

$$\chi(\Phi_i) = \alpha / F[\sigma(\Phi_i), (\sum_{i=1}^m \pi_i(\varphi_i))], \quad (6)$$

где α – коэффициент согласования размерностей, $F(\cdot)$ – функция которая описывает итоговое влияние параметров $\sigma(\Phi_i)$ и $\pi(\varphi_i)$ на величину $\gamma(\Phi_i)$, а m – количество фраз в фрагменте Φ_i . В простейшем случае функция $F(\cdot)$ может означать суммирование значений параметров $\sigma(\Phi_i)$ и $\pi(\varphi_i)$. Приведенное соотношение (6) может быть использовано для количественной оценки величины $\chi(\Phi_i)$ благодаря тому, что численные значения параметров $\sigma(\Phi_i)$ и $\pi(\varphi_i)$ определяются на основе использования элементарных семантических параметров μ и λ , которые, в свою очередь, определены численно.

1. Падучева Е.В. Понятие презумпции в лингвистике/Семиотика и информатика №8, 1977.
2. Мендельсон Э. Введение в математическую логику. М.: Наука, 1971, - 320 с.
3. Гладкий А.В. Формальные грамматики и языки. М.: Наука, 1973.
4. Клини С. Математическая логика. М.: Мир, 1973.

Поступила 2.02.2009р.

УДК 004.832.3

Ю. Р. Валькман, д.т.н., зав. отд., А. Ю. Рыхальский, м. н. с.,
отдел распределенных интеллектуальных систем Международного научно-учебного центра информационных технологий и систем НАН и МОН Украины

КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ ИНДУКТИВНОГО ФОРМИРОВАНИЯ ЗНАНИЙ

Введение. Объектом исследования являются процессы порождения новых знаний в вычислительной среде. Предмет исследования – компьютерные технологии индуктивного формирования знаний. Цель исследования – анализ и разработка обобщенной модели индуктивного формирования знаний в исследованиях сложных систем. Ожидаемые результаты – разработка интеллектуальной компьютерной технологии, поддерживающей синтез и анализ новых знаний об исследуемых процессах и

объектах.

В науке, обыденном мышлении мы идем от незнания к знанию, от неполного знания к более полному. Нам приходится выдвигать и затем обосновывать различные предположения для объяснения явлений и их связи с другими явлениями. И, чаще всего, при этом, явно и неявно, мы используем индуктивные процедуры. Эта работа продолжает наши исследования индуктивных выводов [1-3].

Индуктивное формирование знаний (ИФЗ) на основе эмпирических данных является основным способом получения новых эмпирических знаний в науке и практике. Он заключается в получении общего знания или базы знаний (БЗ) о некотором классе объектов на основании анализа единообразного описания отдельных представителей этого класса - выборки данных. В принципе, это может быть и естественная база знаний, но здесь мы будем иметь в виду более формализованные знания.

И, тогда, БЗ представляет собой предикат, который для одних объектов области определения *истинен*, для других *ложен*.

Объекты, для которых значение БЗ истинно, называются примерами, а их конечная совокупность - *выборкой примеров*.

Объекты, для которых это значение ложно, называются *контрпримерами*, а их конечная совокупность - *выборкой контрпримеров*.

Выборка, на основе которой осуществляется индуктивное формирование БЗ, называется *обучающей*.

Выборка, на основе которой проводится оценка БЗ, называется *контрольной*.

Таким образом, в задаче ИФЗ по конечному числу объектов, описанных обучающей выборкой, состоящей из примеров и контрпримеров, необходимо найти предикат, который истинен на возможно большем числе примеров и ложен на возможно большем числе контрпримеров контрольной выборки.

1. Методы индуктивного формирования знаний и их свойства

Индуктивное формирование знаний как способ познания лежит в основе многих направлений исследований, получивших в англоязычной литературе названия

- *Machine Learning* (машинное обучение),
- *Data Mining* (анализ данных),
- *Knowledge Discovery in Databases* (обнаружение знаний в базах данных),
- *Pattern Recognition* (распознавание образов) и т.д.

Каждое из этих направлений характеризуется собственным подходом к проблеме ИФЗ, собственными постановками задач и методами их решения (см., например, работы Вагина В. Н. [4], Дюка В. К. [5], Загоруйко Н. Г. [6], Клещев А.С. [7], Степашко В. С. [8], Финна В.К.[9]).

Методы ИФЗ можно разделить на

- *проблемно-независимые* (методы, в основе которых лежат аб-

страктные математические модели ИФЗ) и

- *проблемно-ориентированные* (методы, в основе которых лежат модели ИФЗ, существенно использующие зависимости и законы предметной области (ПО) или узкого класса ПО).

При решении прикладных задач возникает необходимость выбора наиболее подходящего метода ИФЗ и формирование выборки достаточного объема. Такой выбор должен осуществляться исходя из условий и ограничений задачи, а также из известных значений свойств методов ИФЗ.

Наиболее важным свойством метода ИФЗ является функция, представляющая зависимость процента дефектов БЗ от объема выборки. Если процент дефектов БЗ с увеличением объема выборки уменьшается, то тогда имеет место сходимостъ метода.

При исследовании качества интерес представляет вид функции, ее асимптотика и то, как она ведет себя на всей области определения. Причина в том, что в разных приложениях и объем выборки может быть разным. Другим важным свойством метода ИФЗ является функция, представляющая зависимость времени ИФЗ от объема выборки. Знание вида этой функции и ее асимптотики важно тогда, когда метод имеет высокую вычислительную сложность. В этом случае необходимо установить предельный объем выборки, на котором время формирования знаний не выйдет за пределы имеющихся у исследователя вычислительных ресурсов.

Устойчивостъ значений свойств метода на выборках различного объема является определяющей характеристикой свойств метода ИФЗ.

Свойство метода устойчиво относительно разных выборок одной и той же ПО, если значения этого свойства близки на разных выборках (этой ПО) одного и того же объема, причем, чем больше объем выборок, тем эти значения ближе.

Свойство метода устойчиво относительно свойств разных ПО, если значения этого свойства близки на разных выборках (взятых из различных ПО) одного и того же объема, причем, чем больше объем выборок, тем эти значения ближе.

Графически устойчивостъ можно представить как некий коридор, обрамляющий значения функции какого-либо свойства метода ИФЗ на разных выборках разного объема. Если этот коридор сужается, то свойство метода устойчиво. Графический образ понятия устойчивости значений свойств метода представлен на рис. 1.

Теоретическое исследование свойств методов ИФЗ обычно неэффективно, поэтому предпочтительным становится их экспериментальное изучение.

Экспериментальные исследования свойств методов ИФЗ проводятся

- либо на реальных данных,
- либо на данных из репозитариев,
- либо на модельных данных.

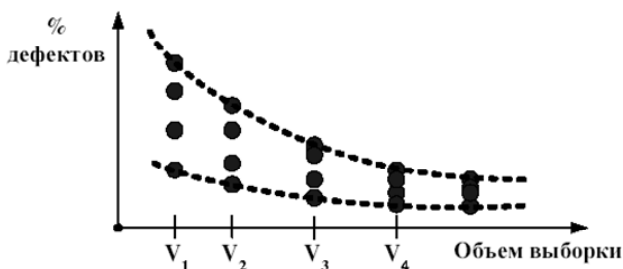


Рис. 1. Устойчивость значений свойств метода ИФЗ на выборках различного объема

Оценки свойств методов ИФЗ строятся из оценок, индуктивно сформированных БЗ (ИФБЗ), полученных с помощью этих методов, - оценок их качества и времени, затраченного на их получение на конкретных выборках.

Таким образом,

- *оценка метода* индуктивного формирования знаний - это *функция*, а
- *оценка ИФБЗ* - значение этой *функции в точке*.

Если известно достаточно много значений функции в точках, то можно понять, какой вид имеет эта функция. Среди оценок значений таких функций в точке выделяют:

- *оценку качества ИФБЗ*,
- *оценку времени*, затраченного на ее получение на конкретных выборках, а также
- *сравнительную оценку*.

1. *Оценки качества ИФБЗ* могут быть двух видов - внешние и внутренние. Они определяются по аналогии с тестированием программ по методу черного и белого ящика.

Внешняя оценка представляет собой оценку ошибок 1-го и 2-го рода, то есть оценку числа дефектов ИФБЗ на контрольной выборке.

Ошибкой 1-го рода считается несоответствие примера из выборки примеров ИФБЗ. Ошибкой 2-го рода считается соответствие контрпримера из выборки контрпримеров ИФБЗ. На рис. 2 представлена условная схема внешней оценке качества индуктивно сформированных БЗ.

Внутренняя оценка является оценкой близости ИФБЗ к "идеальной" БЗ (к знаниям о том, как в действительности устроен мир).

Внутренняя оценка может быть экспертной или автоматической. Если ИФБЗ понятна эксперту ПО, тогда такая оценка применима. Более того, она значительно предпочтительнее формального критерия близости, так как эксперт в первую очередь обращает внимание на важные вещи. При этом его "идеальная" БЗ не обязательно должна быть явно представлена.

2. *Оценка близости* в этом случае - содержательный комментарий: что

в ИФБЗ упущено из-за недостаточно репрезентативной выборки, какие новые знания обнаружил метод ИФЗ и т.д. Если ИФБЗ не понятна эксперту, необходимо сформулировать формальный критерий близости, то есть некий набор правил, проверка которых осуществлялась бы автоматически.

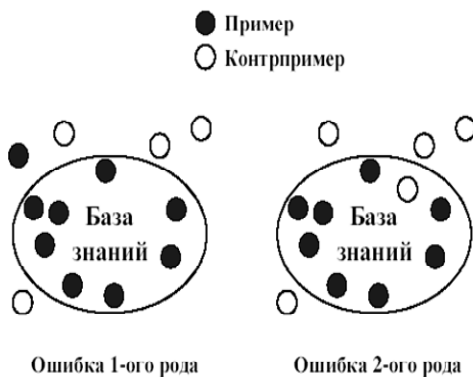


Рис. 2. Внешняя оценка качества ИФБЗ

3. *Сравнительная оценка* - это кортеж внешних оценок нескольких ИФБЗ, полученных с помощью различных методов ИФЗ на одних и тех же выборках. Если для методов от 1 до n известна оценка некоторого свойства метода ИФЗ и разработан некий новый метод ИФЗ i , то сравнительная оценка позволяет найти место его оценки в кортеже.

2. Постановка задачи экспериментального исследования свойств методов ИФЗ

В настоящее время необходимо сделать следующие существенные шаги на пути к решению проблемы экспериментального исследования свойств методов ИФЗ:

- ➔ создать репозитории баз данных, включающие в себя выборки реальных данных из различных ПО, а также
- ➔ разработать несколько универсальных программ генерации модельных данных.

Однако на практике свойства метода ИФЗ чаще всего изучаются на одной сравнительно небольшой выборке реальных данных. Часто это изучение ограничивается лишь оценкой процента ошибок 1-го рода ИФБЗ или сравнительной оценкой свойств методов ИФЗ и их устойчивости на небольшом количестве выборок небольшого объема из какого-либо репозитория.

Таким образом, изучение свойств метода (то есть функций, зависящих от объема выборки) заменяется изучением одного значения лишь одной функции в одной точке, обычно далекой от асимптотики. Важность же таких исследований состоит в том, что, какие бы внешние оценки метода ИФЗ не

были бы получены, у пользователя ИФБЗ без исследования ее внутреннего качества нет полной уверенности в том, что этот метод действительно восстанавливает знания ПО.

Из вышесказанного следует, что существует потребность в общем подходе к экспериментальному исследованию свойств методов ИФЗ. В подобных исследованиях для каждого метода необходимо определить:

✚ *зависимость качества ИФБЗ* от объема и сложности выборки (сложности примеров и контрпримеров), а также от сложности модели ПО (количества классов, признаков и их значений);

✚ *зависимость времени ИФЗ* от объема и сложности выборки, а также от сложности модели ПО для таких методов ИФЗ, для которых на выборках практически важного объема время вычислений может быть велико;

✚ *характеристики устойчивости* значений свойств метода ИФЗ (качества ИФБЗ и времени ИФЗ) относительно выборок одной ПО и относительно свойств разных ПО - для проблемно-независимых методов.

Исходными данными таких исследований должны быть множество выборок различного объема для одной и той же ПО (для проблемно-ориентированных и проблемно-независимых методов ИФЗ) и множество выборок различного объема для разных ПО (для проблемно-независимых методов ИФЗ).

3. Общий подход к исследованию свойств методов ИФЗ

По указанным выше причинам использование реальных данных в полноценном исследовании свойств методов ИФЗ ограничено. Поэтому такие исследования целесообразно проводить на модельных данных, а затем, используя полученную в подобных исследованиях информацию, применять метод ИФЗ к выборкам реальных данных (такого объема, который имеется в распоряжении исследователя) и с учетом его свойств интерпретировать оценки полученных ИФБЗ. На основе информации об устойчивости свойств метода можно сделать вывод о том, какова возможная погрешность оценок ИФБЗ на имеющейся выборке и насколько ИФБЗ может быть близка к "идеальной" БЗ. Если эти оценки ИФБЗ окажутся неприемлемыми, можно прогнозировать, каково будет время работы метода при увеличении объема выборки и насколько этот объем необходимо увеличить для достижения приемлемой оценки. Если результаты, полученные на модельных данных, не соответствуют результатам, полученным на реальных данных, то можно косвенно судить о том, что предположения о ПО, лежащие в основе метода ИФЗ, в данном случае не выполняются.

Перейдем к описанию общего подхода. Введем следующие обозначения.

Пусть M - множество БЗ;

$m \in M$ - некоторая БЗ;

O - множество объектов.

Тогда обозначим $O(m) \subset O$ - множество примеров;

$O \setminus O(m)$ - множество контрпримеров для БЗ $m \in M$.

Пусть V - множество выборок;

$v \in V$ - конкретная выборка (конечное множество объектов - элементов O), $v \subset O$.

В основе любого метода ИФЗ лежит некоторый класс БЗ (моделей) M . Метод ИФЗ решает обратную задачу вычисления значения некоторого отображения $V \rightarrow M$, то есть на основе имеющейся выборки V метод выбирает одну из БЗ $m^*(v)$.

Идея предлагаемого общего подхода состоит в том, чтобы для получения модельных данных, необходимых для экспериментальных исследований свойств метода ИФЗ, явно представить класс БЗ M .

Далее на основе этого явного представления разработать алгоритмы решения следующих прямых задач:

- случайного выбора (или генерации) БЗ m из M ;
- случайной генерации выборок модельных данных $y(m)$ различного объема на основе БЗ m ;
- каждая из таких выборок $y(m)$ представляет собой объединение выборки примеров $v_n(m) \subset O(m)$ и выборки контрпримеров $v_{kn}(m) \subset O \setminus O(m)$.

После этого в процессе экспериментального исследования, применяя метод ИФЗ к сгенерированным выборкам модельных данных $v(m)$ разного объема, можно получать различные ИФБЗ $m^*(v)$ и, используя их оценки, получать оценки свойств метода ИФЗ: *внешнего* и *внутреннего* качества, *времени работы*, а также *устойчивости* в зависимости от объема выборок.

1. При этом *оценка внешнего качества* ИФБЗ - это оценка процента ошибок 1-го рода $m^*(v)$ на контрольной $v_n(m)$ и оценка процента ошибок 2-го рода $m^*(v)$ на контрольной $v_{kn}(m)$.

- Первая представляет собой долю элементов $v_n(m)$, соответствующих ИФБЗ $m^*(v)$, от объема $v_n(m)$,
- вторая - долю элементов $v_{kn}(m)$, не соответствующих ИФБЗ $m^*(v)$, от объема $v_{kn}(m)$.

2. Оценка *внутреннего качества* - это оценка сходства m и $m^*(v)$.

3. Оценка *устойчивости* может быть получена с помощью генерации выборок различного объема и анализа разброса значений свойства метода ИФЗ для этих выборок.

Заключение. Таким образом, предлагаемый подход свободен от трудностей, связанных с экспериментальными исследованиями свойств методов ИФЗ на реальных данных, и позволяет проводить подобные исследования на большом числе выборок различного объема. Результаты экспериментальных исследований методов ИФЗ (значения свойств этих методов на различных объемах выборок) могут быть использованы в качестве

наполнения каталога свойств методов ИФЗ, в котором явно приведены их свойства и значения этих свойств для каждого метода.

1. Валькман Ю. Р., Быков В. С. Дедуктивные и недедуктивные аспекты в моделировании образного мышления. // Моделювання та інформаційні технології, Зб. наук.пр. ПМЕ, Київ, 2006, Випуск 35, - с. 87 - 96.
2. Валькман Ю. Р., Дембовский О. Ю. Процессы порождения и обоснования гипотез: индукция, аналогия, абдукция и дедукция. // Моделювання та інформаційні технології, Зб. наук.пр. ПМЕ, Київ, 2008, Випуск 46, - с. 12 - 22.
3. Валькман Ю. Р. Гипотезы: определения, структура, классификация, жизненные циклы // Моделювання та інформаційні технології, Зб. наук.пр. ПМЕ, Київ, 2008, Випуск 47, - с. 91- 98.
4. Вагин В.Н., Головина Е.Ю., Загорянская А.А. и др. Достоверный и правдоподобный вывод в интеллектуальных системах / Под ред. В.Н. Вагина, Д.А. Поспелова. М.: Физматлит, 2004. – 988 с.
5. Дюк В., Самойленко А. Data Mining: учебный курс. Издательство: Питер, 2001- 368 с.
6. Загоруйко Н. Г. Прикладные методы анализа данных и знаний. Новосибирск: Изд-во Ин-та математики, 1999. - 270 с.
7. Клецев А.С. Задачи индуктивного формирования знаний в терминах непримитивных онтологий предметных областей. Препринт 6 – 2003. Владивосток: ИАПУ ДВО РАН, 2003 - 35 с.
8. Степашко В. С. Алгоритмы МГУА как основа автоматизации процесса моделирования по экспериментальным данным. // Автоматика. - 1988. - №4. - С. 44-45.
9. Финн В.К. Правдоподобные рассуждения в интеллектуальных системах типа ДСМ. Итоги науки и техники. Информатика, Т.15, Москва: 1991.

Поступила 19.11.2008р.

УДК 621.3

М.А.Никулин, Ю.М.Коростиль

АНАЛИЗ СИСТЕМ УПРАВЛЕНИЯ ТРАНСПОРТОМ

Системы управления транспортом достаточно широко используются в различных технических областях и сферах человеческой деятельности. Существует возможность классифицировать такие системы на основании различных признаков. В данном случае воспользуемся следующей классификацией соответствующих систем:

- системы, в которых транспортные средства связаны между собой жесткими связями;
- системы, в которых транспортные средства имеют определённую степень свободы по отношению друг к другу в решении определённой задачи;