

А.А. АЗАРОВ, Е.В. БРОДОВСКАЯ, А.Ю. БУБНОВ, А.А. ГЛАЗКОВ,
Р.Я. ГИБАДУЛИН, О.В. ДМИТРИЕВА, А.М. ДОЛГОРУКОВ,
Р.И. ЖДАНОВ, В.Ф. МУСИНА, В.Д. НЕЧАЕВ, А.С. ОГНЕВ,
А.В. СИРОТКИН, А.В. СУВОРОВА, В.В. ТИТОВ, А.Л. ТУЛУПЬЕВ,
Т.В. ТУЛУПЬЕВА, А.А. ФИЛЬЧЕНКОВ, Р.М. ЮСУПОВ

**PREDICTOR MINING: ПРИМЕНЕНИЕ МЕТОДОВ
ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ
В ЗАДАЧАХ СОЦИАЛЬНОГО КОМПЬЮТИНГА**

Азаров А.А., Бродовская Е.В., Бубнов А.Ю., Глазков А.А., Гибадулин Р.Я., Дмитриева О.В., Долгоруков А.М., Жданов Р.И., Мусина В.Ф., Нечаев В.Д., Огнев А.С., Сироткин А.В., Суворова А.В., Титов В.В., Тулупьев А.Л., Тулупьева Т.В., Фильченков А.А., Юсупов Р.М. Predictor Mining: применение методов интеллектуального анализа данных в задачах социального компьютеринга.

Аннотация. В работе предложено новое направление исследований — интеллектуальный поиск маркеров (Predictor Mining), возникшее на пересечении интеллектуального анализа данных (Data Mining) и социального компьютеринга (Social Computing). Сущность предлагаемого подхода состоит в автоматизированном формировании банка маркеров, характеризующих исследуемую предметную область, а затем интерпретировании полученных показателей с помощью методов социального компьютеринга. В качестве примеров областей исследований, где может быть применен интеллектуальный поиск маркеров, рассмотрены задачи выявления особенностей формирования протестных настроений пользователей Интернет, вычисления количественных характеристик социально-значимого поведения, а также изучение специфики формирования ценностей и стратегий поведения пользователей крупнейших социальных сетей.

Ключевые слова: интеллектуальный анализ данных, интеллектуальный поиск маркеров, социальный компьютеринг, информатика.

Azarov A.A., Brodovskaya E.V., Bubnov A.Y., Glazkov A.A., Gibadulin R.Y., Dmitrieva O.V., Dolgorukov A.M., Zhdanov R.I., Musina V.F., Nechaev V.D., Ognev A.S., Sirotkin A.V., Suvorova A.V., Titov V.V., Tulupyev A.L., Tulupyeva T.V., Filchenkov A.A., Yusupov R.M. Predictor Mining: Data Mining application to Social Computing.

Abstract. The paper introduces a new line of research — intelligent search of predictors of social events (Predictor Mining); this direction lies in the intersection of Data Mining and Social Computing areas of research. The proposed approach includes an automated composition of a predictors set for the social event and then interpretation of the predictors by methods of Social Computing. Detection of protest tones in the Internet, estimation of parameters of socially significant behavior, and investigation of largest social networks members' values and strategies formation specifics are examples of problems that can be analyzed using the Predictor Mining techniques.

Keywords: data mining, predictor mining, social computing, computer science.

1. Введение. Социальный компьютеринг — междисциплинарная область исследований, включающая в том числе изучение социального поведения и социального контекста средствами вычислительных систем (computational systems) [31, 39], а также разработку и использование информационных технологий, которые имеют важный социаль-

ный или политический контекст [1]. Построенные в результате подобных исследований модели социального поведения позволяют как выявлять его отличительные особенности, обусловленные спецификой наблюдения, так и осуществлять прогнозы проявления такого поведения на основе собранных данных и отслеживать эффективность корректирующих воздействий.

Исследования в области социального компьютеринга в значительной степени основаны на анализе интернет-контента в блогах, социальных сетях и других ресурсах сети Интернет. Повышенное внимание к интернет-ресурсам вызвано совокупностью причин. Во-первых, интернет предоставляет уникальную среду для изучения социального поведения, в которой вычислительные системы и информационные технологии являются как средством извлечения знаний, так и средством создания и воспроизводства социального контекста. Во-вторых, современное общество характеризуется тенденцией переноса коммуникации на информационную платформу, создавая через социальные сети, тематические форумы и интернет-мессенджеры альтернативу реальному общению. Вместе с тем, коммуникация в онлайн-среде (online) сопряжена с такими особенностями, не присущими общению в оффлайн (offline) , как анонимность и независимость от физических ограничений (географических, языковых, временных).

Одной из важнейших задач социального компьютеринга является моделирование, анализ и прогнозирование социального поведения индивида и / или группы [11].

Для того, чтобы выбрать элементы информационного поля (документ, текст, изображение, видеозапись), имеющие отношение к исследуемому поведению, необходимо сформировать некоторый набор маркеров, то есть небольших связных элементов (слов, выражений, изображений и т.д.), идентифицирующих принадлежность объекта к исследуемой предметной области, в частности, к вовлеченности в ту или иную форму социального поведения. С точки зрения ряда социальных наук маркеры являются паттернами. Отметим, что в контексте биоинформатики и геномики термин «маркер» имеет более узкое значение и подразумевает признак, выявляемый методами молекулярной биологии на уровне нуклеотидной последовательности ДНК, для определенного гена или для любого другого участка хромосомы при сравнении различных генотипов или особей [10, 37, 41]. Таким образом, в социальном компьютеринге термин маркер имеет более широкое значение, указанное ранее.

Кроме функции классификации элементов коммуникационного пространства в зависимости от соотношения их с предметной областью, маркеры характеризуют отношение к изучаемому событию или

явлению, выраженное в анализируемом источнике. Существует ряд методов [28, 33, 44], позволяющих проводить анализ контента при известном наборе маркеров. Выбор маркеров во многом определяет качество анализа и полученные результаты — например, в ситуации, когда некоторый аспект поведения не соответствует ни одному из маркеров, то данные об этом аспекте не будут учитываться при анализе и, как следствие, результаты не будут отражать возможные взаимосвязи с ним.

Формирование набора маркеров является важным этапом проведения исследования. Как правило, выбор маркеров осуществляется экспертами соответствующей предметной области [28, 33, 44], что требует выполнения значительной части работы вручную с привлечением к исследованию большого числа дополнительных участников. Кроме того, возможно использование банков маркеров, сформированных в результате ранее проведенных исследований по смежным темам. Однако в данном случае возникают риски: могут быть не учтены маркеры, существенные для новой задачи или включены в анализ маркеры, являющиеся характерными только для первоначальной задачи. Вместе с тем, в интеллектуальном анализе данных уже существует ряд методов, позволяющих выявлять закономерности, неизвестные заранее. Например, ряд методов типа LDA (Latent Dirichlet Allocation — латентное размещение Дирихле) [21], позволяют определять тематически близкие документы и присутствующие в них темы без явного изначального задания таких тем.

На наш взгляд, актуальным является направление интеллектуального поиска маркеров или Predictor Mining. Predictor Mining находится на пересечении интеллектуального анализа данных (Data Mining) и социального компьютеринга (Social Computing). В частности, к этому направлению исследований можно отнести анализ интернет-контента с целью выявления маркеров поведения пользователей, которые впоследствии могут быть использованы в социальном прогнозировании, например, для определения уровня удовлетворенности (товаром или услугой, деятельностью государственных органов и т.д.), направленности политических настроений и степени вовлеченности интернет-аудитории в оффлайновые ситуации и события. Такие результаты могут быть достигнуты благодаря интеллектуальному анализу интернет-контента с последующим переосмыслением и переработкой полученных данных. Существенным требованием является полная или частичная автоматизация данного процесса.

Цель данной статьи — представить Predictor Mining в качестве нового направления интеллектуального анализа данных на пересече-

нии интеллектуального анализа данных и социального компьютеринга, а также наметить основные области его применения.

В рамках Predictor Mining для автоматизированного формирования банка маркеров, характеризующих исследуемую предметную область и используемых для дальнейшего анализа интернет-контента, целесообразно применять методы интеллектуального анализа данных. Интеллектуальный анализ данных — междисциплинарная область, развивающаяся на пересечении прикладной статистики, искусственного интеллекта, теории баз данных и др., представляющая собой совокупность методов обнаружения в данных ранее неизвестных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности [23, 24].

Мы предполагаем, что методы интеллектуального анализа данных позволят автоматически или автоматизированно сформировать набор определенных показателей, отображающих, например, мнение пользователей социальной сети по тем или иным вопросам, а затем интерпретировать полученные показатели с помощью методов социального компьютеринга и в результате, с одной стороны, разработать теорию и объяснительные вероятностные модели динамики социального поведения, а с другой стороны, и готовить управленческие решения, позволяющие влиять на исследуемый объект и обеспечить обратную связь через отслеживание изменения показателей.

Кроме того, автоматизация формирования банка маркеров позволит выявить связи между ними, в частности, идентифицировать зависимости между уже известными, полученными аналитическим (экспертным) путем, и новыми, выявленными автоматически. Обнаруженные связи создадут предпосылки для более глубокого и подробного исследования причинно-следственных отношений предметной области, охватывая прежде неочевидные даже специалистам факты.

2. Понятие социального компьютеринга. Широкое распространение глобальной сети Интернет, в частности технологии Web 2.0 и сопутствующих ей сетевых сервисов, блогов, социальных сетей и прочего, обусловило появление своеобразного коммуникационного пространства [35]. В рамках этого виртуального пространства [32] пользователи создают, собирают, обрабатывают, используют, обмениваются информацией самых разных типов: от простого текста до видеозаписей и аудио-подкастов. Наполнение виртуального пространства называют интернет-контентом или просто контентом [34].

Важным фактором развития цифрового социального пространства послужило появление мобильных устройств, которые способствуют стиранию границ между и виртуальными онлайн- и оффлайн-социальными мирами [22]. В наши дни мобильные устройства являют-

ся не только средством виртуальной социальной коммуникации, но и источником информации о физических перемещениях владельца или иной его активности [22].

Таким образом, виртуальное пространство в наши дни является источником огромного количества данных, основную часть которых составляют данные о социальных взаимодействиях в виртуальной коммуникационной среде [34]. Сфера социального компьютеринга стала составной частью области исследования информационных систем, и сейчас является одним из ведущих направлений для развития последней [34].

Существует значительное число определений социального компьютеринга [42].

Согласно [5], «Social Computing (социальный компьютеринг) — это системная интеграция социальных наук и Computer Science (компьютерные науки), основанная на системном подходе (прямые и обратные взаимодействия между социальными науками и компьютерными информационными Web технологиями), реализующая идею Social Intelligence (социального интеллекта). Social Computing представляет собой вычислительные теории, модели, программные и технические средства информационных и коммуникационных технологий, которые учитывают социальную специфику и предназначены для адекватного изучения социальных явлений, процессов, социальных систем и развития социальных взаимодействий в Интернете».

Социальный компьютеринг базируется на пересечении компьютерных и социальных наук [11]. В частности, идеи социального компьютеринга применяются в экономических [34, 43], социально-политических [3, 4], психологических исследованиях [2, 19]. Технологии социального компьютеринга могут использоваться общественными структурами с целью улучшения качества предоставления услуг, корпорациями для определения и мониторинга различных организационных факторов [29, 34] и т.д.

3. Частные задачи. В этом разделе мы остановимся на нескольких частных задачах, рассматриваемых в области социального компьютеринга, на существующих методах их решения и возникающих при этом проблемах и новых задачах, а также на перспективах применения к ним методов Predictor Mining.

3.1. Исследование протестной активности. Примером возможной оптимизации процедур социально-политического исследования посредством инструментов Predictor Mining (далее РМ) может послужить проект, реализованный в 2012–2013 гг. Институтом перспективных гуманитарных исследований и технологий МГГУ им. М.А. Шолохова «Особенности формирования протестных настроений и стратегий

политического поведения молодых пользователей Интернет» [3]. Исследование было направлено на выявление коэволюционной связи между дискурсом политического протеста, транслируемым в пространстве Интернет лидерами мнений, и образом протестной активности, сформировавшимся у молодых пользователей. Под образом протестной активности мы понимаем суммарное представление о политическом протесте, которое создается лидерами мнений и средствами массовой информации в глобальной сети в результате отражения в дискурсе характеристик, оценок, мнений, касающихся протеста.

Достижение цели исследования потребовало решения нескольких задач: во-первых, выявление содержательных и эмоциональных доминант информации об акции протеста 13 января 2013 г., размещаемой в блогах лидеров/сторонников/противников политического протеста; во-вторых, определение модальности оценки рядовыми пользователями протестной акции/протестного поведения; в-третьих, оценка эффективности применения лидерами протестного движения и СМИ технологий мобилизации/демобилизации протеста через вербальные и невербальные средства; в-четвертых, выделение социальных профилей пользователей, выражающих сформированность установки на личное участие в протестных действиях и др. Основные сложности при реализации поставленных задач заключались в следующем:

- необходимость сравнительного анализа деятельности трех акторов: лидеров протеста, СМИ и пользователей, проявивших интерес к протестному дискурсу;

- необходимость одномоментного анализа когнитивных, аффективных, поведенческих составляющих образа протестной активности, транслируемого в виртуальной среде.

Модель социально-политического исследования базировалась на триангуляции таких методов, как интент-анализ (*intention analysis*), ивент-анализ (*event analysis*), психографический анализ [4].

Эмпирическим объектом интент-анализа (*intention analysis*) являлись материалы аккаунтов и блогов лидеров протеста, посвященные акции протеста «Марш против подлецов» (13 января 2013 г.). Фиксировались: число упоминаний категорий «протест», «протестная акция», «Марш против подлецов»; выражения, в которых отражены интенции противопоставления «МЫ» – «ОНИ»; выражения, в которых отражена рационализация (объяснение, обоснование, оправдание) своей позиции в отношении протеста; оценка потенциальных участников акции (количество, социальный состав, мотивы); сопоставление акции протеста с опытом протеста в других странах или в другие исторические периоды; оценка будущего протестного движения в России и др.

Результаты интент-анализа содержания постов оппозиционных блогеров (Б. Немцов, И. Яшин, Г. Каспаров и др.) позволили сформулировать вывод о том, что образы протеста отчасти формируются в коммуникативной среде (преимущественно Twitter, Facebook, Эхо Москвы), построенной по принципам оппозиционно-мобилизационной модели информирования. Подтверждением этому является доминирование в большинстве случаев интенции «самопрезентации», «критики власти, оппонентов и т.д.» (противопоставление «МЫ» – «ОНИ») и «побуждения» присоединиться к протесту.

Ивент-анализ (event analysis) был направлен на исследование массива сообщений новостных агентств об акции протеста «Марш против подлецов» за период с 13 декабря 2012 г. по 18 января 2013 г. (выборка сплошная, т.е. включает все сообщения по теме за обозначенный период). Хронология обусловлена данными киберметрического анализа интенсивности поисковых запросов по теме протестной акции (в исследовании использовались инструменты GoogleTrends). В процессе реализации ивент-анализа была выявлена количественная динамика упоминаний в СМИ события «Марш против подлецов», а также динамика изменения направленности оценки масс-медиа данной акции протеста.

Результаты ивент-анализа новостных сообщений информационных агентств (Интерфакс, Lenta.ru и др.), подкрепленные киберметрическими исследованиями, свидетельствуют о том, что тиражирование идей и лозунгов протеста, сформулированных его лидерами, осуществляется в соответствии с ударно-фоновой моделью информационной кампании, в которой роль «информационного удара» выполняют материалы блогов, а функцию создания и поддержания фона — СМИ.

Психографический анализ рисунков респондентов (анализировались изображения трех типов: «политический протест»; «политический протест в России», «протестующий россиянин») осуществлялся с учетом следующих критериев: уровень когнитивной сложности образа (многообразие смыслов); соотношение деталей (по масштабу), их расположение по отношению друг к другу и в пространстве (композиция); темпоральные характеристики образов (прошлое, настоящее, будущее); наличие смысловой взаимосвязи между образами «протеста», «протеста в России», «протестующего»; субъектность (наличие взаимосвязи изображения с личным опытом) и др. Выборка носила целевой характер и предполагала отбор респондентов в возрасте от 14 до 33 лет по степени информированности, источнику получения информации и статусу в протестном движении. Ограничение возрастной принадлежности респондентов отражает результаты World Internet Project – 2012,

согласно которым ядро активных пользователей в РФ составляют россияне указанной возрастной группы [38].

Результаты интерпретации рисунков респондентов отчасти подтверждают гипотезу о том, что дифференциация образов протеста в восприятии респондентов отражает не традиционные социально-демографические различия, а, прежде всего, разницу в стратегиях коммуникации (источники получения информации/интенсивность вовлеченности в Интернет-коммуникацию). В частности, когнитивная сложность и негативная эмоциональная выраженность образа протеста выше у респондентов, которые являются пользователями и имеют опыт участия в протестных акциях.

Среди недостатков реализации рассматриваемой модели социально-политического исследования можно выделить следующие: 1. сложность верификации взаимосвязи между дискурсом политического протеста, транслируемым в пространстве Интернет лидерами мнений и образом протестной активности, сформировавшимся у молодых пользователей; 2. сложность фиксации перехода онлайн-активности в оффлайн; 3. целевой характер и недостаточный объем выборочной совокупности связанный с отсутствием инструментов автоматизированной обработки массива данных.

Применение инструментов РМ приведет к выявлению маркеров политического дискурса и построению вероятностных моделей, связывающих маркеры с конкретными характеристиками протестного образа, что позволит не только автоматизировать, а, следовательно, существенно увеличить исходный объем анализируемой информации, ускорить процесс ее обработки, но и обосновать систему связей между деятельностью трех групп акторов, вовлеченных в политический протест (лидеры, пользователи, СМИ), оценить их потенциал влияния на развитие протестного движения, детализировать социальные профили групп респондентов, составляющих ядро, структурный резерв и периферию протеста, прогнозировать масштабы и направленность протестных акций в будущем.

3.2. Вычисление количественных оценок характеристик социально-значимого поведения. Следует отметить, что маркер не обязательно представляет собой слово, словосочетание, рисунок и т.д., маркером события может являться, в частности, изменение характеристик определенного вида поведения. Например, важным показателем является изменение интенсивности рассматриваемого поведения, выражаемое в изменении частоты употребления маркеров этого поведения. При этом нас интересуют в большей степени последние изменения интенсивности, то есть необходимо оценивать интенсивность по последним событиям, причем вычислять такие оценки с требуемой

регулярностью. Таким образом, нужен инструментарий для косвенного оценивания интенсивности по нескольким последним эпизодам поведения или, другими словами, по нескольким последним эпизодам употребления маркера данного поведения.

На данный момент предложен математический аппарат оценивания интенсивности поведения по данным об эпизодах поведения [14, 15, 16, 20], разработана его программная реализация [17, 18]. Данный инструментарий предназначен для косвенных измерений интенсивности по сведениям о нескольких последних эпизодах, что позволяет производить мониторинг интенсивности с необходимой регулярностью. Ранее он применялся при проведении опросов респондентов об эпизодах их поведения, но также может быть использован и при обработке данных, полученных из интернет-контента.

Таким образом, инструментарий, основанный на последних эпизодах, является средством оценивания интенсивности употребления маркеров поведения определенного вида, а изменение интенсивности, в свою очередь, можно рассматривать как сложный маркер некоторого события, влияющего на поведение.

С другой стороны, описанный подход предлагает не только инструмент, но и одну из возможных областей применения Predictor Mining. Использование существующего математического аппарата и его программных реализаций для вычисления оценки интенсивности по данным о последних эпизодах предполагает заранее известный спектр действий, определяющих изучаемое поведение. В задаче оценки риска передачи и / или приобретения ВИЧ-инфекции, для решения которой этот подход и был предложен первоначально, определены эпизоды поведения, при которых возможна передача инфекции: употребление инъекционных наркотиков, незащищенный половой контакт, передача от матери к ребенку при беременности, родах, грудном вскармливании [13]. Применение же инструментария для решения других проблем сталкивается с проблемой выявления тех эпизодов поведения, которые можно считать определяющими для рассматриваемой проблемы. Математический аппарат не изменяется, но требуется определить, последние эпизоды какого именно поведения нужно фиксировать. В таком случае методы Predictor Mining позволят выявить маркеры, определяющие изучаемую проблему, и, как следствие, зафиксировать те эпизоды поведения, анализ которых позволит получить необходимые оценки.

3.3. Исследование ценностных ориентаций. Одно из направлений применения Predictor Mining обусловлено бурным развитием Рунета не только как коммуникативной системы, но и как многоуровневого социокультурного пространства. В рамках этого пространства

протекают сложные процессы трансформации «традиционных» и кристаллизации «новых» ценностей и моделей поведения российских граждан. Исследовательская проблема состоит в не изученности влияния содержательных и функциональных характеристик основных типов веб-технологий, представленных в русскоязычном сегменте Интернета на формирование ценностей и стратегий поведения россиян. Предметом такого исследования выступает специфика формирования ценностей и стратегий поведения пользователей крупнейших социальных сетей («В контакте», «Одноклассники» и Facebook), которые на сегодняшний день являются наиболее динамично развивающимся «горизонтальным» сегментом Рунета¹.

Задачи исследования можно представить следующим образом: выявление ключевых ценностей, которые репрезентуют пользователи социальных сетей; структурирование многообразия ценностей, характерных для пользователей социальных сетей; выявление ключевых моделей социального поведения, формирующихся в сетевом пространстве Рунета; диагностика базовых содержательных компонентов и ключевых тенденций трансформации поведенческих моделей в социально-сетевом пространстве Рунета.

Важной прикладной сферой использования результатов исследования является понимание того, как связаны ценностно-психологические профили пользователей страниц и точки перехода «онлайн-оффлайн» активности.

Первый этап исследования — метод качественного анализа сетевого контента, предполагает работу эксперта-кодера по изучению текстовых, графических и аудиальных элементов, присутствующих в анализируемых сетевых сообществах и профайлах пользователей социальных сетей. Ограничения «ручного» метода — малая выборочная совокупность, которая составляет от 500 до 1000 профайлов. Указанная выборочная совокупность является оптимальной в рамках качественного исследования, но не предполагает количественный анализ данных, не позволяет экстраполировать полученные результаты на генеральную совокупность, и направлена исключительно на диагностику наиболее общих, ключевых тенденций формирования ценност-

¹ При реализации исследования были использованы средства государственной поддержки, выделенные в качестве гранта Институтом общественного проектирования по итогам VI Конкурса «Проблемы развития современного российского общества», проведенного в соответствии с распоряжением Президента Российской Федерации от 03 мая 2012 года № 216-рп.

ных ориентаций и стратегий поведения пользователей (например, [12]).

Для перехода ко второму этапу исследования (выявлению системы взаимосвязанных маркеров ценностных репрезентаций и стратегий поведения пользователей социальных сетей Рунета) необходимо сочетание «ручного» экспертного анализа и данных, полученных в фокус-группах, с автоматическим сбором и обработкой данных. В рамках данного этапа должны быть решены следующие задачи: апробация отобранных в ходе качественного этапа маркеров репрезентаций на всей генеральной совокупности; выявление корреляции между типами ценностей и стратегиями поведения пользователей социальных сетей Рунета; кластерный анализ, позволяющий выделить основные ценностно-поведенческие кластеры, представленные в социальных сетях Рунета; выделение и описание базовых тенденций, характеризующих формирование ценностей и стратегий поведения пользователей Рунета.

Важной прикладной подзадачей является определение взаимосвязи маркеров ценностных репрезентаций с социально-демографическими параметрами, что предполагает структуризацию ценностей и стратегий поведения по критерию востребованности у различных социально-демографических групп (характеризующими их носителей: гендер, возраст (принадлежность к социальному поколению), территориальная самоидентификация (регион, город).

4. Перспективы исследований и разработок. Будущее Predictor Mining видится в качестве институционально оформленной гуманитарной технологии, способной обеспечить:

- прогнозирование и дизайн социальных систем в качестве сложным образом взаимодействующих симбиозов (онлайн и офлайн пространств);
- развитие социологических теорий с новой онтологией и методологической «свободой» от устаревшего разделения на качественные и количественные методы исследований;
- сбор, генерацию и анализ данных на основе сложным образом организованных «наблюдений», а не опросниковых методик.

На этой основе в данное направление деятельности будут включены социологи и социальные психологи. Команды исследователей и разработчиков, состоящие из социологов, математиков и программистов начнут создавать новые практики и содержательные теории, позволяющие понимать и прогнозировать динамику смыслов, событий и порождаемых ими ситуаций. На этой основе мы научимся работать с социо-лингвистическими репрезентациями реальности, которая в данном случае представляет собой весьма динамичный гипертекст, и при-

менять их для анализа «повседневного опыта» на основе которого только и можно создать новые объяснительные теории и модели.

Мы начинаем изучать онлайн-пространство, внимательно наблюдая и экспериментируя, а не опрашивая, на основе эффективной современной интерпретации повседневности интернета с помощью сетевых информационных технологий [7]. Для этого надо модифицировать традиционный метод наблюдения в метод, приспособленный наблюдать деятельность, выраженную через контент.

Мы пройдем последовательно через определенные этапы развития методологии:

1. Извлечение латентных и явных паттернов в качестве означающего для действий и ситуаций с помощью методов Data Mining [25, 30].

2. Проведение исследования методом опроса рефлексивных фокус-групп. Это фокус-группы, которым предъявляется для обсуждения материал, порожденный в сети теми, по отношению к кому эти фокус-группы являются репрезентативными.

3. Экспериментальное нарушение привычного хода вещей в информационном пространстве и последующий анализ «объяснений» причин нарушения коммуникации самими насельниками сети. Для этого, этнометодологические техники будут дополнены специально разработанными для данного случая приемами герменевтики (семиотики) [26, 7].

4. Контент-анализ смыслов, а не только (и не столько) устойчивых структур, порождаемых в семантическом пространстве гипертекста. Выше перечисленные методы (но не только они) создают возможность получить данные, которые позволят осмысленно применять процедуры контент-анализа с использованием программ автоматической обработки контента. А это, в свою очередь, создаст надежные основания для количественного обзора больших массивов данных [30].

5. Построение и проверка на эмпирических данных содержательных теорий, учитывающих прагматический аспект — ценности и жизненные ситуации человека, влияющие на его установки и действия. Перспективным является изучение проактивного типа человека, занимающего специфические позиции и играющего определенные роли в узлах сети [25].

6. Построение динамической вероятностной модели, правдоподобно объясняющей связи между означающим и означаемым с учетом жизненного цикла конкретной социальной общности, которая наблюдается, и прагматического контекста в который вовлечены люди [9].

Данная перспектива поможет нам уйти от «уже ставшего и явно-го» к «латентному и нарождающемуся» (рис. 1).

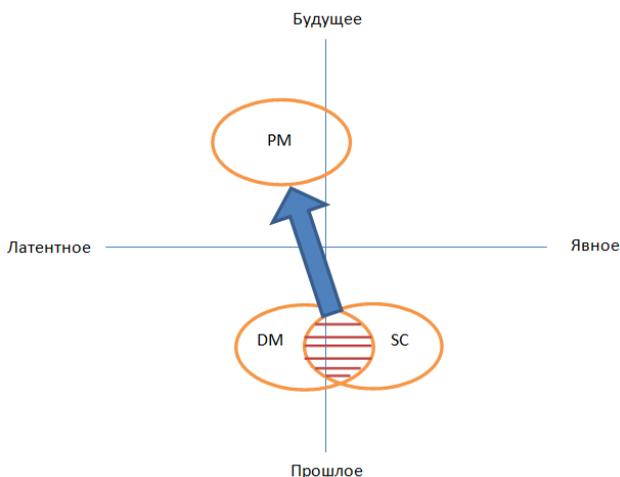


Рис. 1. От «social computing» и «data mining» к «predictor mining».

5. Дискуссия. Сотни тысяч людей выходят каждый день в интернет и совершают рутинные действия, порождая новые коммуникативные практики (часто не осознаваемые). Виртуальная культура только становится; институционально только оформляется, нет явных устойчивых структур, опираясь на которые можно было бы применять опросниковые методики для предсказаний и объяснений. Основная форма «жизни» — спонтанный гипертекст, представляющий собой подвижную и плохо предсказуемую семиосферу без выраженной иерархии.

Установки и поведение людей зависят от «оснований», которые от них самих (за редким исключением) скрыты и поэтому имеют слабую связь с мнениями, которые они высказывают о своих установках и возможном поведении.

Понимание повседневности человеком сильно зависит от контекста, повседневного диалога, который ведет человек, а также от языка, который постоянно «переодевает его мысли».

Можно обоснованно предположить, что все три фактора по-разному проявляются в ситуации онлайн и оффлайн.

Интернет-пространство сегодня представляет собой ризомаморфное образование [6, 7] и изучение его в качестве иерархически устроенной системы (наподобие оффлайн-пространства) является занятием малоэффективным.

Любая социальная сеть — это совокупность узлов, которые, соединяясь к каналам коммуникации, удерживают смыслы социальной

жизни, порождая действия, события и ситуации. Сети динамичны и никогда не остаются теми же самыми с течением времени.

Наряду с институционально оформленными узлами сети и потоками, по привычным каналам информации, с которой «всем понятно», что надо делать, возникают новые неожиданные семиотические потоки. Эти потоки порождают движение идей, денег, людей, упаковочных материалов и т.п. [7, 8].

Двигаясь вне границ каналов и узлов, эти семиотические потоки смещают старые и создают новые смыслы», создают новые идентичности, обесценивают доступ к некоторым узлам сети, создают новые социальные неравенства, порождают и закрывают возможности, вызывают новые желания.

При этом, в оффлайн социальных сетях новые семиотические потоки возникают крайне редко, так как институциональные узлы «строго следят» за тем, чтобы не возникло ничего нового (кроме того, что «нормально и допустимо»). Тогда как семиотические потоки в интернет текучи, непредсказуемы, случаются часто и часто недолговечны, они «лишены» пунктов отправления и назначения, их плотность растёт и падает «мгновенно», «размывают» власть существующих узлов сети, делают идентичности мобильными и непрочными, позволяя экспериментировать с ролями в деле реконструкции себя [36].

Нарождающиеся семиотические потоки «прорывают» привычные русла каналов, непредсказуемо создавая свои незаметные русла и «селевые потоки»; подрывают монополии узлов на создание и поддержание смыслов, окрепнув, закладывают основы для новых будущих узлов и каналов.

Сообщества, функционирующие в социальных сетях, создают социальные поля, существенно и непредсказуемо отличающиеся от социальных общностей реально существующих и изучаемых оффлайн-социологией [8]. Отличие, как минимум, в следующем:

- мобильными становятся идентичности насельников киберполей.
- неустойчивы и плохо определимы границы таких сообществ.

Считается, что в социальных сетях объединены сотни тысяч участников, но такие сети не поддаются «изучению» поскольку «не являются подлинными» [8, 27]. Разумеется, некоторые усреднения и обобщения на таких больших массивах полезны, если они осмысленны.

- отделение визуальной репрезентации от всех остальных, имеющее место в оффлайн-пространстве, превратило «покорение мира как картины» (Хайдеггер) в основной процесс. Фотография, экран, витрина делают все значимым, не зависимо от того, что они отражают. Но в интернет-пространстве властвуют не институциональные узлы и пото-

ки, определяющие, что надо смотреть, а непредсказуемые семиотические потоки [8, 25].

Таким образом, мы можем утверждать, что онлайн-пространство со своего зарождения и до настоящего времени является ризомаморфным социальным образованием без сложившихся структур (социальных форм), без институционально оформленной иерархии, с сильными и частыми новыми семиотическими потоками, с новой, не известной нам, топикой.

Таким образом, можно сделать следующие выводы:

1) такие особенности делают пока малоэффективным применение количественных методов на больших массивах данных;

2) методы имеют смысл, если объекты обследования обладают некоторыми «качествами», например: имеют устойчивые границы (физические, временные, смысловые); обладают устоявшимися структурами внутри себя; упорядочены иерархически и т.п.;

3) время эффективного использования количественных методов на больших выборках придет позже. Тогда, когда будут созданы и проверены содержательные социологические теории, объясняющие, во-первых, как связаны системы и подсистемы маркеров, которые мы способны фиксировать в динамичном гипертексте, с событиями, действиями и вытекающими из этих событий и действий социальными ситуациями; во-вторых, как зарождаются новые смыслы (семиотические потоки), чтобы по «слабым сигналам» обнаруживать новые паттерны и маркировать их; в-третьих, какие инвариантные (устойчивые) связи и отношения могут быть экстраполированы на большие массивы данных.

4) сегодня социология, удовлетворительно объясняющая лишь поведение иерархических социальных общностей (поселения, семьи, общественного движения, этноса, организации (оффлайн-пространство), малоэффективна в изучении интернет.

5) онлайн-пространство может и должно изучаться на основе эффективной современной интерпретации, этнометодологических принципов и методов изучения и объяснения социальной жизни. Поскольку это, с одной стороны, все еще «туземное» интернет-сообщество, а с другой, оно должно изучаться на основе современных сетевых информационных технологий, назовем этот подход «e-Этнометодология» [8] и укажем некоторые его перспективные черты.

Этнометодология — буквально «методы людей» [26]. Она изучает повседневные знания людей, которые они называют «здоровым смыслом». Основные установки этнометодологов можно сформулировать следующим образом: социальная реальность — это эндогенно порождаемое практическое дело, осуществляемое «здесь и сейчас» от кото-

рого нельзя уклониться или отсрочить его. Это, в основном, действия относительно рутинные и не обдуманные.

Важно подчеркнуть, что повседневные разговоры составляют естественный фундамент для других форм межличностных отношений, выливающих в действия. Как можно обнаружить сложные взаимосвязи? Это можно сделать, вызывая к жизни «объяснения». Работая с ними в рефлексивной позиции люди, тем самым, понимают и иногда изменяют свою жизнь, делая избыточными «рекомендации» и внешние предписания. А социологи могут выдвигать правдоподобные гипотезы относительно установок и действий. Ключом для обнаружения и понимания «структур», «культуры» и «ценностей и действий человека» должны выступать оценки самих участников жизненных ситуаций. Задача этнометодолога анализировать как сами «объяснения», так и реакцию других участников ситуации на эти объяснения.

Социологи не должны навязывать этой реальности свои теоретические объяснения. Вместо этого, спровоцировав рефлексивное самоопределение участников «жизненных ситуаций» надо помочь им обнаружить «практически отработанные действия участников» (Шютц)

Не будучи осознаваемыми, именно эти действия порождают «структуры» социального взаимодействия или отдельной личности, «программируя» неожиданные исходы событий.

В связи с этим, социолог должен меньше «спрашивать», но, больше анализировать «разговоры» — естественную жизнь языка повседневности.

Чтобы прогнозировать взаимосвязи «оффлайн- и онлайн-полей» надо разработать и постоянно изучать «словари мотивов» членов онлайн-общностей.

Эффективность исследований придет в том случае, если будут систематически выстраиваться «ситуации понимания» гипертекстовой деятельности в интернет и на этой основе создаваться типологии активностей в Интернет, чтобы знать, что наблюдать.

6. Заключение. В работе предложено новое направление исследований — Predictor Mining, возникшее на пересечении интеллектуального анализа данных (Data Mining) и социального компьютеринга (Social Computing). Благодаря применению методов автоматического извлечения закономерностей из больших объемов данных для анализа и прогнозирования социальных процессов и их динамики, Predictor Mining создаст основу для развития нового теоретического и прикладного инструментария исследования социального поведения. Ядром такого инструментария станут автоматически и автоматизировано выявленные маркеры поведения.

Заметим, что в литературе [40] использовалось лишь словосочетание Predictive Data Mining, которое, хотя и обозначает некую предсказательность (а точнее, методы построения прогнозов), но не выражает в достаточной мере существо излагаемого нами подхода.

Мы видим миссию Predictor Mining как институционально оформленной гуманитарной технологии в обеспечении автоматизированного построения множества систем маркеров, которые сделают преимущественно рутинными следующие направления деятельности: прогнозирование и дизайн социальных систем в качестве сложным образом взаимодействующих симбиозов; развитие социологических теорий с новой онтологией и методологической «свободой» от устаревшего разделения на качественные и количественные методы исследований; сбор, генерацию и анализ данных на основе сложным образом организованных «наблюдений».

Литература

1. 1-я Международная конференция «Социальный компьютинг: Основы, Технологии развития, Социально-гуманитарные эффекты» / Конференции.RU. URL: <http://konferencii.ru/info/88360> (дата обращения: 10.03.2013)
2. *Азаров А.А., Ванюшичева О.Ю.* Синтез агрегированных оценок вероятности реакций пользователя в ответ на социо-инженерное атакующее действие. // VI Международная научно-техническая конференция молодых специалистов, аспирантов и студентов «Математическое и компьютерное моделирование естественнонаучных и социальных проблем» (МК-70-912) (Пенза, 21-24 мая 2012 г.). П: Приволжский Дом Знаний. 2012. С. 117-120.
3. *Бродовская Е.В.* «Марш против подлецов»: образ протеста в пространстве интернет // Модернизационные процессы в обществе: проблемы теории и практики. Материалы Международной научно-практической конференции. Тверь: «СФК-офис», 2013. С. 33–44
4. *Бродовская Е.В.* Формирование образа протестного поведения в Рунете: опыт прикладного исследования // Дни науки философского факультета – 2013: Международная научная конференция (Киев, 2013). Киевский национальный университет им. Т. Шевченко, 2013. С. 16–18
5. *Давыдов А.А.* Системная социология: social computing. Электронный ресурс. URL: http://www.isras.ru/index.php?page_id=1016 (доступ 12.03.2013)
6. *Делез Ж., Гваттари Ф.* Тысяча плато: Капитализм и шизофрения. У-Фактория. Астель. 2010
7. *Долгоруков А.М.* е-Этнометодология: от «social computing» к «computable sociology» // Social Computing. Scientific journal. Выпуск 2, 2013
8. *Долгоруков А.М.* е-Этнометодология – новая социологическая перспектива? // Материалы Первой Международной научно – практической конференции «Социальный компьютинг: основы, технологии развития, социально-гуманитарные эффекты» (ISC-12): 11 статей и тезисов. М.: Изд-во МГГУ им. М.А. Шолохова, 2012. Стр. 115–120
9. *Долгоруков А.М.* Социальный дизайн: создание жизнеспособных систем. М.: РИЦ МГГУ им. М.А. Шолохова, 2012
10. *Календарь Р.Н., Глазко В.И.* Типы молекулярно-генетических маркеров и их применение // Физиология и биохимия культурных растений. 2002. № 34 (4). С. 141-156.

11. Материалы Первой Международной научно-практической конференции «Социальный компьютеринг: основы, технологии развития, социально-гуманитарные эффекты» (PISC-12): 11 статей и тезисов. М.: МГГУ, 2012. 248 с.
12. Молодежь Москвы: ценностные приоритеты, стратегии поведения и перспективы развития: 11 информационно-аналитических материалов/ под редакцией доктора политических наук В.Д. Нецаева. Москва, 2012
13. Риск передачи ВИЧ // STOP СПИД. URL: http://www.stopspid.ru/info/risk_transmission/ (дата обращения: 03.03.2013).
14. Степанов Д.В., Мусина В.Ф., Суворова А.В., Тулупьев А.Л., Сироткин А.В., Тулупьева Т.В. Функция правдоподобия с гетерогенными аргументами в идентификации пуассоновской модели рискованного поведения в случае информационного дефицита // Труды СПИИРАН. 2012. Вып. 23. С. 157–184.
15. Суворова А.В., Тулупьев А.Л., Пащенко А.Е., Тулупьева Т.В., Красносельских Т.В. Анализ гранулярных данных и знаний в задачах исследования социально значимых видов поведения // Компьютерные инструменты в образовании. №4. 2010. С. 30–38
16. Суворова А.В., Тулупьева Т.В., Тулупьев А.Л., Сироткин А.В., Пащенко А.Е. Вероятностные графические модели социально-значимого поведения индивида, учитывающие неполноту информации // Труды СПИИРАН. 2012. Вып. 22. С. 101–112.
17. Тулупьев А.Л., Суворова А.В., Пащенко А.Е. Программа для расчёта нечётких оценок интенсивности угрозообразующего поведения и риска, с ним связанного, Fuzzy Risk-&Rate Calculator, Version 2(F.R.-&R.C. v.2) // Роспатент. Свид. о гос. рег. прогн. для ЭВМ № 2010614267 от 30.06.2010.
18. Тулупьев А.Л., Суворова А.В., Пащенко А.Е. Программа для учёта неточных сведений об угрозообразующем поведении Fuzzy Data Register for Risky Behavior, Version 1 (F.D.R.R.V. v. 1) // Роспатент. Свид. о гос. рег. прогн. для ЭВМ № 2010613161 от 14.05.2010.
19. Тулупьев А.Л., Тулупьева Т.В., Азаров А.А., Григорьева О.Ю. Психологические особенности персонала, предрасполагающие к успешной реализации социоинженерных атак // Научные труды Северо-Западного института управления РАН-ХиГС. 2012. Т. 3, вып. 3(7). С. 256–266.
20. Тулупьева Т.В., Пащенко А.Е., Тулупьев А.Л., Красносельских Т.В., Казакова О.С. Модели ВИЧ-социально-значимого поведения в контексте психологической защиты и других адаптивных стилей. СПб.: Наука, 2008. 140 с.
21. Blei D.M., Ng A.Y., Jordan M.I. Latent Dirichlet Allocation // Journal of machine Learning Research. 2003. N.3, P. 993–1022.
22. Capra L., Quercia D. Middleware for social computing: a roadmap // Journal of Internet Services and Applications. 2012. Vol. 3, No 1. Pp. 117–125.
23. Data Mining от Oracle: настоящее и будущее. URL: http://citforum.ru/database/oracle/data_mining_solutions/ (дата обращения: 10.03.2013).
24. Han J., Kamber M., Pei J. Data mining: concepts and techniques. Morgan Kaufmann, 2006. 743 p.
25. Eco U. A Theory of semiotics. Indiana University Press, 1979
26. Francis D., Hester S. An Invitation to Ethnomethodology. SAGE Publications, 2004
27. Kadushin Ch. Understanding Social Networks: theories, concepts and findings. Oxford University Press, 2012.
28. Krippendorff K. Content analysis: An introduction to its methodology. SAGE Publications, Incorporated, 2004. 413 p.
29. Kwai Fun IP R., Wagner C. Weblogging: A study of social computing and its impact on organizations // Decision Support Systems. 2008. Vol. 45, No 2. P. 242–250.
30. Leetaru K. Data Mining Methods for the Content Analyst: An Introduction to the Computational Analysis of Content (Routledge Communication Series)
31. Liu H., Salerno J., Young M. Social Computing and Behavioral Modeling. Springer, 2009. 280 p.

32. *Messinger P. R., Stroulia E., Lyons, K., Bone M., Niu R. H., Smirnov K., Perelgut S.* Virtual worlds—past, present, and future: New directions in social computing // *Decision Support Systems*. 2009. 47(3). P. 204–228.
33. *Neuendorf K. A.* The content analysis guidebook. SAGE Publications, Incorporated, 2002. 301 p.
34. *Parameswaran M., Whinston A. B.* Research issues in social computing // *Journal of the Association for Information Systems*. 2007. Vol. 8, No 6. P. 336–350.
35. *Parameswaran M., Whinston A. B.* Social Computing: an Overview // *Communications of the Association for Information Systems*. 2007. Vol.19. 762–780.
36. *Plant S.* Zeros and Ones. L.: Fourth Estate, 1997
37. *Southern E.M.* Detection of specific sequences among DNA fragments separated by gel electrophoresis // *Journal of Molecular Biology*. 1974. Vol. 98 (3). P. 503-517.
38. Typical features of Internet activity of Russian people // *European Applied Studies: modern approaches in scientific researches*, 2nd International scientific conference. (February 18–19, 2013, Stuttgart, Germany). Volume 1. Stuttgart: ORT Publishing, 2013. P. 189-191.
39. *Tuzhilin A.* Social Recommender Systems: Combining Recommender Systems and Social Computing into an Integrated Platform // *Материалы Первой Международной научно-практической конференции «Социальный компьютеринг: основы, технологии развития, социально-гуманитарные эффекты» (ИПС-12): 11 статей и тезисов*. Москва: Изд-во МГТУ им. М.А. Шолохова, 2012. С. 14
40. *Velickov S., Solomatine D.* Predictive Data Mining: Practical Examples // *Artificial Intelligence in Civil Engineering*. Proc. 2nd Joint Workshop, March 2000, Cottbus, Germany.
41. *Villanueva E., Maciel C.D.* Modeling associations between genetic markers using Bayesian networks // *Bioinformatics*. 2010. Vol. 26, Issue 18. P. i632–i637.
42. *Wang F. Y., Carley K. M., Zeng D., Mao W.* Social computing: From social informatics to social intelligence // *Intelligent Systems, IEEE*. 2007. 22(2). P. 79–83.
43. *Wash R., Rader E.* Public bookmarks and private benefits: An analysis of incentives in social computing // *Proceedings of the American Society for Information Science and Technology*. 2007. Vol. 44, No 1. P. 1–13.
44. *Weber R.P.* Basic content analysis. Sage Publications, Incorporated, 1990. Issue 49. 96 p.
45. *Witten I. H., Frank E.* Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2005. 525 p.

Азаров Артур Александрович — м.н.с., лаборатория теоретических и междисциплинарных проблем информатики, СПИИРАН. Область научных интересов: защита информации, анализа защищенности информационных систем. Число научных публикаций — 42. artur-azarov@yandex.ru; СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450.

Azarov Artur Alexandrovich — junior researcher, Laboratory of Theoretical and Interdisciplinary Computer Science, SPIIRAS. Research interests: information protection, information system's protection analysis. The number of publications — 42. artur-azarov@yandex.ru; SPIIRAS, 39, 14-thLine V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450.

Бродовская Елена Викторовна — д. полит. н., проф; заведующая кафедрой политологии, заместитель директора Института политики, права и социального развития МГТУ им. М.А. Шолохова. Область научных интересов: теория политики, политическая аксиология, трансформационные процессы на посткоммунистическом пространстве, политическая система России, геополитический статус России, прикладная политология, политическая психология, психология рекламы, политический компьютеринг. Число публика-

ций — 126. brodovskaya@inbox.ru. 109240, Москва, ул. Верхняя Радищевская, д. 16-18, тел. (495)647-4477.

Brodovskaya Elena Victorovna — PhD in Politics, Dr. of Sc. Professor; Head of Institute for Political Sciences Faculty, Deputy Director of Politics, Law and Social Development SMSUH. Research area: theory of Politics, political axiology, transformational processes in post-communist area, Russia's political system, Russia's geopolitical status, applied Political Sciences, psychology of Politics, advertising psychology, political computing. Number of publications — 126. brodovskaya@inbox.ru; 16-18, Verkhnyaya Radishchevskaya street, Moscow 109240; office phone (495)647-4477.

Бубнов Александр Юрьевич — к. филос. н., доц.; заместитель директора Института гуманитарных технологий в сфере социального компьютеринга МГГУ им. М.А. Шолохова. Область научных интересов: политическая философия, сравнительное религиоведение, история русской философии, политические идеологии, политическая лингвистика, интернет-технологии в политике, социальный компьютеринг, теория геополитики. Число публикаций — 55. alexandr-bubnov@mail.ru. 109240, Москва, ул. Верхняя Радищевская, д. 16-18, тел. (495)647-4477.

Bubnov Alexander Yuryevich — PhD in Philosophy, Deputy Director of Institute for High-Hume Technologies in Social Computing, SMSUH. Research area: Political Philosophy, comparative Religious Studies, history of Russian Philosophy, political ideologies, political linguistics, Internet technologies in Politics, Social Computing, Theory of Geopolitology. Number of publications — 55. alexandr-bubnov@mail.ru; 16-18, Verkhnyaya Radishchevskaya street, Moscow 109240; office phone (495)647-4477.

Гибадулин Рустем Яхьевич — директор Института перспективных гуманитарных исследований и технологий МГГУ им. М.А. Шолохова. Область научных интересов: компьютерная лингвистика, социальный компьютеринг. Число публикаций — 5. rias@mgu-sh.ru, info@bookseller.ru. 109240, Москва, ул. Верхняя Радищевская, д. 16-18, тел. (495) 647-44-77 (доб. 12-51).

Gibadulin Rustem Yakhyevich — Director of Russian Institute for Advanced Study in Humanities and Technology, SMSUH. Research area: computer linguistics, Social Computing. Number of publications — 5. rias@mgu-sh.ru, info@bookseller.ru. 16-18, Verkhnyaya Radishchevskaya street, Moscow 109240; office phone (495)647-4477 (12-51).

Глазков Андрей Александрович — к.э.н.; проректор по науке, инновациям и стратегическому развитию МГГУ им. М.А. Шолохова. Область научных интересов: показатели эффективности деятельности организации, инновационные технологии в образовании, социальный компьютеринг. Число публикаций — 30. a.a.glazkov@gmail.com. 109240, Москва, ул. Верхняя Радищевская, д. 16-18, тел. (495) 683-31-07.

Glazkov Andrey Alexandrovich — PhD in Economics, pro-rector in Sciences, Innovations and Strategical Development SMSUH. Research area: indicators of performance of organizations, innovative technologies in education, Social Computing. Number of publications 30. a.a.glazkov@gmail.com. 16-18, Verkhnyaya Radishchevskaya street, Moscow 109240; office phone (495) 683-31-07.

Дмитриева Оксана Владимировна — директор Института гуманитарных технологий в сфере социального компьютеринга МГГУ им. М.А. Шолохова. Область научных интересов: социальное моделирование и прогнозирование, социальная психология, социальный

компьютинг. Число публикаций — 2. dmitrieva.oksana@gmail.com. 109240, Москва, ул. Верхняя Радищевская, д. 16–18, тел. (495) 647-44-77.

Dmitrieva Oksana Vladimirovna — Director of Institute for High-Hume Technologies in Social Computing, SMSUH. Research area: social modeling and prognostics, social psychology, Social Computing. Number of publications — 2. dmitrieva.oksana@gmail.com. 16-18, Verkhnyaya Radishchevskaya street, Moscow 109240; office phone (495)647-4477.

Долгоруков Александр Михайлович — к.филос.н., проф.; директор научно-образовательного центра «Социология инноваций» МГУ им. М.А. Шолохова. Область научных интересов: социология, социальный дизайн; е-этнометодология, управленческое консультирование. Число публикаций — 50. dolgorukov.am@gmail.com, adolg@mail.ru 109240, Москва, ул. Верхняя Радищевская, д. 16–18, тел. (495)683-31-07.

Dolgorukov Alexander Mikhailovich — PhD in Philosophy, Head of scientific-educational center “Sociology of Innovations”, SMSUH. Research area: sociology, sociological design, etnomethodology, administrative consulting. Number of publications — 50. dolgorukov.am@gmail.com, adolg@mail.ru; 16-18, Verkhnyaya Radishchevskaya street, Moscow 109240; office phone (495)683-31-07.

Жданов Ренат Ибрагимович — д.х.н., проф.; главный научный сотрудник Института перспективных гуманитарных исследований и технологий МГУ им. М.А. Шолохова, главный научный сотрудник Химического Института им. А.М. Бутлерова, профессор кафедры фундаментальной и клинической фармакологии Биолого-почвенного факультета КФУ. Область научных интересов: биологически активные парамагнетики, невирусные системы для направленного генного переноса в целях генотерапии; специфичность липид-нуклеиновых взаимодействий и комплексов. Число публикаций — 390. zrenad@gmail.com, rias@mnggu-sh.ru. 109240, Москва, ул. Верхняя Радищевская, д. 16–18, тел. 8 (495) 683-31-07.

Zhdanov Renad Ibragimovich — PhD in Chemistry, Dr. of Sc., Prof. Principal researcher of Russian Institute for Advanced Study in Humanities and Technology, SMSUH; principal researcher of Butlerov Institute for Chemistry, professor of Fundamental and Medical Pharmacology, Faculty of Biology and Social Sciences, KFU. Research area: biologically active paramagnetics, non-viral systems for directed transmission of genes for genotherapy, specificity of lipid-nucleic interactions and complexes. Number of publications — 390. zrenad@gmail.com, rias@mnggu-sh.ru, 16-18, Verkhnyaya Radishchevskaya street, Moscow 109240; office phone (495) 683-31-07.

Мусина Валерия Фуатовна — младший научный сотрудник лаборатории теоретических и междисциплинарных проблем информатики СПИИРАН, студент магистратуры экономического факультета СПбГУ. Область научных интересов: вероятностное и статистическое моделирование, моделирование риска, биостатистика, вероятностные графические модели. Число научных публикаций — 15. valery.musina@gmail.com, www.tulupyev.spb.ru; СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450.

Musina Valeriya Fuatovna — junior research fellow Theoretical and Interdisciplinary Computer Science Laboratory, SPIIRAS, graduate student of Faculty of Economics at Saint Petersburg State University. Research area: probabilistic and statistic modelling, risk modelling, biostatistics, probabilistic graphical models. Number of publications — 15. valery.musina@gmail.com www.tulupyev.spb.ru; SPIIRAS, 14-th line V.O., 39, St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450.

Нечаев Владимир Дмитриевич — д.полит.н., проф.; ректор МГГУ им. М.А. Шолохова, научный руководитель Центра киберполитики и прикладных политических исследований, научный руководитель World Internet Project – 2012 в РФ, член Совета Российской Ассоциации Политической Науки (РАПН) по политологическому (профессиональному) образованию. Область научных интересов: теория политического мифа; региональные политические системы в постсоветской России; теория эффективной децентрализации; местное самоуправление в РФ; социальный компьютеринг. Число публикаций — 130. vdnechaev@rambler.ru; 109240, Москва, ул. Верхняя Радищевская, д. 16–18, тел. 8 (495) 647 44 77.

Nechaev Vladimir Dmitrievich — PhD in Politics, Dr. of Sc., Prof. Principal of SMSUH, Research Advisor of Center for Cyberpolitics and Applied Political Studies, Scientific Advisor of World Internet Project – 2012 in Russia, member of Russian Association for Political Science in political education. Research area: theory of political myth, local political systems in post-Soviet Russia, theory of effective decentralization, local authorities in Russia, Social Computing. Number of publications— 130. vdnechaev@rambler.ru. 16-18, Verkhnyaya Radishchevskaya street, Moscow 109240; office phone 8 (495) 647 44 77.

Огнев Александр Сергеевич — д.пс.н., проф.; проректор по науке, послевузовскому образованию и общественным связям МГГУ им. М.А. Шолохова. Область научных интересов: психология самоорганизации личности и социальной группы; психотехнологии персонального успеха, благополучия и процветания; психотехнологии организации конструктивного диалога и противодействие манипулятивному влиянию; инструментальная психодиагностика субъектного потенциала личности; когнитивные аспекты достижения мудрости как высшего уровня нравственно-интеллектуального уровня развития человека. Число публикаций — 119. altognev@mail.ru; 109240, Москва, ул. Верхняя Радищевская, д. 16–18, тел. 8 (495) 647 44 77.

Ognev Alexander Sergeevich — PhD in Psychology, Dr. of Sc., Prof.; Pro-rector in Science, Postgraduate education, and public relations SMSUH. Research area: psychology of group and personal self-organization, psychotechnologies of personal success, well-being and prosperity, psychotechnologies of constructive dialog management and resistance to manipulative influence, instrumental psychodiagnostics of subjective person's potential, cognitive aspects of achieving wisdom as the highest level of moral-intelligent person's development. Number of publications— 119. altognev@mail.ru; 16-18, Verkhnyaya Radishchevskaya street, Moscow 109240; office phone 8 (495) 647 44 77.

Сироткин Александр Владимирович — к.ф.-м.н., научный сотрудник лаборатории теоретических и междисциплинарных проблем информатики, СПИИРАН. Область научных интересов: алгебраические байесовские сети: вычислительные аспекты логико-вероятностного вывода в условиях неопределенности. Число научных публикаций — 100. avs@iias.spb.su; СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450.

Sirotkin Alexander Vladimirovich — PhD, researcher, Theoretical and Interdisciplinary Computer Science Laboratory, SPIIRAS. Research interests: algebraic Bayesian networks, algorithms of probabilistic-logic inference under uncertainty. The number of publications — 100. avs@iias.spb.su, www.tulupyev.spb.ru; SPIIRAS, 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450.

Суворова Алена Владимировна — младший научный сотрудник лаборатории теоретических и междисциплинарных проблем информатики СПИИРАН, аспирант математико-

механического факультета Санкт-Петербургского государственного университета (СПбГУ). Область научных интересов: математическая статистика, теория вероятности, применение методов математического моделирования в эпидемиологии. Число научных публикаций — 53. suvalv@gmail.com; СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450. Научный руководитель — А.Л. Тулупьев.

Suvorova Alena Vladimirovna — junior researcher, Laboratory of Theoretical and Interdisciplinary Computer Science, SPIIRAS, PhD student, Faculty of Mathematics and Mechanics of St. Petersburg State University (SPbSU). Research interests: mathematical statistics, probability theory, application of mathematical modeling in epidemiology. The number of publications — 55. suvalv@gmail.com; SPIIRAS, 39, 14th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450. Scientific advisor — A.L. Tulupiev.

Титов Виктор Валерьевич — к.полит.н., доцент кафедры политологии МГГУ им. М.А. Шолохова. Область научных интересов: политическая культура, политическая идентичность и социокультурные трансформации в постсоветской России, политическая психология, политическая культура Рунета. Число публикаций — 30. titov-msu@mail.ru. 109240, Москва, ул. Верхняя Радищевская, д. 16–18, тел. 8 (495) 647 44 77.

Titov Victor Valeryevich — PhD in Politics, Ass. Prof.; Of Department for Political Studies, SMSUH. Research area: political culture, political identity and sociocultural transformations in post-Soviet Russia, political psychology, political culture of Runet. Number of publications — 30. titov-msu@mail.ru. 16-18, Verkhnyaya Radishchevskaya street, Moscow 109240; office phone 8 (495) 647 44 77.

Тулупьев Александр Львович — д.ф.-м.н., профессор; заведующий лабораторией теоретических и междисциплинарных проблем информатики СПИИРАН, доцент кафедры информатики математико-механического факультета С.-Петербургского государственного университета (СПбГУ). Область научных интересов: представление и обработка данных и знаний с неопределенностью, применение методов математики и информатики в социокультурных исследованиях, применение методов биostatистики и математического моделирования в эпидемиологии, технология разработки программных комплексов с СУБД. Число научных публикаций — 250. ALT@iias.spb.su; СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450.

Tulupye Alexander Lvovich — PhD in Computer Science, Dr. of Sc. Professor; Head of Theoretical and Interdisciplinary Computer Science Laboratory, SPIIRAS, Associate Professor of Computer Science Department, SPbSU. Research area: uncertain data and knowledge representation and processing, mathematics and computer science applications in socio-cultural studies, biostatistics, simulation, and mathematical modeling applications in epidemiology, data intensive software systems development technology. Number of publications — 250. ALT@iias.spb.su; SPIIRAS, 14-th line V.O., 39, St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450.

Тулупьева Татьяна Валентиновна — доцент, канд. психол. наук; с. н. с. Лаборатории теоретических и междисциплинарных проблем информатики, СПИИРАН. Область научных интересов: применение методов математики и информатики в гуманитарных исследованиях, информатизация организации и проведения психологических исследований, применение методов биostatистики в эпидемиологии, психология личности, психология управления. Число научных публикаций — 90. TVT@iias.spb.su; СПИИРАН, 14-я

линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450.

Tuluyeva Tatiana Valentinovna — associate professor, PhD in Psychology; senior researcher, Theoretical and Interdisciplinary Computer Science Laboratory, SPIRAS. Research interests: application of mathematics and computer science in humanities, informatization of psychological studies, application of biostatistics in epidemiology, psychology of personality, management psychology. The number of publications — 90. TVT@iias.spb.su; SPIRAS, 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450.

Фильченков Андрей Александрович — аспирант кафедры информатики математико-механического факультета С.-Петербургского государственного университета (СПбГУ), младший научный сотрудник лаборатории теоретических и междисциплинарных проблем информатики СПИИРАН. Область научных интересов: автоматическое обучение вероятностных графических моделей. Число научных публикаций — 100. aaafil@mail.ru, СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450. Научный руководитель — А.Л. Тулупьев.

Filchenkov Andrey Alexandrovich — PhD student of Computer Science Department, SPbGU, junior researcher, Theoretical and Interdisciplinary Computer Science Laboratory, SPIRAS. Research area: machine learning of probabilistic graphical models. The number of publications — 100. aaafil@mail.ru, SPIRAS, 14-th line V.O., 39, St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450. Scientific advisor — A.L. Tuluyev.

Юсупов Рафаэль Мидхатович — член-корреспондент РАН, д.т.н., профессор, заслуженный деятель науки и техники РФ; директор СПИИРАН. Область научных интересов: теория управления, информатика, теоретические основы информатизации и информатизационного общества, информационная безопасность. Число научных публикаций — 390. E-mail: yusupov@iias.spb.su; www.spiiras.nw.ru. СПИИРАН, 14-я линия, д. 39, Санкт-Петербург, 199178, РФ; тел. +7(812) 328–3311, +7(812) 328–3411, факс +7(812) 328–4450.

Yusupov Rafael Midkhatovich — Corresponding Member of the Russian Academy of Sciences (RAS), Doctor of Sciences (Tech), Professor, Director SPIRAS, Honored scientist of Russian Federation. Research interests: control theory, informatics, theoretic basics of informatization and information society, information security. Number of research publications — 390. E-mail: yusupov@iias.spb.su; www.spiiras.nw.ru. SPIRAS, 14th Line, 39, St.Petersburg, 199178, Russia; office phone +7(812)328–3311, +7(812)328–3411, fax +7(812) 328–4450.

Поддержка исследований. Статья содержит материалы исследований, частично поддержанных грантами РФФИ 09-01-00861-а, 12-01-00945-а, 12-01-31202-мол_а; грантами для молодых ученых и кандидатов наук Комитета по науке и высшей школе Правительства Санкт-Петербурга; грантом СПбГУ на 2011–2013 гг. проект № 6.38.72.2011; грантом Института общественного проектирования по итогам VI Конкурса «Проблемы развития современного российского общества», проведенного в соответствии с распоряжением Президента Российской Федерации от 03 мая 2012 года № 216–рп.

Рекомендовано ТИМПИ СПИИРАН, зав. лаб. д. физ.-мат. н., доцент А.Л. Тулупьев.
Статья поступила в редакцию 10.03.2013.

РЕФЕРАТ

Азаров А.А., Бродовская Е.В., Бубнов А.Ю., Глазков А.А., Гибадулин Р.Я., Дмитриева О.В., Долгоруков А.М., Жданов Р.И., Мусина В.Ф., Нечаев В.Д., Огнев А.С., Сироткин А.В., Суворова А.В., Титов В.В., Тулупьев А.Л., Тулупьева Т.В., Фильченков А.А., Юсупов Р.М. **Predictor Mining: применение методов интеллектуального анализа данных в задачах социального компьютеринга.**

В работе предложено новое направление исследований — интеллектуальный поиск маркеров или Predictor Mining, возникшее на пересечении интеллектуального анализа данных (Data Mining) и социального компьютеринга (Social Computing).

Предполагается, что методы интеллектуального анализа данных позволят автоматически или автоматизированно сформировать набор определенных показателей (маркеров), отображающих, например, мнение пользователей социальной сети по тем или иным вопросам, а затем интерпретировать полученные показатели с помощью методов социального компьютеринга и в результате, с одной стороны, разработать теорию и объяснительные вероятностные модели динамики социального поведения, а с другой стороны, и готовить управленческие решения, позволяющие влиять на исследуемый объект и обеспечить обратной связью через отслеживание изменения показателей. Автоматизация формирования банка маркеров позволит выявить связи между ними, в частности, идентифицировать зависимости между уже известными, полученными аналитическим (экспертным) путем, и новыми, выявленными автоматически. Обнаруженные связи создадут предпосылки для более глубокого и подробного исследования причинно-следственных отношений предметной области, охватывая прежде неочевидные даже специалистам факты.

В качестве примеров областей исследований, где может быть применен интеллектуальный поиск маркеров, рассмотрены задачи выявления особенностей формирования протестных настроений и стратегий политического поведения пользователей Интернет, вычисления количественных характеристик социально-значимого поведения, а также изучение специфики формирования ценностей и стратегий поведения пользователей крупнейших социальных сетей.

Благодаря применению методов автоматического извлечения закономерностей из больших объемов данных для анализа и прогнозирования социальных процессов и их динамики, Predictor Mining создаст основу для развития нового теоретического и прикладного инструментария исследования социального поведения.

SUMMARY

Azarov A.A., Brodovskaya E.V., Bubnov A.Y., Glazkov A.A., Gibadulin R.Y., Dmitrieva O.V., Dolgorukov A.M., Zhdanov R.I., Musina V.F., Nechaev V.D., Ognev A.S., Sirotkin A.V., Suvorova A.V., Titov V.V., Tulupyev A.L., Tulupyeva T.V., Filchenkov A.A., Yusupov R.M. **Predictor Mining: Data Mining application to Social Computing.**

The paper introduces a new line of research — intelligent search of predictors of social events (Predictor Mining); this direction lies in the intersection of Data Mining and Social Computing.

We suppose that Data Mining methods can be used to compose the set of certain identifiers (predictors) automatically, that can, for example, form an image of the opinion of social network's users on some topics. Then these predictors can be analyzed by Social Computing techniques. The theory of probabilistic models of social behavior dynamics can be constructed as result of the intersection of mentioned areas of research. This theory, in its turn, can be used in management decision making associated with considered object. The influence of decisions can be traced in the dynamics of the mentioned predictors. Automated process of predictors' formation will allow revealing connections among the predictors including interactions between ones identified by experts and ones identified automatically. These connections will be the base for deep and detailed research in the causal structure of the area, as they can reveal facts unknown to the experts.

Detection of protest tones in the Internet, estimation of parameters of socially significant behavior, and investigation of largest social networks members' values and strategies formation specifics are examples of problems that can be analyzed using the Predictor Mining techniques.

Automated methods for extraction connections between variables in big data can be successfully applied for analysis and prognostics of social dynamics. Predictor Mining can form the base for development of the new theoretical and applied instruments of social behavior investigation.