

УДК 004.67; 004.94

**Д. В. Ландэ, Я. А. Калиновский, Ю. Е. Бояринова**  
Институт проблем регистрации информации НАН Украины  
ул. Н. Шпака, 2, 03113 Киев, Украина

## **Гиперкомплексный подход к моделированию репутации в социальных сетях**

*Рассмотрена одна из характеристик агентов в социальных сетях — репутация, а также возможность использования гиперкомплексного представления данных для моделирования и оценки репутации.*

***Ключевые слова:** репутация, социальная сеть, гиперкомплексная числовая система.*

### **Введение**

В последнее время в рамках теории анализа социальных сетей большое внимание уделяется оценке репутации отдельных субъектов (агентов, узлов социальных сетей) и уровня доверия между ними [1–3].

Формально социальная сеть представляет собой граф, в котором множество вершин — это совокупность агентов, субъектов — индивидуальных или коллективных, а множество ребер представляет собой совокупность отношений совокупности социальных связей между агентами.

При моделировании социальных сетей возникает необходимость учета динамики социальных связей — взаимного влияния агентов.

Влияние в данном случае рассматривается как процесс и результат изменения индивидом (субъектом влияния) поведения другого субъекта — объекта влияния, его установок и оценок в ходе взаимодействия [4]. Таким образом, влияние — это способность воздействовать на чьи-либо представления или действия [5]. Различают направленное и ненаправленное влияние [4]. Направленное влияние используется в качестве механизмов воздействия на другого человека убеждение и внушение. При этом индивид — субъект влияния — ставит перед собой задачу добиться определенных результатов от объекта влияния. Ненаправленное (нецеленаправленное, косвенное) влияние — это влияние, при котором индивид не ставит перед собой задачу добиться определенных результатов от объекта влияния.

Целенаправленное влияние участников социальной сети (или субъектов, не входящих в сеть, но использующих ее в качестве инструмента информационного воздействия) является частным случаем информационного управления, заключаю-

© Д. В. Ландэ, Я. А. Калиновский, Ю. Е. Бояринова

щегося в формировании у управляемых субъектов такой информированности [6], чтобы принимаемые ими на ее основе решения были наиболее выгодны для управляющего субъекта [7].

Возможности влияния одних участников социальной сети на других ее участников существенно зависят от репутации первых. Репутация — «создавшееся общее мнение о достоинствах или недостатках кого-либо, чего-либо, общественная оценка» [6]. Репутацию можно рассматривать как «весомость» мнения сообщества об отдельном агенте или группе агентов, определяемую его взглядами и деятельностью (активностью). При этом репутация может быть как индивидуальной, так и коллективной.

Репутация возрастает, если выбор агента (ответы на некоторые ключевые вопросы) совпадает с тем, чего от него ожидает сообщество, и понижается в противном случае.

**Целью** работы является построение модели определения репутации в социальной сети с использованием гиперкомплексного представления данных

### Анализ существующих решений

Дадим некоторое формальное определение информации, соответствующее приведенному в [4].

Пусть  $\{a_1, a_2, \dots, a_n\}$  — множество агентов — узлов социальной сети, которые влияют друг на друга. Матрицу влияния запишем как  $A = \|a_{ij}\|_{i=1, n}^{j=1, n}$  ( $a_{ij} \geq 0$  обозначает степень доверия  $i$ -го агента  $j$ -му). При этом очевидно, что если  $i$ -й агент влияет на  $j$ -го, а  $j$ -й влияет на  $k$ -го, то это означает следующее:  $i$ -й агент косвенно влияет на  $k$ -го (транзитивность), что позволяет строить цепочки косвенных влияний.

Предположим, что у каждого агента в начальный момент времени имеется мнение по некоторому ключевому вопросу. Пусть мнение сообщества агентов сети отражает вектор начальных мнений  $b^0$  размерности  $n$ . Мнение каждого агента меняется под влиянием мнений других агентов социальной сети.

Будем считать, что мнение  $i$ -го агента в момент времени  $t$  равно

$$b_i^t = \sum_{j=1}^n a_{ji} b_j^{t-1}.$$

В [2] показано, что при многократном обмене мнениями, мнения агентов сходятся к результирующему вектору мнений  $B = \lim_{t \rightarrow \infty} b^t$ . Таким образом, справедливо соотношение  $B = Ab$ .

Обозначим  $r_i$  — параметр, описывающий репутацию  $i$ -го агента в социальной сети (сообществу), которую можно определить как нормированную сумму его влияний на всех остальных агентов социальной сети (предполагается, что  $a_{ij} \geq 0$ ,  $i, j = 1, \dots, n$ ), т.е.

$$r_i = \frac{\sum_{i \neq j} a_{ij}}{R}, \quad j = 1, \dots, n.$$

Здесь  $R = \sum_k \sum_{j \neq k} a_{kj}$ ,  $k, j = 1, \dots, n$  — суммарное взаимное влияние друг на друга всех членов социальной сети.

В соответствии с приведенным выражением агент  $i$  имеет тем более высокую репутацию, чем выше его влияние на всех остальных членов социальной сети.

### Использование гиперкомплексных числовых систем для моделирования в социальных сетях

Моделирование с использованием гиперкомплексных числовых систем (ГЧС) позволяет применять развитый инструментарий из этой области математики [8–10].

В рамках модели, основанной на использовании ГЧС, каждый субъект (узел социальной сети) характеризуется своим отношением к ряду важных (ключевых) вопросов (пусть их количество равно  $N$ ). Тогда, по аналогии с [8], субъекту  $A$  можно поставить в соответствие гиперкомплексное число с базисом размерности  $2N$ :

$$A = e_1 w_1^+ + e_2 w_1^- + \dots + e_{2N-1} w_N^+ + e_{2N} w_N^-.$$

При этом каждому вопросу приписываются весовые значения  $w_i^+$  и  $w_i^-$ , которые соответствуют уровню положительного отношения субъекта к данному вопросу ( $w_i^+$ ) или отрицательного ( $w_i^-$ ), что является естественным расширением приведенного выше подхода. Оба значения могут быть в интервале  $[0,1]$ , в отдельных случаях можно предположить, что  $w_i^+ + w_i^- = 1$ .

Предлагается использовать ГЧС размерности  $2N$  с базисом  $\{e_1, e_2, \dots, e_{2N}\}$  и законом умножения, который можно представить в виде таблицы:

	$e_1$	$e_2$	$e_3$	$e_4$	...	$e_{2N-1}$	$e_{2N}$
$e_1$	$e_1$	$e_2$	0	0	...	0	0
$e_2$	$e_2$	$e_1$	0	0	...	0	0
$e_3$	0	0	$e_3$	$e_4$	...	0	0
$e_4$	0	0	$e_4$	$e_3$	...	0	0
...	...	...	...	...	...	...	...
$e_{2N-1}$	0	0	0	0	0	$e_{2N-1}$	$e_{2N}$
$e_{2N}$	0	0	0	0	0	$e_{2N}$	$e_{2N-1}$

Модель субъекта социальной сети в данном случае рассматривается как гиперкомплексное число вида:

$$D = e_1 w_1^+ + e_2 w_1^- + e_3 w_2^+ + e_4 w_2^- + \dots + e_{2N-1} w_N^+ + e_{2N} w_N^-.$$

Можно рассмотреть оценку близости мнений двух субъектов  $Est(A, B)$  (соответствующими гиперкомплексными числами)  $A = e_1 a_1^+ + e_2 a_1^- + \dots + e_{2N-1} a_N^+ + e_{2N} a_N^-$  и  $B = e_1 b_1^+ + e_2 b_1^- + \dots + e_{2N-1} b_N^+ + e_{2N} b_N^-$ :

$$Est(A, B) = Norm\left(\frac{1}{N} \left(\sum_{i=1}^N (e_{2i-1} a_i^+ + e_{2i} a_i^-)(e_{2i-1} b_i^+ + e_{2i} b_i^-)\right)\right),$$

где  $Norm()$  — функция нормы гиперкомплексного числа:  $Norm(e_{2i-1}) = e_1$ ;  $Norm(e_{2i}) = -e_1$ .

Отношение большей части участников социальной сети (общества) к выбранным вопросам также представляется в виде гиперкомплексного числа  $Q = q_1 e_1 + q_2 e_2 + q_3 e_3 + \dots + q_{2N} e_{2N}$ , как и отдельный субъект  $D$ . Чем больше значение  $Est(Q, D)$ , тем субъект более лояльный, «релевантный» обществу.

Можно ввести и другую оценку близости между гиперкомплексными числами, по аналогии с нормой различий между обычными векторами в векторном пространстве:

$$Est_1(A, B) = Norm\left(\frac{1}{N} \left(\sum_{i=1}^N (e_{2i-1} a_i^+ - e_{2i-1} b_i^+)^2 (e_{2i} a_i^- - e_{2i} b_i^-)^2\right)\right).$$

В этом случае субъект будет более лояльный по отношению к обществу, если оценка  $Est_1(Q, D)$  будет меньшей.

Вместе с этим, вторая оценка по содержанию менее пригодна для задач выявления взаимного влияния субъектов. Например, при сравнении отношения субъекта к обществу с запросом, даже при полностью нулевых значениях весовых коэффициентов, относящихся к значению всей социальной сети (обществу), сумма в приведенном для выражения  $Est_1(Q, D)$  не будет нулевой, т.е. полностью зависит от субъекта. Поэтому ограничимся применением первой оценки.

Рассмотрим для примера некоторые упрощенные частные случаи, при которых анализируются отношения общества и субъекта к одному вопросу.

1. Пусть значение, соответствующее обществу, имеет вид:  $Q = \frac{1}{2} e_1 + \frac{1}{2} e_2$ , т.е. отношение к выбранному вопросу в обществе может быть как позитивным, так и негативным, с равной вероятностью. Пусть отношение субъекта к этому же вопросу однозначно позитивное, а именно:  $D = e_1$ . В этом случае  $Est(Q, D) = \frac{1}{2} - \frac{1}{2} = 0$ , что соответствует полной неопределенности.

2. Пусть значение, соответствующее обществу, как и в предыдущем случае, имеет вид:  $Q = \frac{1}{2} e_1 + \frac{1}{2} e_2$ . Пусть отношение субъекта к этому же вопросу также

имеет вид:  $D = \frac{1}{2}e_1 + \frac{1}{2}e_2$ . В этом случае  $Est(Q, D) = \frac{1}{4} + \frac{1}{4} - \frac{1}{2} = 0$ , что, как и в предыдущем случае, соответствует полной неопределенности.

3. Пусть значение, соответствующее обществу, имеет вид:  $Q = e_1$ , а отношение субъекта к этому же вопросу:  $D = \frac{4}{5}e_1 + \frac{1}{5}e_2$ , тогда  $Est(Q, D) = \frac{4}{5} - \frac{1}{5} = \frac{3}{5}$ .

Следует отметить, что не все нулевые элементы приведенной выше «идеальной» таблицы в реальности могут быть нулевыми, однако предполагается, что данная таблица будет разреженной. Редкие ненулевые элементы в ней могут характеризовать взаимосвязь различных вопросов.

Значение коэффициентов при базисных элементах образов субъектов социальной сети могут соответствовать вероятностям позитивного (или негативного) отношения субъектов к соответствующим вопросам. В этом случае путем перенумерации базисов таблицу умножения ГЧС можно представить в следующем виде:

	$e_1$	$e_2$	$e_3$	$e_4$	...	$e_{4N-3}$	$e_{4N-2}$	$e_{4N-1}$	$e_{4N}$
$e_1$	$B_1$								
$e_2$									
$e_3$									
$e_4$									
...									
...									
$e_{4N-3}$						$B_N$			
$e_{4N-2}$									
$e_{4N-1}$									
$e_{4N}$									

где блок  $B_1$  (а в общем случае и любой ненулевой блок) будет иметь вид, дополненный коэффициентами  $w_i^j$ , вычисляемыми в процессе обучения модели:

	$e_1$	$e_2$	$e_3$	$e_4$
$e_1$	$w_1^1 e_1$	$w_1^2 e_2$	$w_2^1 e_1$	$w_2^2 e_1$
$e_2$	$w_1^2 e_2$	$w_1^1 e_1$	$w_2^2 e_1$	$w_2^1 e_1$
$e_3$	$w_2^1 e_1$	$w_2^2 e_1$	$w_3^1 e_3$	$w_3^2 e_4$
$e_4$	$w_2^2 e_1$	$w_2^1 e_1$	$w_3^2 e_4$	$w_3^1 e_3$

При этом  $w_1^1$  — вес позитивного отношения к вопросу  $t_1$ ;  $w_1^2$  — вес негативного отношения к вопросу  $t_1$ ;  $w_2^1, w_2^2$  — веса взаимосвязей наличия противоречивых одновременных позитивных и негативных отношений к вопросу  $t_1$ .

Тогда для агента  $A = a_1e_1 + a_2e_2 + a_3e_3 + a_4e_4$  и агента  $B = b_1e_1 + b_2e_2 + b_3e_3 + b_4e_4$  оценка близости будет иметь следующий вид:

$$\begin{aligned} Sim(A \cdot B) = Norm & \left( \frac{1}{2} (e_1 (w_1^1 (a_1b_1 + a_2b_2) + w_2^1 (a_1b_3 + a_2b_4 + a_3b_1 + a_4b_2)) + \right. \\ & + w_2^2 (a_1b_4 + a_2b_3 + a_3b_2 + a_4b_1)) + e_2 (w_1^2 (a_1b_2 + a_2b_1)) + \\ & \left. + e_3 (w_3^1 (a_3b_3 + a_4b_3)) + e_4 (w_3^2 (a_3b_4 + a_4b_4)) \right). \end{aligned}$$

При этом функция *Norm* соответствует той ГЧС, таблица которой используется для поиска. Следует учитывать, что предложенная таблица может быть составной частью таблицы большей размерности. Переход от заполненной таблицы умножения к слабозаполненной (разреженной) можно осуществить изоморфным переходом [10], что значительно сократит количество операций при вычислении функции близости между субъектами.

В общем случае уровень доверия (близость) между субъектами  $A = e_1a_1^+ + e_2a_1^- + \dots + e_{2N-1}a_N^+ + e_{2N}a_N^-$  и  $B = e_1b_1^+ + e_2b_1^- + \dots + e_{2N-1}b_N^+ + e_{2N}b_N^-$ , которую можно трактовать как степень доверия, также характеризуется функцией, приведенной выше с учетом возможного наличия ненулевых недиагональных элементов таблицы:

$$Sim(A, B) = Norm \left( \frac{1}{N} (e_1a_1^+ + e_2a_1^- + \dots + e_{2N-1}a_N^+ + e_{2N}a_N^-) \right) + (e_1b_1^+ + e_2b_1^- + \dots + e_{2N-1}b_N^+ + e_{2N}b_N^-).$$

Репутация субъекта  $A_i = e_1a_{i1}^+ + e_2a_{i1}^- + \dots + e_{2N-1}a_{iN}^+ + e_{2N}a_{iN}^-$  в рамках всей социальной сети (т.е. по отношению к обществу) при этом определяется как нормированная сумма уровней доверия со всеми остальными субъектами:

$$Trust(A_i) = \frac{\sum_{j \neq i} Sim(A_i, A_j)}{\sum_{k \neq j} Sim(A_k, A_j)}.$$

Для оценки уровня взаимного влияния субъектов социальной сети могут также использоваться другие различные методы, среди которых можно выделить: расчет меры взаимной информации (mutual information), расчет модифицированного коэффициента Dice (modified Dice coefficient), вхождение правдоподобия (log likelihood ratio), оценку  $\chi^2$  (Chi-square test) [11–14]. Вместе с тем, без специальных модификаций никакой из них не позволяет учитывать одновременно уровень позитивного и негативного отношения одного субъекта к ключевым вопросам, взаимную зависимость ключевых вопросов, вплоть до смысловой синонимии.

## ВЫВОДЫ

Применение модели определения репутации в социальных сетях на базе использования ГЧС может обеспечивать: возможность обучения системы, учет некоторой смысловой синонимии, омонимии в ключевых вопросах (под омонимией здесь понимается различные по смыслу ключевые вопросы, заданные примерно одним набором слов) на уровне расширения таблиц умножения соответствующих ГЧС; возможность применения имеющихся наработок в области гиперкомплексных числовых систем, в том числе изоморфных ГЧС, более пригодных для быстрых вычислений.

1. *Расторгуев С.П.* Информационная война. Проблемы и модели. Экзистенциальная математика / С.П. Расторгуев. — М.: Гелиос АРВ, 2006. — 240 с.
2. *Губанов Д.А.* Модели репутации и информационного управления в социальных сетях / Д.А. Губанов, Д.А. Новиков, А.Г. Чхартишвили // Математическая теория игр и ее приложения. — 2009. — № 2. — С. 14–37.
3. *Ермаков Н.С.* Модели репутации и норм деятельности / Н.С. Ермаков, А.А. Иващенко, Д.А. Новиков. — М.: ИПУ РАН, 2005. — 67 с.
4. *Ландэ Д.В.* Интернетика: Навигация в сложных сетях: модели и алгоритмы / Д.В. Ландэ, А.А. Снарский, И.В. Безсуднов. — М.: Либроком (Editorial URSS), 2009. — 264 с.
5. *Ландэ Д.В.* Поиск знаний в Internet / Д.В. Ландэ. — М.: Диалектика-Вильямс, 2005. — 262 с.
6. *Новиков Д.А.* Теория управления организационными системами / Д.А. Новиков. — М.: Физмалит, 2007.
7. *Новиков Д.А.* Теория управления организационными системами / Д.А.Новиков, А.Г. Чхартишвили. — М.: Синтег, 2002. — 227 с.
8. *Lande D.V.* The Model of Information Retrieval Based on the Theory of Hypercomplex Numerical Systems / D.V. Lande, Ya.A. Kalinovskiy, Yu.E. Boyarinova // Preprint Arxiv 1205.3031. — On-line: <http://arxiv.org/abs/1205.3031>
9. *Гіперкомплексні* числові системи: основи теорії, практичні використання, бібліографія / [М.В. Синьков, Ю.Є. Боярінова, Я.О. Каліновський та ін.] — К.: Ін-т проблем реєстрації інформації, 2009. — 44 с. — (Препринт / НАН України, Ін-т проблем реєстрації інформації).
10. *Калиновский Я.А.* Высокоразмерные изоморфные гиперкомплексные числовые системы и их использование для повышения эффективности вычислений / Я.А. Калиновский, Ю.Е. Бояринова. — К.: Инфодрук, 2012. — 183 с.
11. *Manning C.D.* An Introduction to Information Retrieval / C.D. Manning, P. Raghavan, H. Schutze // Cambridge University Press, England. — 2009. — 544 p.
12. *Query Term Disambiguation for Web Cross-Language Information Retrieval using a Search Engine* / Akira Maeda, Fatiha Sadat, Masatoshi Yoshikawa, and Shunsuke Uemura Query. — On-line: <http://www.dl.is.ritsumei.ac.jp/~amaeda/pub/IRAL00-e.pdf>
13. *Myung-Gil Jang.* Using Mutual Information to Resolve Query Translation Ambiguities and Query Term Weighting / Myung-Gil Jang, Sung Hyon Myaeng, Se Young Park. — On-line: <http://acl.ldc.upenn.edu/P/P99/P99-1029.pdf>
14. *Габидулин Э.М.* Лекции по теории информации / Э.М. Габидулин, Н.И. Пилипчук. — М.: МФТИ, 2007. — 214 с.

Поступила в редакцию 13.08.2013