

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ К ЗАДАЧАМ В ОБЛАСТИ МЕДИЦИНЫ

Молодченков А.И., Руденкова Ю.С.

Федеральное государственное бюджетное учреждение науки Институт системного анализа
Российской академии наук, aim@isa.ru

Федеральное государственное бюджетное образовательное учреждение высшего
профессионального образования «Российский университет дружбы народов»,
julia.rudenkova@gmail.com

Аннотация. В работе представлен обзор методов машинного обучения и их применение для решения различных задач в области медицины.

Ключевые слова: искусственный интеллект, машинное обучение, медицина.

Введение

Последние достижения в области информационных технологий (ИТ) создали новые возможности как для врачей, так и для пациентов. Именно эти возможности привели к фундаментальной перестройке клинических процессов благодаря интеграции электронных средств связи на всех уровнях. Медицинские ИТ создают возможность превращения пациентов из пассивных реципиентов медицинской помощи в активных, информированных участников процессов принятия клинических решений. Задача новых информационных технологий заключается в улучшении и обогащении взаимодействия врачей и пациентов.

С развитием как ИТ, так и медицины, изменились и требования к системам, предназначенным для поддержки медицинских технологических процессов. Если лет десять назад было достаточно, если компьютерная система делает прогноз или осуществляет поддержку принятия решения в некоторый момент времени выполнения процесса лечения, то современные системы должны осуществлять поддержку сложных и чаще всего мультидисциплинарных медицинских технологических процессов. Одной из задач при разработке подобных систем является разработка алгоритмов их обучения. Основными источниками знаний для систем поддержки медицинских технологических процессов являются эксперты (практикующие врачи), медицинские базы данных, клинические руководства.

Целью этой работы является обзор существующих методов машинного обучения, использующих различные источники знаний и чаще всего применяемых для решения различных медицинских задач, и анализ их применения в области медицины.

Работа выполнена в рамках проекта РФФИ 13-07-12165.

Методы машинного обучения

Методы машинного обучения классифицируются следующим образом:

1) по способам обработки данных:

- статистические (Байесовский классификатор, Метод опорных векторов SVM, Метод K ближайших соседей (K-nearest neighbors), регрессии);
- логические (Метод покрытый AQ, Деревья решений ID3, ДСМ-метод порождения гипотез);
- эволюционные (Нейронные сети, генетические алгоритмы).

2) По предоставляемым данным:

- С учителем (Supervised).
- Без учителя (Unsupervised).

Байесовский классификатор.

Широкий класс алгоритмов классификации, основанный на теореме, утверждающей, что если плотности распределения каждого из классов известны, то искомым алгоритм можно выписать в явном аналитическом виде. Для классифицируемого объекта вычисляются функции правдоподобия каждого из классов, по ним вычисляются апостериорные вероятности классов. Объект относится к тому классу, для которого апостериорная вероятность максимальна. Более того, этот алгоритм обладает минимальной вероятностью ошибок. Байесовский классификатор относится к методам машинного обучения "с учителем", но в отличие от перцептрона и других подобных моделей, он не требует длительной процедуры обучения [1].

Нейронные сети.

Появление искусственных нейронных сетей индуцировано биологией: они состоят из элементов, функциональные возможности которых напоминают функциональные возможности биологических нейронов. Эти элементы затем организируются по способу, который, с некоторой точки зрения, может соответствовать анатомии мозга. Практически все разновидности нейронных сетей могут быть реализованы на обычном универсальном компьютере.

Наиболее распространенной областью применения искусственных нейронных сетей является распознавание образов [2]. Сложные конфигурации искусственных нейронных сетей (когнитроны) позволяют распознавать сильно зашумленный и трансформированный образ. Сети ДАП [двунаправленная ассоциативная память] можно использовать для сжатия входного сигнала и последующего однозначного восстановления.

Алгоритм ID 3.

В самом простом виде дерево решений - это способ представления правил в иерархической, последовательной структуре. Основа такой структуры - ответы "Да" или "Нет" на ряд вопросов.

Один из наиболее ранних алгоритмов обучения деревьев решений рекурсивно разбивает подмножества в узлах дерева по одному из выбранных атрибутов. Разбиение начинается с корня дерева, в котором содержатся все примеры обучающего множества. Для разделения в нем выбирается один из атрибутов, и для каждого принимаемого им значения строится ветвь, и создается дочерний узел, в который распределяются все содержащие его записи. Процедура повторяется рекурсивно до тех пор, пока в узлах не останутся только примеры одного класса, после чего они будут объявлены листьями и ветвление прекратится. Наиболее проблемным этапом здесь является выбор атрибута, по которому будет производиться разбиение. Классический алгоритм ID3 использует для этого критерий увеличения информации или уменьшения энтропии.

Метод опорных векторов SVM.

Основная идея метода [3] — перевод исходных векторов в пространство более высокой размерности и поиск разделяющей гиперплоскости с максимальным зазором в этом пространстве. Две параллельных гиперплоскости строятся по обеим сторонам гиперплоскости, разделяющей наши классы. Разделяющей гиперплоскостью будет гиперплоскость, максимизирующая расстояние до двух параллельных гиперплоскостей. Алгоритм работает в предположении, что чем больше разница или расстояние между этими параллельными гиперплоскостями, тем меньше будет средняя ошибка классификатора [3].

Применение методов машинного обучения в медицине

Методы машинного обучения нашли свое применение в различных областях медицины: онкология [4-8], патология печени [9], прогнозирование [6], урология [6, 7], диагностика [8-15], ревматология [11], кардиология [12] и др. Все эти алгоритмы и

методы предназначены для решения широкого круга задач. Например, для решения задач мониторинга, предупреждения, напоминания используются методы, основанные на правилах; моделирования и прогнозирования – калькуляторы, статистическое моделирование; уточнения, поиска информации – методы анализа текстов, навигаторы; фрейминга и принятия решений – алгоритмы принятия решений, логические модели решений (например, нейронные сети, байесовский классификатор), анализа медицинских данных и изображений – нейронные сети, методы интеллектуального анализа данных.

Чаще всего существующие методы машинного обучения работают только с медицинскими данными, хранящимися в базах данных. Однако для современных систем поддержки медицинских технологических процессов этого недостаточно. Необходимо использовать знания экспертов и информацию, изложенную в клинических руководствах.

В современных медицинских системах клинические руководства используются путем их формализации. Эта формализация заключается в том, что клинические руководства размечаются и создаются системы, позволяющие врачу быстро находить интересующую его информацию по сделанной разметке. Для извлечения знаний от эксперта чаще всего используется метод «интервью» [13]. При использовании этого подхода для работы с экспертом привлекается инженер по знаниям. Однако существенным недостатком существующих систем машинного обучения, применяемых в медицине, является то, что они решают очень узкий круг задач и нет систем, позволяющих осуществлять поддержку сложных и мультидисциплинарных медицинских технологических процессов.

Выводы

Задача создания методов позволяющих разработать компьютерную систему, осуществляющую поддержку широкого числа этапов процесса лечения, является актуальной. Для решения этой задачи необходимо провести адаптацию различных методов искусственного интеллекта, позволяющих осуществлять обучение компьютерных систем и поддержку как простых, так и сложных медицинских технологических процессов в различных областях медицины и здравоохранения.

Литература

1. Осипов Г.С. Методы искусственного интеллекта. – М.: ФИЗМАТЛИТ, 2011. – 296 с.
2. Талалаев А.А., Тищенко И.П., Фраленко В.П., Хачумов В.М. Анализ эффективности применения искусственных нейронных сетей для решения задач распознавания, сжатия и прогнозирования. – Искусственный интеллект и принятие решений, №2, 2008, с.24-33.
3. Osuna E., Freund R., Girosi F. Support vector machines: Training and applications: Tech. Rep. AIM-1602: 1997.
4. Bratko I., Mulec P., An Experiment in Automatic Learning of Diagnostic Rules, *Informatica*, Ljubljana, Vol. 4, No. 4, 1980, pp. 18-25.
5. Zwitter M., Bratko I., Kononenko I., Rational and irrational reservations against the use of computer in medical diagnosis and prognosis, *Proc. 3rd Mediterranean Conf. on Medical and Biological Engineering*, Portoroz, Slovenia, Sept. 5-9 1983.
6. Kononenko I., Bratko I., Roskar E., Experiments in automatic learning of medical diagnostic rules, *International School for the Synthesis of Expert's Knowledge Workshop*, Bled, Slovenia, August 1984
7. Bratko I., Kononenko I., Learning Rules from Incomplete and Noisy Data, in B. Phelps (ed.), *Interactions in Artificial Intelligence and Statistical Methods*, Hampshire: Technical Press, 1987.

8. Elomaa T., Holsti N., An Experimental Comparison of Inducing Decision Trees and Decision Lists in Noisy Domains, *Proc. 4th European Working Session on Learning*, Montpellier, Dec. 4-6 1989, pp. 59-69.
9. Lesmo L., Saitta L., Torasso P., Learning of Fuzzy Production Rules for Medical Diagnoses, In Gupta M.M., Sanchez E.(eds.) *Approximate Reasoning in Decision Analysis*, North-Holland, 1982.
10. Igor Kononenko, Ivan Bratko and Matjaz Kukar *Application of Machine Learning to Medical Diagnosis / Machine Learning and Data Mining: Methods and Applications*. Edited by R.S. Michalski, I. Bratko, and M. Kubat, 1997, John Wiley & Sons Ltd, pp 16-23.
11. Kononenko I., Jauk A., Janc T., Induction of reliable decision rules, *International School for the Synthesis of Expert's Knowledge Workshop*, Udine, Italy, 10-13 Sept. 1988.
12. Bratko I., Mozetic I., Lavrac N., *KARDIO: A Study in Deep and Qualitative Knowledge for Expert Systems*, Cambridge, MA: MIT Press, 1989.
13. Г.С. Осипов. Приобретение знаний интеллектуальными системами: Основы теории и технологии. – М.: Наука. Физматлит, 1997, - 112 с.

APPLICATION OF MACHINE LEARNING METHODS FOR TASKS IN MEDICINE

Molodchenkov A.I., Rudenkova J.S.

Institution of Russian Academy of Sciences Institute for systems analysis, aim@isa.ru

People's Friendship University of Russia, julia.rudenkova@gmail.com

Abstract. The paper presents a review of machine learning methods and their application to solve different tasks in medicine.

Key words: artificial intelligence, machine learning, medicine.