

**О наследуемости диагностических заключений
при пополнении обучающей выборки новыми
эмпирическими данными**

М.И.Забежайло
ФИЦ ИУ РАН



Начальная постановка задачи

- База Фактов – прецедентов (примеров и контрпримеров) диагностируемого эффекта\явления
- Новые прецеденты для диагноза
- Интерполирующие текущую **БФ** Эмпирические Зависимости (**ЭЗ**) и их экстраполируемость на новые прецеденты



Две критически значимые особенности решаемой задачи интерполяции-экстраполяции

- Постоянная *раширяемость* новыми эмпирическими данными используемой **БФ** (учет вновь поступающих прецедентов – примеров и контрпримеров)
- Формирование «*наследуемых*» управляющих воздействий (например, - терапевтических мероприятий, назначаемых по результатам проведенной диагностики),
в частности, -
с учетом результатов выделения ЭЗ, наследуемых при расширении текущей **БФ**



Наследуемость ЭЗ и анализ причинности возникновения диагностируемого эффекта

- Предположение об *«уточнении»* деталей описания диагностируемого эффекта за счет использования все более *«полной»* версии **БФ**
- Возможности формирования **«управляющих» воздействий**, ориентируясь на факторы причинного влияния, которыми обусловлено возникновение изучаемого эффекта



Эвристика выделения факторов причинного влияния в описаниях прецедентов

- **Сходство** в описаниях примеров (прецедентов наличия изучаемого эффекта) \implies описания *комбинаций факторов влияния* как комбинаций определенных значений определенных параметров, входящих в описания примеров
- *Запрет на вложимость* таких комбинаций факторов влияния в описания *контрпримеров*: если это – причина возникновения эффекта, то ее не должно быть в описаниях контрпримеров (они же представляют ситуации, в которых нет исследуемого эффекта)



Схема алгебраической формализации предлагаемой эвристики

- **Сходство** описаний прецедентов как *бинарная алгебраическая операция*
- **Отношение сходства** (по *непустоте* результата вычисления операции сходства)
- **Классы сходства**
- **Классы эквивалентности**
- Классы эквивалентности примеров и контрпримеров. **Запрет на КонтрПримеры. ЗКП-классы эквивалентности прецедентов**
- «Диагностика» целевого эффекта на новых прецедентах **отнесением к ЗКП-классам эквивалентности прецедентов**



- Пусть $\Omega = \{ O_1, O_2, \dots, O_m \}$. Построим по Ω и \otimes множество $Dom(\Omega)$:

(i) $\Omega = \{ O_1, O_2, \dots, O_m \} \subset Dom(\Omega)$.

(ii) $\{ [A \in Dom(\Omega)] \ \& \ [B \in Dom(\Omega)] \ \& \ [(A \otimes B) \neq \emptyset] \} \rightarrow [(A \otimes B) \in Dom(\Omega)]$,

(iii) Других элементов в $Dom(\Omega)$ нет.

- Отношение **сходства** \mathbf{R}^\otimes :

$$O_{i1} \mathbf{R}^\otimes O_{i2} \text{ (т.е. } \langle O_{i1}, O_{i2} \rangle \in \mathbf{R}^\otimes) \quad \text{iff} \quad O_{i1} \otimes O_{i2} \neq \emptyset$$

- **Классы сходства** (толерантности) на Ω :

$$\mathbf{T}(O_{i1}) = \{ O_{i2} \mid O_{i1} \mathbf{R}^\otimes O_{i2} \}$$

- Фиксируя каждый конкретный результат $V=V_0$ вычисления операции \otimes сходства на элементах множества $Dom(\Omega)$:

$$A \otimes B = V_0$$

выделим соответствующие подклассы (**эквивалентности**) сформированных классов сходств:

$$\mathbf{E}_{V_0} = \{ \langle O_{i1}, O_{i2} \rangle \mid O_{i1} \otimes O_{i2} = V_0 \}$$

- **Запрет на контрпримеры** (условие **ЗКП**):

$$(\forall V_0) [(\mathbf{E}_{V_0}^+ = \{ \langle O_{i1}^+, O_{i2}^+ \rangle \mid (O_{i1}^+ \otimes O_{i2}^+ = V_0) \}) \ \& \ (O_{i1}^+ \in \Omega^+) \ \& \ (O_{i2}^+ \in \Omega^+)] \supset$$

$$\supset \neg (\exists O_{i0}^-) \{ (V_0 \otimes O_{i0}^- = V_0) \ \& \ (O_{i0}^- \in \Omega^-) \}$$



Характеристические Функции на последовательностях расширяющихся БФ

- Характеристическая Функция (ХФ), принимает
 - значение «*истина*» на всех фактах **ф** (**примерах**) текущей базы фактов **БФ**, характеризующих наличие анализируемого целевого свойства и
 - значение «*ложь*» на всех фактах **ф** (**контрпримерах**) текущей базы фактов **БФ**, характеризующих наличие анализируемого целевого свойства
- Характеристические Функции как (*каузально-ориентированный*) подкласс семейства *частичных функций*, интерполирующих выборки прецедентов – примеров и контрпримеров диагностируемого явления. *Пополнение* текущей **БФ** описаниями новых прецедентов обеспечивает *до-определение* некоторых ранее не определенных значений ХФ
- Представление о наследуемости ХФ при расширении БФ описаниями новых прецедентов. Особая роль контрпримеров



Некоторые свойства процедурной конструкции ХФ

- Понятие *каузальной репрезентативности* текущей БФ

Утверждение 1

Оценка *каузальной репрезентативности* анализируемой БФ может быть выполнена алгоритмом полиномиальной вычислительной сложности

Утверждение 2

Задача о числе характеристических функций, формируемых на произвольной базе фактов БФ, принадлежит классу #PC перечислительно полных переборных задач



Проблема экстраполируемости уже построенных ХФ на конкретные расширения текущей БФ

- БФ как «хранилище» *постоянно накапливаемого* эмпирического материала. «*Плюрализм*» диагностируемых *эффектов* и возможности модификации используемого языка представления данных и знаний
- Возможности формальной *экстраполируемости* на вновь анализируемый прецедент таких ХФ, которые представляют **разные** (в т.ч. – содержательно *несовместные*) эффекты.

Пример: *ПсП* vs. *СОО*



Задачи об анализе эффектов *наследуемости*

Два типа анализируемых ситуаций:

- *раширение* текущей $\mathbf{БФ}_1$ новыми эмпирическими данными $\Delta\mathbf{БФ}$ при переходе к новой $\mathbf{БФ} = \mathbf{БФ}_1 \cup \Delta\mathbf{БФ}$ (учет вновь поступающих прецедентов – примеров и контрпримеров)
- переход от *экстраполируемости* характеристических функций из множества $X\Phi(\mathbf{БФ})$ на заданный новый прецедент φ к *экстраполируемости* на него же характеристических функций из множества $X\Phi(\mathbf{БФ}_1 \cup \Delta\mathbf{БФ})$

Анализ наследуемости:

- свойства *репрезентативности* текущей $\mathbf{БФ}$ при заданном ее расширении $\Delta\mathbf{БФ}$
- *экстраполируемости* $X\Phi$ на новый прецедент φ при переходе от заданной $\mathbf{БФ}_1$ к заданной $\mathbf{БФ}_2 = \mathbf{БФ}_1 \cup \Delta\mathbf{БФ}$
- $X\Phi$, *экстраполируемых* на заданный новый прецедент φ (выделение *наследуемых* при таком переходе $X\Phi$)



Некоторые факты об экстраполируемости ХФ

- Пример 1

Переход от $\mathbf{БФ}_1$ к $\mathbf{БФ}_2$: и $\mathbf{БФ}_1$, и $\mathbf{БФ}_2$ - каузально репрезентативны, более того – и $\mathbf{ХФ}(\mathbf{БФ}_1)$ и $\mathbf{ХФ}(\mathbf{БФ}_2)$ имеют *exp-быстро* растущие размеры, однако, в $\mathbf{ХФ}(\mathbf{БФ}_2)$ **нет** ни одной ХФ, представляющей собою «продолжение» какой-либо ХФ из $\mathbf{ХФ}(\mathbf{БФ}_1)$, – **классы зависимостей различны, а результат диагностики один и тот же**

- Пример 2

Экстраполируемость ХФ на новый прецедент φ при переходе от $\mathbf{БФ}_1$ к $\mathbf{БФ}_2$: и $\mathbf{БФ}_1$, и $\mathbf{БФ}_2$ - каузально репрезентативны, более того – и $\mathbf{ХФ}(\mathbf{БФ}_1)$ и $\mathbf{ХФ}(\mathbf{БФ}_2)$ имеют *exp-быстро растущие размеры*, дополнительно, в $\mathbf{ХФ}(\mathbf{БФ}_2)$ **нет** ни одной ХФ, представляющей собою «продолжение» какой-либо ХФ из $\mathbf{ХФ}(\mathbf{БФ}_1)$, однако и $\mathbf{ХФ}(\mathbf{БФ}_1)$, и $\mathbf{ХФ}(\mathbf{БФ}_2)$ содержат «свои» ХФ, экстраполируемые на новый прецедент φ , - **классы зависимостей различны, а результат диагностики один и тот же**

- Утверждение

Задача о наличии в $\mathbf{ХФ}(\mathbf{БФ}_2)$ такой $\mathbf{ХФ}_2$, которая является «*продолжением*» некоторой $\mathbf{ХФ}_1$ из $\mathbf{ХФ}(\mathbf{БФ}_1)$ и при этом также, как и $\mathbf{ХФ}_1$, экстраполируется на заданный новый прецедент φ , **полиномиально** быстро разрешима



Заключение

- Особый статус задач диагностического типа:

Каузальность \neq **интерполируемость**

- Идентифицированные эффекты – своего рода комбинаторная «плата» за **точность интерполяции** в задачах **диагностического** типа
- Отдельных эмпирических зависимостей (*Характеристических Функций*), позволяющих вести результативную диагностику, может оказаться **слишком много** для их исчерпывающего перебора. \implies
Можно оперировать «*однородными*» (по результатам экстраполяции но вновь диагностируемый прецедент) *классами* таких зависимостей
- В общем случае размеры таких классов растут *exp*-быстро с линейны ростом размеров исходной **Базы Фактов** (обучающей выборки прецедентов). Однако, задача о **непустоте** таких классов оказывается *полиномиально* разрешимой.



Забейсайло М.И. О емкости семейств характеристических функций, обеспечивающих корректное решение задач диагност. типа // 19 Всероссийская конференция «Мат. методы распознавания образов» (ММРО-2019), 25- 28 октября 2019 г. , Москва. - Тез. док. — с. 305-306.

Грушо А.А., Забейсайло М.И., Тимонина Е.Е. О каузальной репрезентативности обучающих выборок прецедентов в задачах диагностического типа. // Инфор. и ее применения., №1, 2020 — С. 80-86.

Забейсайло М.И. Об оценках размеров семейств характеристических функций, обеспечивающих корректное решение задач диагностического типа // ЖВМ и МФ. – 2021. – Т.61, №5 (в печати)

Забейсайло М.И. О некоторых оценках сложности вычислений при прогнозировании свойств новых объектов средствами характеристических функций //НТИ, сер.2.- 2020.-№12 (в печати)



Спасибо за внимание



m.zabezhailo@yandex.ru

