

© 2024 г. М.М. ЗУЕВА (m.zueva@hse.ru),
С.О. КУЗНЕЦОВ, д-р физ. мат. наук (skuznetsov@hse.ru)
(Научно-исследовательский университет “Высшая школа экономики”, Москва)

ИНДЕКСЫ ИНТЕРЕСНОСТИ ДЛЯ ПОСТРОЕНИЯ НЕЙРОННЫХ СЕТЕЙ НА ОСНОВЕ РЕШЕТОК ПОНЯТИЙ¹

Трудность интерпретации результатов работы нейронных сетей является насущной проблемой, решению которой уделяется много внимания. Нейронные сети, основанные на решетках понятий, представляют собой перспективное направление в данной области. Отбор понятий для построения нейронной сети ключевым образом влияет на качество ее работы. Средством отбора понятий могут являться индексы интересности, когда для построения нейронной сети используются понятия с наибольшими показателями определенного индекса. В статье исследуется влияние выбора индекса интересности как средства отбора формальных понятий на качество работы нейронной сети.

Ключевые слова: архитектура нейронной сети, анализ формальных понятий, индексы интересности, нейронные сети на основе решеток понятий.

DOI: 10.31857/S0005231024030044, EDN: TWSZEY

1. Введение

Сложность интерпретации результатов при работе с нейронными сетями является важной проблемой, которой в последнее время посвящено много научных работ. Одним из возможных решений является построение нейронной сети на основе решеток понятий (concept lattice). В [1] была представлена нейронная сеть с архитектурой, построенной на основе решетки понятий для повышения устойчивости классификации. В [2] был предложен метод построения нейронной сети из решеток понятий, которые строились как на основе монотонных, так и антимонотонных соответствий Галуа.

Так как количество понятий растет экспоненциально с размером входных данных, важной задачей является возможность уменьшения количества понятий для построения нейронной сети без потери качества (ее) работы. Это можно сделать двумя способами: за счет отбора наиболее значимых признаков (предобработка) и за счет отбора наиболее важных (“интересных”) понятий (постобработка). В [3] были рассмотрены различные методы отбора “интересных” понятий, основанных на “индексах интересности”. В [4] меры интересности понятий сравнивались по таким аспектам, как эффективность вычисления и возможность их применимости к зашумленным данным.

В данной работе проведено исследование четырех индексов интересности (*basic level*, *target entropy*, Δ -*stability* и *lift*) в качестве критериев отбора инте-

¹ Работа была выполнена при поддержке Российского научного фонда (проект № 22-11-00323).

ресных понятий для построения нейронной сети и классификации объектов. Статья состоит из следующих разделов:

- в разделе 2 приведены основные определения анализа формальных понятий (АФП);
- раздел 3 посвящен теоретическим сведениям об изучаемых индексах интересности;
- в разделе 4 даны постановка задачи и формальное описание эксперимента;
- в разделе 5 представлен метод разработки архитектуры нейронной сети;
- в разделе 6 обсуждаются результаты экспериментов;
- раздел 7 содержит выводы, полученные по результатам работы.

2. Анализ формальных понятий

Обратимся к главным определениям из анализа формальных понятий (АФП) [5]. В АФП исследуется множество G объектов, множество M признаков и бинарное отношение $I \subseteq G \times M$ такое, что $(g, m) \in I$ тогда и только тогда, когда объект g имеет признак m . Такая тройка $K = (G, M, I)$ называется *формальным контекстом*. Используя *операторы Галуа*, определяемые для $A \subseteq G$, $B \subseteq M$ как

$$A' = \{m \in M \mid gIm \text{ для всех } g \in A\},$$

$$B' = \{g \in G \mid gIm \text{ для всех } m \in B\},$$

формальное понятие контекста K определяется как пара (A, B) такая, что $A \in G$, $B \in M$, $A' = B$, $B' = A$. При этом A называется *объемом*, а B называется *содержанием* понятия (A, B) . Формальные понятия частично упорядочены отношением $\geq$:

$$(A_1, B_1) \leq (A_2, B_2) \iff A_1 \subseteq A_2,$$

которое задает полную (алгебраическую) решетку на множестве понятий, называемую *решеткой понятий* $L = (G, M, I)$.

Отношение покрытия, соответствующее частичному порядку $\leq$ (если оно существует), обозначается знаком $\prec$:

$$(A_1, B_1) \prec (A_2, B_2) \iff (A_1, B_1) \leq (A_2, B_2),$$

и не существует понятия (A_3, B_3) такого, что $(A_1, B_1) < (A_3, B_3) < (A_2, B_2)$.

3. Индексы интересности

Приведем формальное описание изучаемых в статье индексов интересности.

3.1. Базовый уровень (Basic Level)

Впервые общее определение базового уровня понятия было представлено в [6]. Неформально связностью понятия называется мера сходства всех пар

объектов из содержания понятия. Согласно идее Э. Роша, формализованной в [6], понятие (A, B) принадлежит базовому уровню, если оно удовлетворяет следующим условиям:

- (BL_1) (A, B) обладает высокой связностью;
- (BL_2) (A, B) обладает большей связностью, чем его верхние соседи (т.е. понятия, покрывающие понятие (A, B) в смысле отношения покрытия $\prec$);
- (BL_3) (A, B) обладает лишь чуть меньшей связностью, чем его нижние соседи (т.е. понятия, покрываемые понятием (A, B) в смысле отношения покрытия $\prec$).

В другом виде:

$$(1) \quad BL(A, B) = \mathcal{C}(\alpha_1(A, B), \alpha_2(A, B), \alpha_3(A, B)),$$

где $\mathcal{C}(\alpha_1, \alpha_2, \alpha_3) = \alpha_1 \otimes \alpha_2 \otimes \alpha_3$; $\otimes$ – t -норма.

В расчетах данного индекса предлагается использовать любое из двух следующих известных определений сходства множеств sim_Y :

$$(2) \quad sim_{SMC}(B_1, B_2) = \frac{|B_1 \cap B_2| + |Y - (B_1 \cup B_2)|}{|Y|},$$

$$(3) \quad sim_J(B_1, B_2) = \frac{|B_1 \cap B_2|}{|B_1 \cup B_2|}.$$

Далее вводятся два индекса связности формального понятия:

$$(4) \quad coh^\varnothing(A, B) = \frac{\sum_{\{x_1, x_2\} \subseteq A, x_1 \neq x_2} sim(x_1, x_2)}{|A| \cdot (|A| - 1)/2}$$

– среднее сходство двух объектов, входящих в объем данного формального понятия;

$$(5) \quad coh^m(A, B) = \min_{x_1, x_2 \in A} sim(x_1, x_2)$$

– наименьшая степень сходства двух объектов, входящих в объем данного формального понятия.

Так как в [7] авторы заключают, что показатель на основе индекса связности $coh^\varnothing(A, B)$ дает лучшие результаты отбора интересных понятий, в данной работе будут использованы только два вида показателя *базового уровня*, основанных на данном индексе: BL_{ees} – с использованием sim_{SMC} и BL_{eeJ} – с использованием sim_J .

В этих показателях

$$(6) \quad \alpha_1^\varnothing = coh^\varnothing(A, B),$$

$$(7) \quad \alpha_2^\varnothing = 1 - \frac{\sum_{c \in \mathcal{UN}(A, B)} coh^\varnothing(c) / coh^\varnothing(A, B)}{|\mathcal{UN}(A, B)|},$$

$$(8) \quad \alpha_3^\varnothing = \frac{\sum_{c \in \mathcal{LN}(A, B)} coh^\varnothing(A, B) / coh^\varnothing(c)}{|\mathcal{LN}(A, B)|}.$$

3.2. Целевая энтропия (*Target Entropy*)

Целевая энтропия формального понятия определяется как дисперсия значений целевого признака, соответствующих содержанию данного формального понятия.

3.3. Δ -устойчивость (Δ -stability)

Устойчивость формального понятия является широко применяемой характеристикой, однако сложность алгоритма ее нахождения экспоненциально растет с увеличением количества признаков в содержании понятия. Поэтому в [8] была введена оценка устойчивости – Δ -устойчивость.

$$(9) \quad \Delta(p) = \min(\Delta(p, q)), q < p,$$

$\Delta(p, q)$ – оценка устойчивости сверху. Данная величина является минимальной разницей между размером объема понятия и размером объема ближайшего снизу понятия.

3.4. Подъем (*Lift*)

Согласно [9] *lift* определяется как отношение наблюдаемой совместной вероятности X и Y к их ожидаемой совместной вероятности, если бы они были статистически независимы.

В [10] приводится формула расчета индекса интересности *lift* формального понятия, для этого можно рассматривать только содержание формального понятия и общее множество признаков:

$$(10) \quad lift(A, B) = \frac{\prod_{b \in B} Pr(b)}{Pr(B)}, \text{ где } Pr(\cdot) = \frac{|(\cdot)'|}{|G|}.$$

4. Постановка задачи

Выше были рассмотрены четыре индекса интересности понятий:

- 1) *Basic Level* (в данной работе были использованы BL_{ees} и BL_{eeJ});
- 2) Δ -stability;
- 3) *target entropy*;
- 4) *lift*.

Задача заключается в исследовании влияния выбора индекса для отбора понятий (когда формальные понятия уже получены). Исследование проводилось в следующей последовательности:

- бинаризация и подготовка датасета к обработке;
- построение формального контекста на основе набора данных;
- вычисление множества понятий на основе формального контекста;

- вычисление каждого индекса интересности для каждого формального понятия;
- сортировка понятий на основе величины изучаемого индекса;
- отбор k -лучших понятий для построения нейронной сети.

5. Архитектура нейронной сети

После отбора интересных понятий нейронная сеть строится на основе отношения покрытия на отобранных понятиях. Архитектура нейронной сети на основе решетки понятий выглядит следующим образом [2] (рисунок):

— входной слой (*Input Layer*) состоит из нейронов, связанных с признаками $m \in M$ контекста $K = (G, M, I)$;

— скрытые слои (*Hidden Layer_i*). Каждое формальное понятие может быть однозначно представлено своим содержанием. Признаки из множества признаков M итеративно соединяются в скрытых слоях таким образом, чтобы в последнем скрытом слое были получены нейроны, соответствующие отобранным формальным понятиям;

— выходной слой (*Output layer*). Число нейронов в данном слое соответствует числу целевых классов.

Для построения понятий из формального контекста были использованы инструменты библиотеки FCApy (<https://pypi.org/project/fcapy/>). Функции для расчета индексов BL_{ees} и BL_{eeJ} , $lift$ были написаны согласно определениям и формулам из раздела 3. Для расчета индексов $target\ entropy$ и $\Delta-stability$ были использованы встроенные возможности библиотеки FCApy.

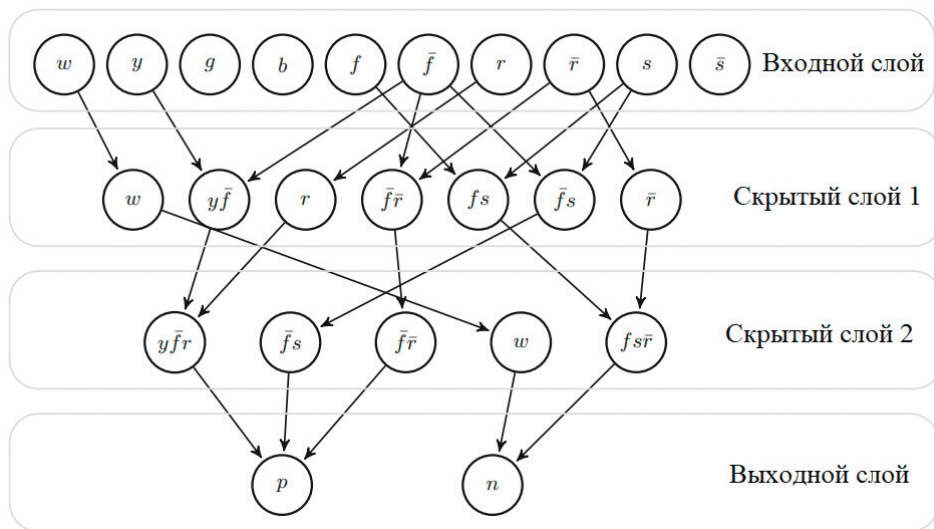


Рис. 1. Схема архитектуры нейронной сети на основе решетки понятий.

При выборе количества понятий использовался следующий критерий: наименьшее подмножество понятий, покрывающее все множество объектов:

$$\{(A_1, B_1), (A_2, B_2), \dots, (A_n, B_n)\} : A_1 \cup A_2 \cup \dots \cup A_n = G.$$

После расчетов индексов интересности для каждого индекса выбиралось k понятий с наибольшим значением данного индекса. Далее на основе данного множества понятий строилась нейронная сеть (использовались возможности библиотеки `neural_lib`, построенной на основе описания из [2]). Эта библиотека работает на основе пакета PyTorch).

Ее основные параметры: функция активации ReLU; оптимизатор Adam.

Наборы данных были разделены в отношении 70% и 30% на тренировочную и тестовую выборки. Проводились эксперименты с различным количеством генераций, лучшие результаты представлены в табл. 3–6.

5.1. Описание наборов данных

Для анализа были взяты четыре набора данных из библиотеки UCI (<http://archive.ics.uci.edu/ml/>) и предварительно бинаризованы. Названия и основные характеристики использованных наборов данных приведены в табл. 1.

Таблица 1. Характеристики наборов данных

Название набора	Количество объектов	Количество признаков	Количество классов
Heart Disease	303	33	2
House Votes	232	16	2
Car Evaluation	1727	21	4
Iris	150	16	3

Все используемые в работе наборы данных являются сбалансированными, кроме Car Evaluation.

5.2. Эксперименты с различными методами МО

Перед проведением основных экспериментов ряд базовых моделей были применены для анализа взятых наборов данных (табл. 2). Как видно из таблицы, лучшие результаты модели показывают на наборах данных House Votes и Iris, при этом для всех наборов данных лучшее качество получает модель XGBoost и случайный лес (Random Forest).

Таблица 2. Результаты State Of The Art моделей (метрика – Accuracy)

Название набора	Метод ближайшего соседа	Случайный лес	Наивный Байес	XGBoost	SVM
Heart Disease	0,83	0,85	0,81	0,81	0,79
House Votes	0,96	0,96	0,94	0,96	0,97
Car Evaluation	0,88	0,95	0,81	0,96	0,91
Iris	0,94	0,94	0,94	0,94	0,92

5.3. Сравнение результатов работы нейронной сети для разных индексов интересности

В табл. 3–6 приведены результаты экспериментов с индексами интересности. **Выделенным цветом** показаны результаты, сравнимые с качеством, полученным с использованием базовых моделей для тех же наборов данных.

Таблица 3. Результаты применения индексов интересности для набора данных Heart Disease

	BL_{ees}	BL_{eeJ}	target entropy	Δ -stability	lift
# генераций	8000	6000	8000	6000	7000
Recall	0,88	0,91	0,89	0,96	0,85
F1	0,84	0,80	0,88	0,95	0,84
Аккуратность	0,82	0,76	0,72	0,94	0,83
# понятий	7	7	20	7	7

Таблица 4. Результаты применения индексов интересности для набора данных House Votes

	BL_{ees}	BL_{eeJ}	target entropy	Δ -stability	lift
# генераций	5000	2000	3000	2000	3000
Recall	0,85	0,94	0,94	0,97	0,94
F1	0,88	0,91	0,95	0,95	0,95
Аккуратность	0,88	0,91	0,95	0,95	0,95
# понятий	7	7	20	7	7

Таблица 5. Результаты применения индексов интересности для набора данных Car Evaluation

	BL_{ees}	BL_{eeJ}	target entropy	Δ -stability	lift
# генераций	5000	5000	5000	5000	5000
Recall	0,44	0,45	0,25	0,47	0,25
F1	0,40	0,41	0,20	0,43	0,20
Аккуратность	0,82	0,84	0,68	0,87	0,68
# понятий	7	7	20	7	7

Таблица 6. Результаты применения индексов интересности для набора данных Iris

	BL_{ees}	BL_{eeJ}	target entropy	Δ -stability	lift
# генераций	5000	3000	7000	5000	3000
Recall	0,95	0,95	0,87	0,95	0,95
F1	0,95	0,95	0,86	0,95	0,95
Аккуратность	0,95	0,95	0,86	0,95	0,95
# понятий	7	7	20	7	7

Стоит отметить:

— Результаты качества с использованием индекса Δ -устойчивости в качестве критерия отбора понятий для всех четырех наборов данных оказались сравнимыми с эталонными моделями (метод ближайшего соседа, случайный лес, наивный Байес, XGBoost, SVM), тогда как индекс *target entropy* показал сопоставимые результаты только для набора данных House Votes (табл. 4).

— Индекс *lift* был успешен во всех экспериментах, кроме набора Car Evaluation (табл. 5).

— Индексы BL_{ees} и BL_{eeJ} показали близкие результаты, но для набора Heart Disease (табл. 3) индекс BL_{ees} оказался более успешен и сравним с эталонными моделями в отличие от BL_{eeJ} .

— Наихудшие результаты были получены для набора Car Evaluation (табл. 5), что можно объяснить его несбалансированностью при наличии четырех значений целевого признака.

— Самые высокие показатели качества были получены для наборов House Votes (табл. 4) и Iris (табл. 6). Это сбалансированные наборы данных со сравнительно небольшим количеством признаков в отличие от остальных использованных наборов данных.

— Индекс Δ -устойчивость во всех случаях показал более высокие показатели по сравнению с другими индексами интересности для тех же наборов данных.

6. Заключение

По полученным результатам можно сделать следующие выводы:

1) с помощью использования индексов интересности можно получить качество классификации, сравнимое с работой эталонных моделей;

2) индекс интересности *target entropy* показал наихудшие результаты относительно остальных индексов интересности;

3) индекс *lift* показал хорошие результаты, но не справился с классификацией несбалансированного набора данных с несколькими целевыми признаками;

4) индексы интересности *Basic Level* - BL_{ees} и BL_{eeJ} справились с классификацией в наборах данных с небольшим количеством признаков;

5) Индекс Δ -устойчивости в качестве средства отбора понятий показал хорошие результаты как на наборах данных с бинарным целевым признаком, так и при классификации с несколькими целевыми классами, в отличие от остальных индексов, исследованных в работе. Соответствующие показатели качества обучения превосходят полученные с помощью других индексов.

В дальнейшем планируется исследование других индексов интересности в качестве критериев отбора интересных понятий для построения нейронных сетей на их основе.

СПИСОК ЛИТЕРАТУРЫ

1. *Tsopze N., Nguifo E.M., Tindo G.* CLANN: Concept lattice-based artificial neural network for supervised classification // The Fifth International Conference on Concept Lattices and Their Applications. 2007. P. 24–26.
2. *Kuznetsov S.O., Makhazhanov N., Ushakov M.* On neural network architecture based on concept lattices // ISMIS 2017. P. 653–663.
3. *Kuznetsov S.O., Makhalova T.P.* Concept interestingness measures: a comparative study // Proceedings of the Twelfth International Conference on Concept Lattices and Their Applications. 2015. P. 59–72.
4. *Kuznetsov S.O., Makhalova T.P.* On interestingness measures of formal concepts // Inf. Sci. 442. 2018. P. 202–219.
5. *Ganter B., Wille R.* Contextual attribute logic / International Conference on Conceptual Structures. 1999. P. 377–388.
6. *Rosch E.* Basic objects in natural categories // Cognitive Psychology 8. 1976. P. 382–439.
7. *Belohlavek R., Trnečka M.* Basic level of concepts in formal concept analysis // ICFCA 2012. P. 28–44.
8. *Buzmakov Al., Kuznetsov S.O., Amedeo Napoli.* Scalable Estimates of Concept Stability // ICFCA 2014. P. 157–172.
9. *Zaki M.J., Meira W., Jr.* Data Mining and Analysis: Fundamental Concepts and Algorithms // Cambridge University Press. 2014. P. 339.
10. *Makhalova T.* Interesting Measures of Closed Patterns for Data Mining and Knowledge Discovery // HSE University, Moscow, Russia. 2020. P. 25.

Статья представлена к публикации членом редколлегии А.А. Галяевым.

Поступила в редакцию 08.07.2023

После доработки 16.10.2023

Принята к публикации 20.01.2024