

Индексы интересности как инструмент отбора формальных понятий для построения нейронной сети на основе решётки формальных понятий

М.М. Зуева, С.О. Кузнецов, факультет компьютерных наук,
департамент анализа данных и искусственного интеллекта,
НИУ ВШЭ, Покровский бульвар 10, Москва, Россия
skuznetsov@hse.ru, mmzueva@edu.hse.ru, <https://www.hse.ru>

Аннотация

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

Ключевые слова Архитектура нейронной сети - Анализ формальных понятий - Индексы интересности - Нейронные сети на основе решеток формальных понятий

1. Введение

Сложность интерпретации результатов при работе с нейронными сетями является важной проблемой, которой в последнее время посвящено много научных работ. Одним из предложенных решений является построение нейронной сети на основе решеток формальных понятий. В [1] была представлена нейронная сеть с архитектурой, построенной в соответствии с решеткой понятий исходного набора данных для повышения стабильности классификации. В [2] было предложено построение нейронной сети на основании решетки формальных понятий, где использовались формальные понятия на основе монотонного и антимонотонного соответствия Галуа.

Однако, так как количество формальных понятий для данного

набора данных растет экспоненциально с размером входных данных, важной задачей является возможность уменьшения количества формальных понятий для построения нейронной сети без потерь качества ее работы. Можно сделать это двумя способами - за счет отбора наиболее значимых признаков (предобработка), и за счет отбора наиболее важных формальных понятий (постобработка). В [3] были рассмотрены различные методы выбора интересных формальных понятий, основанных на их индексах интересности. В [4] меры интересности понятий сравнивались по таким аспектам, как эффективность нахождения и возможность их применимости к зашумленным данным.

В данной работе проведено исследование 4 индексов интересности: *basic level*, *target entropy*, Δ -*stability* и *lift* в качестве критериев отбора интересных формальных понятий для построения нейронной сети и классификации объектов. Статья организована следующим образом:

- в разделе 2 приведены основные определения теории анализа формальных понятий (АФП);
- раздел 3 посвящен теоретическим сведениям об изучаемых индексах интересности;
- раздел 4 посвящен постановке задачи и формальному описанию эксперимента;
- раздел 5 посвящен архитектуре нейронной сети;
- раздел 6 посвящен результатам экспериментов и их обсуждению;
- раздел 7 посвящен выводам, полученным по результатам работы.

2. Анализ формальных понятий

Обратимся к главным определениям из анализа формальных понятий [5]. В нем рассматривается 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 *дифференциальные операторы*, определенные для $A \subseteq G$, $B \subseteq M$ с помощью

$$A' = \{m \in M \mid gImg \in A\},$$

$$B' = \{g \in G \mid gImt \in B\},$$

мы можем определить *формальное понятие контекста* K как пару (A, B) , такую, что $A \in G$, $B \in M$, $A' = B$, $B' = A$. A называется *объемом*, B называется *содержанием* понятия (A, B) . Эти понятия, упорядоченные по

$$(A_1, B_1) \geq (A_2, B_2) \iff A_1 \supseteq A_2$$

образуют полную решетку, называемую *решеткой понятий* $L = (G, M, I)$.

3. Индексы интересности

Далее приведено формальное описание изучаемых индексов интересности:

3.1 Basic Level

Впервые общее определение базового уровня понятия было представлено в [6]. Согласно [6], понятия принадлежат базовому уровню, если объекты в понятиях данного уровня схожи друг с другом, тогда как объекты высших понятий гораздо менее схожи, а объекты низших понятий только немного более схожи.

Авторы в [7] предлагают формальное описание базового уровня формального понятия в решетке формальных понятий и формальное описание степени принадлежности формального понятия к базовому уровню как индекс интересности.

Они предлагают включить три показателя в расчет данного индекса:

(BL1) – показатель, что $\langle A, B \rangle$ имеет высокую связность;

(BL2) – показатель, что $\langle A, B \rangle$ имеет гораздо более высокую связность, чем его верхние соседи;

(BL3) – показатель, что $\langle A, B \rangle$ имеет чуть менее высокую связность, чем его нижние соседи;

В другом виде:

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

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 к использованию любая из 2 следующих известных формул схожести множеств sim_Y :

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

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

Далее вводятся две формулы расчета связности формального понятия:

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

- средняя схожесть двух объектов покрытых данным формальным понятием.

С другой стороны,

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

- наименьшая степень схожести двух объектов покрытых данным формальным понятием.

Так как в [7] авторы заключают, что первая функция схожести является более перспективной для отбора лучших формальных понятий, в данной работе будут использованы только 2 вида данного индекса *basic level*, основанных на данной функции - BL_{ees} - с использованием sim_{SMC} и BL_{eeJ} - с использованием sim_J .

В них:

$$\alpha_1^{\emptyset} = coh^{\emptyset}(A, B)$$

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

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

3.2 Target Entropy

Так как данный индекс является распространенной характеристикой формального понятия, было решено включить его в список исследуемых индексов.

Целевая энтропия данного формального понятия рассчитывается как дисперсия целевых переменных объектов из объема формального понятия.

3.3 Δ -stability

Устойчивость формального понятия является его широко применяемой характеристикой. Однако сложность алгоритма ее нахождения экспоненциально растет с увеличением количества признаков. В качестве критерия формального понятия, удобного для расчета, в [8] была введена оценка устойчивости - Δ -stability.

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

$\Delta(p, q)$ - оценка устойчивости сверху.

Данная величина является минимальной разницей в поддержках паттерна и его ближайших детей.

3.4 Lift

Согласно [9], lift определяется как отношение наблюдаемой совместной вероятности X и Y к их ожидаемой совместной вероят-

ности, если бы они были статистически независимы.

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

$$lift(A, B) = \frac{\prod_{b \in B} Pr(b)}{Pr(B)}, \text{ where } Pr(\cdot) = \frac{|\cdot|}{|G|}$$

4. Постановка задачи

Выше были рассмотрены 4 индекса интересности формальных понятий:

- *Basic Level* (в данной работе были использованы BL_{ees} и BL_{eej})
- Δ -*stability*
- *target entropy*
- *lift*

Была поставлена задача изучить влияние выбора данных индексов как критериев при отборе формальных понятий. Данное действие осуществляется при постобработке (когда формальные понятия уже получены)

Поэтому данная работа проводилась по следующему алгоритму:

- Бинаризация и подготовка датасета к обработке
- Построение формального контекста на основе набора данных
- Вычисление множества формальных понятий на основе формального контекста
- Вычисление каждого индекса интересности для каждого формального понятия
- Сортировка формальных понятий на основе величины изучаемого индекса
- Отбор k -лучших формальных понятий для построения нейронной сети

5. Архитектура нейронной сети

После выбора лучших формальных понятий нейронная сеть строится на основе отношения покрытия верхней части решетки формальных понятий. Архитектура нейронной сети на основе решетки формальных понятий организована следующим образом [2] (рис.1):

- 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 скры-

тых слоях таким образом, чтобы в последнем скрытом слое были получены нейроны, соответствующие отобранному формальному понятию.

-Выходной слой *Output layer*. Число нейронов в данном слое соответствует числу целевых классов.

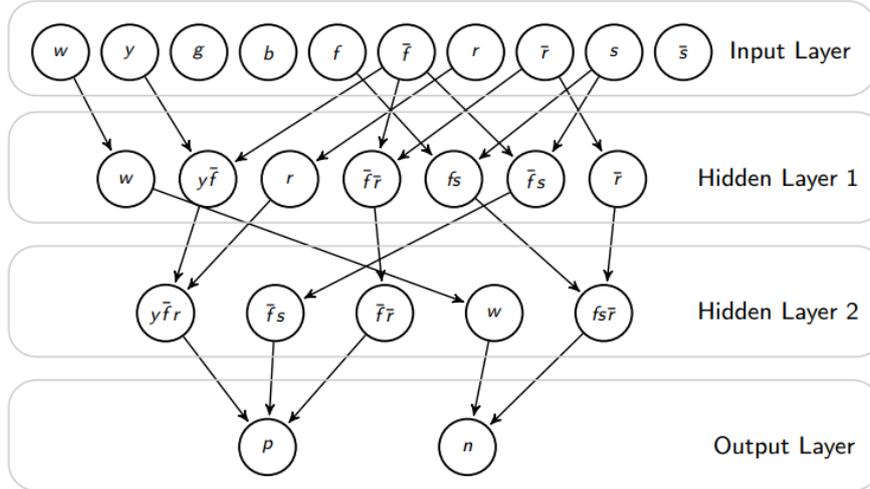


Рис. 1: Схема архитектуры нейронной сети на основе решётки формальных понятий.

6. Эксперименты

Для построения формальных понятий из формального контекста были использованы инструменты библиотеки FCApy. Функции для расчета индексов BL_{ees} и BL_{eeJ} , $lift$ были написаны согласно определениям и формулам из основной части. Для расчета индексов $target\ entropy$ и $\Delta-stability$ были использованы встроенные возможности библиотеки FCApy.

При выборе количества формальных понятий использовались следующий критерий: наименьший набор формальных понятий, покрывающий 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 выбирались 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 работы [2]). Данная библиотека работает на основе пакета PyTorch).

Ее основные параметры: метод инициализации - ReLU; оптимайзер: Adam.

Предварительно датасеты были разделены в отношении 70% и 30% на тренировочную и тестовую выборки. Проводились эксперименты с различным количеством генераций, лучшие результаты представлены в таблицах.

6.1 Описание наборов данных

Для анализа были взяты 4 набора данных из библиотеки UCI (<http://archive.ics.uci.edu/ml/>) и предварительно бинаризованы. Названия и основные характеристики использованных датасетов приведены в таблице 2.

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

Из использованных датасетов все являются сбалансированными, кроме набора данных Car Evaluation.

6.2 Эксперименты с различными методами МО

Перед проведением основных экспериментов ряд моделей State-Of-The-Art были применены для анализа взятых наборов данных (таблица 3). 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.

6.3 Сравнение результатов работы нейронной сети для разных индексов интересности

В таблицах 4,5,6,7 приведены результаты экспериментов с ин-

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

дексами интересности. Выделенным цветом показаны результаты, сравнимые с качеством, полученным с использованием State-Of-The-Art моделей для тех же наборов данных.

Таблица 3: Результаты применения индексов интересности для отбора понятий датасета 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
Ассурасу	0,82	0,76	0,72	0,94	0,83
# понятий	7	7	20	7	7

Таблица 4: Результаты применения индексов интересности для отбора понятий датасета 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
Ассурасу	0,88	0,91	0,95	0,95	0,95
# понятий	7	7	20	7	7

Таблица 5: Результаты применения индексов интересности для отбора понятий датасета 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
Ассурасу	0,82	0,84	0,68	0,87	0,68
# понятий	7	7	20	7	7

Таблица 6: Результаты применения индексов интересности для отбора понятий датасета 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
Ассурасу	0,95	0,95	0,86	0,95	0,95
# понятий	7	7	20	7	7

Можно заметить, что результаты качества с использованием индекса Δ -stability в качестве критерия отбора понятий во всех четырех датасетах оказались сравнимыми с эталонными моделями, тогда как индекс *target entropy* показал сопоставимые результаты только в наборе House Votes (таблица 5).

Индекс lift был успешен во всех экспериментах кроме набора Car Evaluation (таблица 6).

Индексы BL_{ees} и BL_{eeJ} показали близкие результаты, но для набора Heart Disease (таблица 4) индекс BL_{ees} оказался более успешен и сравним с эталонными моделями в отличие от BL_{eeJ} .

Наиболее низкие результаты были получены для датасета Car Evaluation (таблица 6), что можно объяснить его несбалансированностью при наличии 4 значений целевого признака.

Самые высокие показатели качества вышли у наборов House Votes (таблица 5) и Iris (таблица 7). Это сбалансированные датасеты со сравнительно небольшим количеством признаков, в отличие от остальных использованных наборов данных.

Также стоит отметить, что индекс Δ -stability во всех случаях показал более высокие показатели, чем те, что показали другие индексы интересности для тех же датасетов.

7. Заключение

По полученным результатам можно сделать следующие выводы:

- С помощью использования индексов интересности можно получить качество классификации сравнимое с работой эталонных моделей.
- Индекс интересности *target entropy* показал наиболее низкие результаты относительно остальных использованных индексов интересности.
- Индекс *lift* показал хорошие результаты, но не справился с классификацией несбалансированного датасета с несколькими целевыми признаками.
- Выбранные индексы интересности категории Basic Level - BL_{ees} и BL_{eeJ} справились с классификацией датасетов с небольшим количеством признаков.
- Можно заключить, что Δ -*stability* в качестве критерия отбора формальных 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. Его показатели качества превосходят полученные с помощью других индексов. Можно предположить, что Δ -*stability* является наиболее перспективным для использования в качестве критерия отбора интересных формальных понятий.

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

Список литературы

1. Norbert Tsopze, Engelbert Mephu Nguifo, Gilbert Tindo, CLANN: concept lattice-based artificial neural network for supervised classification, in: The Fifth International Conference on Concept Lattices and Their Applications, 2007, pp. 24–26.
2. Sergei O. Kuznetsov, Nurtas Makhazhanov, Maxim Ushakov, On neural network architecture based on concept lattices, in: ISMIS 2017, 2017, pp. 653–663.
3. S.O. Kuznetsov, T.P. Makhalova, Concept interestingness measures: a comparative study, in: Proceedings of the Twelfth International Conference on Concept Lattices and Their Applications, 2015, pp. 59–72

4. Kuznetsov, S.O., Makhalova, T.P.: On interestingness measures of formal concepts. *Inf. Sci.* (2017)
5. B. Ganter, R. Wille, Contextual attribute logic, in: W.M. Tepfenhart, W. Cyre (Eds.), *Conceptual Structures: Standards and Practices*, Lecture Notes in Computer Science, 1640, Springer Berlin Heidelberg, 1999, pp. 377-388.
6. Rosch, E., et al.: Basic objects in natural categories. *Cognitive Psychology* 8, 382–439 (1976)
7. Belohlavek Radim, Trnecka Martin. Basic level of concepts in Formal Concept Analysis // International Conference on Formal Concept Analysis / Springer. — 2012, pp. 28–44.
8. Aleksey Buzmakov, Sergei O. Kuznetsov, Amedeo Napoli: Scalable Estimates of Concept Stability. *ICFCA 2014*: 157-172
9. M.J. Zaki, W. Meira Jr, W. Meira, *Data Mining and Analysis: Fundamental Concepts and Algorithms*, Cambridge University Press, 2014, p. 302
10. Tatiana Makhalova: *Interesting Measures of Closed Patterns for Data Mining and Knowledge Discovery*. HSE University, Moscow, Russia, 2020, p. 25