ОБЗОР ОНЛАЙНОВЫХ СИСТЕМ РЕПУТАЦИИ/ДОВЕРИЯ

Д.А. Губанов

(Институт проблем управления РАН, Москва, DimaGubanov@mail.ru)

1. Введение

В Интернете все чаще появляются **виртуальные сообщества**, в которых люди обмениваются мнениями по широкому кругу вопросов и взаимодействуют между собой. В таких сообществах для регулирования деятельности существуют **нормы** поведения, взаимодействия, общения и т.п., которые устанавливаются сообществом и могут динамично изменяться со временем при возникновении новых ситуаций. В качестве системы санкций при несоблюдении норм используется **репутация**, которая одновременно является сигналом для сообщества. Т.е. формируется система репутации/доверия для упорядочения взаимодействия и стимулирования порядка, поскольку обычные механизмы нормативного регулирования и права не работают.

Возьмем пример **электронной торговли**, в которой участвуют продавцы (в общем случае провайдеры) и покупатели (соответственно потребители). Нормами здесь могут быть, например: «Описание товара должно соответствовать ему», «При получении денег продавец должен выслать товар» и т.п. Зачастую покупатель товара имеет недостаточно информации о продавце, а также о предоставляемых товарах и услугах. Покупатель вынужден идти на риск, оплачивая товары и услуги до их оценки и получения, в то время как продавец обычно точно знает, что он получит деньги. Поскольку заранее определить качество товара трудно, для покупателя нет смысла платить высокую цену за качество – нет никакой гарантии, что товар ей соответствует. Соответственно, и продавцу незачем продавать хороший товар. В результате такой ситуации, электронный рынок деградирует – на нем торгуется только товар низкого качества, с низким уровнем сервиса (теория о «рынке лимонов» [7]). Такая информационная асимметрия может быть смягчена репутацией и доверием. Идея заключается в том, что даже если покупатель не может опробовать товар или услугу заранее, он уверен в том, что он покупает, если доверяет продавцу с высокой репутацией (например, получен сигнал на основании 1000 положительных отзывов – «всем известно, что продавец X продает товар высокого качества, присылает его вовремя и т.п.»). Соответственно продавец не будет продавать товар плохого качества, поскольку, в этом случае никто больше покупать у него ничего не будет.

При введении онлайновых систем репутации и доверия возникает большое количество вопросов и проблем. Важны исследования новых социальных механизмов обратной связи, которые могут оказать сильное влияние на бизнес, политику, общество, чтобы понять их выгоду и правильно использовать. Вот некоторые вопросы: В каких областях применимы механизмы репутации и доверия? Насколько они будут и экономически эффективны и социально справедливы? И для кого? Какие существуют риски при использовании таких механизмов, возможные злоупотребления стратегическими агентами и как от них защититься? Как спроектировать хорошую систему репутации и доверия для конкретной области (как представить все факторы и сигналы, используемые людьми в реальной жизни для вынесения доверия, как собрать свидетельства для принятия решения о доверии незнакомым партнерам, как должны понимать и воспользоваться сигналами системы люди при принятии решения и т.п.)?

Проблемам доверия и репутации в онлайновых системах посвящено большое количество исследований [14] в области экономики (формирование репутации и социальное обучение), компьютерной науки (вычислительные модели доверия и репутации, вопросы масштабируемости, распределенности и безопасности вычислений), социологии и психологии (отмечающих ограниченную рациональность людей, важность эмоциональных и когнитивных факторов), науке о методах управления (рассматривающей и учитывающей влияние репутации/доверия в маркетинге, создании бренда и т.п.), а также политологии (влияние репутации на общественное мнение). Для начала попробуем определиться с понятиями доверия и репутации.

2. Доверие, репутация и их связь

В литературе термину «доверие» дают самые разные трактовки в зависимости от его использования и, соответственно, применяемой модели.

В психологии и социологии: **Доверие** [19] – *это открытые и положительные взаимоотношения между людьми, содержащие уверенность в порядочности и доброжелательности другого человека, с которым доверяющий находится в тех или иных отношениях*. Отмечаются отношения, открытость (а значит и уязвимость), уверенность, готовность положиться и делегировать выполнение задач. Также отмечается разноплановый характер доверия в зависимости от играемых социальных ролей и социальных отношений. И утверждается то, что доверие является основой всех социальных институтов. Все это подробнее рассматривается в теории Social Network Analysis (SNA).

В своей широко цитируемой работе [27] McKnight и Chervany выделяют следующие ключевые качества, значимые при принятии решений о доверии: **компетентность** (способность выполнить действие с должным уровнем качества), **доброжелательность** (готовность тратить усилия), **честность** (честное поведение), и **предсказуемость** (есть доказательства, подтверждающие, что желаемый результат будет иметь место). Т.е. составляющими доверия являются компетентность, доброжелательность, честность и предсказуемость.

Другой пример набора качеств: предсказуемость, добросовестность, соответствие (соответствие действий заявлениям), открытость (предоставление информации), признание (уважение) и чуткость (уделение внимания).

Экономисты и представители компьютерных наук используют более формальное и конструктивное определение в рамках теоретико-игрового подхода:

Доверие [*28*] *– это субъективная вероятность со стороны А выполнения данного действия стороной В, которое А еще не может наблюдать и которое влияет на действия А. Действие В повлияет на благосостояние А, его выгоду.*

Т.е. под доверием подразумевается **надежность** (вероятность), отмечается субъективность доверия, зависимость (влияние на действия) и выгода (агент стремится максимизировать свой выигрыш). Ожидается, что взаимодействия между агентами имеют повторяющийся характер (например, многошаговая игра Дилемма заключенного).

Доверие [*28*] *– это субъективное ожидание агентом А будущего поведения В на основе истории их взаимодействий.* Отмечается история взаимодействий, субъективное ожидание как результат прошлых взаимодействий.

Mui также отмечает [28] (ссылаясь на теорию эволюции) тот факт, что сотрудничество и кооперация способствовали выживанию человека как вида. Вводится социальная норма *«взаимности/обоюдности»* (reciprocity – взаимный обмен действиями): «ты мне, я тебе», «око за око». Отмечается, что при взаимодействии агенты субъективно ожидают взаимность. Происходит это следующим образом:

1. Репутация В влияет на степень доверия А к В.
2. Увеличение доверия А к В приводит к увеличению вероятности того, что А будет ожидать от В позитивной «взаимности» в действиях.
3. Выполнение «взаимности» агентом В приведет к росту его репутации и увеличению общей выгоды.

Однако следует отметить, что люди обладают ограниченной рациональностью, их доверие может быть совершенно иррациональным (пример – влияние эмоций), а также основываться на различных когнитивных факторах. То есть определение доверия у представителей экономических наук излишне упрощено. В [27] предлагается следующее определение:

Доверие – это мера готовности стороны А положиться на кого-то или что-то в данной ситуации с некоторой относительной уверенностью, несмотря на возможные негативные последствия.

Несмотря на некоторую размытость данного определения, оно позволяет выделить аспекты субъективности, зависимости, надежности, полезности, меры риска, си-

туационного контекста (например, наличия тех или законов, права и.п.) в контексте **принятия решения** стороной A.

Castelfranchi и *Falcone* [12] критикуют подход экономистов и адептов теоретико-игрового подхода, отмечая непрозрачность, примитивность определения доверия как «субъективной вероятности», и предлагают свою когнитивную модель доверия. С когнитивной точки зрения: *Доверие – это сложная структура убеждений и целей, требующая наличия у доверяющего «теории мышления» доверяемого субъекта.*

Такая структура убеждений определяет «степень доверия» и оценку риска, а на его основе и на основе персональных порогов риска определяет решения о возможности взаимодействия сторон. Структура убеждений хорошо соответствует когнитивной теории агентов (BDI подход: убеждение (belief), желание (desire) и намерение (intention)), где важная роль уделяется нормам, ожиданиям, ролям и т.д., а также социально-обусловленной теории агентов в сообществах.

Исследователи доказывают важность и полезность когнитивной модели, ментальных компонентов доверия в электронной торговле, там, где агенты должны выявлять источники и причины для доверия и рациональной основы для решений. Аргументируют это тем, что экономическая сфера существует в социальной сфере (институты, нормы, культура). Также экономические субъекты в полной мере являются социальными, и они действуют в экономических транзакциях в силу своего характера, со всеми их мотивами, имеющими место быть взаимоотношениями и т.д. Так что доверие не сводится к «субъективной вероятности» (непонятному для пользователя магическому числу), а представляет собой следствие из убеждений агента и его моделей о мире (внешних аспектов и обстоятельства) и мышлении других агентов (внутренних аспектов: убеждений, веры, ожиданий в отношении другого агента, его желаний, мотивов, уверенности в себе, настойчивости, моральных принципов, общности целей). Для чего это надо? Во-первых, потому что иначе мы не сможем объяснить или предсказать восприятие риска и принятие решений агентом. Во-вторых, потому, что без явной и четкой когнитивного модели доверия любая теория доверия, убеждения, влияния, обмана, репутации и т.д. является просто пустой. Итак:

Доверие – это сложные ментальные отношения/позиции доверяющего когнитивного агента X (*ментальное состояние*)*, характеризующие его мышление, по отношению к выбранной сущности/агенту Y по поводу ожидаемого поведения/действия* α*, имеющего значение для достижения цели G* (*конкретное состояние мира, необходимое и желаемое X*)*. Агент X в сущности делегирует выполнение* α*.*

Каковы ментальные компоненты доверия X к Y? Это следующие убеждения (belief):

- Вера в компетентность: агент X должен верить, что Y действительно может выполнить задачу и выдать ожидаемый результат, необходимый для достижения цели;
- Вера в намерение (disposition): агент Y не только может выполнить задачу, но и выполнит ее. Формируется на основе еще двух убеждений:
  - Вера в готовность (willingness): агент X верит в то (моделирует мышление Y), что Y решился и намерен выполнить α.
  - Вера в стойкость (persistence): агент X верит в то, что Y стабилен в своих намерениях (intentions) сделать α, если Y имеет предсказуемый характер и у него нет расхождений по поводу α.
- Вера в зависимость (dependence): агент X считает, что Y необходим для выполнения задачи (строгая зависимость) или, что лучше полагаться на него, чем не полагаться (слабая зависимость).
- Вера в самоуверенность: X должен верить в то, что У знает, что он может сделать α. Трудно доверять кому-то, кто не доверяет себе сам.

И, наконец, последний компонент доверия, следующий из остальных:

- Вера в выполнение (fulfillment): X верит, что G будет достигнуто (благодаря Y), поэтому он не отказывается от цели G, не ищет альтернативу агенту Y и достигает G через Y.

Степень доверия – есть субъективная уверенность убеждений, количественная оценка зависит от количественной оценки составляющих. В работе [11] исследователи подробно описывают математические соотношения между компонентами доверия.

2.1. Аспекты доверия

При моделировании доверия возникает необходимость учета факторов (эффектов), имеющих место в социальных системах:

- Надежность/Готовность положиться/Делегировать выполнение;
- Дискретность/немонотонность доверия (для человека естественно делить его на дискретные уровни);
- Субъективность (доверие персонально) и асимметричность (если мы кому-то доверяем, это не значит, что нам так же доверяют);
- Транзитивность доверия (в общем случае оно не транзитивно);
- Неопределенность доверия (сложно четко определить доверие, можно указать границы);
- Многофакторность (доверие состоит из многих когнитивных компонентов: компетентность и т.п.);
- Зависимость от контекста (обстоятельства и область доверия);
- Зависимость от рекомендаций и репутации (указано выше, подробнее ниже);
- Связь с концепцией взаимности («око за око»);
- Динамика (доверие может изменяться со временем с приобретением опыта, так и без его приобретения);
- Доверие непосредственно связано с риском (уязвимость) и учитывается при принятии решения;
- Доверие может основываться на истории взаимодействий.

Отдельно рассмотрим кратко транзитивность доверия, социальные сети и связанные с доверием понятия риска, порога риска и принятия решения.

2.1.1. Транзитивность доверия

В общем случае доверие не транзитивно [25]. Мы не обязательно должны доверять тем, кому доверяют те, кому мы доверяем: A доверяет B, B доверяет C, но A имеет полное право не доверять C (или если A не доверяет B, а B не доверяет C, то это не значит, что A не доверяет C). Но при некоторых условиях доверие может действовать по цепочке (имеющей ограничение на длину): по крайней мере, ранее неизвестному лицу, рекомендованному нам кем-то, кому мы доверяем, доверяем больше чем незнакомцу (например, кто-то может доверять писателю из-за издателя, а издателю может доверять только потому, что его рекомендовал друг). Здесь важен контекст доверия, и мы рассматриваем доверие в смысле надежности.

2.1.2. Социальные сети

Агенты объединяются для эффективной обработки информации, в том числе для обмена информацией о доверии и репутации. Агенты образуют так называемые **сети доверия**.

2.1.3. Риск, порог риска и принятие решений

Как уже отмечалось, агент обычно определяет степень своего доверия, оценивает риск, а на его основе и на основе персональных порогов риска определяет решение о возможности взаимодействия с потенциальным партнером.

Агент должен довериться (даже если есть неопределенность доверия) и принять некоторую вероятность неудачи, т.е. принять риск. Риск определяет возможные негативные последствия принятия решения. При этом недостаточно производить некоторую положительную оценку доверия, нужна оценка порога «приемлемого» ущерба. Стоимость ущерба может оказаться слишком высокой для агента, даже независимо от вероятности неудачи (возможно очень низкой) и получаемой выгоды (возможно очень большой) при успехе. Т.е. опасность слишком высока.

Кроме того, нужно отметить, что в доверии могут быть иррациональные, необоснованные оценки компонентов доверия (слепая вера и т.п.), учет этого и следствия подробно рассмотрены в работе *Falcone* [12].

Перейдем к понятию репутации, которое уже затрагивалось нами ранее.

2.2. Репутация

С репутацией немного проще, чем с доверием. Итак:

Репутация [1] – *сформировавшееся общественное мнение о качествах, достоинствах и недостатках того или иного индивида*

Репутация напрямую связана с социальными сетями, участники которой и формируют мнение о ее индивидах. Как видно доверие связано с репутацией, но отличается от нее: мы можем доверять кому-либо благодаря репутации, однако можем доверять и вопреки ней (возможно у нас есть какие-то свои персональные знания, следующие из личного опыта и личных связей).

А как определяют репутацию исследователи компьютерной науки? Да, в общем-то, также:

Репутация [28] – *восприятие об агенте, сложившееся на основе его прошлых действий, о его намерениях и нормах.* Количественная оценка, рассчитываемая на основе действий данного агента, наблюдений других агентов в социальной сети. Можно сказать, что репутация с вычислительной точки зрения – это общественная мера надежности, основанная на рекомендациях или оценках членов сообщества. В то время как субъективное доверие – комбинация личного опыта (обычно более значимого), полученных рекомендаций и других факторов.

Можно говорить и о **коллективной репутации** – репутации группы, которая моделируется как среднее всех значений репутации ее членов или как среднее мнений других участников социальной сети, воспринимающих группу как одно целое. Обзор и ряд оригинальных теоретико-игровых (в том числе – рефлексивных) моделей репутации приведен в [3] (см. также теоретико-игровые модели индивидуальной и коллективной репутации в [1, 9, 16, 17, 18, 24, 33]).

Наше доверие может основываться не только на репутации, а на основе каких-либо удостоверений (например, при приеме на работу я доверяю незнакомому человеку с дипломом о высшем образовании больше, чем человеку без оного).

3. Общая модель онлайновой системы доверия

Онлайновая система использует механизмы доверия и репутации для защиты своего функционирования в интересах экономической эффективности и социальной справедливости. Система состоит из следующих элементов:

3.1. Среда взаимодействия.

Среда, в которой происходит взаимодействие участников, в которой реализуются механизмы репутации и доверия. Это могут быть электронные рынки, пиринговые сети, онлайновые социальные сети и т.п. В настоящее время происходит образование глобальной социальной сети, в которой появляются конкретные узкоспециализированные сервисы и агенты, при помощи новых технологий Web 2.0 и Semantic Web.

3.2. Субъект доверия. Тот, кто доверяет: агент (пользователь, сервис и т.п.).

3.3. Объект доверия. То/т, чему/кому доверяют. Варьируется в зависимости от поставленной проблемы. В системах контроля доступа объектом доверия являются пользователи. В коммуникационных сетях объектом доверия становятся агенты или пользователи, использующие канал связи (или сами каналы). При поиске надежного сервиса объектом доверия становятся агенты или Web-сервисы сети. При поиске информации в Интернете объектом доверия может стать агент (сайт тоже может быть), предоставляющий контент, или даже сам контент.

Таким образом каждая ситуация предъявляет разные требования к доверию. Пользователи, программные агенты и сервисы – все должны заслуживать доверия в

различных приложениях и ситуациях. Поэтому в онлайновых системах можно различать виды доверия, необходимые для их функционирования.

3.4. Виды доверия

Определим виды доверия:

Доверие на предоставление услуг (provision trust) описывает доверие доверяющей стороны в оказание качественных услуг провайдером услуг или ресурсов (то, что мы рассматриваем).

Доверие делегирования (delegation trust) описывает доверие в агента (представителя), действующего и выносящего решения от имени доверяющей стороны. Как частный случай provision trust.

Доверие доступа (access trust) описывает доверие доверяющей стороны (провайдера) к агентам, которым предоставляется доступ к ресурсам. Это – контроль доступа.

Доверие к подлинности описывает убеждение в заявленную подлинность агента. Используется в системах аутентификации.

Контекстное доверие описывает меру веры участника в необходимые системы и институциональные механизмы, поддерживающие транзакции и обеспечивающие безопасность сети, в том случае, если что-то пойдет не так (страхование, правовая система, правоохранительные органы – тоже рассматривались ранее как ситуационный контекст доверия).

Далее мы ограничимся доверием на предоставление услуг и доверием делегирования (хотя все вышеперечисленные виды доверия, конечно, взаимосвязаны).

3.4.1. Виды доверия: методы защиты в онлайновых системах

Цель методов – обеспечение защиты системы от злонамеренных агентов. Для каждого вида доверия (см. выше) используются свои методы.

Для доверия подлинности и доверия доступа используются механизмы жесткой безопасности: шифрование канала связи, схемы криптографической аутентификации и авторизации, политики для предоставления полномочий. Используются цифровые подписи и сертификаты, выданные доверенной всеми третьей стороной (может использоваться свойство транзитивности доверия) и т.д. Эти традиционные методы обеспечения безопасности не будут рассматриваться нами далее.

Для доверия к предоставлению услуг и доверия делегирования используются механизмы мягкой безопасности. Провайдеры услуг могут предоставлять ложную или вводящую в заблуждение информацию, и традиционные механизмы обеспечения безопасности не могут защитить пользователей от этого вида угроз. А могут механизмы социального контроля. Системы доверия и репутации могут обеспечить защиту пользователя от таких угроз, более того они могут защитить саму систему (так называемые надежные системы, Trusted systems).

В соответствии с видом доверия определяется

3.5. Предмет доверия: То, на что направлено внимание доверяющего субъекта (область отношения доверия). Например, Y «является первоклассным программистом» и, следовательно, ему можно доверить написание программы (provision trust).

3.6. Действия.

Собственно действия, предпринимаемые агентами в транзакциях, исходя из доверия к партнерам (различного рода сущностям). Например: «Купить ноутбук», «Принять информацию» и т.п.

После завершения транзакции (взаимодействия) агенты могут оценить действия друг друга (**обратная связь**). Для этого используются подходящие меры, значения которых используются для расчета репутации и доверия.

3.7. Обратная связь. Типы мер для измерения доверия:

Семантика мер может быть описана в терминах *конкретность-общность* (конкретный аспект доверия – среднее всех аспектов) и *субъективность-объективность* (субъективное мнение – объективная оценка по формальным критериям).

Субъективные и конкретные меры используются в анкетных опросах, в которых люди выражают свое мнение по конкретным вещам («Оцените, пожалуйста, работу сайта X» по шкале оценок «Ужасно, Плохо, Средне, Хорошо, Отлично»), тем самым формируя субъективный вектор доверия.

Субъективные и общие меры (пример – eBay). Проблема таких систем, что они не позволяют оценить некоторые аспекты; например, клиент поставил низкую оценку при сделке, т.к. не получил вовремя товар, но на самом деле была виновата служба доставки.

Объективные и конкретные меры используются, например, в технических тестах продукта, где качество продукта измеряется объективно (например, по потреблению энергии, шуму и т.п.).

Объективные и общие меры могут быть примером вычислений на векторе объективных и конкретных мер.

А дальше происходит повторение цикла. Обратная связь влияет на доверие, доверие влияет на действия агента в следующей транзакции и т.п. (можно применить схему: репутация, доверие, «взаимность», доверие, польза).

Способы вычисления доверия (вычисления доверия: вычислительные модели), вопросы комбинации источников сигналов доверия, представление сигналов, вопросы сбора таких сигналов и хранения, оценка риска и принятие решений, а также вопросы манипулирования такими системами являются предметом многих исследований, некоторые из них рассматриваются далее.

4. «Исследовательские» аспекты онлайновых систем доверия

В онлайновых системах необходимо учитывать и как-то представлять многие сигналы, используемые нами в обычной жизни для принятия решения о доверии. С другой стороны, как правило, в обычной жизни коммуникация и обмен информацией, связанной с доверием и репутацией, является нелегким делом и ограничивается пределами локальных сообществ, тогда как использование информационных технологий в Интернете может привести к разработке чрезвычайно эффективных систем для обмена и сбора такой информации в глобальных масштабах.

Таким образом, задачами исследований в области онлайновых систем репутации/доверия являются: 1. Поиск адекватных представлений традиционным сигналам, используемых нами в обычной жизни, выявление информационных элементов (специфичных для конкретных онлайновых приложений), пригодных для получения мер репутации/доверия. 2. Использование преимуществ информационных технологий в Интернете для создания эффективных систем сбора этой информации и получения мер доверия/репутации с целью поддержки принятия решений и улучшения качества работы онлайновых сервисов (например, электронных рынков).

Возникают также и сопутствующие вопросы: Какие элементы информации являются наиболее подходящими для получения мер доверия и репутации в том или ином приложении? Как их можно получить? Каковы оптимальные принципы проектирования таких систем с теоретической точки зрения и с точки зрения удобства использования? Могут ли они быть устойчивыми к атакам манипуляции стратегических агентов? Как пользователи должны использовать при принятии решения информацию, предоставляемую такими системами? Какую роль могут играть эти системы в бизнес-модели коммерческих компаний? И могут ли они действительно улучшить качество услуг? Ответы на эти вопросы определяют возможности систем доверия и репутации в Интернет-среде.

Соответственно, онлайновые системы доверия/репутации должны:

1. Моделировать доверие (с использованием метрик, затрагивающих аспекты представления и вычислений значений доверия)

2. Управлять доверием (фокусироваться на сборе фактов и оценке риска)
3. Поддержать принятие решения пользователем.

4.1. Моделирование доверия (*метрики*)

4.1.1. Представление доверия.

Значение доверия определяется:

1. *Доменом*. Значения могут быть *бинарными* («доверие»/»недоверие»), *дискретными* (метки, обозначающие множество натуральных чисел, естественны для понимания человеком) и *непрерывными* (хорошо поддерживаются известными математическими теориями, в зависимости от семантики значений доверия). Внимания заслуживает представление значения с семантикой «не знаю» и «не доверяю». Так, в некоторых моделях с непрерывным доменом значения доверия значения «не доверяю» нет вовсе: [0; 1], где 0 – это «не знаю», а 1 – доверяю полностью; в других 0 означает полное недоверие, 0.5 – «не знаю» , 1 – полное доверие; в третьих диапазон [-1; 1] и т.п.

2. *Размерностью*. Некоторые модели представляют доверие одним значением, а другие несколькими (например, <b, d, u> – значение доверия, значение недоверия, значение неопределенности или представление доверия интервалом). Отдельного рассмотрения требует связанная с доверием *мера надежности репутации/доверия* (иногда для окончательного принятия решения необходимо знать насколько надежно полученное значение доверия/репутации).

3. *Семантикой*. В некоторых моделях значение доверия представляется рейтингом (прямо указывает на степень надежность, например, «очень надежный»), в других – рангом (относительная величина, не указывает прямо, является основой для сравнения), вероятностью (ожидания), верой (belief), нечетким значением.

Сюда же относятся сопутствующие вопросы представления истории взаимодействий, политик доверия, протоколов взаимодействия и т.п. в виде правил, записей, онтологий (Semantic Web).

4.1.2. Вычисление доверия

Доверие может рассчитываться количественно по-разному. В некоторых подходах, в том числе Semantic Web, используются дискретные значения доверия (например, доверие, недоверие, или нейтральность), в то время как в других используется непрерывный диапазон (см. выше). Алгоритмы для вычислений могут варьироваться от простого среднего значения до вычисления собственных значений по матрицам смежности соответствующего графа. Многие подходы не учитывают изменение доверия со временем. В тех случаях, когда информации для расчета доверия требуется много или информация непрерывно меняется, обычно предлагается использовать локальное вычисление доверия, а не глобальное. Подробнее эти модели рассматриваются в разделе «Вычислительные модели»

Следующие факторы определяют различия вычислительных моделей:

Концептуальная модель. Вычислительные модели моделируют доверие либо с когнитивной точки зрения как функцию основополагающих убеждений (как результат ментального состояния агента), либо с теоретико-игровой как субъективную вероятность (как результат прагматичной игры).

Рассмотрим **теоретико-игровой подход.** В социальной сети происходят *повторяющиеся* игры *N* агентов, у каждого из которых существует *неопределенность* относительно структур полезности других агентов. На каждом шаге агент должен смоделировать действия других агентов (основываясь на своих ожиданиях их действий), максимизирующие их функции полезности. Для решения игры используется концепция равновесия Нэша, в которой одностороннее отклонение не выгодно ни одному из агентов. Необходимо учесть, что при выборе агентом тех или иных действий ожидания других игроков могут измениться и соответственно изменится его репутация (обратная

связь). Поэтому изменится поведение агентов на следующих этапах игры («зуб за зуб»), что приведет к новым равновесиям.

Недостатки такого подхода:

- Рассматриваются игроки, играющие в течение длительного периода (а в больших онлайновых сообществах многократные взаимодействия между игроками редки – однако вероятно можно применить в узкопрофессиональных тематических сообществах);
- Люди ограниченно рациональны;
- Взаимодействия одних агентов рассматриваются отдельно от других взаимодействий;
- Практически не учитываются проблемы механизмов репутации;
- Сведения доверия и репутации к вероятности при возрастании сложности модели недостаточно.

Источники информации для расчета репутации/доверия. Для расчета репутации/доверия используются как традиционные источники: непосредственный опыт и косвенная информация (от свидетелей), так и менее распространенные, например, информация, связанная с социальными аспектами поведения агентов. Правильное комбинирование этих источников может повысить надежность и точность рассчитываемого значения доверия/репутации, хотя и увеличивает сложность модели и требует умных агентов для обработки предоставленной информации.

Непосредственный опыт. Наиболее надежный источник информации. Включает как опыт, накопленный в ходе непосредственного взаимодействия с партнером, так и опыт, основанный на наблюдении взаимодействий других пользователей (в современных моделях используется редко, в рамках сценария системы).

Косвенная информация (*от свидетелей*). Свидетельская информация (также называемая информацией, передаваемой из уст в уста, или косвенной информацией) является информацией, поступающей от других пользователей (непосредственный опыт одного из агентов). Хотя такая информация часто используется, ее сложнее использовать в моделях доверия и репутации из-за неопределенности того, как была получена такая информация (может быть сокрыта или искажена свидетелями в своих собственных интересах).

Социологическая информация. Основу этих знаний составляют социальные отношения между агентами (торговля, конкуренция, сотрудничество и т.п.) и играемые ими роли в обществе. И социальные отношения, и роли агента влияют на его поведение и взаимодействие с другими агентами. Используются методы *social network analysis* (*SNA*) для анализа социальных структур и их реляционных (затрагивающих отношения) аспектов. Лишь немногие модели используют данный вид информации для улучшения расчета значений доверия и репутации, поскольку современные системы практически не содержат информацию подобного рода. Однако в будущем увеличение сложности мультиагентных систем, обогащение их разного рода сложными отношениями приведет к повышению значимости данного вида информации.

Контекстная зависимость. Очевидно, что доверие/репутация в общем случае контекстно зависимы, то есть зависимы от ситуации. Чаще в вычислительных моделях рассматривается контекст предметной области: если мы доверяем преподавателю русского языка в вопросах правописания, то это не значит, что мы должны доверять ему в вопросах организационного управления. В многоконтекстных моделях доверия/репутации с каждым пользователем для каждого из контекстов связаны значения репутации/доверия. Обычно информации в системах не хватает, поэтому хорошая многоконтекстная система должна правильно использовать ее в разных контекстах. Введение нескольких контекстов усложняет систему и на данном этапе развития современные системы используют только один контекст в связи с решением ограниченных, конкретных задач (все действия пользователя происходят в одном контексте).

Предположения о поведении агента. Способность иметь дело с манипулирующими агентами. Возможны три вида моделей:

1. Модель явно не учитывает таких агентов, считается, что большое количество агентов, предоставляющих достоверную информацию, нейтрализует недостоверную.

2. Модель предполагает, что агенты могут скрыть информацию или завысить/занизить оценку, но они никогда не лгут.

3. Модель использует специальные механизмы для борьбы с лжецами.

«Дискриминация». Для расчета доверия/репутации модели иногда используют механизмы, выделяющие агентов в пользующиеся определенной репутацией группы по некоторым признакам (например, поведению).

Использование глобальных или локальных значений. Значение репутации **глобально** (одно для каждого пользователя), значение доверия **персонально** (для каждой пары пользователей). В первом случае значение репутации рассчитывается исходя из мнений пользователей, взаимодействовавших с данным в прошлом. Эта величина является доступной для всех пользователей и обновляется каждый раз при появлении нового мнения. Во втором случае устанавливается персональное значение пользователя *x* с точки зрения пользователя *y* исходя из непосредственного опыта, косвенной информации, полученной от других агентов, известных отношений между пользователями и т.п.

Глобальные значения используются сегодня в большинстве онлайновых систем, предназначенных для сценариев с тысячами или даже миллионами пользователей. Размер этих сценариев делает маловероятными повторные взаимодействия между одними и теми же пользователями, и, следовательно, снижает стимулы для пользователей к сотрудничеству для построения выгодных отношений. Надежность этих систем зависит от количества мнений по данному пользователю, большое число мнений сводит риск к минимуму [14].

В простых вопросах применение глобальных значений допустимо, но не для сложных и субъективных вопросов (то, «что русскому хорошо, то немцу – смерть»). Персональные/локальные значения используется в небольших и средних по количеству пользователей системах, где часты взаимодействия и устанавливаются прочные связи между пользователями. Данное соображение еще более важно в контексте социальных сетей.

4.2. Управление доверием и поддержка принятий решений

Сущности, которые определяют доверие, могут варьироваться в зависимости от применения/приложения. Во многих традиционных системах централизованная служба (*центр*) выступает в качестве доверенной третьей стороны для установления доверия между двумя неизвестными агентами или пользователями. Но во многих случаях (опять же в зависимости от приложения, а также соображений безопасности центр может действовать в своих интересах) такой подход неудобен. В децентрализованных подходах отдельные агенты могут принимать решения о доверии без участия центра, и агенты могут использовать информацию так называемых **рефералов**, имеющих предшествующий опыт с неизвестным лицом (также рассматривается контекст доверия, разрешение проблемы противоречивости полученной информации и того, как изменяется репутация рефералов).

Существуют системы доверия, которые включают элементы систем репутации, и наоборот, так что не всегда ясно, как должны быть классифицированы такие системы. Говорить что-либо, в общем для всех систем, уже становится сложнее, поэтому перейдем к видам систем, основывающихся на доверии.

5. Системы, основывающиеся на доверии

5.1. Системы репутации

Системы репутации должны обладать тремя свойствами:

1. Агенты должны быть долгоживущими, так чтобы при выполнении каждого взаимодействия ожидались взаимодействия в будущем. Поэтому агенту должно быть трудно изменить свой «псевдоним», чтобы избавиться от истории взаимодействий.

2. Оценки текущих взаимодействий должны получаться и распространяться. Это обеспечивается протоколом в системе, и для распределенных систем в отличие от централизованных систем это является проблемой. Для работы системы участники должны быть готовы предоставлять оценки, для этого должны быть разработаны соответствующие механизмы стимулирования.

3. Оценки прошлых взаимодействий должны учитываться при принятии решений о текущих взаимодействиях. Это зависит от удобства использования такой системы.

Укажем также на главные различия между системами репутации и доверия следующие:

1. Системы доверия рассчитывают значения, отражающие субъективное мнение участника о надежности агента, а системы репутации выводят значения репутации агента на основании информации всего сообщества.

2. Транзитивность является непосредственным компонентом в системах доверия, а системы репутации, как правило, лишь косвенно принимают во внимание транзитивность.

3. Системы доверия, как правило, получают на вход субъективные и общие меры (надежность) доверия, тогда как в системы репутации – информацию или оценки о конкретных (объективных) событиях, таких как транзакции.

Архитектуры сетей репутации. Архитектура определяет, как оценки и значения репутации передаются между участниками системы репутации.

Централизованные системы репутации. В такой системе собираются оценки действий данного участника, данные другими участниками сообщества, имеющих прямой опыт. Имеется центр, который собирает и публикует их, рассчитывает значение репутации. В дальнейшем участники могут использовать эту информацию для решения того вступать во взаимодействие или нет. Мы имеем дело здесь со следующими аспектами: 1. Централизованные протоколы коммуникации, предоставляющие возможность как предоставить оценки участниками о партнерах по транзакции центру, так и получить от центра значение репутации потенциального партнера по транзакции. 2. Модель вычисления репутации, используемая центром для выведения значений репутации каждого участника, основывающаяся на полученных оценках и возможно другой информации.

Распределенные репутационные системы. В распределенной системе нет какого-либо центра сбора оценок и получения значений репутации. Вместо этого используются распределенные хранилища оценок или даже каждый участник может хранить у себя свои мнения об опыте с другими участниками и по запросу отправлять эту информацию. Агент для принятия решения об участии в транзакции находит эти хранилища или получает оценки от членов сообщества, имеющих непосредственный опыт с потенциальным партнером. После чего агент вычисляет значение репутации потенциального партнера, основываясь на полученных оценках и своем прямом опыте. Укажем на следующие аспекты: 1. Распределенный коммуникационный протокол, позволяющий участникам получать оценки от других членов сообщества. 2. Метод вычисления репутации, используемый каждым агентом на основе полученных оценок и возможно другой информации.

5.2. Системы совместного фильтрования

Системы совместного фильтрования (такие как, например, Last.Fm) имеют сходство с системами репутации в том, что собирают оценки членов сообщества. Однако они также имеют принципиальные различия. Предположение таких систем заключается в том, что разные люди имеют разные вкусы и оценивают вещи по-разному. Если два пользователя оценивают множество вещей одинаково, то, значит, они имеют аналогичные вкусы, и называются *соседями*. Эта информация может быть использована для рекомендации вещей, нравящихся одному участнику, его соседям. Реализации этого метода часто называют *рекомендующими системами*. Их не следует путать с системами репутации, основанных на противоположном предположении того, что все

члены сообщества должны **согласованно** судить об эффективности взаимодействий агентов или качестве товаров или услуг.

Системы совместного фильтрования учитывают индивидуальные вкусовые оценки, тогда как системы репутации нет. Системы совместного фильтрования полагаются на надежность и честность участников, системы репутации априори полагаются на их ненадежность.

6. Вычислительные модели (метрики)

6.1. Простое суммирование или среднее значение оценок

Значение репутации является суммой положительных и отрицательных откликов. Пример – eBay. Такой метод подсчета примитивен и значение репутации получается грубым, и, тем не менее, представляются важными следующие преимущества данного метода: прозрачность и понятность для пользователя. Более сложные схемы используются в Epinions и Amazon, в которых производится взвешивание оценок в зависимости от репутации, времени оценки, расстояния и т.п.

6.2. Marsh

В последнее время широко цитируется работа [25], считающаяся первой из известных комплексных вычислительных моделей доверия. *Marsh* задался вопросами понимания доверия, а также вопросами использования доверия в литературе и в повседневной жизни. В своей работе *Marsh* моделирует только прямое доверие. Он предлагает множество переменных и их способ объединения для получения одного значения доверия в диапазоне [-1; 1] (хотя по его утверждению не бывает полного доверия или недоверия). Каждая из переменных доверия зависит от контекста и времени. Marsh определил три типа доверия: **базовый**, во всех контекстах T_x^t; **общий,** между двумя людьми во всех контекстах $T_x(y)^t$, и **ситуационный**, между двумя людьми в конкретных условиях $T_x(y,a)^t = U_x(a)^t \cdot I_x(a)^t \cdot \overline{T}_x(y)^t$, где $U_x(a)^t$ – полезность ситуации *a* для агента *x*, $I_x(a)^t$ – важность ситуации *a* для агента *x*, $\overline{T}_x(y)$ – оценка общего доверия для всех ситуаций s, близких к *a* (одного класса) и имеющих место быть в прошлом, т.е. «усреднение» $T_x(y,s)^T$ во временном окне $\theta < T < t$ («усреднение» – это либо max, либо min, либо среднее арифметическое).

Эти значения доверия используются для расчета риска (который зависит также от затрат и выигрыша), связанного с данной ситуацией и предполагаемой компетенции целевого агента (доверяемого), для того, чтобы помочь агенту принять решение о взаимодействии с другим агентом на основе некоторого порога. Сотрудничество возможно, если ситуационное доверие выше порога. Также принятие решения расширяется понятием «взаимности»: «ты мне, я тебе» (для модификации значения доверия, т.е. если агент x помог y в прошлом, а y в ответ нет, то значение доверия будет снижено).

6.3. Advogato Trust Metric (модель потока)

Алгоритм *Advogato* Trust Metric [6] лег в основу **блога** http://www.advogato.org и позволил защитить сообщество от таких негативных социальных явлений как, например, **спам**, **троллинг**. Это алгоритм позволяет выявить участников сообщества, пользующихся его доверием. Сообщество доверяет участникам, если им доверяет ядро. «Ядро» сообщества (или «ядро доверия») формируется из нескольких предопределенных участников, обладающих наивысшим доверием.

Отношения доверия в сообществе моделируются графом, вершинами которого являются участники сообщества, а ребра строятся на основании сертификации каждым участником сообщества тех участников, которым он доверяет. В простейшем случае, доверие со стороны ядра к другим участникам можно было бы определять наличием пути на графе доверия от вершин ядра к вершине каждого из участников (используется транзитивность доверия), но алгоритм *Advogato* Trust Metric использует

более сложный подход, основанный на расчете «потока доверия» через граф. Основная идея алгоритма – найти максимальный поток через граф, после чего все те участники, вершины которых получают на вход ненулевой поток доверия, будут доверенными:

Каждой вершине исходного графа присваивается пропускная способность $capacity_i = cap(d_{s,i})$, где $d_{s,i}$ – кратчайшее расстояние между данной вершиной i и вершиной «ядра доверия» s ($cap(\cdot)$ – табличная функция). Исходный граф преобразуется к специальному виду: вводится вершина «супер-сток» («истоком» является виртуальная вершина, состоящая из всех вершин «ядра доверия»). Каждая вершина i расщепляется на две вершины $i-$ и $i+$. Все входящие ребра устанавливаются на $i-$, а исходящие ребра – на $i+$ (их пропускная способность равна бесконечно большому числу). Пропускная способность нового ребра между $i-$ и $i+$ становится равной $capacity_i - 1$; кроме того, добавляется новое ребро между $i-$ и «супер-стоком» с единичной пропускной способностью. В таком модифицированном графе используется алгоритм *Форда-Фалкерсона* для поиска максимального потока через граф. Вершины $i-$, через которые проходит поток к «супер-стоку», являются доверенными.

В *Advogato* узлы делятся на три категории: «ядро доверия» (по определению нельзя ввести в заблуждение) и «хорошие узлы» (они доверяют только хорошим узлам), «плохие» узлы (злоумышленники), пораженные (confused) узлы (те, которые доверят плохим узлам). Утверждается, что такой подход очень надежен к шуму и даже нападению на сеть доверия. Плохие узлы не получат много доверия. Однако по утверждению [32] сеть все же уязвима.

Осталось только добавить, что, вероятно, алгоритм применим не только для использования в сообществе, но и для любых агентов, вычисляющих персональные списки доверенных агентов (в этом случае сам агент составляет ядро доверия).

6.4. Математические модели, основанные на бета-функции распределения

6.4.1. Модели, использующие Байесов подход. В [28] предлагается вычислительная модель, основывающаяся на правилах Байеса и бета-функции распределения вероятности для подсчёта и распространения репутации. Апостериорные значения репутации вычисляются как комбинация априорных значений новых оценок (рейтингов). Значение репутации может быть представлено в форме математического ожидания бета-функции, заданного параметрами a и b, где a – количество положительных оценок участника, а b – количество отрицательных:

$$beta(p \mid a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\cdot\Gamma(b)} \cdot p^{a-1} \cdot (1-p)^{b-1}, \text{ где } 0 \le p \le 1,\ a,\ b > 0.$$

Математическое ожидание такого распределения будет $E(p) = \frac{a}{a+b}$. Для новичка $a = 1$ и $b = 1$, при таких параметрах бета-распределение вырождается в равномерное распределение. После получения r позитивных и s негативных оценок параметрами апостериорного распределения будут $a = r + 1$ и $b = s + 1$. Распределение отражает вероятность успешных взаимодействий с данным пользователем в будущем.

6.4.2. Модели субъективной логики. Модель доверия, названная «субъективной логикой» и представленная *Josang* в [22], сочетает в себе элементы Байесовой теории вероятностей и теории убеждений/веры (belief theory). Подход для работы с неопределенностью называется теорией убеждений/веры. В [22] мнение выражается тройкой (b, d, u), где b – убеждение/вера, d – неверие/недоверие, и u – неопределенность. Параметры связаны между собой уравнением $b + d + u = 1$. Josang предоставляет следующие уравнения:

$$b = \frac{r}{r+s+2},\ d = \frac{s}{r+s+2},\ u = \frac{2}{r+s+2}.$$

Кроме того, он определяет операторы для объединения (оператор консенсуса) и конкатенации (оператор дисконтирования/снижения значимости) мнений. Модель поддерживает также операторы конъюнкции, дизъюнкции и отрицания (алгебра высказываний).

В некоторых исследованиях показано как «субъективная логика» может быть использована для моделирования доверия в криптографических системах с открытым ключом.

6.5. Abdul-Rahman и Hailes

Модель [5] разработана для применения в онлайновых сообществах (электронная коммерция и искусственные автономные агенты). Понятие доверия основывается на субъективной вероятности. В модели обсуждается прямое доверие (доверие к агенту на основе непосредственного опыта) и доверие к свидетелю (доверие в способность свидетеля давать хорошие рекомендации).

Значения прямого доверия для агента представлены дискретными именованными уровнями: «очень надежный» (vt), «надежный» (t), «ненадежный» (u) и «очень ненадежный» (vu). Агент ведет историю взаимодействий для каждого партнера по каждому из контекстов в виде записей (c_{vt}, c_t, c_u, c_{vu}). Максимальное значение накопленного опыта в записи определяет персональный уровень доверия. Например, значение «надежный» (t) доверия агента *X* к агенту *A* в контексте с будет представлено в записи (0, **5**, 2, 1). Модель снимает возможную неопределенность с помощью таблицы ситуаций, охватывающей возможные случаи (например, случай, когда существует более чем одна позиция с максимальным значением в записи).

Интересно то, что эта модель может разрешить проблему использования разными агентами одних и тех же меток, но с разной *субъективной семантикой*. Предположим, что агент *A* помечает агента *C* меткой «Надежный», основываясь на своем непосредственном опыте, и в то же время знает, что агент *B* помечает агента *C* меткой «Очень надежный». Несовпадение между этими метками может быть вычислено как «*семантическое расстояние*». Оно может быть использовано для коррекции агентом *A* дальнейших рекомендаций *B* при расчете доверия. Однако непонятно как различать агентов, которые думают альтернативно, от агентов, которые лгут.

При объединении рекомендаций более значимыми признаются рекомендации агентов с аналогичной точкой зрения в данном контексте (такие рекомендации не требуют коррекции). Т.е. значением доверия является взвешенная сумма рекомендаций. Важно отметить то, что непосредственный опыт не используется для расчета доверия, а только для вычисления «семантического расстояния» (для коррекции свидетельской информации, как это уже было указано выше). Также неясно как модель работает с рекомендациями рекомендаций. Неясно как оценивается риск, и как осуществляется принятие решений.

6.6. Richardson, Agrawal и Domingos

В своей работе [29] исследователи утверждают, что Интернет в скором будущем станет семантическим (Semantic Web). По замыслу создателей Semantic Web приведет к появлению распределенной глобальной базы знаний, основанной на онтологиях, правилах и фактах, и в итоге к релевантному и качественному поиску информации. Однако в силу закономерного отсутствия централизованного контроля источников информации качество полученной пользователем информации может быть невысоким. Исследователи предполагают, что каждый пользователь в Semantic Web явно определяет пользователей, которым он доверяет. Такая связная сеть доверия (**web of trust**) может использоваться для рекурсивного вычисления доверия пользователя к другим пользователям сети. С помощью полученных значений доверия пользователь сможет персонально контролировать качество поступающей к нему информации (фильтрация информации на основе сети доверия).

В предложенной модели предполагается, что информация в Semantic Web представлена в форме логических утверждений. Предлагается способ вычисления веры

(belief) агента (потребителя информации) в утверждение, высказанное агентами Semantic Web (источниками информации), которое в дальнейшем может использоваться для вывода новых утверждений с помощью логического/вероятностного исчисления. Как «вера» связана с сетью доверия? Авторы исходят из того, что вера агента в утверждение должна быть функцией его доверия к источникам данного утверждения. Для вычисления доверия к неизвестному источнику используется свойство транзитивности доверия в сети: *Если A доверяет B со степенью u, а B доверяет C со степенью v, то A доверяет C с некоторой степенью доверия t, являющейся функцией f*(*u, v*). Утверждается, что модель поддерживает многие виды комбинационных функций, правда на них накладываются определенные ограничения.

Имеется *N* агентов, высказавших *M* утверждений. Каждое утверждение рассматривается отдельно.

Вера. Каждый агент i верит в утверждение со степенью $b_i \in [0,1]$. Чем надежнее, по мнению агента, утверждение, тем больше значение b_i. Множество мнений всех агентов представляется в виде вектора **b**.

Доверие. Агент *i* может определить свое персональное доверие $t_{ij} \in [0,1]$ к агенту *j.* Если j-ый агент, по мнению i-ого, надежен и предоставляет достоверную информацию, то значение t_{ij} велико. Множество персональных доверий представляется в виде матрицы **T** ($\mathbf{t}_i$ – вектор доверия i-ого агента к другим агентам сети) размерности *N* x *N*.

Обобщение. С помощью сети доверия для любого агента можно вычислить его степень *обобщенного доверия* ***T*** к агентам сети и степень *обобщенной веры* ***b*** в утверждение. Требуется найти вектор ***b*** и матрицу ***T***.

Вычисления. Обобщенное значение веры зависит от совокупности путей между данным агентом и другими агентами, имеющими персональные значения веры в рассматриваемое утверждение. Для ациклических графов обобщенное значение будет вычисляться следующим образом:

1. Перечислить все пути между данным агентом и каждым агентом с персональным значением веры в утверждение.
2. Вычислить значение веры в утверждение для каждого пути с помощью функции конкатенации доверия в пути и персонального значения веры финального узла.
3. Комбинировать полученные значения веры функцией «усреднения».

Обозначения операций: конкатенация ○ (например, умножение и минимум) и агрегация ◊ (например, сложение и максимум). Приведем пример: $t_{ik} \circ t_{kj}$ – значение доверия агента *i* к *j-ому* посредством *k-ого*. Совокупное доверие $\Diamond(\forall k: t_{ik} \circ t_{kj})$. Вводится матричная операция $\boldsymbol{C}=\boldsymbol{A} \bullet \boldsymbol{B}$ такая, что $c_{ij} = \Diamond(\forall k: a_{ik} \circ b_{kj})$.

Вычисление обобщенных значений веры. Обобщенное значение веры в утверждение для данного агента рассчитывается с помощью алгоритмов транзитивного замыкания (например, Уоршалла):

1. $\boldsymbol{b}^{(0)} = \mathbf{b}$;
2. $\boldsymbol{b}^{(n)} = \mathbf{T} \bullet \boldsymbol{b}^{(n-1)}$, или $\boldsymbol{b}_i^{(n)} = \Diamond(\forall k: t_{ik} \circ \boldsymbol{b}_k^{(n-1)})$;
3. Повторить шаг 2, если $\boldsymbol{b}^{(n)} \neq \boldsymbol{b}^{(n-1)}$.

Где $\boldsymbol{b}^{(i)}$ – значение обобщенной веры на *i*-ой итерации.

Вычисление обобщенного доверия. Если ◊ – коммутативная и ассоциативная операция, а ○ – ассоциативная и дистрибутивная относительно ◊, то можно локально комбинировать значения доверия для вычисления глобального обобщенного доверия: $\boldsymbol{T}^{(0)} = \mathbf{T}$, $\boldsymbol{T}^{(n)} = \mathbf{T} \bullet \boldsymbol{T}^{(n-1)}$, повторяя до тех пор, пока $\boldsymbol{T}^{(n)} \neq \boldsymbol{T}^{(n-1)}$ (для вычислений агент должен знать только значения обобщенного доверия соседей).

Для упрощения расчетов лучше вычислить значения обобщенного доверия и использовать их для вычисления обобщенных значений веры, поскольку, если ◊ – коммутативная и ассоциативная операция, и ○ – ассоциативная и дистрибутивная относительно ◊, то $\mathbf{T} \bullet \boldsymbol{b} = \boldsymbol{T} \bullet \mathbf{b}$. Для циклических графов все останется в силе, если операции ◊ и ○ нейтральны к циклам.

6.7. EigenTrust

Существенной проблемой для p2p-сети (сеть, в которой участники равноправны и занимаются обменом информацией) является возможность обмана со стороны её участников, выкладывающих файлы, содержимое которых не соответствует их описанию, или устанавливающих такие параметры подключения, при которых файл невозможно скачать. Алгоритм EigenTrust [23] позволяет осуществить распределенное «голосование» участников сети относительно репутации друг друга с помощью оценки успешности передачи файла.

Такая система репутации обладает следующими свойствами:

1. Саморегулируемость (этика поддерживается самими узлами сети).
2. Поддержка анонимности (узел имеет анонимный идентификатор).
3. Отсутствие преимуществ для новичков (для предотвращения влияния на сеть злонамеренных узлов).
4. Надежность (не позволяет злонамеренным коллективам узлов влиять на сеть).
5. Вычислительная разрешимость.

Каждый узел хранит историю взаимодействия с другими узлами и оценивает доверие как разность между успешными и неуспешными попытками передачи файлов, например, *i-ый* агент оценивает *j-ого* следующим образом: $s_{ij}=sat(i,j)-unsat(i,j)$, где *sat* указывает число успешных транзакций, а *unsat* – число неуспешных. Алгоритм EigenTrust основывается на предположении о транзитивности доверия к участникам: участники, не связанные непосредственно, доверяют друг другу в том случае, если между ними есть цепочка участников, доверяющих друг другу. Идея транзитивности приводит к системе, в которой глобальное доверие к участникам определяется как собственный вектор матрицы нормированных локальных значений s_{ij}.

Нормированное локальное значение c_{ij} (одинаковое как для пользователей с неизвестной репутацией, так и для пользователей с плохой репутацией) определяется как:

(1) $c_{ij}=max(s_{ij},0)/(\Sigma_j max(s_{ij},0))$.

Как происходит агрегирование мнений? Узел i опрашивает мнения знакомых (т.е. тех, для которых установлена оценка) о *k-ом* узле и оценивает доверие как $t_{ik}=\sum_j c_{ij}c_{jk}$.

Пусть *C* – матрица значений $[c_{ij}]$, а $\vec{t}_i$ – вектор значений t_{ik}, тогда $\vec{t}_i = C^T \cdot \vec{c}_i$. Произведя транзитивное замыкание (т.е. учитывая мнения знакомых наших знакомых и т.п.), получим $t = \left(C^T\right)^n \cdot c_i$ для достаточно больших *n* (удовлетворяются предположения о том, что матрица не редуцируема и апериодична). Для больших *n* $\vec{t}_i$ сходится к одному и тому же вектору для любого i-ого узла – левому собственному вектору матрицы *C*, иными словами $\vec{t}$ – глобальный вектор доверия в модели, элементы которого и являются оценкой доверия (репутацией) узлов сети.

Следовательно, вместо $\vec{c}_i$ можно взять любой другой нормированный вектор. Алгоритм расчета вектора глобальных значений доверия будет следующим:

$\vec{t}^{(0)} = \vec{e}$;
repeat
$\vec{t}^{(k+1)} = C^T \vec{t}^{(k)}$;
$d = \left\| \vec{t}^{(k+1)} - \vec{t}^{(k)} \right\|$;
until $d < e$

Удобно взять вектор априорного распределения доверия $\vec{p}$ (вместо $\vec{e}$), для которого $p_i = \frac{1}{|P|}$, если $i \in P$ (малое множество особо доверенных узлов, например, соз-

дателей сети), и $p_i = 0$ иначе. С помощью этого вектора находится вектор t при небольших значениях n.

С другой стороны, в (1) значение может быть не определено, если узел не скачивал никакую информацию, в этом случае он доверяет изначально доверенным узлам:

$$c_{ij} = \begin{cases} \dfrac{\max(s_{ij},0)}{\sum_j \max(s_{ij},0)}, если \sum_j \max(s_{ij},0) \neq 0 \\ p_j, иначе \end{cases}.$$

В сети может быть коллектив злоумышленников, оценивающих друг друга высоко, а остальных - низко. Для решения этой проблемы предлагается модифицировать вектор мнений $\vec{c}_i = (1-a)\vec{c}_i + a\vec{p}_i$, где a – константа меньшая 1 (например, 0.05), т.е. мы усиливаем доверие узла i к изначально доверенным узлам P, ослабляя тем самым доверие к злоумышленникам. Для данной задачи существует алгоритм распределенного расчета.

6.8. Sporas и Histos

Sporas и *Histos* [35] – развитые математические модели репутации (предназначенные для применения в e-commerce, а, в общем, для совершения транзакций) построены на следующих принципах:

1. Пользователь не может **мимикрироваться** (т.е. использовать чужой идентификатор), но его идентификатор может быть анонимным для сохранения приватности.
2. Значение репутации пользователя не может быть меньше значения репутации новичка (минимальное значение), в противном случае при плохой репутации у него появится стимул сменить идентификатор.
3. Злоумышленники не могут (или для них нет стимулов) объединяться и совершать «пустые транзакции», положительно оценивая друг друга и тем самым повышая собственную репутацию.
4. Злоумышленник не может создавать «пустые» идентификаторы для повышения репутации своего идентификатора (для этого вес оценки должен зависеть от репутации).
5. Значение репутации ограничено сверху, в противном случае обладатель высокой репутации может не заботиться о мнениях других участников.
6. Значение репутации является субъективным ожиданием и зависит от прошлой деятельности агента (память), значение обновляется после каждой оценки транзакции (обратная связь).
6. Вводится временное окно для расчета репутации, поскольку людям свойственно изменять поведение со временем;

В модели Sporas подсчитывается только самая последняя оценка между двумя пользователями. Еще одной важной характеристикой является то, что для пользователей с высокой репутацией ее значение после оценки изменяется намного меньше, чем для пользователей с низкой репутацией. Также внедряется мера надежности репутации пользователя, основанная на стандартном отклонении значения репутации.

Модель Histos разработана как развитие модели *Sporas.* Первая может работать как с прямой (непосредственной) информацией, так и с косвенной (свидетельской). В отличие от Sporas в данной модели рассчитывается персональное значение репутации. Влияние прямого взаимодействия в этой модели репутации ограничивается использованием самых последних оценок взаимодействия агента. Однако применимость этой модели зависит от связности графа: если граф слабосвязный, то лучше использовать модель *Sporas*.

Недостатком этой модели (как и многих) является использование значения репутации агента, установленного свидетелем (исходя из его надежности). Однако если агент является хорошим продавцом товаров, это вовсе не означает, что он может быть надежным свидетелем.

6.9. Schillo

Децентрализованная модель подсчёта репутации и доверия, предложенная *Schillo* [31], предназначена для сценариев, в которых результат взаимодействия между двумя агентами (с точки зрения доверия) представлен бинарным значением «хорошо» или «плохо».

Агенты участвуют в многоэтапной игре, на каждом этапе есть фаза выбора партнера. Каждый агент получает информацию о результатах игры, в которую он играет, а также о результатах игр, играемых некоторым подмножеством игроков (его соседей). Результатом взаимодействия является оценка честности партнера (выполнил ли партнер то, что «обещал» (утверждал) в фазе выбора или нет) и принятие решения о сотрудничестве. Модель основана на теории вероятностей. Формулой для расчета доверия того, что агент *Q* достоин агента *A* (вероятности того, что агент будет честным в следующем взаимодействии) является $p(A, Q) = e / n$, где n – общее число наблюдаемых ситуаций, а e – количество ситуаций, в которых агент был честным.

Дополнительно агент может опросить других агентов. Каждый агент использует ориентированный граф TrustNet, в котором он представлен корневой вершиной, свидетели представлены дочерними вершинами, а ребра несут информацию о наблюдениях свидетелей. Очевидно, что свидетельская информация может быть недостоверной по различным причинам. Однако исследователи предполагают, что свидетелям будет невыгодно предоставлять ложные сведения, поскольку опрашивающий агент обладает набором наблюдений (а не их единым агрегированным значением) и может легко обнаружить ложность предоставленной информации (сложно обманывать, зная то, что возможны и другие показания, и то, что тебя могут обвинить в лжесвидетельстве). То есть модель предполагает, что агенты никогда не лгут, но могут скрыть положительную информацию. Если предположить, что негативная информация будет всегда предоставлена свидетелями, проблема заключается в том, как учесть сокрытие положительной информации. Для этого сокрытие информации моделируется как случайный процесс, в котором агент решает сообщить позитивную информацию о другом агенте с вероятностью P и не сообщать с вероятностью $(1 – P)$. Таким образом, можно использовать теорию вероятностей для оценки скрытого количества положительной информации. Т.е. агент может восстановить информацию – понять, что свидетели сказали бы, если бы они были полностью честными. При этом неясно как сочетать прямой опыт со свидетельской информацией, и опять-таки значение доверия не зависит от контекста.

6.10. Yu и Singh

В модели, предложенной *Yu* и *Singh* [34], информация о прямых взаимодействиях, хранимая агентом, представляет собой набор значений, отражающих качество этих взаимодействий (качество обслуживания QoS – Quality of Service). Каждый агент определяет верхний и нижний пороги, которые определяют границы между надежным QoS, неопределенным QoS и ненадежным QoS. Затем, используя информацию об истории взаимодействий, с помощью теории свидетельств *Демпстера-Шейфера* агент может вычислить вероятность отнесения взаимодействия к каждой из этих групп. Если разность между вероятностями того, что QoS принадлежит к первой и последней группе, превышает пороговый уровень для надежности, то агент считается надежным. Свидетельская информация передается по цепочке. Для агрегирования информации от различных свидетелей используется правило комбинации *Демпстера*. Эта модель не объединяет прямую информацию со свидетельской: если есть прямая информация, то используется только она, если ее нет, то только тогда модель обращается к свидетельской информации.

6.11. Carter

Согласно *Carter* [10] репутация агента основывается на выполнении роли, предписанной ему обществом: если общество считает, что агент выполняет свою функцию, то его репутация растет, иначе – падает. Поскольку каждое общество имеет свой собственный набор ролей, то репутация имеет смысл только в контексте данного конкретного общества. Т.е. невозможен универсальный расчет репутации.

Авторы формализуют множество ролей (набор функций в пределах общества) и предлагают методы расчета степени исполнения обязательств каждой роли. Общество является множеством агентов, обменивающихся значимой информацией в ответ на запросы пользователей. Определяется пять ролей:

1. *Провайдер социальной информации*: пользователь общества должен регулярно вносить новые знания о своих знакомых в общество. Каждая конкретная рекомендация $g_u^i(t)$ (в момент времени *t* может быть *n* рекомендаций), предоставленная пользователем *u* в момент времени *t*, имеет ассоциированный с ней вес – **силу рекомендации** $w_u^i(t) = e^{-a(t-t_0)} R_u(t)$, *0<a<1,* которая зависит от времени и репутации пользователя. Степень удовлетворенности Γ данной ролью данного пользователя рассчитывается как нормированная сумма всех этих весов $W_u = \sum_{i=0}^{n} w_u^i(t)$: $\Gamma = f(W_u) = \frac{1}{1+e^{bW_u}}$, где бета - параметр.

2. *Активность*: пользователи должны регулярно использовать систему, иначе система становится бесполезной. Значение удовлетворенности рассчитывается как число операций пользователя в течение определенного периода времени, разделенное на общее количество операций, выполняемых в системе в течение того же периода.

3. *Провайдер контента*: пользователи должны предоставлять обществу свои (экспертные) знания. Значение удовлетворенности ролью отражается качеством информационных сообщений, предоставляемых пользователем. Качество информационного сообщения определяется близостью его предмета с интересами автора сообщения, т.е. пользователь, создающий информационные сообщения в своей области знаний, дает контент более высокого качества.

4. *Административная роль*: пользователи должны предоставлять обратную связь о качестве работы системы (определяется простотой в использовании, скоростью, стабильностью и качеством информации).

5. *Роль долгожителя* (longevity): необходимо, чтобы пользователи поддерживали репутацию во времени в целях содействия долговечности системы.

Таким образом, общая репутация пользователя рассчитывается как взвешенная сумма степени выполнения каждой его роли. Значение веса зависит от конкретного общества. Значение репутации для каждого агента рассчитывается централизованным механизмом, контролирующим систему. Это значение является глобальным и используемым всеми агентами.

6.12. ReGreT

Модель доверия и репутации ReGreT [30] предназначена для использования в системах электронной торговли, в которых социальные отношения между индивидами играют важную роль. ReGreT учитывает не только прямую и свидетельскую информацию, но и информацию о социальных отношениях (конкуренция, сотрудничество и торговля) взаимодействующих сторон. Для расчета доверия, прежде всего, берутся результаты непосредственных взаимодействий (т.е. сравнивается обещанная в контракте и итоговая цена, дата поставки, качество и т.п.). Если же взаимодействия в прошлом отсутствовали или рассчитанное значение доверия ненадежно, то для расчета доверия используется репутация целевого агента. Модель репутации включает три специализированных вида репутации в зависимости от источника информации:

- *Репутация от свидетелей*, которая основывается на информации (например, мнениях) о целевом агенте, поступающей от свидетелей (которые взаимодействовали в прошлом с целевым агентом). Проблемы: свидетели могут лгать, скрывать информацию или свидетели могут высказывать свое мнение на основе только одного и того же события или свидетели могут испытывать влияние друг друга. Поэтому выбор свидетелей происходит следующим образом: находится множество всех свидетелей; на этом множестве строится граф отношений (из социограммы); выявляются максимально связные компоненты графа; для каждого компонента находятся «мосты» или, если их нет, узлы с максимальной степенью (смысл в том, чтобы выбрать репрезентативных представителей каждой группы). Далее запрашивается информация относительно репутации целевого агента у выбранных узлов (представителей). Доверие к представленной информации относительно целевого агента зависит от социальных отношений между свидетелем и целевым агентом, например, для расчета доверия используются следующие нечеткие множества и нечеткое правило: ЕСЛИ Сотрудничество(свидетель, целевой агент) - *высокое*, ТО Доверие(данный агент, свидетель, целевой агент) – *слабое*.

- *Репутация от соседей*, которая основывается на информации о соседях целевого агента и социальных отношениях между целевым агентом и его соседями. Например, если у соседа целевого агента репутация мошенника, и они находятся в отношениях кооперации, то вероятно, что целевой агент тоже мошенник.

- *Системная репутация*, которая основывается на информации об институциональных структурах и о роли, играемой целевым агентом в соответствующей институциональной структуре. Например, если i-ый агент принадлежит к компании A, в которой компания B пользуется плохой репутацией, то для i-ого агента j-ий агент из B будет по умолчанию обладать плохой репутацией.

Также учитывается онтологическая структура доверия и репутации, обеспечивающая необходимую информацию для расчета значений сложных аспектов доверия в зависимости от простых.

Система ReGreT является модульной. Примеры модулей: модуль прямого доверия, использующий результаты непосредственных оценок для расчета доверия; модуль, оценивающий надежность свидетелей и надежность свидетельской информации. Совместная работа модулей позволяет пользователю рассчитывать доверие по наиболее полной модели доверия. Вместе с тем модульный подход к разработке системы позволяет пользователю определить следует ли ему использовать те или иные модули для расчета доверия к другому пользователю. Еще одним преимуществом модульного подхода является то, что пользователь может работать с системой, даже не имея достаточного количества информации. При накоплении пользователем информации о других членах сообщества и социальных отношениях между ними система начинает подключать дополнительные модули для повышения точности расчета значений доверия и репутации.

В ReGreT каждое значение доверия и репутации имеет связанное значение меры надежности. С помощью этой меры агент может принять решение о целесообразности использования рассчитанного значения доверия и репутации. Нужно отметить, что в данной работе не оценивается риск, не рассматривается подробно принятие решения пользователем.

6.13. Golbeck

Golbeck [21] представляет децентрализованную модель расчета доверия в социальных сетях и предлагает использовать онтологии для явного представления социальных отношений между участниками сети.

Расчет доверия. *Golbeck* представляет социальную сеть в виде графа, вершиной которого является пользователь сети, а отношения доверия между пользователями моделируются направленными и взвешенными ребрами. Значение доверия к неизвестному пользователю рассчитывается исходя из того, насколько доверяют ему друзья и друзья друзей. При этом используются такие предположения:

1. Значения доверия по любому пути не превосходит значения любого ребра в пути.
2. Длина пути должна сказываться на значении рассчитываемого доверия, поскольку люди верят больше своим непосредственным друзьям.
3. Недоверие не является транзитивным. Например, если А не доверяет Б по некоторому вопросу, а Б не доверяет С, то это не значит, что А не доверяет С (скорее всего даже доверяет). Следовательно, нужно предоставить возможность пользователям определить свои метрики доверия.

Для расчета доверия используется рекурсивный алгоритм расчета по всем достижимым путям: $t_{is} = \frac{\sum_{j=0}^{n} \left\{ \begin{matrix} (t_{js} \cdot t_{ij}) \; если \; t_{ij} \geq t_{js} \\ (t_{ij}^2) \; если \; t_{ij} < t_{js} \end{matrix} \right\}}{\sum_{j=0}^{n} t_{ij}}$ (учитывается то, что мы не можем доверять кому-то больше, чем посреднику). Согласно законам Small World средняя длина путей в графе социальной сети увеличивается логарифмически, поэтому данный алгоритм хорошо масштабируется.

Онтологии. *Golbeck* [21] предлагает явно представить отношения доверия с помощью расширения понятий и отношений онтологии FOAF. Доверие задается отношениями trustHighly, ..., trustLow, …, distrustHighly (по 9-бальной шкале от абсолютного недоверия до абсолютного доверия), т.е. пользователи (Person) в социальной сети могут явно указать то, насколько они доверяют своим друзьям в общем (Джо очень сильно доверяет Сью):

<Person rdf:ID=«Joe»>
<mbox rdf:resource=«mailto:bob@example.com»/>
<trustsHighly rdf:resource=«#Sue»/>
</Person>

И указать то, насколько они доверяют им в какой-либо области (Боб доверяет Дэну в области вычислительной техники, но не в вопросах маркетинга):

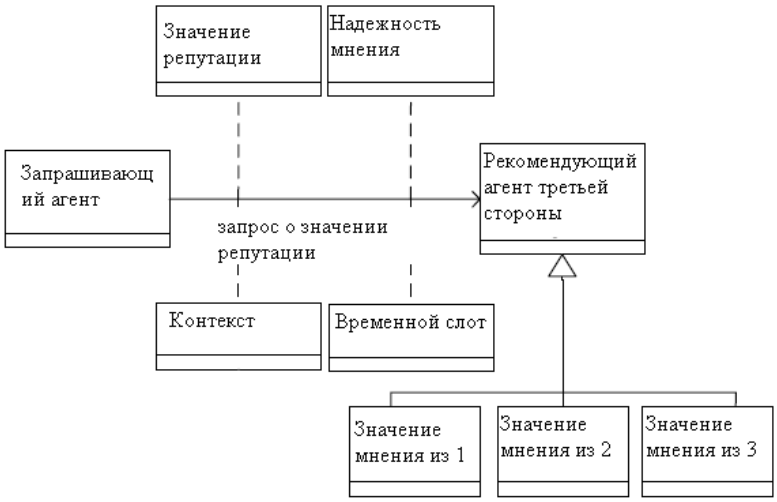
<Person rdf:ID=«Bob»> Персона Bob
<mbox rdf:resource=«mailto:joe@example.com»/>
<trustsHighlyRe> отношения высокого доверия
<TrustsRegarding> доверие
<trustsPerson rdf:resource=«#Dan»/> к персоне Dan
<trustsOnSubject
rdf:resource=«http://example.com/ont#ComputerScience «/> вычислительная техника
</TrustsRegarding>
</trustsHighlyRe>
<distrustsAbsolutelyRe> отношения абсолютного недоверия
<TrustsRegarding> доверие
<trustsPerson rdf:resource=«#Dan»/> к персоне Dan
<trustsOnSubject
rdf:resource=«http://example.com/ont#Marketing»/> в маркетинге
</TrustsRegarding>
</distrustsAbsolutelyRe>
</Person>

Golbeck также предлагает веб-сервис для расчета доверия между любыми пользователями сети, воспользоваться услугами которого могут интеллектуальные агенты.

6.14. Chang

Chang [13] заявляет об исключительной важности в XXI веке репутационных механизмов в распределенных веб-системах и предлагает рассмотреть сеть, в которой существуют агенты двух типов: клиент и агент, предоставляющий товар или услуги

(причем агентом может быть как человек, так и целая коммерческая организация). Такие системы сделают прозрачной оценку качества продуктов и услуг для обеспечения гарантий сторон и исключения мошенничества, а также предоставят средства для создания отношений доверия с клиентами с помощью оценивания мнений клиентов, изучения реакции рынка на товары и услуги. Для таких систем *Chang* предлагает онтологию репутации (онтология запроса), в которой представлены следующие базовые сущности: *Агент, запрашивающий информацию о значении репутации*, *Агент третьей стороны*, *Значение репутации*, *Контекст*, *Временной слот*, *Мнение (из первых рук, из «вторых рук», из «третьих рук»)* и отношение *Запрос значения репутации*. Значение репутации рассчитывается как U(Значение рекомендации*Значение надежности*{Значение мнения из первых рук, Значение мнения из вторых рук, Значение мнения из третьих рук}*Фактор времени).



Дальше последовательно рассматривается расширение онтологии репутации, в которое добавляются дочернее понятие *Аспект агента* для понятия *Контекст* и дочернее понятие *Критерий* понятия *Аспект*.

Наиболее критичным является расчет достоверности мнения (сам метод расчета *Chang* не предлагает), поэтому Chang предлагает онтологию мнений, в которой представлены понятия: *Получатель*, *Обозреватель*, *Обратная связь*, *Критерий оценки*, *Временной слот*, *Достоверность критерия оценки*.

7. Некоторые примеры действующих системы доверия и репутации

Системы электронной торговли. Интернет-магазин *eBay*, *Amazon Auctions* и OnSale Exchange являются хорошими примерами виртуальных рынков, использующих механизмы репутации. Все эти модели рассматривают репутацию как глобальное свойство и используют одно значение, не зависящее от контекста. Источником свидетельской информации для вычисления репутации являются агенты, которые ранее взаимодействовали с целевым агентом. Эти системы не предоставляют четкие механизмы борьбы с пользователями, предоставляющими ложную информацию. Единственный путь к повышению надежности значения репутации – большое количество мнений, «размывающее» ложную или предвзятую информацию.

eBay является одним из крупнейших в мире Интернет-магазинов с сообществом, насчитывающим десятки миллионов зарегистрированных пользователей. Основа применяемого механизма – отзывы/оценки, которые пользователи (продавец и покупатель) оставляют после завершения сделки. Пользователь может задать три возможных значения: позитивное (1), отрицательное (-1) или нейтральное (0). Значение репутации вычисляется как сумма этих оценок в течение последних шести месяцев.

Другими примерами являются **сайты экспертов, сайты обзоров продуктов** (*Epinions, BizRate, Amazon* и др.), **дискуссионные тематические форумы** (*Slashdot, Kuro5in, Хабрахабр* и др.).

Литература

1. Глоссарий. http://glossary.ru
2. Голован С.В. Эффект забывания в теории коллективной репутации. – М.: Российская экономическая школа, 1999. – 38 с.
3. Ермаков Н.С., Иващенко А.А., Новиков Д.А. Модели репутации и норм деятельности. – М.: ИПУ РАН, 2005. – 67 с.
4. Новиков Д.А., Чхартишвили А.Г. Прикладные модели информационного управления. М.: ИПУ РАН, 2004. – 130 с.
5. Abdul-Rahman A., Hailes S. Supporting trust in virtual communities // In: Proc. of Hawaii International Conference on System Sciences. 2000.
6. Advogato Trust Metric. http://www.advogato.org/trust-metric.html
7. Akerlof G. A. The market for «lemons»: Quality uncertainty and the market // The Quarterly Journal of Economics. 1970. Vol. 84. No. 3. P. 488-500.
8. Artz D., Gil Y. A Survey of Trust in Computer Science and the Semantic Web // Web Semantics. 2007. Vol. 5. No. 2. P. 58-71.
9. Beaufils B., Branouy O. Reputation games and the dynamics of exchange network. Lille: University of Science and Technology, 2004 (forthcoming). – 22 p.
10. Carter J. Reputation Formalization for an Information-Sharing Multi-Agent Sytem // Computational Intelligence. Vol. 18 (2). P. 515-534.
11. Castelfranchi C., Falcone R. Principles of Trust for MAS: Cognitive Anatomy, Social Importance, and Quantification // Proceedings of the International Conference on Multi-Agent Systems. 1998. P. 72-79.
12. Castelfranchi C., Falcone R. Trust is much more than subjective probability: Mental components and sources of trust. 2000. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.105.8007
13. Chang E., Hussain F.K., Dillon T. Reputation Ontology for Reputation Systems. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.93.9031
14. Dellarocas C., Resnick P. Online Reputation Mechanisms A Roadmap for Future Research. 2003. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.15.514
15. Ding Li. Modeling and Evaluating Trust Network Inference // Seventh International Workshop on Trust in Agent Societies at AAMAS 2004. 2004.
16. Fudenberg D., Kreps D. Reputation in the simultaneous play of multiple opponents // Review of Economic Studies. 1987. N 4. P. 541 – 568.
17. Fudenberg D., Levine D. Reputation and equilibrium selection in games with a single patient player // Econometrica. 1989. Vol. 57. P. 251 – 268.
18. Fudenberg D., Tirole J. Sequential bargaining with incomplete information // Review of Economic Studies. 1983. Vol. 50. N 2. P. 221 – 247.
19. Gibb J.R. Trust: A New View of Personal and Organizational Development. Guild of Tutors Press, 1978. 320 p.
20. Golbeck J. Computing and Applying Trust in Web-Based Social Networks. University of Maryland at College Park, 2005. 199 p.
21. Golbeck J., Parsia B., Hendler J. Trust Networks on the Semantic Web // Cooperative Information Agents VII. 2003. P. 238-249.
22. Jøsang A., Ismail R., Boyd. C. A Survey of Trust and Reputation Systems for Online Service Provision // Decision Support Systems. 2007. Vol. 43. P.618-644.
23. Kamvar S.D., Schlosser M.T., Garcia Molina H.. The EigenTrust Algorithm for Reputation Management in P2P Networks // Proceedings of the 12th international conference on World Wide Web. 2003. P. 640-651.
24. Kreps D., Wilson R. Reputation and imperfect information // Journal of Economic Theory. 1982. Vol. 27. P. 253 – 279.
25. Marsh S. Formalising Trust as a Computational Concept. 1994. Ph.D. dissertation, University of Stirling.
26. Marsh S. Formalising Trust as a Computational Concept. 1994. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.47.6243

27. McKnight D.H., Chervany N.L. The Meanings of Trust. 1996. http://www.misrc.umn.edu/workingpapers/fullPapers/1996/9604_040100.pdf
28. Mui L., Mohtashemi M., Halberstadt A. A computational model of trust and reputation // System Sciences. 2002. P. 2431-2439.
29. Richardson M., Agrawal R., Domingos P. Trust management for the semantic web // International Semantic Web Conference. 2003. P. 351-368.
30. Sabater J., Sierra C. Reputation and social network analysis in multi-agent systems // Proceedings of the first international joint conference on Autonomous agents and multiagent systems. 2002. P. 475-482.
31. Schillo M., Funk P., Rovatsos M. Using trust for detecting deceitful agents in artificial societies // Applied Artificial Intelligence, 14. 2000. P. 825-848.
32. The Advogato trust metric is not attack-resistant. http://www.squarefree.com/2005/05/26/advogato/
33. Tirole J. A theory of collective reputation (with applications to the persistence of corruption and to firm quality) // Review of Economic Studies. 1996. Vol. 63. P. 1 – 22.
34. Yu B., Singh, M. P. An evidential model of distributed reputation management // Proceedings of the first international joint conference on Autonomous agents and multiagent systems. 2002. P. 294-301.
35. Zacharia G. Trust management through reputation mechanisms // Applied Artificial Intelligence. 2000. Vol. 14. P. 881-907.