

Применение искусственных нейронных сетей для классификации раковых клеток

Лисица Е.В.¹; Яцков Н.Н.¹; Апанасович Т.В.²; Апанасович В.В.¹.

1) Кафедра системного анализа и компьютерного моделирования
Белорусский государственный университет
Минск, Республика Беларусь

2) Thomas Jefferson University, Jefferson Medical College, Philadelphia, USA
e-mail: ylisitsa@gmail.com, yatskou@bsu.by, Tatiyana.Apanasovich@jefferson.edu

Аннотация—В работе исследуются алгоритмы классификации на основе искусственных нейронных сетей с использованием смоделированных данных на устойчивость алгоритмов к размеру обучающей выборки и уровню шума.

Ключевые слова: классификация; нейронные сети; моделирование

III. ВВЕДЕНИЕ

Методы обработки и анализа данных люминесцентной микроскопии получили широкое распространение в цитологических методах диагностики [1], [2], [3]. Однако анализ люминесцентных изображений, получаемых в результате исследования, проводится только интегрально, что не позволяет достигнуть необходимой точности исследования [5]. Внедрение автоматических методов анализа усложняется различными особенностями проведения эксперимента, такими как метод изготовления образца [6], способ регистрации изображения [7][8], методы обработки изображения [9].

Подобное разнообразие независимых факторов проявляющихся одновременно в одном эксперименте не позволяет отобрать наиболее устойчивые алгоритмы к определённым, самым часто-встречаемым воздействиям [10]. Однако данная проблема может быть успешно решена, используя имитационную модель. Имитационное моделирование позволяет создавать упрощенные наборы данных максимально приближенные к экспериментальным результатам и нигилировать наименее важные экспериментальные признаки, усилить влияние наиболее интересных факторов [11]. Второе преимущество в использовании имитационной модели – объективная заранее известная информация независимая от экспертов о структуре данных. Использование модели позволяет задать точное количество клеток и их тип, и их параметры [12]. В третьих, компьютерная имитация позволяет исследовать условия, которые не достижимы в эксперименте [12].

Методы статистической обработки информации можно использовать в цитологии для решения задачи классификации, когда разбиение объектов на группы проходит с учителем (тренировочной заранее известной выборкой экспериментальных данных), и кластеризации (обучение без учителя) [13]. Преимущества использования искусственных нейронных сетей (ИНС) для решения задачи

статистической обработки данных перечислены ниже: а) решение задач при неизвестных закономерностях; б) устойчивость к шумам во входных данных; в) приспособляемость к изменениям окружающей среды; г) потенциальное сверхвысокое быстродействие [14].

Однако несмотря на такое количество преимуществ, сам процесс обучения нейронной сети скрыт, таким образом использование ИНС не является безопасным и требует предварительного детального исследования.

IV. ТЕОРЕТИЧЕСКОЕ ОПИСАНИЕ

Искусственная нейронная сеть – это математическая модель, построенная по принципу организации и функционирования биологических нейронных сетей, т.е. ИНС представляют собой соединённую систему простых процессов [14]. Наиболее популярные и изученные архитектуры – это многослойный перцептрон, нейронная сеть с общей регрессией, нейронные сети Кохонена и другие. Для обучения сети часто используется метод обратного распространения ошибки (Error Back Propagation) [14]. В работе исследованы следующие виды нейронных сетей:

1. Многослойный перцептрон – нейронная сеть без обратных и латеральных (между элементами одного слоя) связей. В каждом из слоев вычисляется взвешенная сумма сигналов предыдущего слоя нейронов (или входных сигналов x_i) [15].
2. Слой Кохонена – представляет собой один слой адаптивных линейных сумматоров, работающих по принципу WTA – Winner Takes All, или «Победитель забирает все». Т.е. нейрон, имеющий наибольший сигнал на входном векторе, идентифицирует класс, к которому нейронная сеть относит этот входной вектор. Если требуется разбить множество m -мерных векторов на k кластеров, то нужно построить сеть с m входными нейронами и k выходными нейронами, затем обучить ее на предложенном множестве [16],[17].
3. Радиально-базисная нейронная сеть (РБФ-сеть), используется в качестве активационных функций радиально-базисную функцию вида

$$f(x) = \exp(-n^2) \quad (1)$$

и обладает следующим свойствами: а) единственный скрытый слой; б) только нейроны скрытого слоя имеют нелинейную активационную функцию; в) синаптические

веса связей входного и скрытого слоев равны единице [18].

V. ПОСТАНОВКА ЭКСПЕРИМЕНТА

Для исследования устойчивости методов классификации был сгенерирован набор из 5 независимых изображений с различным количеством объектов-клеток на каждом из них – (222, 369, 705, 1276, 1773, рис. 1).

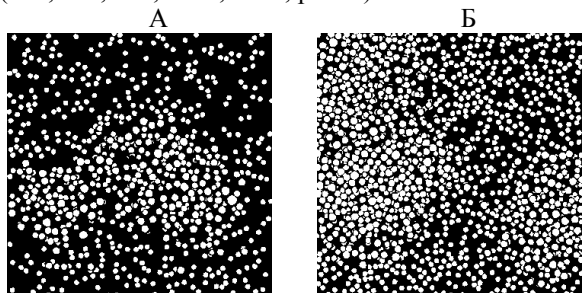


Рис.1. Маски ядер клеток смоделированных изображений А)1276, Б) 1773

Для смоделированных изображений с помощью методов сегментации были построены маски ядер и цитоплазм. С каждого ядра были сняты характеристики подробно описанные в работе [18].

Для изучения воздействия шума к данным добавлялся белый шум, со среднеквадратичным отклонением, рассчитанным по формуле (2) [8]:

$$SNR = \frac{mean}{std} \quad (2)$$

Соотношение сигнал к шуму варьировалось от $5 \cdot 10^{-3}$ до 1. Для оценки качества работы методов использовались матрицы ошибок (1, 2 рода и общая ошибка) для различных уровней шума и различных размеров тренировочных выборок, где ошибка первого рода характеризует, что «здоровая» клетка была отнесена к классу больных клеток, а ошибка второго рода показывает, когда больные клетки были проклассифицированы как здоровые.

VI. РЕЗУЛЬТАТЫ

В ходе проведения исследования установлено, что наилучшие результаты получены для радиально базисной нейронной сети (рис. 2). Результаты классификации с использованием слоя Кахонена и RBF-сети сопоставимы, но время классификации у RBF-сети значительно меньше. Наихудшие результаты классификации получены в случае

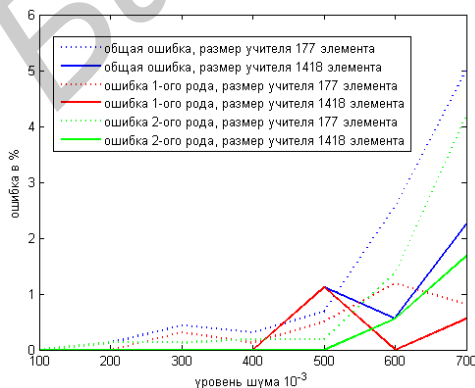


Рис.2. Результаты тестирования.

применения многослойной нейронной сети. Согласно полученным результатам точность классификации увеличивается при снижении шума в данных и при увеличении размера обучающего множества.

- [1] Абламейко С.В., Лагуновский Д.М. Обработка изображений: технология, методы, применение. Мн.: «Амалфея». 2000. стр. 304.
- [2] Материалы международной научно-практической конференции «Информационные технологии, электронные приборы и системы» (ITEDS'2010) 6–7 апреля 2010 г. Минск. БГУ.
- [3] Demeter J, Beauheim C, Gollub J, Hernandez-Boussard T, Jin H, Maier D, Matese JC, Nitzberg M, Wymore F, Zachariah ZK, et al.
- [4] The Stanford Microarray Database: implementation of new analysis tools and open source release of software // *Nucleic Acids Res.* -2007. -С. D766-770.
- [5] Hoheisel JD Microarray technology: beyond transcript profiling and genotype analysis. *Nat Rev Genet.* -2006. -№7(3). -С. 200-210.
- [6] Alberts B., Johnson A., Lewis J., Raff M., Roberts K., Walter P. *Molecular Biology of the Cell.*- Garland Science.-2007-С. 1392
- [7] Феофанов А. В. Спектральная лазерная сканирующая конфокальная микроскопия в биологических исследованиях. *Успехи биологической химии.* -2007. -№. 47. -С. 371-410.
- [8] Bushberg J.T., Leidholdt E.M., Boone J.M. *The essential physics of medical imaging.* - Philadelphia: Lippincott Williams & Wilkins, 2001. - 933 с.
- [9] Гонсалес Р., Вудс Р., Эддинс С. Обработка изображений в среде MATLAB. -М.: Техносфера. -2006. –616 с.
- [10] Ronneberger O., Baddeley D., Scheipl F., Vermeer P. J., Burkhardt H., Cremer C., Fahrmeir L., Cremer T. and Joffe B. Spatial quantitative analysis of fluorescently labeled nuclear structures: Problems, methods, pitfalls// *Chromosome Research.* -2008. -№16. -С.523-562
- [11] A. Lehmussola, P. Ruusuvoori, J. Selinummi, H. Huttunen, O. Yli-Harja, “Computational framework for simulating fluorescence microscope images with cell populations”. 2007, *IEEE transactions on medical imaging*, pp. 1010-1016.
- [12] Novikov E, Barillot E. An algorithm for automatic evaluation of the spot quality in two-color DNA microarray experiments. *BMC Bioinformatics.* -2005. № 6:293.
- [13] Лагутин М.Б. Наглядная математическая статистика. Учебное пособие. М.: БИНОМ. Лаборатория знаний, 2007, с. 472.
- [14] Лутковский В.М. Нейронные сети. Минск, БГУ, 2003. – 99 с.
- [15] Newman A.M, Cooper J.B., AutoSOME: a clustering method for identifying gene expression modules without prior knowledge of cluster number. *BMC Bioinformatics.* -2010–Vol.11:117.
- [16] Hasan S, Shamsuddin SM., Multistrategy self-organizing map learning for classification problems.// *Comput Intell Neurosci* – 2011.
- [17] Jrad N, Grall-Maës E, Beuseroy P. Gene-based multiclass cancer diagnosis with class-selective rejections// *J Biomed Biotechnol.* -2009
- [18] Хайкин С. Нейронные сети. Полный курс. 2-е изд., испр.: М.: ООО «И.Д.Вильямс», 2006. – 1104 с.
- [19] Е.В. Лисица, Н.Н. Яцков, Т.В. Апанасович, Х. Ри, В.В. Апанасович. *Разработка метода оценки формы границ сегментированных объектов по люминесцентным изображениям живых клеток.* Материалы конференции молодых учёных и студентов «Современные проблемы математики и вычислительной техники». Том 2. 2011 Брест.