

Применение марковской модели для анализа влияния участников интернет-сообществ

Федянин Денис Николаевич
dfedyanin@inbox.ru

Институт проблем управления им. В. А. Трапезникова РАН

Содержание работы

В работе исследуется взаимное влияние участников социальной сети (агентов) друг на друга в рамках так называемой марковской модели. Основной задачей проводимого исследования являлась проверка гипотез о зависимости вычисляемой влиятельности агентов от используемых методов идентификации модели, а также непротиворечивость результатов получаемых в результате использования различных методов. В исследовании используются данные только о связях между агентами, но игнорируются данные о написанных ими друг другу сообщениях.

Используемое определение термина влияние

Под влиянием понимается процесс и результат изменения субъектом (субъектом влияния) поведения другого субъекта (индивидуального или коллективного объекта влияния), его установок, намерений, представлений и оценок (а также основывающихся на них действий) в ходе взаимодействия с ним. Как показывают наблюдения психологов в социальной сети агенты часто не имеют достаточной для принятия решения информации или не могут самостоятельно обработать её, поэтому их решения могут основываться на наблюдаемых ими решениях и/или представлениях других агентов (социальное влияние).

Об используемых для анализа данных

- Анализ был проведен по трем сообществам Живого Журнала. Сайт Живого Журнала (<http://www.livejournal.com>) состоит из блогов – последовательностей сообщений, называемых постами.
- Википедия [<http://ru.wikipedia.org>] дает следующее определение “Блог (англ. blog, от web log — интернет-журнал событий, интернет-дневник, онлайн-дневник) — веб-сайт, основное содержимое которого — регулярно добавляемые записи (посты), содержащие текст, изображения или мультимедиа. Хранение данных осуществлялось в трех таблицах: список сообществ, список блогеров и список связей между блогерами. Количество участников в сообществах: 964, 2960, 6587.

Формат хранения данных об агентах

Идентификатор связи между агентами	Идентификатор читающего агента	Идентификатор агента, блог которого читает	Идентификатор сообщества, к которому принадлежат оба агента
1	1	2	1
2	3	2	1
3	4	1	2

Особенности блогов

- Для блогов характерны недлинные записи временной значимости, отсортированные в обратном хронологическом порядке (последняя запись сверху). Отличия блога от традиционного дневника обуславливаются средой: блоги обычно публичны и предполагают сторонних читателей, которые могут вступить в публичную полемику с автором (в комментарии к блогозаписи или своих блогах)”. Авторы таких постов называются блогерами. Большинство постов доступно для чтения и комментирования другим блогерам. Живой Журнал также предоставляет возможность блогерам объединяться в сообщества и подписываться на чтение блогов. Блогер может состоять в нескольких сообществах одновременно.

Краткий обзор известных моделей

- В литературе существует несколько подходов к описанию взаимодействия участников в социальной сети: марковская модель (или модель Де Гроота), модель с порогом (Liner Threshold Model), модель независимых каскадов (Independent Cascade Model), модель просачивания и заражения, модель Изинга, модель клеточных автоматов и другие. Модели исследованы с точки зрения различных аспектов, в том числе учитывающих: условия сходимости мнений членов социальной сети), динамическое изменение влияния, скорость сходимости, условие единственности итогового мнения.

Краткий обзор известных методов ранжирования агентов

- Методы ранжирования агентов включают в себя индексы влияния: индекс Хёде-Баккера (Houde-Bakker), индекс Банцафа, вычисление импакт-фактора научных журналов, ранжирование страниц в интернете алгоритмами PageRank , а также упорядочение по параметрам "промежуточность" , "центральность" , "коэффициент кластеризации" и другие.

Подробнее о марковской модели

- Суть модели состоит в представлении взаимодействия агентов в виде последовательности линейных преобразований в пространстве мнений этих агентов, где мнение агентов в сети представляется в виде вектора, в котором i -ая координата – мнение i -го агента.
- Матрица линейного преобразования называется матрицей прямого влияния, а матрица линейного преобразования, являющегося результатом бесконечного применения таких линейных преобразований, называется матрицей результирующего влияния. Для того чтобы построенная модель являлась марковской необходимо, чтобы матрица прямого влияния была стохастической.

Формула влиятельности

Влиятельностью называется сумма элементов транзитивного замыкания матрицы прямого влияния по столбцам

$$w_j = \sum_i a_{ij}^{\infty}$$

Можно ее вычислить также для исходной стохастической матрицы прямого влияния. Для того чтобы различать эти влиятельности, будем называть влиятельность определенную по матрице прямого влияния **прямой влиятельностью**, а определенную по матрице транзитивного замыкания **результатирующей влиятельностью**.

Простейший метод идентификации

- Обычный метод идентификации агентов заключается в том, что матрица прямого влияния формируется из матрицы смежности по формуле

$$a_{ij} = \frac{b_{ij}}{\sum_i b_{ij}}$$

- где a_{ij} – веса в матрице прямого влияния, b_{ij} – элементы матрицы смежности

Модификации метода

- В некоторых случаях можно попытаться учесть зависимость влияния от авторитетности агента. Можно рассмотреть случаи, когда авторитетность линейно или нелинейно пропорциональна количеству друзей агента, то есть

$$a_{ij} = \frac{f_j(b_{ij})}{\sum_i f_j(b_{ij})}; f_j(x) = (\sum_i b_{ij})^\beta$$

- где f_j имеет смысл авторитетности i -го агента.

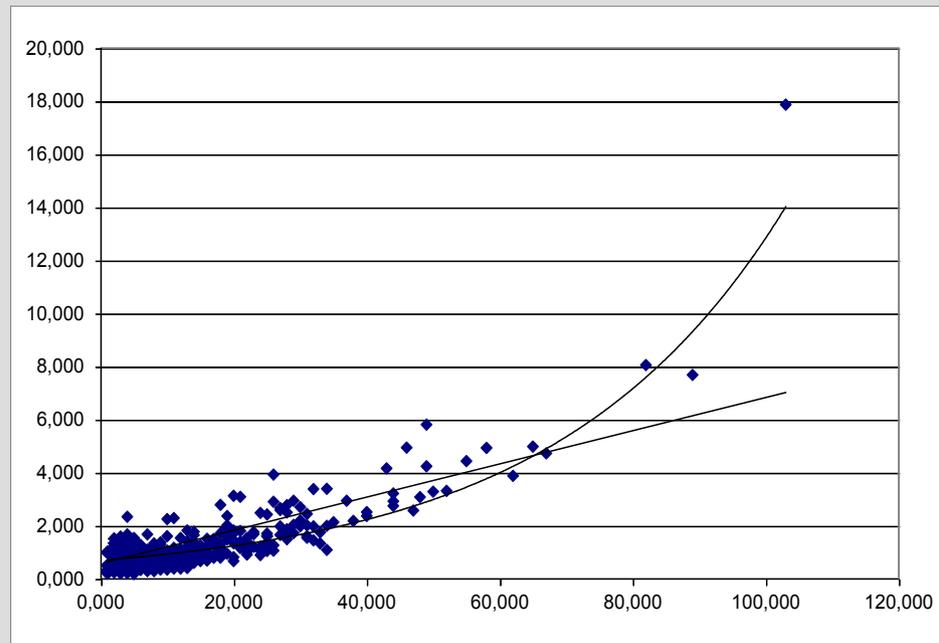
Исследуемые гипотезы

- **Гипотеза 1.** Прямая влиятельность зависит от количества друзей агента, и зависимость между ними близка к некоторой степенной функции.
- **Гипотеза 2.** Количество друзей не коррелирует с результирующей влиятельностью оценки влиятельности.
- **Гипотеза 3.** Есть корреляция между результирующими влиятелями агентов, построенных методами с различным учетом авторитетности агентов.
- **Гипотеза 4.** Прямая влиятельность агента не коррелирует с его результирующей влиятельностью.

Результаты анализа сети.

Гипотеза 1: прямая влиятельность зависит от количества друзей

- Численное значение линейной корреляции равно 0,75.

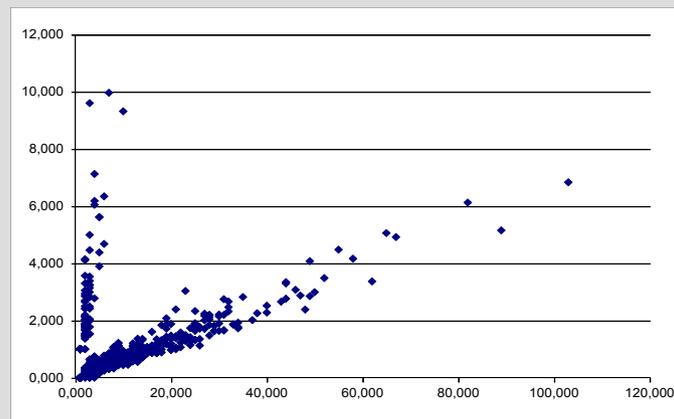


- Прямая влиятельность отложена по вертикальной оси, количество друзей — по горизонтальной

Результаты анализа сети.

Гипотеза 2: количество друзей не коррелирует с результирующей влиятельностью оценки влиятельности.

- Обращает на себя внимание почти вертикальный “хвост”, его наличие означает, что существует некоторое количество агентов, обладающих небольшим количеством друзей, однако оказывающим существенное влияние. Численное значение корреляции равно 0,45.

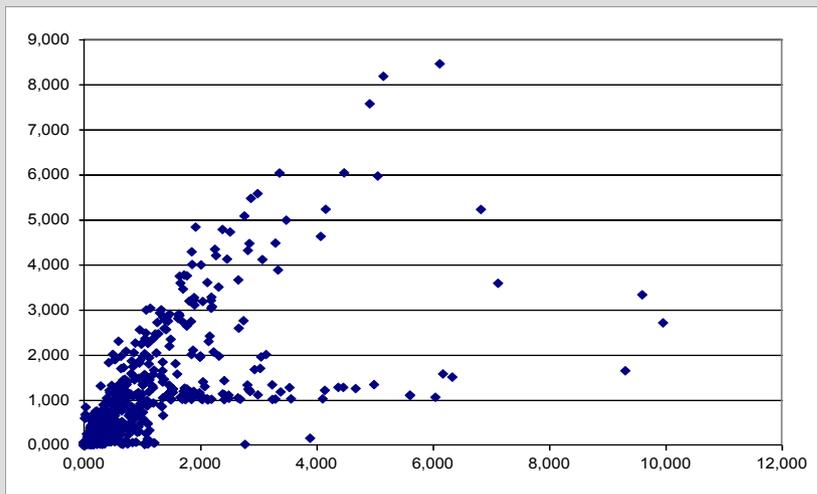


Влиятельность отложена по вертикальной оси, количество друзей — по горизонтальной

Результаты анализа сети.

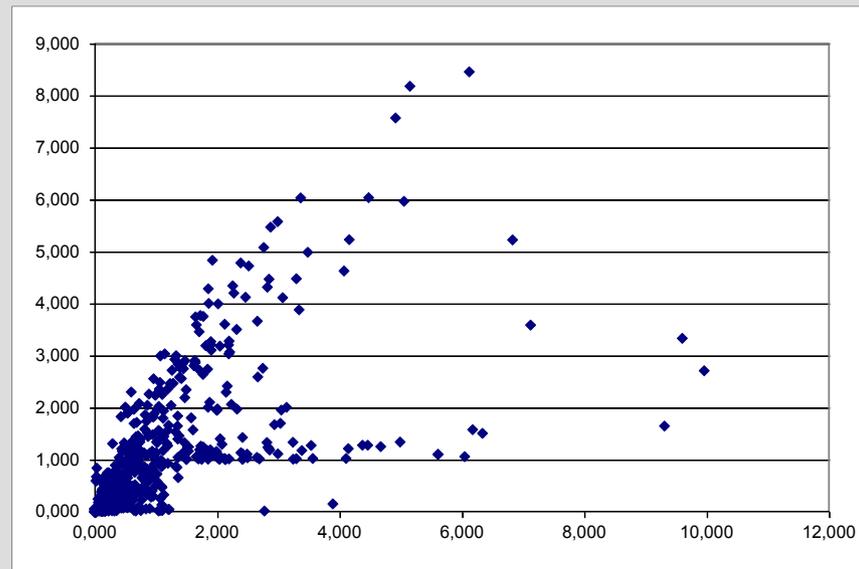
Гипотеза 3: наличие корреляции между результирующими влиятельностью агентов, построенных методами с различным учетом авторитетности агентов. 1-ый слайд

- Численное значение линейной корреляции равно 0,64.
- Зависимость транзитивной влиятельности агента, вычисленной без учета его авторитетности (по горизонтали) от транзитивной влиятельности, вычисленной с учетом влиятельности ($\beta=4$) (по вертикали).



Результаты анализа сети.

Гипотеза 3: наличие корреляции между результирующими влиятельностью агентов, построенных методами различным учетом авторитетности агентов 2-ой слайд.

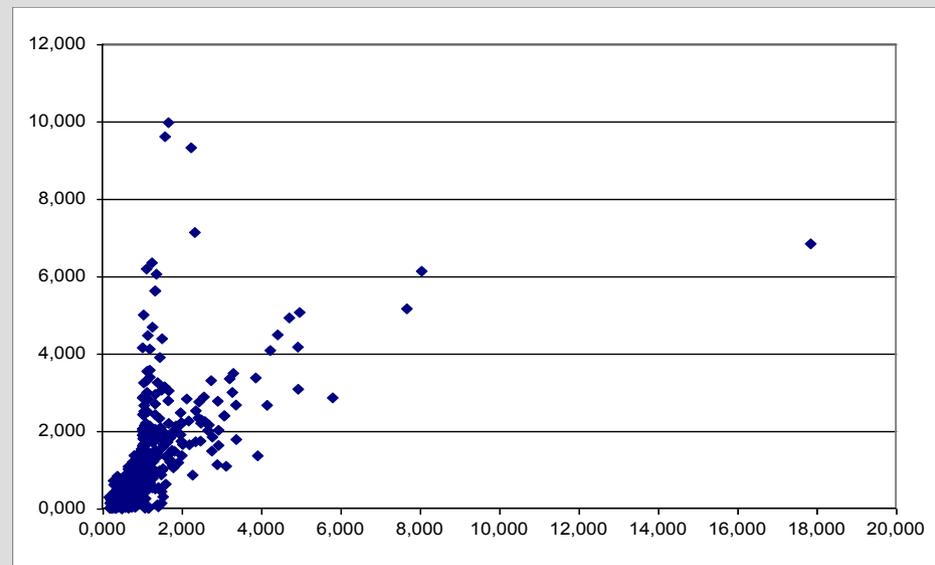


- На рисунке видно, что корреляция отсутствует. Это факт является важным в силу того, что делая предположения о значимости количества друзей агента на его авторитетности, можно получить, вообще говоря, заметно отличающиеся результаты.

Результаты анализа сети.

Гипотеза 4: прямая влиятельность агента не коррелирует с его результирующей влиятельностью. 1-ый слайд

- Зависимость результирующей влиятельности агента (по вертикали) от его прямой влиятельности (по горизонтали), вычисленных без учета авторитетности агентов. Численное значение линейной корреляции равно 0,60.



Результаты анализа сети.

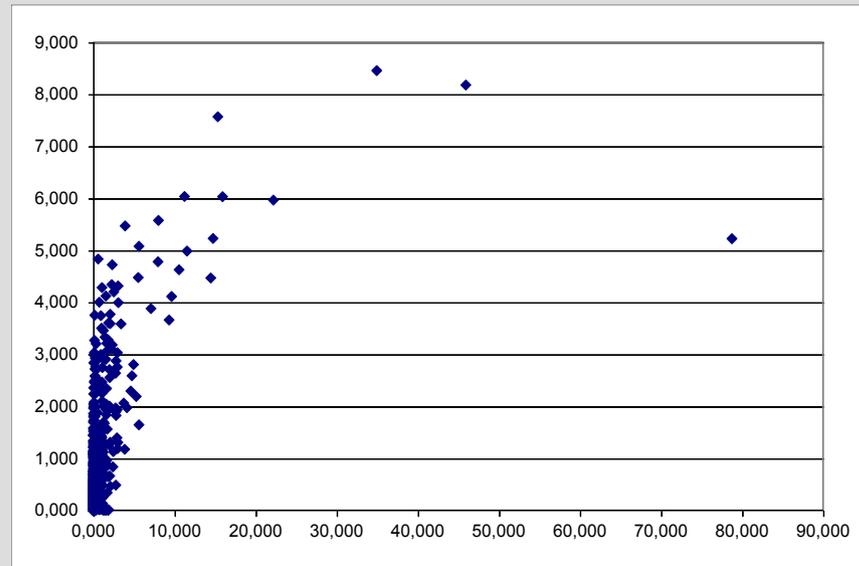
Гипотеза 4: прямая влиятельность агента не коррелирует с его результирующей влиятельностью. 2-ой слайд

На рисунке видно, что линейная корреляция между прямой и результирующей влиятелями отсутствует. Это имеет также и содержательное применение, так как, если мы полагаем, что не все агенты могут принимать решения, основанные на вычислении результирующей влиятельности, и поэтому вынуждены использовать для этого прямую влиятельность.

Тогда мы приходим к очевидному выводу, что в реально существующих сетях такие агенты будут находиться в заблуждении, однако, важно подчеркнуть, что они будут ошибаться лишь в отношении меньшего количества агентов. Более чем в 50% случаев они будут правы.

Результаты анализа сети.

Гипотеза 4: прямая влияние агента не коррелирует с его результирующей влиятельностью. 3-ий слайд



- Зависимость результирующей влиятельности агента (по вертикали) от его прямой влиятельности (по горизонтали), вычисленных с учетом авторитетности агентов. Численное значение линейной корреляции равно 0,56

Выводы

- Исследование показало наличие некоторого количества аномалий и эффектов, которые необходимо учитывать при планировании идентификации марковских моделей по экспериментальным данным.
- Выяснилось, что авторитетность агентов оказывает существенное влияние на результат вычисления влияния агентов.
- Оказывается, что существует специфическая зависимость результирующей влияния от прямой влияния.
- Подтвердилось, что существует аномальный кластер агентов, обладающих небольшим количеством друзей, но имеющих большое значение результирующей влияния.

Перспективные направления дальнейших исследований

- Интересным может быть включение в анализ возможность учета обмена сообщениями между агентами.
- Представляется также перспективным построить корреляционные зависимости между ранжированием агентов методом вычисления пока-зателя альфа-центральность, ранжированием алгоритмом PageRank, а также другими популярными и широко используемыми методами и прямыми и результирующими влиятельностью агентов.

Литература

- DeGroot M.H. Reaching a Consensus // Journal of American Statistical Association. **1974**. №69, P.118-121.
- Jackson M. Social and Economic Networks. — Princeton: Princeton Uni-versity Press, **2008**.
- Губанов Д.А., Новиков Д.А., Чхартишвили А.Г. Социальные сети: модели информационного влияния, управления и противоборства. - М.: Физматлит, **2010**. - 228 с.
- Зуев А.С., Федянин Д.Н. Модели управления мнениями агентов в социальных сетях / Проблемы управления. № 1. М.: ИПУ РАН, **2011**. С. 37-45.