.
- [4] Дубровский Р., Теплицкий В. Организация мониторинга наркоситуации // Наркология. 2009. №10. С. 43-49.
- [5] Воронцов К.В. Алгоритмы кластеризации и многомерного шкалирования. Курс лекций. МГУ, 2007.

- [6] Иванов С. В. Web-ориентированный производственно-исследовательский центр "Социодинамика" / Иванов С. В., Болгова Е. В., Каширин В. В., Якушев А. В., Чугунов А. В., Бухановский А. В. // Приборостроение. 2011. №10. С. 65-72.
- [7] Пятницкая И.Н. Общая и частная наркология: руководство для врачей. М.: ОАО "Издательство "Медицина". 2008. 640 с.
- [8] Сербенюк С.Н. Картография и геоинформатика – их взаимодействие. М., 1990. 159 с.
- [9] Сингапур и Петербург. Во второй день работы V Инновационного форума глава Комитета по информатизации и связи администрации Санкт-Петербурга Иван Громов представил долгосрочную целевую программу «Наш город - Санкт-Петербург». URL: http://forum.spbinno.ru/i2011/press_center/news/e2424/ (дата обращения: 27.09.2012).
- [10] Сочевко Д.В. Психодинамика. М.:МПЦИ. 2007. 431 с.
- [11] Татаркин А. И., Куклин А.А. Комплексная методика диагностики качества жизни в регионе. Екатеринбург: Институт экономики УрО РАН, 2010. 136 с.
- [12] Фридман Л.С., Флеминг Л.Ф., Робертс Д.Х., Хайман С.Е. Наркология: Пер. с англ. 2-е изд. испр. М.; СПб.; "Издательство БИНОМ" – "Невский диалект", 2000. 320 с.

Approaches to analysis of social networks and messages from people in geographic information systems for monitoring the drug situation in the region

P. Mityagin, S. Bolbin, S. Mityagin

In the present article the review of approaches to the joint analysis of information on the social phenomena that are hidden from direct observation by the example of drug addiction. As the availability of information resources for the analysis of the population is considering communications on the facts of drug addiction in the area of residence and research findings of social networks for drug consumption.