



3. Arefyev, N.V. Evaluating Three Corpus based Semantic Similarity Systems for Russian / N.V. Arefyev, A.I. Panchenko, A.V. Lukanin // Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference Dialogue 2015 (Moscow, RGGU) . – 2015. – Vol. 2. – P. 106–118.

4. A. Panchenko A. The First Workshop on Russian Semantic Similarity // Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference Dialogue 2015 (Moscow, RGGU). – 2015. – Vol. 2. – P. 89–105.

А.И. Иванов

### СИНТЕЗ НОВЫХ ИНТЕГРО-ДИФФЕРЕНЦИАЛЬНЫХ СТАТИСТИЧЕСКИХ КРИТЕРИЕВ И ЭКВИВАЛЕНТНЫХ ИМ ИСКУССТВЕННЫХ НЕЙРОНОВ ДЛЯ МАЛЫХ ВЫБОРОК

(АО «Пензенский научно-исследовательский электротехнический институт»)

К сожалению, созданные в 20 веке статистические критерии для проверки гипотез нормальности или равномерности данных, ориентированы на большие выборки. В биометрии, медицине, экономике, во многих случаях, приходится работать с малыми выборками, например, в 16 опытов. В связи с этим возникла задача параллельного использования множества разных статистических критериев [1, 2] при анализе одной малой выборки. Каждому классическому критерию создается эквивалентный нейрон. Чем больше статистических критериев (эквивалентных искусственных нейронов) используется, тем выше достоверность, принимаемых решений, обобщающей сетью искусственных нейронов. При этом желательно синтезировать новые статистические критерии. Для синтеза может быть использован классический критерий Джинни, тогда новые критерии создаются дифференцированием входных случайных данных.

Таблица 1. Классический критерий Джинни и три его новых дифференциальных аналога

1	$D = \int_{-\infty}^{\infty}  \tilde{P}(x) - P(x)  dx$	$P_{EE} = 0.423$	$\text{corr}(D, dD) = -0.029$ $\text{corr}(D, d^2D) = -0.0047$ $\text{corr}(D, d^3D) = 0.0367$ $\text{corr}(dD, d^2D) = 0.894$ $\text{corr}(dD, d^3D) = 0.805$ $\text{corr}(d^2D, d^3D) = 0.888$
2	$dD = \int_{-\infty}^{\infty}  \tilde{p}(x) - p(x)  dx$	$P_{EE} = 0.039$	
3	$d^2D = \int_{-\infty}^{\infty} \left  \frac{d(\tilde{p}(x))}{dx} - \frac{d(p(x))}{dx} \right  dx$	$P_{EE} = 0.047$	
4	$d^3D = \int_{-\infty}^{\infty} \left  \frac{d^2(\tilde{p}(x))}{(dx)^2} - \frac{d^2(p(x))}{(dx)^2} \right  dx$	$P_{EE} = 0.082$	
	.....		



Из таблицы 1 видно, что новые дифференциальные аналоги уменьшают вероятности ошибок первого и второго рода  $P_1=P_2=P_{EE}$  (наблюдаем снижение от 10 до 5 раз по сравнению с их критерием-родоначальником).

Программная реализации на языке MathCAD первого дифференциального аналога - dD и результаты численного моделирования приведены на рисунке 1.

Технический прием по переходу от сравнения теоретической функций вероятности с экспериментальной функцией к сравнению их плотностей распределения и их производных более высоких порядков применим и для других статистических критериев. В таблице 2 размещены данные, отражающие характеристики дифференциальных вариантов критерия Ватсона.

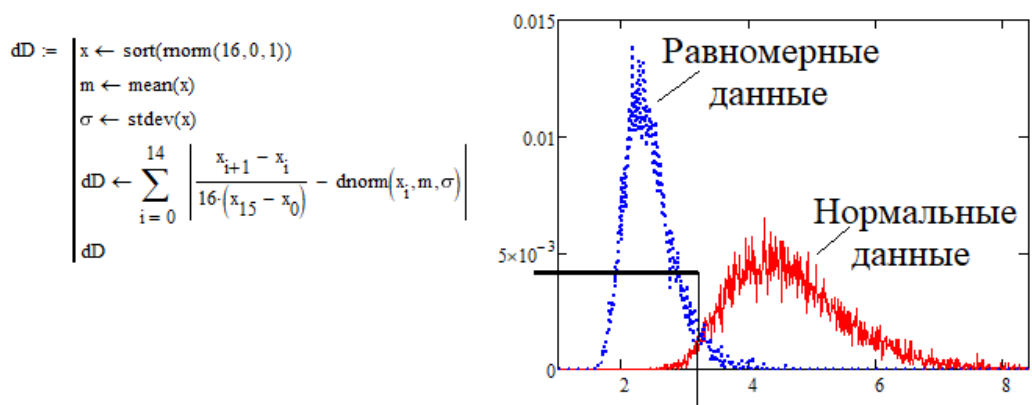


Рис. 1. Дифференциальный аналог - dD классического интегрального критерия Джинни

Таблица 2. Классический критерий Ватсона и три его новых дифференциальных аналога

1	$W = \int_{-\infty}^{\infty} \left\{ \tilde{P}(x) - P(x) - \int_{-\infty}^x \{ \tilde{P}(x_1) - P(x_1) \} dx_1 \right\}^2 dx$	$P_{EE}=0.167$
2	$dW = \int_{-\infty}^{\infty} \left\{ \tilde{p}(x) - p(x) - \int_{-\infty}^x \{ \tilde{p}(x_1) - p(x_1) \} dx_1 \right\}^2 dx$	$P_{EE}=0.037$
3	$d^2W = \int_{-\infty}^{\infty} \left\{ \frac{d(\tilde{p}(x))}{dx} - \frac{d(p(x))}{dx} - \int_{-\infty}^x \left\{ \frac{d(\tilde{p}(x_1))}{dx_1} - \frac{d(p(x_1))}{dx_1} \right\} dx_1 \right\}^2 dx$	$P_{EE}=0.043$
4	$d^3W = \int_{-\infty}^{\infty} \left\{ \frac{d^2(\tilde{p}(x))}{d^2x} - \frac{d^2(p(x))}{d^2x} - \int_{-\infty}^x \left\{ \frac{d^2(\tilde{p}(x_1))}{d^2x_1} - \frac{d^2(p(x_1))}{d^2x_1} \right\} dx_1 \right\}^2 dx$	$P_{EE}=0.064$
	.....	

Таким образом, число сегодня применяемых статистических критериев может быть существенно увеличено за счет перехода от обычных критериев к использованию их интегро-дифференциальных аналогов.

### Литература

1. Иванов А.И., Банных А.Г., Безяев А.В Искусственные молекулы, собранные из искусственных нейронов, воспроизводящих работу классических



статистических критериев. // Вестник пермского университета. Серия: Математика. Механика. Информатика. 2020 № 1 (48), с.26-32

2. Иванов А.И. Искусственные математические молекулы: повышение точности статистических оценок на малых выборках (программы на языке MathCAD): препринт // Пенза, из-во «Пензенского государственного университета», 2020 г., 36 с. ISBN 978-5-907262-42-3.

Сардор Каримов Илхом угли

## МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ УРОЖАЙНОСТИ С ИСПОЛЬЗОВАНИЕМ СПУТНИКОВЫХ ИЗОБРАЖЕНИЙ SENTINEL-2

(Санкт-Петербургский государственный университет)

**Аннотация.** Прогнозирование изменчивости урожая в пределах поля может помочь фермерам принимать правильные решения в различных ситуациях. Текущие достижения в области дистанционного зондирования и доступность изображений высокого разрешения, высокой частоты и бесплатных изображений Sentinel-2 улучшают внедрение точного земледелия для более широкого круга фермеров.

**Ключевые слова.** Deep learning, XGBoost, LightGBM, зондирования, Sentinel

### **Введение**

Цель этой статьи — создать модель, способную оценить урожайность полей в Восточной Азии. Дан временной ряд изображений Sentinel 2 и климатических переменных. Модель сможет оценить пространственную изменчивость урожайности зерна кукурузы в тоннах на акр.

Данные Sentinel-2 открывают новые возможности для регионального, а также глобального сельскохозяйственного мониторинга, позволяя просматривать Землю в 12 спектральных диапазонах с пространственным разрешением 10–20 м, с глобальным охватом и 5-дневной периодичностью повторных посещений и совместимыми с текущие и исторические миссии Landsat. Мониторинг свойств почвы и состояния посевов, наряду с картированием обработки почвы, помогает исследователям и фермерам оценивать землепользование, прогнозировать урожай, отслеживать сезонные изменения и помогать в реализации политики устойчивого развития. С ростом числа доступных источников спутниковых данных, многие из которых можно использовать бесплатно, потенциал огромен (Sentinel-hub, 2021) [1].