

Т.И. Горбунова**ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ
В МЕДИЦИНСКОЙ ДИАГНОСТИКЕ**

Процесс принятия решений довольно сложный процесс, особенно в области медицинской диагностики. Приходится учитывать множество различных факторов, начиная с внешних симптомов больного и заканчивая его полным обследованием, чтобы установить точный диагноз. В работе рассматривается построение лингвистической модели, которая направлена на решение поставленной задачи.

Медицинская диагностика, лингвистическая модель, лингвистическая переменная, терм, построение модели, обучающая выборка, контрольная выборка

T.I. Gorbunova**INTELLIGENT DECISION SUPPORT SYSTEM IN MEDICAL DIAGNOSTICS**

Making decision is quite a difficult process especially when you are working with medical diagnostics. You should take into account enormous different factors starting with outward symptoms a patient has and finishing with the full medical examination of the patient. In this report is explained how to create a linguistic model that will allow diagnosing correct diagnoses to patients.

Medical diagnostic, linguistic model, linguistic variable, term, model creation, learning sample, check sample

В работе представлена информационная система, направленная на диагностику заболеваний по заданному набору симптомов.

Выбор заболевания – это сложный процесс, в т.ч. для квалифицированного врача. Приходится учитывать множество различных факторов, начиная с внешних симптомов (температура тела, кашель, ...) и заканчивая полным обследованием больного (анализ крови, рентген, томография, ...). Также очень важно учитывать заболевания, перенесенные больным, группу крови больного, наследственность, условия и образ жизни, реакцию на негативные воздействия [1]. Учет этих факторов позволит добиться более точного диагностирования заболевания и выбрать соответствующее лечение. В связи с этим период выздоровления больного будет намного сокращен и не будет нанесен дополнительный вред организму, вызванный неправильным лечением.

На данный момент существуют диагностические системы, предназначенные для идентификации заболевания по проявленным симптомам, т.е. по симптомам, которые испытывает больной в данный момент [2]. Однако, данные системы не учитывают дополнительные факторы, которые могут существенно влиять на корректную диагностику заболевания.

Исходные данные для медицинской диагностики заболевания можно представить следующим образом: каждому заболеванию $z_k \in Z$ соответствует набор показателей $x_i \in X$, таких что $X = \bigcup x_i$, $x_i \cap x_j \neq \emptyset$. К показателям относятся симптомы заболеваний, результа-

ты анализов крови, аппаратные исследования, характеристики состояния здоровья пациента, такие как перенесенные ранее заболевания, наследственность, образ и условия жизни. Диагностика усложняется тем, что одни показатели могут соответствовать разным заболеваниям. Задача интеллектуальной диагностической системы заключается в определении заболевания z_k , которому соответствует определенный набор показателей:

$$\bigcap_{i=1}^m x_i \longrightarrow z_k \in Z. \quad (1)$$

Предлагается использовать лингвистическую модель для построения модели диагностики заболевания, в которой зависимости входных показателей и заболевания связаны совокупностью нечетких правил: «ЕСЛИ конъюнкция показателей, часть из которых принимает нечеткие значения, ТО наиболее вероятное заболевание».

Выбор лингвистической модели в качестве модели диагностирования обусловлен в первую очередь тем, что входные показатели представляют разнотипные переменные. Кроме того, определенный набор показателей может соответствовать нескольким заболеваниям. Например, высокая температура может соответствовать как воспалению легких, так и гриппу. Однако степень данного соответствия будет различной. В связи с этим невозможно использовать логические модели, которые ставят четкое соответствие между симптомами и диагнозом.

Моделируемую лингвистическую модель можно представить в следующем виде:

$$\bigcap_{i=1}^n \mu_{a_{ji}}(x_i) \longrightarrow \mu_{b_{jk}}(y), \quad j_i = 1, \dots, l_i^x, \quad j_k = 1, \dots, l_k^y, \quad (2)$$

где n – число показателей из множества X ; a_{ji} – j -й терм i -го входного показателя; l_i^x – число термов i -го входного показателя; b_{jk} – j -й терм k -й выходной переменной – заболевания; l_k^y – число термов k -й выходной переменной; $\mu_c(z)$ – обозначение функции принадлежности значения базовой переменной z терму c .

Как отмечалось ранее, определенный набор показателей может соответствовать нескольким заболеваниям. Однако следует понимать, что выявление более сложного заболевания (например, у которого могут быть сложные последствия) на раннем этапе диагностирования является наилучшим результатом. В связи с этим предлагается упорядочить элементы множества заболеваний Z по важности диагностирования определенного заболевания. Упорядочивание заболеваний позволит моделируемой системе не только быстро выявить искомое заболевание, но и определить заболевание, риск осложнений у которого максимален.

Процедура построения модели по заданным входным параметрам, представляет собой последовательное выполнение следующих шагов:

1. Выделяется обучающая выборка данных у которой определенный набор показателей соответствует определенному заболеванию, диагностированным квалифицированным врачом.

2. На основе обучающей выборке данных происходит разбиение входных и выходных показателей на термы. Разбиение на термы осуществляется одним из методов кластеризации, описанном в [3].

3. В качестве настроечного критерия для формирования лингвистической модели вычисляется информационная мера на основе дивергенции Кульбака и её дисперсии:

$$d(c_{js}) = q_{js} \cdot (1 - \tilde{q}_{js}), \quad (3)$$

где q_{js} , $\tilde{q}_{js}$ – оценки вероятностей, соответственно, различения и неразличения конъюнкцией

ей $c_{js} = \bigwedge_{i=1}^n a_{ki}$ модели выходного терма b_j .

4. Оценивается степень соответствия сформированной модели реальной системе, представленной в обучающей выборке, на основе информационного критерия:

$$J(e_M) = \ln(st(\delta)) - \frac{c_M}{N} \sum_{i=1}^N \sum_{j=1}^{m^o} \ln \left(\frac{|k_{ji} - \tilde{k}_{ji}|}{m^o} \right), \quad k_{ji} = \rho(b_{ji}), \quad \tilde{k}_{ji} = \rho(\tilde{b}_{ji}), \quad (4)$$

где $st(\delta)$ – показатель устойчивости оценок критерия; k_{ji} , $\tilde{k}_{ji}$ – расстояния в заданной метрике $\rho(\cdot)$ действительного b_{ji} и оцениваемого $\tilde{b}_{ji}$ терма выходной переменной от начала отсчета; c_M – параметр регуляризации модели, предназначенный для повышения её обусловленности; m^o – количество термов выходной переменной.

5. Адекватность работы сформированной модели проверяется на основе контрольной выборке данных, которая не участвовала в обучении.

6. Если степень соответствия построенной модели к реальной системе, представленной в контрольной выборке, мала, то происходит изменение параметров кластеризации входных и выходных лингвистических переменных и перерасчет информационных критериев.

7. При получении необходимой степени соответствия разрабатываемой модели и реальной системы, делается вывод о соответствии входного набора показателей определенному заболеванию.

Анализ работоспособности спроектированной модели проведен нами на основе пяти медицинских карточек больных, которым был поставлен конечный диагноз врачом. На основе карточек было выявлено, что врач смог поставить конечный диагноз для первого больного через 7 дней (после его первого посещения), для второго больного через 5 дней, для третьего больного через 7 дней, для четвертого больного через 3 дня и для пятого больного врач поставил конечный диагноз сразу при первом посещении. На постановку конечного диагноза для первого и третьего больного сильное влияние оказал тот факт, что больные являлись заядлыми курильщиками. Однако этот дополнительный параметр выявлен только через несколько дней после первого посещения. В связи с этим постановка конечного диагноза заняла много времени.

На основе медицинских карточек больных сформированы симптомы, которые введены в спроектированную модель. Кроме того, все дополнительные параметры, такие как “курит больной” или “нет”, как “часто занимается спортом”, “наличие близких родственников, больных сахарным диабетом”, также поданы на вход модели.

Экспериментальные проверки доказали работоспособность разработанной модели: во всех пяти случаях диагноз, спрогнозированный моделью совпал с диагнозом, поставленным квалифицированным врачом. При этом процесс диагностирования занял сравнительно меньше времени.

ЛИТЕРАТУРА

1. Амосов Н.М. Раздумья о здоровье. М.: АСТ, Астрель, 2005. 287 с.
2. Газарян В.А., Нагорный Ю.М., Пытьев Ю.П. О теоретико-возможностном методе медицинской диагностики // Математические методы распознавания образов : сб. докл. в 13-я Всерос. конф. М.: МАКС Пресс, 2007. 668 с.
3. А.К. Jain, M.N. Murty, P.J. Flynn. Data Clustering: A Review.
4. Токарев В.Л. Основы теории обеспечения рациональности решений. Тула: ТулГУ. 120 с.
5. Андрейчиков А.В., Андрейчикова О.Н. Интеллектуальные и информационные системы. М.: Финансы и статистика, 2004. 424 с.
6. Борисов В.В., Круглов В.В., Федурлов А.С. Нечеткие модели и сети. М.: Горячая линия-Телеком, 2007. 284 с.

Горбунова Татьяна Игоревна –

аспирант кафедры «Электронные вычислительные машины» Тульского государственного университета

Статья поступила в редакцию 7.03.12, принята к опубликованию 12.03.12