

Извлечение экспертных знаний при построении систем прогнозной диагностики

В. А. Суздальцев¹, Э. Г. Тахавова, С. А. Зарайский, В. В. Мокшин

Казанский национальный исследовательский технический университет им. А.Н. Туполева-КАИ

¹svlant@mail.ru

Аннотация. Рассматривается способ формирования нечетких правил для проведения прогнозной диагностики экспертной системой. Предлагается методика, позволяющая построить систему нечетких правил при опросе эксперта и сокращающая количество задаваемых вопросов. Допустимость рассмотренного подхода получения нечетких правил подтверждается практическим опытом его применения при построении экспертных систем прогнозирования возникновения патологий беременности и новорожденных.

Ключевые слова: экспертная система; прогнозная диагностика; извлечение знаний; нечеткая модель; нечеткие множества; лингвистические переменные

Экспертные системы (ЭС) представляет собой программные средства имитирующие решение задач требующих мышления и профессионального опыта, к которым относятся задачи прогнозной диагностики (ПД) в областях техники [1], медицины, и т.д. При решении задач ПД необходимо установить возможное наличие особых ситуаций (нештатных ситуаций) при протекании процессов в будущем времени и требующих своевременного вмешательства специалиста.

При построении подобных ЭС в случае отсутствия фактических данных о наступлении штатных ситуаций (НС) приходится прибегать к опросу экспертов для установления соответствия между типичными текущими ситуациями и риском возникновения НС в будущем времени. Основными проблемой таких опросов является его трудоемкость, которую удастся снизить при соблюдении определенных условий. Рассмотрим один из таких случаев.

Образует лингвистическую переменную (ЛП) [2] с именем β_0 («Частота наступления штатной ситуации»):

$$\langle \beta_0, T_0, Y \rangle,$$

где $T_0 = \{ \alpha_{10}, \alpha_{20} \}$ – двух элементное множество значений ЛП с элементами:

α_{10} – высокий риск возникновения штатной ситуации;

α_{20} – низкий риск возникновения штатной ситуации;

$Y = [0,1]$ – базовое множество ЛП, определяющее количественную меру частоты наступления штатной ситуации.

Элементы множества T_0 являются именами нечетких переменных:

$$\langle \alpha_{i0}, Y, \tilde{M}_{i0} \rangle, i=1,2.$$

$\tilde{M}_0 = \{ \langle \mu_0(y), y \rangle / y \in Y \}$ – нечеткое множество с функцией принадлежности $\mu_0(y)$ и определяющее значение ЛП с лингвистическим именем

$$\alpha_{i0}, i=1,2.$$

Построим лингвистические переменные: $\langle \beta_j, T_j, X_j \rangle$ с именами $\beta_j (j=1,n)$, каждая из которых соответствует признаку штатной ситуации, $T_j = \{ \alpha_{ij} / i=1,m_j \}$, α_{ij} является именем нечеткой переменной $\langle \alpha_{ij}, X_j, \tilde{M}_{ij} \rangle$, где \tilde{M}_{ij} – нечеткое множество с функцией принадлежности $\mu_{ij}(x_j)$ и определяющее нечеткое значение признака с лингвистическим именем α_{ij} :

$$\tilde{M}_{ij} = \{ \langle \mu_{ij}(x_j), x_j \rangle / x_j \in X_j \}.$$

Обозначим через B множество имен лингвистических переменных:

$$B = \{ \beta_j / j=1,n \}.$$

Определим множество возможных типовых нечетких ситуаций объекта прогнозной диагностики, как декартово произведение множеств $T_j (j=1,n)$:

$$S = T_1 \times T_2 \times \dots \times T_n.$$

Для построения правил прогнозной диагностики при опросе эксперта необходимо установить соответствие

между каждой нечеткой ситуацией из S и степенью риска выраженного элементами множества Y , т.е. определить сюръекцию:

$$\kappa: S \rightarrow Y.$$

Количество различных типовых ситуаций и вопросов к эксперту о риске возникновения нештатной ситуации (L) определяется выражением:

$$L = \prod_{j=1}^n m_j.$$

Опыт показывает, что даже при небольшом количестве ЛП и их нечетких значений количество вопросов, которые необходимо задать эксперту, достигает больших величин (например, при количестве ЛП равных 7 и количестве нечетких значений – 3 количество вопросов задаваемых эксперту равно 2187, и при среднем времени ответа на один вопрос равным три минуты превышает 100 часов).

Уменьшить количество задаваемых эксперту вопросов можно только, имея дополнительную информацию о свойствах отображения κ .

Для задачи прогнозной диагностики будем использовать следующее предположение:

На множестве значений T_j каждой лингвистической переменной j ($j=1, n$) определено отношение порядка R_j такое, что $(\alpha_{sj}, \alpha_{tj}) \in R_j$, если значение α_{tj} более характерно для возникновения нештатной ситуации, чем значение α_{sj} и построено отношение доминирования R на множество нечетких ситуаций S :

$$R = \{ (s_v, s_w) / (s_v, s_w) \in S \times S \ \& \ \forall j=1, n, (\alpha_{vj}, \alpha_{wj}) \notin R_j \ \& \ \exists k=1, n \ \& \ (\alpha_{vk}, \alpha_{wk}) \in R_k \}.$$

Ситуация s_w более характерна для возникновения нештатной ситуации чем ситуация s_v , если значение каждого признака α_{vj} ($j=1, n$) в ситуации s_v не является более характерным для возникновения нештатной ситуации, чем значение α_{wj} того же признака у ситуации s_w и существует хотя бы один признак k ($k=1, n$), значение которого α_{wk} в ситуации s_w является более характерным для возникновения нештатной ситуации, чем значение α_{vk} того же признака в ситуации s_v .

Тогда справедливо следующее утверждение:

$$\forall s \in S \ \forall s_w \in S ((s_v, s_w) \in R \rightarrow (\kappa(s_w) \geq \kappa(s_v))),$$

т.е., если ситуация s_w доминирует над ситуацией s_v , то и степень риска y_w ($y_w = \kappa(s_w)$) возникновения нештатной

ситуации в ситуации s_w будет не меньше, чем степень риска y_v ($y_v = \kappa(s_v)$) возникновения нештатной ситуации в ситуации s_v .

Обозначим границы вариации значений степени риска возникновения нештатной ситуации в ситуации s через y_s^- и y_s^+ :

$$y_s^- \leq y_s \leq y_s^+.$$

В начале опроса предположим, что для каждой ситуации из S возможно любое значение степени риска и $y_s^- = 0$, а $y_s^+ = 1$.

При опросе эксперта границы значений y_s^+ и y_s^- будут изменены. Пусть эксперт оценил риск возникновения нештатной ситуации в текущей ситуации s и оценил значение риска в y_s . Тогда для ситуаций s_w , которые являются доминирующими по отношению к ситуации s нижняя граница может быть увеличена до значения y_s ($\delta_{ws} = y_s - y_w^-$), а для ситуаций s_v , для которых ситуация s является доминирующей верхняя граница может быть уменьшена до величины y_s ($\delta_{vs} = y_v^+ - y_s$). Изменения вариаций значений степени риска возникновения нештатной ситуации для всех ситуаций из S , если эксперт установил степень риска НС для ситуации s , определяется значением выражения:

$$Q_s = \sum_{v \in S} \delta_{vs}.$$

Очевидно, что при очередном опросе эксперта, необходимо выбирать такую ситуацию s ($s \in S$), при которой величина Q_s была бы максимальной:

$$s = \arg(\max_{w \in S} Q_w)$$

Однако перед опросом эксперта значение y_s является неизвестной величиной с известными границами вариаций:

$$y_s^- \leq y_s \leq y_s^+.$$

Для оценки риска нештатной ситуации используется оценочное значение y_s^* :

$$y_s^* = (y_s^+ - y_s^-) / 2.$$

Для подтверждения применения предложенной методики перед практическим использованием были проведены эксперименты с имитацией [3] ответов эксперта. Результаты проведенных экспериментов приведены в таблице:

ТАБЛИЦА I РЕЗУЛЬТАТЫ ПРОВЕДЕННЫХ ЭКСПЕРИМЕНТОВ

Количество признаков	Количество типовых ситуаций	Требуемая точность, $y_s^+ - y_s^-$, $s \in S$	Среднее количество вопросов, заданных эксперту	Процент количества задаваемых вопросов, %
5	243	0.33	195	80
5	243	0.5	88	36
6	729	0.33	461	63
6	729	0.5	115	16
7	2187	0.33	1288	58
7	2187	0.5	232	11

Результаты опроса эксперта позволяют построить обучающую выборку, которую можно использовать для формирования нечетких правил прогнозной диагностики [4], [5].

Рассмотренная методика использовалась при построении экспертных систем определения групп риска новорожденных и прогнозирования патологий беременности [6].

СПИСОК ЛИТЕРАТУРЫ

[1] Tutubalin P.I., Mokshin V.V. The Evaluation of the cryptographic strength of asymmetric encryption algorithms / 2017 Second Russia and

Pacific Conference on Computer Technology and Applications (RPC). IEEE. 2017, P.180 – 183.

- [2] D. Rutkowska, M. Pilinski, L. Rutkowska. Sieci neuronowe, algorytmy genetyczne i systemy rozmyte. Wydawnictwo Naukowe PWN, Warszawa Lodz, 1999.
- [3] Yakimov, A. Kirpichnikov, V. Mokshin, Z. Yakhina, R. Gainullin. The comparison of structured modeling and simulation modeling of queueing systems. (Communications in Computer and Information Science, CCIS Vol. 800. Springer. 2017. P. 256-267
- [4] Suzdaltsev V.A., Chermoshencev S.F., Suzdaltsev I.V., Bogula N.Y. Formation of expert system decision explanation in the object classification process. В сборнике: Proceedings of the 19th International Conference on Soft Computing and Measurements, SCM 2016 19. 2016. P. 319-320.
- [5] Suzdaltsev V.A., Suzdaltsev I.V., Bogula N.Yu. Fuzzy rules formation for the construction of the predictive diagnostics expert system. В сборнике: Proceedings of 2017 20th IEEE International Conference on Soft Computing and Measurements, SCM 2017 20. 2017. P. 481-482
- [6] З.И. Аскарова, В.А. Суздальцев В.А., И.Х. Сабиров, А.А. Хасанов, «Прогнозирование группы риска беременных женщин по развитию аномалий сократительной деятельности матки». «Информационные технологии в науке, образовании и производстве: Материалы Всероссийской научной конференции. 30-31 мая 2007 года. Казань: Из-во Казан. Гос. Техн. Ун-та, 2007.»