

УДК 621.3 : 658.5.

О.Н. Поляруш, А.В. Витько

Харьковский национальный университет радиоэлектроники, Харьков

ПРИМЕНЕНИЕ БАЙЕСОВСКИХ ВЕРОЯТНОСНЫХ СЕТЕЙ ДЛЯ АНАЛИЗА СОЦИАЛЬНЫХ СЕТЕЙ

В данной статье проводится анализ основных проблем стандартных методов исследования социальных сетей; рассматриваются преимущества байесовских сетей и реальные примеры их использования при анализе социальных сетей. Также описываются наиболее общие классы аналитической информации, которую можно получить на основе исследований, и подаются новые варианты применения байесовских сетей в рамках социальных исследований.

Ключевые слова: анализ данных, байесовские вероятностные сети, социальные сети.

Введение

В настоящее время распространенность и популярность социальных сетей очень высока, и они уже успели накопить в себе гигабайты информации. Но эта информация практически никак автоматически не обрабатывается, знания из нее не извлекаются и не используются. Существует множество социальных сетей с большим объемом данных (Facebook, Одноклассники, Вконтакте, Orkut, LinkedIn и др.), которые представляют собой большой интерес для аналитиков. Анализ таких социальных сетей представляет собой трудоемкую задачу и основывается на изучении связей между людьми, неявных данных, которые могут быть обнаружены в ходе исследований, а также на информации, которая известна о данных людях.

События, происходящие в социальных сетях, редко могут быть описаны как прямые следствия строго детерминированных причин. На практике в таких случаях широко применяется вероятностное описание явлений [1]. Обоснований тому несколько: и наличие неустранимых погрешностей в ходе сбора данных, и невозможность полного описания структурных сложностей изучаемой системы, и неопределенности вследствие конечности объема наблюдений.

Однако, на пути вероятностного моделирования встречаются определенные сложности, которые можно условно разделить на две группы:

- технические (вычислительная сложность, “комбинаторные взрывы” и т.п.);
- идейные (наличие неопределенности, сложности при постановке задачи в терминах вероятностей, недостаточность статистического материала).

Данные проблемы можно решить с применением вероятностных байесовских сетей, которые представляют собой графовые модели причинно-следственных отношений. В байесовских вероятностных сетях могут органически сочетаться эмпири-

ческие частоты появления различных значений переменных, субъективные оценки “ожиданий” и теоретические представления о математических вероятностях тех или иных следствий из априорной информации. Это является важным практическим преимуществом и отличает байесовские сети от других методик информационного моделирования [2].

В данной статье рассматриваются примеры использования байесовских вероятностных сетей при анализе социальных сетей, а также особенности их применения.

Целевой аудиторией данной статьи являются аналитики социальных сетей, задача которых состоит в извлечении наиболее ценной информации из большого объема данных.

В разделе 1 рассматриваются основные проблемы и ограничения стандартных методов, которые используются для анализа социальных сетей; описываются трудности при сборе и анализе исходных данных, а также проблемы построения социальных сетей. В разделе 2 рассматриваются варианты потенциального применения байесовских вероятностных сетей для анализа социальных сетей; описываются классы задач, которые могут быть решены при помощи такого подхода.

1. Основные проблемы анализа социальных сетей

Несмотря на то, что стандартные методы анализа социальных сетей могут быть успешно использованы для понимания социальных сетей, они ограничены в силу нескольких причин. Прежде всего, такие методы подразумевают наличие хорошо организованной социальной сети. Более того, реальные методы сбора информации не могут гарантировать того, что социальная сеть, полученная в результате, будет содержать все необходимые данные.

Стандартные методы анализа социальной сети сосредоточены, прежде всего, на существовании

связей между узлами сети, но не на свойствах/атрибутах этих отношений или узлов, которые позволяют отображать смысл исследуемых данных. Более того, неопределенность атрибутов в узлах или связях не рассматривается такими методами, тогда как в реальных социальных сетях это встречается довольно часто.

Наконец, алгоритмы теории графов, которые используются стандартными методами анализа, имеют тенденцию к фокусировке на однородной выборке сущностей и связей, что делает сложным проведение анализа сети, которая включает разнородные выборки узлов, связанных различными типами связей.

1.1. Работа с исходными данными. Традиционные методы анализа зависят от информации об исследуемой социальной сети, которая задается с некоторым уровнем определенности. Однако существует множество источников неопределенности, которые встречаются при сборе информации. Знание источников этой неопределенности, вместе с методами, которые способны работать с неопределенностью, могут повысить качество производимого анализа социальной сети.

Традиционные методы сбора данных обычно основываются на различных опросах, позволяющих определить те элементы сети, которые являются наиболее важными для решения поставленной задачи. Однако, нюансы традиционных методов, такие как контекст, используемый вопросниками, могут приводить аналитиков в замешательство.

Еще одной проблемой является объективность взглядов, так как каждый человек имеет собственное восприятие своей социальной сети. Так, к примеру, человеку свойственно воспринимать себя как центр сети, и, несмотря на то, что существуют методы устранения этих эгоистических склонностей, это делает сложным проведение объективного анализа сети. Множество экспериментов также показывают, что человек может отобразить свою социальную сеть со средней долей точности, а восприятие социальной сети меняется с течением времени. Существует также вероятность измерительных ошибок, а также вероятность того, что люди могут быть не совсем искренним при проведении исследований.

Поскольку социальная сеть предоставляет информацию не только об одном человеке, обработка таких данных становится достаточно интересной для аналитиков. Вместе с тем, сбор данных, достаточно полных для предоставления выводов, требует значительных усилий. Это приводит к экспоненциальному увеличению задач по идентификации и классификации объектов сети.

Для тех методов анализа, которые базируются на наблюдениях за электронными показателями, появление таких задач становится сложной пробле-

мой. К примеру, в одном из известных экспериментов социальная сеть отношений и зависимостей между виндсёрферами строилась на основе реальных наблюдений. Исследования производились на пляже в одно и то же время каждый день на протяжении 31 дня [3]. Следовательно, вполне возможно, что наблюдатели могли попасть под влияние временной зависимости.

Также очевидно, что многочисленное количество существующих методов по сбору данных для анализа социальных сетей отображает сложность получения релевантных данных. Более того, в реальных социальных сетях невозможно избежать неопределенности в анализируемой информации. Неопределенность информации о связях между объектами социальной сети ограничивает количество применяемых методов. Именно поэтому для повышения качества проводимого анализа следует рассматривать методы, которые способны обрабатывать такого рода информацию.

Байесовские сети используются в условиях неопределенности, когда суть приобретаемого знания состоит в понимании, влияет ли полученная информация на наши ожидания относительно других событий [2].

Исчисление вероятностей формально не требует, чтобы использованные вероятности базировались на теоретических выводах или представляли собой пределы эмпирических частот. Числовые значения в байесовских сетях могут быть также и субъективными, личностными оценками ожиданий экспертов по поводу возможности осуществления событий.

Априорные вероятности могут также назначаться экспертами данной предметной области. Если необходим учет мнения экспертов о возможности наступления события, к которому неприменимо понятие повторяемости, а также невозможно его описание в терминах совокупности элементарных событий, использование субъективных ожиданий является единственной альтернативой на практике. Именно это делает байесовские вероятностные сети наиболее подходящими для анализа социальных сетей.

1.2. Построение сети. Однородные узлы сети и типы связей. Анализ социальных сетей может быть также рассмотрен как задача исследования зависимостей внутри этой сети. Для этого производится анализ графовой модели социальной сети.

Традиционные графо-теоретические алгоритмы объединяют анализ различных узлов и связей, в то время как последние склонны быть однородными в рамках сети (например, рассматривая отдельный узел или один тип связи в рамках анализа), нежели быть неоднородными (например, множество типов связей и множество типов узлов). Так, граф социальной сети может представлять разные концепты с

помощью связей, таких как оценка (А нравится В, А уважает В), поведенческое взаимодействие, перемещение, принадлежность, физическая связь и биологические отношения [4]. Граф может также предоставлять информацию о формальных отношениях, таких как соавторство [1].

Более того, графо-теоретические алгоритмы обычно не рассматривают атрибуты связей или узлов (например, связи с заданной мощностью или надежностью или узлы с заданными атрибутами). В то время как различные связи могут также означать мощность определенного концепта. К примеру, социальная сеть, которая оценивает мощность дружбы: «не общаются», «редкие беседы и кофе», «взаимопомощь», «близкие отношения» [5].

2. Применение байесовских вероятностных сетей

Вероятностные байесовские сети могут быть применимы для анализа социальных сетей с целью получения данных, которые не могут быть получены при помощи традиционных методов анализа социальных сетей.

Результаты такого анализа используются в разных предметных областях. Так, к примеру, бизнес-консультанты используют методы анализа социальных сетей для определения наиболее эффективных связей между работниками, которые позволяют повысить коэффициент производительности [6]. Данные методы также используются для анализа террористических и криминальных сетей [7]. Так, к примеру, задержать Садама Хусейна помогли результаты исследований социальных сетей [8].

Далее мы выделяем два основных класса задач, которые можно решить, применяя байесовские вероятностные сети при исследовании социальных сетей.

2.1. Поиск объекта социальной сети. Применяя вероятностные методы анализа при анализе социальной сети, пользователь может найти нужного человека в социальной сети. Это является практически полезным, когда пользователь работает с большим объемом данных (например, при обработке почтового трафика в многомиллионной компании) и хочет найти узел сети, отвечающий заданным требованиям. В другом случае, к примеру, пользователь может быть заинтересован в поиске индивидуума, который может стать потенциальным лидером в организации. Эта задача отличается от поиска по таким атрибутам узла, как «имя» или «возраст», потому что понятие «потенциальный лидер» является психосоциальным фактором, базирующемся на комбинации других атрибутов и связей, которые не могут быть заданы простыми возможностями поиска. Некоторые из таких атрибутов могут быть заданы с определенной долей вероятности.

В данном случае пользователь может создать вероятностную байесовскую сеть как показано на рис.1.



Рис. 1. Байесовская вероятностная сеть для определения «Потенциального лидера организации»

Такая байесовская сеть строится на основании суждений эксперта. Именно он определяет то, что с помощью данного набора атрибутов, взаимосвязей между атрибутами и узлами сети можно найти ответ на поставленную задачу.

Данная сеть использует такое понятие как «важность» (алгоритмическое вычисление центрированности и узел, который представляет общую достоверность всех данных), а также такие дополнительные параметры, как «предыдущий опыт лидерства», «курсы по приобретению лидерских навыков». Пользователь может применять данную байесовскую сеть для каждого индивидуума сети и анализировать список результатов. Для этого производится проецирование байесовской сети на социальную сеть. Узлы в байесовской вероятностной сети представляют собой концепты, которые связываются с атрибутами узлов или связями в социальной сети.

Сортируя полученный список результатов, можно также определить тех людей, которые имеют наибольшую вероятность стать лидером.

Поскольку числовые значения в байесовской сети могут быть субъективными, а априорные вероятности могут назначаться экспертами данной предметной области, это позволяет объединять информацию, полученную из социальной сети, с информацией, полученной из других источников. Так для борьбы с правонарушителями можно объединить данные любой социальной сети и данные, доступные правоохранительным органам.

2.2. Анализ связей. На основе проведенного анализа социальной сети с некой долей достоверности могут быть также определены новые связи между объектами сети. Также во время анализа могут быть обнаружены связи, менее интересные для стандартных шаблонов. Однако, именно такие связи могут представлять наибольший интерес для аналитика.

К примеру, прочно установленные связи данных (например, демографические данные от исследовательских центров) могут быть использованы для получения выводов о дополнительных возможных связях (к примеру, вероятность того, что двое людей знают друг друга, исходя из демографических данных и примерно одинакового социально-экономического положения). Анализ членов группы в рамках социальной сети может помочь в определении возможных связей [9].

Выводы

Анализ социальных сетей является одной из наиболее актуальных проблем аналитиков социальных сетей в силу важности задач, которые можно разрешить с его помощью. Большой объем информации, которая динамически меняется, может быть наполнена неопределенностью, проблемы со сбором достоверных и полных данных – всё это приводит к уменьшению количества методов, которые можно применять.

Байесовские вероятностные сети являются наиболее подходящим методом анализа социальных сетей при заданных условиях.

На данных момент существует большое количество готовых программных продуктов, которые позволяют построить байесовскую вероятностную

сеть для проведения анализа. Всё это делает байесовские вероятностные сети не только удобным средством для проведения анализа, но и наиболее эффективным качественным методом улучшения анализа социальных сетей.

Список литературы

1. Goldenberg A. *Scalable Graphical Models for Social Networks* / A. Goldenberg. – PA.: CMU, 2007 – 105 с.
2. Терехов С.А. Введение в байесовы сети / С.А. Терехов // *Лекции по нейроинформатике*. – М.: МИФИ, 2003. – № 1. – С. 149-184.
3. Freeman L.C. *On Human Social Intelligence* / L.C. Freeman, S.C. Freeman, A.G. Michaelson // *Journal of Social and Biological Structures*. – 1988. – С. 415-425.
4. Wasserman S. *Social Network Analysis: Methods and Applications* / S. Wasserman, K. Faust. – Cambridge University Press., 1994. – 819 с.
5. Heran F. *Comment Les Francais Voisinent* / F. Heran // *Economie et Statistique*. – 1987. – С. 43-60.
6. Ehrlich K. *Inside Social Network Analysis* / K. Ehrlich, I. Carboni // *IBM Technical Report 05-10*. – 2005. – 16 с.
7. Krebs V.E. *Uncloaking Terrorist Network* [Электронный ресурс] / V.E. Krebs. – Режим доступа к ресурсу: http://131.193.153.231/www/issues/issue7_4/krebs/index.html.
8. Hougham V. *Sociological Skills Used in the Capture of Saddam Hussein* [Электронный ресурс] / V. Hougham. – Режим доступа к ресурсу: <http://www.asanet.org/footnotes/julyaugust05/fn3.html>.
9. *Stochastic Link and Group Detection*. In *Proceedings of AAAI Workshop on Link Analysis* / J. Kubica, A. Moore, J. Schneider, Y. Yang. – 2002. – 7 с.

Поступила в редколлегию 15.09.2010

Рецензент: д-р техн. наук, проф. Е.В. Бодянский, Харьковский национальный университет радиоэлектроники, Харьков.

ЗАСТОСУВАННЯ БАЄСОВСКИХ ЙМОВІРНІСНИХ МЕРЕЖ ДЛЯ АНАЛІЗУ СОЦІАЛЬНИХ МЕРЕЖ

О.М. Поляруш, О.В. Вітько

У даній статті проводиться аналіз основних проблем стандартних методів дослідження соціальних мереж; розглядаються переваги баєсовських мереж і реальні приклади їх використання при аналізі соціальних мереж. Також описуються найбільш загальні класи аналітичної інформації, яку можна отримати на основі досліджень, та подаються нові варіанти застосування баєсовських мереж у межах соціальних досліджень.

Ключові слова: аналіз даних, байесовські ймовірнісні мережі, соціальні мережі.

USAGE OF BAYESIAN BELIEF NETWORKS FOR SOCIAL NETWORK ANALYSIS

O.M. Poliarush, A.V. Vitko

In this paper we make an analysis of main problems of standard methods that are used for social network investigation; we observe advantages of Bayesian networks and provide real examples how they could be used for social network analysis. In addition we describe the most general types of analytical information that can be received and present new variants of Bayesian belief network usage in the context of social researches.

Keywords: analysis of data, Bayesian probabilistic networks, social networks.