

Применение обобщенного алгоритма обработки слабоформализованной информации для оценивания и повышения производительности труда программистов

А.А. Копыльцов

*Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»
имени В.И. Ульянова (Ленина)*

Аннотация: Рассматривается применение обобщенного алгоритма обработки слабоформализованной информации к задаче повышения производительности труда программистов, т. е. работников, которые много времени проводят в сети Интернет. Предполагается, что сотрудник компании в течение рабочего дня выходит в Интернет и работает с различными сайтами. Сайты могут, как относиться к работе, так и не иметь отношения к деятельности компании. Требуется оценить, какое количество времени сотрудник проводит на сайтах, относящихся к работе компании и на других сайтах. Трудность заключается в том, что нет достаточно объективных критериев оценки сайтов, т. е. эта информация является слабоформализованной. Поэтому, предлагается применить обобщенный алгоритм для обработки такой информации. Применение этого алгоритма позволяет выявить стереотип поведения сотрудника на работе, т. е. оценить, сколько времени он проводит на различных видах сайтов, насколько это достоверно, насколько устойчиво такое его поведение, т. е. есть ли зависимость от каких-либо видов сайтов и т.д. Если такая зависимость устойчиво проявляется в течение длительного времени и сайты не имеют отношения к работе компании, то руководству компании нужно обратить внимание на подобных сотрудников.

Ключевые слова: слабоформализованная информация, обработка информации, алгоритм, производительность труда, управление

Введение

Особенностью настоящего времени является то, что в компаниях занимающихся производством программных продуктов сотрудники много времени проводят в Интернете. Они могут находиться на сайтах относящихся непосредственно к работе компании, на игровых и прочих развлекательных сайтах, а также на иных сайтах. Сложность состоит в классификации сайтов на сайты, которые относятся к работе компании и не являются таковыми. Информация о сайтах является слабоформализованной и часто ее недостаточно для отнесения сайта к тому или иному классу [1 – 7]. Для обработки слабоформализованной информации предлагается обобщенный алгоритм [8 – 13], который позволяет выявить стереотип поведения

сотрудников на работе, т. е. оценить, сколько времени они проводят на различных сайтах, насколько это достоверно, насколько устойчиво такое их поведение, т. е. есть ли зависимость от каких-либо видов сайтов и т.д. Если такая зависимость устойчиво проявляется в течение длительного времени и сайты не имеют отношения к работе компании, то руководству компании нужно обратить внимание на подобных сотрудников.

Повышение производительности труда

Рассмотрена задача повышения производительности труда работников компании, занимающейся производством программных продуктов. Особенность такой компании – сотрудники много времени проводят в сети Интернет. Рассмотрена модель обработки информации, поступающей от компьютера, подключенного к сети Интернет. Предполагается, что сотрудник компании выходит в Интернет и подключается к различным сайтам. Сайты различаются, например, по следующим видам:

- сайты, относящиеся к работе;
- сайты, относящиеся к развлечениям (например, игры);
- прочие сайты (например, новости, личные увлечения и др.).

Предположим, что сотрудник в течение рабочего дня выходил в Интернет:

- на сайты, относящиеся к работе, N раз и оставался на них в течение $t_{11}, t_{12}, \dots, t_{1N}$ часов, соответственно;
- на сайты, относящиеся к развлечениям, K раз и оставался на них в течение $t_{21}, t_{22}, \dots, t_{2K}$ часов, соответственно;
- на иные сайты L раз и оставался на них в течение $t_{31}, t_{32}, \dots, t_{3L}$ часов, соответственно.

Тогда, просуммировав, получим, что время, затрачиваемое на каждый из трех видов сайтов, равно

$$t_1 = \sum_{i=1}^N t_{1i}, t_2 = \sum_{i=1}^K t_{2i}, t_3 = \sum_{i=1}^L t_{3i},$$

или, нормируя за 8 часовой рабочий день, получим

$$t_1^N = \frac{t_1}{8}, t_2^N = \frac{t_2}{8}, t_3^N = \frac{t_3}{8}.$$

Можно построить зависимость нормированного времени (за 8 часов) от видов сайтов, например, относящихся к работе (A), относящихся к развлечениям (B) или других сайтов (C). Например, если сотрудник 4 часа провел на сайтах, относящихся к работе, 1,6 часа на сайтах относящихся к развлечениям и 2,4 часа на иных сайтах, то можно вычислить $A1=4/8=0,5$, $B1=1,6/8=0,2$, $C1=2,4/8=0,3$ и построить график (рис. 1).

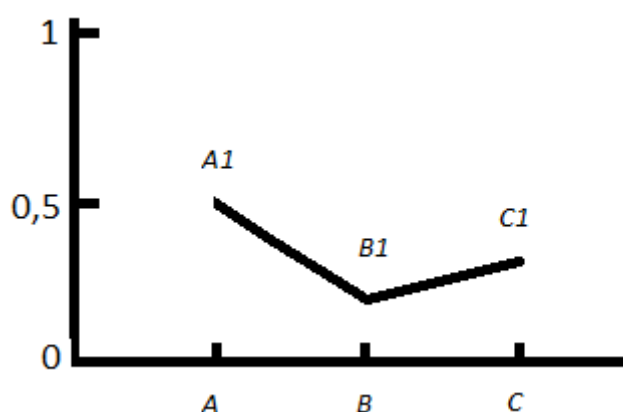


Рис. 1. – Зависимость нормированного времени (ось ординат) от видов сайтов (ось абсцисс).

Можно, аналогично, построить усредненные зависимости за неделю, месяц, квартал или год. Если точка $A1$ ниже точки $B1$ или $C1$, то руководству компании следует задуматься о том, на что сотрудник тратит свое рабочее время.

Применение обобщенного алгоритма

Обобщенный алгоритм обработки слабоформализованной информации включает 14 шагов [8 – 13].

1. Сбор информации о выходе сотрудников в сеть Интернет и подключении к различным сайтам.

2. Распознавание сайтов, к которым подключаются сотрудники.

3. Классификация информации по n классам. В нашем случае $n=3$, поскольку сайты классифицируются по трем видам: сайты, относящиеся к работе, развлечениям и другие сайты.

4. В каждом из трех выделенных классов за определенный промежуток времени (например, рабочий день, неделя, месяц, год) вычисляем [8]:

- сколько раз сотрудник выходил на конкретные сайты и сколько времени на них находился (чем больше времени проводил на каком-либо сайте, тем большим приоритетом этот сайт обладает);
- свертку, т. е. среднее количество выходов на сайты и среднее время пребывания на них за рабочий день, а также дисперсию и среднее квадратичное отклонение;
- общее время пребывания за рабочий день на сайтах каждого из трех видов сайтов.

5. Оценивание достоверности осуществляется следующим образом. Предполагается, что ранее были проведены эксперименты, в результате которых было получено, что средняя продолжительность времени $H_{ср_i}$ нахождения на сайтах одного из трех видов за рабочий день $0 \leq H_{ср_i} \leq H_{max_i}$, где i – номер вида сайта ($i=1, 2, 3$), $N_{сайт}=3$ – всего видов сайтов, H_{max_i} – наибольшее значение, определяемое на основании ранее проведенных экспериментов. Если для новой поступающей информации имеем $H_{новая_i} \leq H_{max_i}$ для всех i , то информация считается достоверной. Если это выполняется для NI видов сайтов, то достоверность равна $NI/N_{сайт}$.

6. Оценивание безопасности информации осуществляется путем проверки $N_{сайт} - NI$ видов сайтов, где $H_{новая_i} > H_{max_i}$. Если это неравенство справедливо и $H_{новая_i}$ удовлетворяет старому проверенному тестовому

варианту, то H_{max_i} корректируется. Если же это несправедливо, то этот конкретный сайт отнесен не к тому виду сайтов, т. е. его классификация неправильная. Выявляем число неправильно классифицированных сайтов N_2 и тогда безопасность равна $(N_{сайт} - N_2) / N_{сайт}$.

7. Установление связей между вновь поступающей новой информацией и ранее полученной информацией осуществляется следующим образом. Новая информация, т. е. набор усредненных значений времени проводимого сотрудником на каждом из видов сайтов за рабочий день, сравнивается с ранее полученными наборами, которые хранятся в памяти компьютера, и ищется наиболее близкий по минимуму среднего квадратичного отклонения. В случае задания некоторого интервала, может оказаться, что в этот интервал попадает несколько ранее полученных вариантов. Подсчитываем количество таких вариантов (M).

8. Оценивание вероятности, т. е. вероятности с которой можно доверять поступающей новой информации в каждом из выбранных трех классов осуществляется следующим образом. Если $M \geq M^*$, где M^* – некоторое заранее заданное число, то $P=1$. Если это неравенство не выполняется, то осуществляется переход на второй шаг. Если, например, $M^*=1$, и $M=2$, то $M \geq M^*$ и, следовательно, $P=1$.

9. Поддержка принятия решений в трех выделенных классах осуществляется следующим образом. В случае выполнения следующих условий поступающая новая информация записывается в хранилище для данного класса. Перечислим эти условия:

- $ИДПИ > ИДПИ^*$ для некоторого $ИДПИ^*$, где $ИДПИ$ – достоверность поступающей слабоформализованной информации;
 - $ИБПИ > ИБПИ^*$ для некоторого $ИБПИ^*$, где $ИБПИ$ – безопасность поступающей слабоформализованной информации;
 - $M > M^*$ для некоторого M^* , где M – число связей;
-

- $P > P^*$ для некоторого P^* , где P – вероятность, с которой можно доверять полученной слабоформализованной информации.

Нормировка числа связей M осуществляется следующим образом

$$M^N = \frac{M}{M_{\max}},$$

где M^N – нормированное, а M_{\max} – максимальное значение M .

На основании выше изложенного имеем следующие неравенства

$$0 \leq \text{ДПИ} \leq 1, 0 \leq \text{БПИ} \leq 1, 0 \leq M^N \leq 1, 0 \leq P \leq 1.$$

Если выше перечисленные условия и условие $M^N > M^{N*}$, где M^{N*} – некоторое число, выполнены, то новая поступающая информация может быть записана в хранилище для данного класса. Поскольку в нашем случае все условия выполняются, то набор средних значений промежутков времени проводимого сотрудником на каждом из видов сайтов за рабочий день, записывается в хранилище.

10. Обобщенная поддержка принятия решений проводится на основе предыдущего шага, т. е. набор средних значений промежутков времени проводимого сотрудником на каждом из видов сайтов за рабочий день, записывается в хранилище.

11. Оценивание правильности принятых решений проводится следующим образом. Если вновь поступающая информация записывается в хранилище, то осуществляется и ее свертка, в результате которой получаем, например, C_N [8]. Сравнение C_N со свертками C_1, C_2, \dots, C_k ранее полученной информации и записанными в хранилище проводим следующим образом. Если для некоторых $i \in \{1, 2, \dots, k\}$ и $\varepsilon > 0$, имеем $\delta_i = |C_N - C_i| < \varepsilon$, то считается, что вновь поступающая информация с номером N и ранее записанная в хранилище информация с номером i связаны между собой. Подсчитываем число таких связей (MM). Вероятность того, что принятое решение правильное $P=1$, если $MM > MM^*$ для некоторого MM^* , а в

противном случае $P < 1$. На основе выше изложенного в нашем случае имеем $P = 1$.

12. Выработка устойчивой реакции на поступающую информацию происходит следующим образом. Если среднее квадратичное отклонение между вновь поступившим набором средних значений промежутков времени проводимого сотрудником на каждом из видов сайтов за рабочий день и несколькими ранее полученными наборами меньше ε , для некоторого $\varepsilon > 0$, то считается, что вырабатывается устойчивая реакция на поступающую информацию. Эти наборы записываются в хранилище с индексом «устойчивая»: *IUI*.

13. Генерация решений осуществляется следующим образом. Рассмотрим фрагменты информации с индексом «устойчивая» и связи между ними. Если $C_{ij} > C^*$, где C^* – некоторое число, C_{ij} – число связей между фрагментами информации, имеющими номера i и j , то полагаем, что фрагменты информации с номерами i и j связаны и уровень связности равен C_{ij} .

14. Генерация новой информации осуществляется следующим образом. Новая информация – это информация с номером I_k о том, что фрагменты информации с номерами I_i и I_j связаны. Эту информацию можно направить на первый шаг. Таким образом, если информация с номером I_i характеризуется показателями i_1, i_2, \dots, i_{i1} , а информация с номером I_j – показателями j_1, j_2, \dots, j_{j1} , то информация с номером $I_k = I_{i \cup j}$ характеризуется показателями $i_1, i_2, \dots, i_{i1}, j_1, j_2, \dots, j_{j1}$, которые являются объединением показателей I_i и I_j . Свертка на 4 шаге делается по объединенным показателям $i_1, i_2, \dots, i_{i1}, j_1, j_2, \dots, j_{j1}$. Применительно к нашему случаю делаем объединение двух экспериментов, а затем вычисляем для трех видов сайтов средние значения. Если это так, то полученный вариант средних значений тоже может быть. Это является новой информацией, которая может быть передана на

обработку на первый шаг. Можно вычислить средние значения по всем видам сайтов, их дисперсии, средние квадратичные отклонения и т.д.

В итоге, выявляется стереотип поведения сотрудника (сколько времени он проводит на различных видах сайтов, насколько это достоверно, насколько устойчиво такое его поведение, т. е. есть ли зависимость от каких-либо видов сайтов и т.д.) Окончательное решение принимает руководитель компании. Например, если в течение рабочего дня сотрудник первые два часа находился на сайтах связанных с работой, следующий час находился на сайтах развлечений, следующие 2 часа находился на других сайтах, а последние три часа – на сайтах связанных с развлечениями, то руководству компании нужно подумать о целесообразности привлекать к работе такого сотрудника.

Заключение

Предлагается применение обобщенного алгоритма обработки слабоформализованной информации к задаче повышения производительности труда работников, много времени проводящих в сети Интернет. Предполагается, что сотрудник компании по роду своей работы выходит в Интернет и подключается к различным сайтам. Сайты могут быть различными, как относящимися к работе, так и не относящимися к работе, например, развлекательные, игровые и др. Требуется оценить, сколько времени сотрудник проводит на сайтах, относящихся к работе и на сайтах, не относящихся к работе. А также выявить стереотип поведения сотрудника на работе, т. е. оценить, сколько времени он проводит на различных видах сайтов, насколько это достоверно, насколько устойчиво такое его поведение, т. е. есть ли зависимость от каких-либо видов сайтов и т.д. Недостаток слабоформализованной информации состоит, например, в том, что сайты обычно довольно сложно квалифицировать по признаку относятся ли они к работе или нет. Аналогичная ситуация наблюдается, например, при чтении



литературы, т. е. не всегда с уверенностью можно сказать, что какая-либо книга или статья относится к работе или нет.

Литература

1. Cherpakov A.V., Soloviev A.N., Gricenko V.V., Mohanty S.C., Parshin D.Y., Butenko U.I., Bocharova O.V. Comparison of finite element modeling and analytical approach results for oscillating rod structure with crack // Инженерный вестник Дона, 2014, №4 URL: ivdon.ru/ru/magazine/archive/n4y2014/2785
2. Синявская Е.Д. Оптимизация на основе вероятностного подхода нечетких моделей управления производственными объектами управления // Инженерный вестник Дона, 2014, №3 URL: ivdon.ru/ru/magazine/archive/n3y2014/2462
3. Rutkowski L. Metody i techniki sztucznej inteligencji. Warszawa: Wydawnictwo Naukowe PWN. 2005. 520 p.
4. Moon F.C. Chaotic Vibrations. New York: A Wiley-Interscience Publication John Wiley & Sons. 1990. 312 p.
5. Feder J. Fractals. New York: Plenum Press. 1991. 260 p.
6. Murray J.D. Lectures on Nonlinear Differential Equations: Models in Biology. Oxford: Clarendon Press. 1977. 398 p.
7. Kandel E.R. In Search of Memory. New York: A Wiley-Interscience Publication John Wiley & Sons. 2006. 736 p.
8. Копыльцов А.А. Применение обобщенного алгоритма обработки слабоформализованной информации для управления неравновесной химической реакцией // Инженерный вестник Дона, 2015, №1, ч.2 URL: ivdon.ru/ru/magazine/archive/n1p2y2015/2812
9. Копыльцов А.А. Модель классификации информации и алгоритм ее предварительной обработки для статических и динамических объектов // Известия СПбГЭТУ «ЛЭТИ» (известия государственного

электротехнического университета), серия “Информатика, управление и компьютерные технологии”. 2013. №6. С. 134-139.

10. Копыльцов А.А., Копыльцов А.В. Алгоритм обработки слабоформализованной информации, поступающей от технических систем // Известия СПбГЭТУ “ЛЭТИ” (известия государственного электротехнического университета), серия “Информатика, управление и компьютерные технологии”. 2012. №8. С. 30-36.

11. Копыльцов А.А. Алгоритм коррекции связей между фрагментами слабоформализованной информации и генерация новой информации // Вестник Нижневартковского государственного университета. 2014. №3. С. 28-34.

12. Копыльцов А.А., Копыльцов А.В. Обработка слабоформализованной информации, поступающей от технических систем // Вестник Нижневартковского государственного гуманитарного университета. 2013. №1. С. 32-36.

13. Копыльцов А.А., Копыльцов А.В. Обобщенный алгоритм обработки слабоформализованной информации и его применение // Вестник Нижневартковского государственного университета. 2014. №35. С. 35-44.

References

1. Cherpakov A.V., Soloviev A.N., Gricenko V.V., Mohanty S.C., Parshin D.Y., Butenko U.I., Bocharova O.V. Inzhenernyj vestnik Dona (Rus), 2014, №4. URL: ivdon.ru/ru/magazine/archive/n4y2014/2785

2. Sinyavskaya E.D. Inzhenernyj vestnik Dona (Rus), 2014, №3. URL: ivdon.ru/ru/magazine/archive/n3y2014/2462

3. Rutkowski L. Metody i techniki sztucznej inteligencji [Methods and technologies of an artificial intelligence]. Warszawa: Wydawnictwo Naukowe PWN. 2005. 520 p.



4. Moon F.C. Chaotic Vibrations. New York: A Wiley-Interscience Publication John Wiley & Sons. 1990. 312 p.
5. Feder J. Fractals. New York: Plenum Press. 1991. 260 p.
6. Murray J.D. Lectures on Nonlinear Differential Equations: Models in Biology. Oxford: Clarendon Press. 1977. 398 p.
7. Kandel E.R. In Search of Memory. New York: A Wiley-Interscience Publication John Wiley & Sons. 2006. 736 p.
8. Kopyltsov A.A. Inženernyj vestnik Dona (Rus), 2015, №1, part 2. URL: ivdon.ru/ru/magazine/archive/n1p2y2015/2812
9. Kopyltsov A.A. Journal of Saint-Petersburg State Electrotechnical University «LETI» (Rus). 2013. №6. pp. 134-139.
10. Kopyltsov A.A., Kopyltsov A.V. Journal of Saint-Petersburg State Electrotechnical University «LETI» (Rus). 2012. №8. pp. 30-36.
11. Kopyltsov A.A. Journal of Nizhnevartovsk State University (Rus). 2014. №3. pp. 28-34.
12. Kopyltsov A.A., Kopyltsov A.V. Journal of Nizhnevartovsk State University. 2013, (Rus). №1. pp. 32-36.
13. Kopyltsov A.A., Kopyltsov A.V. Journal of Nizhnevartovsk State University (Rus). 2014. №35. pp. 35-44.