

## ГИПОТЕЗЫ ПРИ РАССЛЕДОВАНИИ АВИАЦИОННЫХ ПРОИСШЕСТВИЙ

д. т. н. Ю. В Попов, П. А. Комардин  
Межгосударственный авиационный комитет

*В статье обсуждаются проблемы выдвижения гипотез при расследовании авиационных происшествий. Приводится общая постановка проблемы. Рассмотрена основная задача построения алгоритма выдвижения гипотезы на основе кластерного анализа.*

В науке по расследованию авиационных происшествий (АП), как и в других науках, гипотеза занимает видное место. В обычном употреблении гипотеза имеет несколько значений. При расследовании АП гипотезой называют предположение о причине известной совокупности явлений, недоступной в настоящее время обнаружению, однако недоступной только в силу случайных обстоятельств, так что причина эта в любой момент может быть обнаружена и может стать предметом наблюдения [1].

Гипотеза – это всеобщая необходимая для любого познавательного процесса форма развития знаний. Там, где есть поиск новых идей или фактов, закономерных связей или причинных зависимостей, там всегда присутствует гипотеза. Поэтому после АП всегда возникают различного рода гипотезы.

Построение гипотезы всегда сопровождается выдвижением предположений о природе исследуемых явлений, которые являются логической сердцевиной гипотезы и формулируются в виде отдельного суждения или системы взаимосвязанных суждений. Гипотеза всегда содержит в себе нуждающееся в проверке вероятное знание. На рис. 1 для наглядности приведена схема, которая позволяет понять место гипотезы при расследовании АП.

К гипотезе предъявляются следующие требования:

- она не должна включать в себя слишком много положений: как правило, одно основное, редко больше;
- в нее нельзя включать понятия и категории, не являющиеся однозначными, не уясненные самим исследователем;
- при формулировке гипотезы следует избегать ценностных суждений, гипотеза должна соответствовать фактам, быть проверяемой;
- требуется безупречное стилистическое оформление, логическая простота, соблюдение преемственности.

В процессе проведения расследования АП гипотезы различаются по своим познавательным функциям и по объекту исследования.

По функциям в познавательном процессе гипотезы различают

описательные и объяснительные.

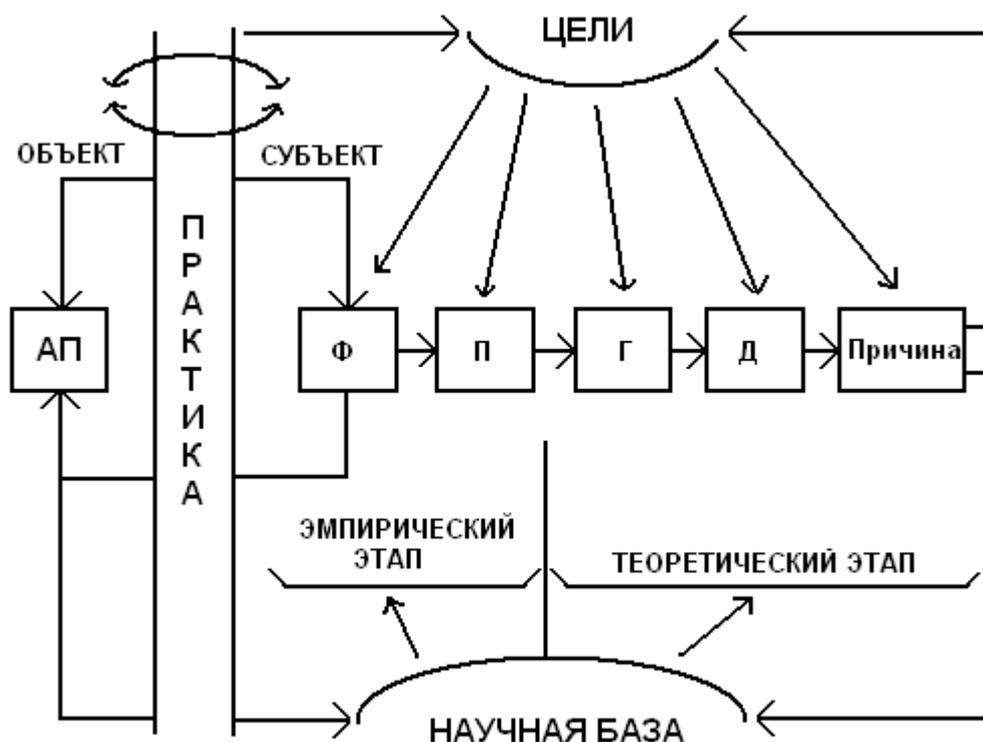


Рис. 1. Место гипотезы при расследовании АП

Описательная гипотеза – это предположение о присущих особенностях исследуемого АП. Оно обычно отвечает на вопрос: «Что представляет собою данное АП?» или «Какими признаками обладает данное АП?».

Описательные гипотезы выдвигаются с целью выявления состава или структуры АП, раскрытия механизма или процедурных особенностей его возникновения и причины, определения действующих характеристик при АП и дальнейших действий по установлению причин АП.

Объяснительная гипотеза - это предположение о причинах возникновения АП. Такие гипотезы обычно выясняют: «Почему произошло АП?» или «Каковы причины АП?».

Объяснительные гипотезы фиксируют возможные следствия из определенных причин, а также характеризуют условия, при которых эти следствия обязательны, т.е. объясняется, в силу каких признаков и условий возможно данное следствие, каков механизм их проявления.

Как правило, по логической структуре гипотезы при расследовании АП одинаковы, но они различаются по своему содержанию и выполняемым функциям. На рис. 2 представлена классификация гипотез.

Общая гипотеза - это вид гипотезы, объясняющей причину явления или группы явлений в целом.

Частная гипотеза - это разновидность гипотезы, объясняющая какую-либо отдельную сторону или отдельное свойство явления или события.

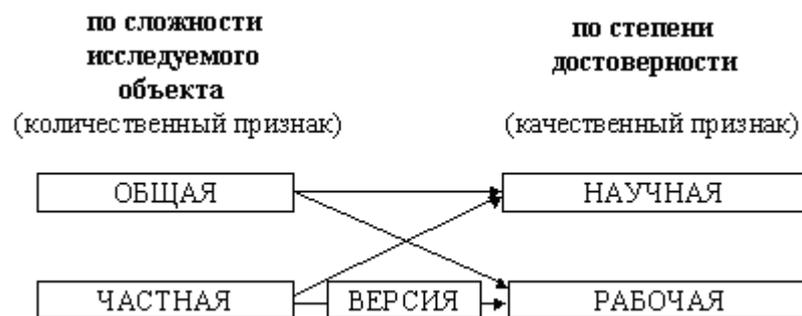


Рис. 2. Классификация гипотез

При этом необходимо иметь в виду, что деление гипотезы на общую и частную имеет смысл, когда мы соотносим одну гипотезу с другой. Это деление не является абсолютным, гипотеза может быть частной по отношению к одной гипотезе и общей по отношению к другим гипотезам.

Научная гипотеза – это предложение о закономерностях развития природы, общества или мышления, т. е. о явлениях, носящих общий характер и служащих предметом научного исследования.

Рабочая гипотеза - это временное предположение или допущение, которым пользуются при построении гипотезы. Рабочая гипотеза выдвигается, как правило, на первых этапах исследования. Она непосредственно не ставит задачу выяснить действительные причины исследуемых явлений, а служит лишь условным допущением, позволяющим сгруппировать и систематизировать результаты наблюдений и дать согласующееся с наблюдениями описание явлений. Рабочие гипотезы выдвигаются для расследования АП.

Таким образом, использование гипотез при расследовании АП имеет большое значение. Гипотеза формирует сведения и определения причины АП. С точки зрения логической структуры она не сводится к какой-то одной форме мышления: понятию, суждению или умозаключению, а включает в свой состав все эти формы.

Гипотеза представляет собой процесс развития мысли. Безусловно, дать общий образец построения гипотезы для всех случаев жизни не представляется возможным. Это связано с тем, что условия для разработки гипотезы зависят от своеобразия практической деятельности, а также от специфики рассматриваемой проблемы. Тем не менее, можно определить общие границы этапов, которые проходят мыслительный процесс в гипотезе. Основными этапами разработки гипотезы являются:

- выдвижение гипотезы;
- развитие гипотезы;
- проверка гипотезы.

Чтобы выдвинуть гипотезу, необходимо располагать некоторой совокупностью фактов, относящихся к АП, которые бы обосновывали вероятность определенного предположения, объясняли неизвестное. Поэтому

построение гипотезы связано, в первую очередь, с собиранием фактов, имеющих отношение к АП, которое расследуется.

На основании собранных фактов высказывается предположение о том, что представляет собой расследуемое АП, т.е. формулируется гипотеза в узком смысле слова.

Развитие гипотезы связано с выведением из нее логических следствий. Предполагая выдвинутое положение истинным, из него дедуктивным путем выводят ряд следствий, которые должны существовать, если существует предполагаемая причина.

Логические следствия, выводимые из гипотез, нельзя отождествлять со звеньями причинно-следственной цепи явлений, всегда хронологически следующими за вызвавшей их причиной. Под логическими следствиями понимаются мысли не только об обстоятельствах, вызванных АП, но и об обстоятельствах, предшествующих ему по времени, сопутствующих и последующих, а также об обстоятельствах, вызванных иными причинами, но находящихся с АП в какой-либо связи.

Проверка гипотезы на практике, превращение ее в достоверное знание есть процесс сложный и длительный. Поэтому проверку истинности гипотезы нельзя сводить к какому-то одному логическому действию. При проверке гипотезы используются различные логические формы и способы доказательства или опровержения (рис. 3).

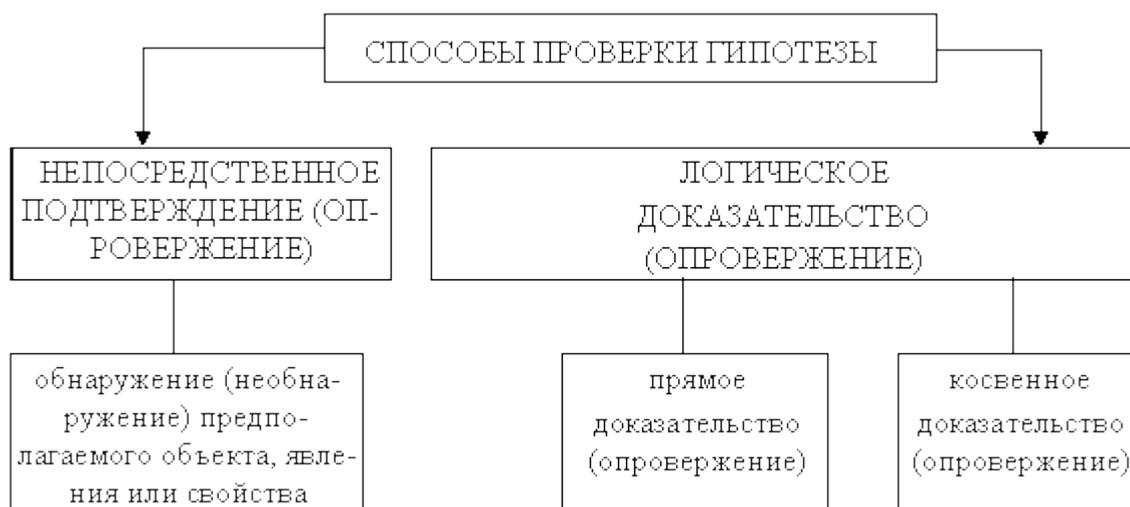


Рис. 3. Способы проверки гипотез

Мы предлагаем алгоритм выдвижения рабочей гипотезы при расследовании АП, который основан на методе кластерного анализа [2]. Кластерный анализ — математическая процедура многомерного анализа, позволяющая на основе множества показателей, характеризующих ряд объектов, сгруппировать их в классы (кластеры) таким образом, чтобы объекты, входящие в один класс, были более однородными, сходными по сравнению с объектами, входящими в другие классы.

Задачи, решаемые с использованием кластерного анализа, делятся следующим образом [3]:

- разработка классификации;
- исследование схем группирования объектов;
- порождение гипотез на основе исследования данных;
- отнесение собранных данных к тому или иному объекту или гипотезе.

Процедура кластеризации в общем виде может быть представлена последовательностью следующих шагов:

- формирование множества элементов, подлежащих делению на кластеры;
- определение множества признаков, по которым должны оцениваться элементы множества;
- определение меры сходства между элементами множества;
- деление элементов множества на кластеры.

Исходными данными для составления классификации гипотез являются факты, относящиеся к различным АП, которые могут быть представлены виде матрицы наблюдений:

$$X = \begin{matrix} & \text{Ж} & x_{11} & x_{12} & x_{13} & \cdots & x_{1j} & \text{Ц} \\ & & \begin{matrix} \text{З} \\ \text{З} \end{matrix} & x_{21} & x_{22} & x_{23} & \cdots & x_{2j} & \text{Ч} \\ & & \begin{matrix} \text{З} \\ \text{З} \end{matrix} & x_{31} & x_{32} & x_{33} & \cdots & x_{3j} & \text{Ч} \\ & & \begin{matrix} \text{З} \\ \text{З} \end{matrix} & \vdots & \vdots & \vdots & \ddots & \vdots & \text{Ч} \\ & & \begin{matrix} \text{З} \\ \text{И} \end{matrix} & x_{i1} & x_{i2} & x_{i3} & \cdots & x_{ij} & \text{Ш} \end{matrix},$$

где  $j=1, 2, \dots, n$  – номер признака;

$i=1, 2, \dots, m$  – номер наблюдения;

$x_{ji}$  – значение  $j$  признака у  $i$  наблюдения.

Абсолютное большинство методов кластеризации [2 - 6] базируется на анализе квадратной и симметричной относительно главной диагонали матрицы  $D$  коэффициентов сходства (расстояния, сопряженности, совместимости, корреляции и т.д.) между фактами исходной матрицы наблюдений:

$$D = \begin{matrix} & \text{Ж} & 0 & d_{12} & d_{13} & \cdots & d_{1p} & \text{Ц} \\ & & \begin{matrix} \text{З} \\ \text{З} \end{matrix} & d_{21} & 0 & d_{23} & \cdots & d_{2p} & \text{Ч} \\ & & \begin{matrix} \text{З} \\ \text{З} \end{matrix} & d_{31} & d_{32} & 0 & \cdots & d_{3p} & \text{Ч} \\ & & \begin{matrix} \text{З} \\ \text{З} \end{matrix} & \vdots & \vdots & \vdots & \ddots & \vdots & \text{Ч} \\ & & \begin{matrix} \text{З} \\ \text{И} \end{matrix} & d_{p1} & d_{p2} & d_{p3} & \cdots & 0 & \text{Ш} \end{matrix}$$

Матрица коэффициентов сходства организовывается по заданному набору признаков с использованием различных формул для меры расстояния, выбираемых исследователем. Наиболее прямой путь вычисления расстояний между признаками матрицы наблюдений состоит в вычислении евклидовых расстояний. Эта мера является реальным геометрическим расстоянием между признаками в матрице наблюдений. Однако алгоритм объединения не «заботится» о том, являются ли «предоставленные» для этого расстояния

настоящими или некоторыми другими производными мерами расстояния, что более значимо для исследователя. Задачей исследователей является подобрать правильный метод для специфических применений.

Мера расстояния по Евклиду является геометрическим расстоянием в многомерном пространстве и вычисляется следующим образом:

$$d_{ij} = e \sqrt{(x_{ij} - x_{ij})^2}$$

Заметим, что евклидово расстояние (и его квадрат) вычисляется по исходным, а не по стандартизованным данным. Это обычный способ его вычисления, который имеет определенные преимущества (например, расстояние между двумя объектами не изменяется при введении в анализ нового признака, который может оказаться выбросом). Тем не менее, на расстояния могут сильно повлиять отличия между осями, по координатам которых вычисляются эти расстояния.

Для придания отдаленным друг от друга фактам больше веса используется мера квадрата Евклидова расстояния. Это расстояние вычисляется следующим образом (см. также замечания в предыдущем пункте):

$$d_{ij} = e (x_{ij} - x_{ij})^2$$

Расстояние городских кварталов (манхэттенское расстояние) является просто средним разностей по координатам. В большинстве случаев эта мера расстояния приводит к таким же результатам, как и для обычного расстояния Евклида. Однако отметим, что для этой меры влияние отдельных больших разностей (выбросов) уменьшается (так как они не возводятся в квадрат). Манхэттенское расстояние вычисляется по формуле:

$$d_{ij} = e |x_{ij} - x_{ij}|$$

Расстояние Чебышева может оказаться полезным, когда желают определить два признака как «различные», если они различаются по какой-либо одной координате (каким-либо одним измерением). Расстояние Чебышева вычисляется по формуле:

$$d_{ij} = \max |x_{ij} - x_{ij}|$$

Наиболее общей формулой для подсчета расстояния в  $m$ -мерном признаковом пространстве между признаками  $x_{ij}$  и  $x_{ij}$  является мера Минковского (степенное расстояние):

$$d_{ij} = e \left| x_{ij} - x_{ij} \right|^p$$

где  $r$  и  $p$  – параметры, определяемые исследователем, с помощью которых можно прогрессивно увеличить или уменьшить вес, относящийся к переменной  $i$ , по которой соответствующие признаки наиболее отличаются.

Параметр  $p$  ответственен за постепенное взвешивание разностей по отдельным координатам, параметр  $r$  определяет прогрессивное взвешивание больших расстояний между объектами. Если оба параметра  $r$  и  $p$ , равны двум, то это расстояние совпадает с расстоянием Евклида.

Собственно кластерный анализ включает в себя набор различных алгоритмов классификации, сутью которых является группировка данных в наглядные структуры (таксоны). На первом шаге, когда каждый признак представляет собой отдельный кластер, расстояния между этими признаками определяются выбранной мерой. Однако когда связываются вместе несколько признаков, возникает вопрос, как следует определить расстояния между кластерами? Другими словами, необходимо правило объединения или связи для двух кластеров. Пусть исходные данные - матрица сходства  $D$ , где  $d_{ij}$  - некоторая мера близости между каждой парой классифицируемых признаков  $x_{ij}$  и  $x_{ji}$ . Хорошо известно, что для любого заданного разбиения признаков на группы и любого  $\epsilon > 0$  можно указать метрику, такую, что расстояния между признаками из одной группы будут меньше  $\epsilon$ , а между признаками из разных групп - больше  $1/\epsilon$ . Тогда любой разумный алгоритм кластеризации даст именно заданное разбиение.

Наиболее часто применяется так называемый агломеративный иерархический алгоритм «Дендрограмма», отдельные версии которого отличаются правилами вычисления расстояния между кластерами.

Использование агломеративных иерархических алгоритмов для определения кластеров имеет следующие преимущества:

- последовательно объединяет наиболее сложные признаки по матрице сходства;
- последовательность объединения кластеров можно представить визуально в виде древовидной диаграммы (дендрограммы);
- для полной кластеризации признаков требуется  $N-1$  шаг ( $N$  - размер матрицы сходства);
- дает правило, указывающее, каким образом, исходя из матрицы сходства, признаки могут объединяться в кластеры.

Агломеративные иерархические методы различаются главным образом по правилам построения кластеров. Существует много различных правил группировки, каждое из которых порождает специфический иерархический метод. Известно, по крайней мере, двенадцать различных методов группировки, четыре из них наиболее распространенные: одиночной связи, полной связи, средней связи и метод Уорда [4].

Наиболее простым методом группировки является метод одиночных связей, в котором используется матрица сходства. В этом методе кластер образуется по следующему правилу: признак будет присоединен к уже существующему кластеру, если, по крайней мере, один из элементов кластера находится на том же уровне сходства, что и признак, претендующий на включение. Таким образом, присоединение определяется лишь наличием единственной связи между признаком и кластером. Главное преимущество этого метода заключается в его математических свойствах: результаты,

полученные по этому методу, инвариантны к монотонным преобразованиям матрицы сходства; применению метода не мешает наличие совпадений в данных. Первые из этих свойств особенно важны по той причине, что все другие агломеративные иерархические методы такими свойствами не обладают. Это означает, что метод одиночной связи является одним из немногих методов, результаты применения которых не изменяются при любых преобразованиях данных, оставляющих без изменения относительное упорядочение элементов матрицы сходства.

В методе одиночных связей правила группировки определяются формулой:

$$d(h, k) = 0,5 \left( d(h, i) + d(k, j) + \left| d(h, i) - d(k, j) \right| \right),$$

где  $d(h, k)$  - различие (расстояние) между кластерами  $h$  и  $k$ , причем кластер  $k$  является результатом объединения кластеров (признаков)  $i$  и  $j$ , в ходе агломеративного шага ( $i = 1 \dots k, j = 1 \dots h$ ).

Кроме агломеративных иерархических алгоритмов классификации также большое распространение получили различные итерационные процедуры, которые пытаются найти наилучшее разбиение, ориентируясь на заданный критерий оптимизации, не строя при этом полного дерева (метод  $K$ -средних Мак-Кина, алгоритмы «Форель», «Медиана», «Краб» и т.д.). Итерационный процесс начинается, как правило, с  $K$  случайно выбранных кластеров, а затем изменяется принадлежность объектов к ним, чтобы:

- минимизировать изменчивость внутри кластеров;
- максимизировать изменчивость между кластерами.

Для этих алгоритмов важной является «проблема остановки»: завершится ли процесс улучшения положения центра кластера через конечное число шагов или же он может быть бесконечным.

Нами предложена методика по автоматизации решения задачи по выдвижению гипотезы при расследовании АП. Для ее осуществления необходимо сформировать признаковое пространство. Задача построения признакового пространства для проведения кластерного анализа для выдвижения гипотез при расследовании АП является довольно сложной. В признаковое пространство необходимо включать все признаки, относительно которых известно, что они, в принципе, в той или иной мере характеризуют АП. В него не следует включать те признаки, относительно которых отсутствует априорная информация о связях между этими признаками и причинами АП. Построенное таким образом признаковое пространство будем называть априорным. В этой статье не предполагалось осветить методы построения признакового пространства. Данной задачи необходимо посветить отдельную статью.

В заключение следует сказать, что предложенная методика позволит сократить сроки проведения расследований АП и установления причины, а, следовательно, повысит безопасность полетов.

## Литература

1. *Попов Ю. В.* Диагностика при расследовании авиационных происшествий// Проблемы безопасности полетов № 9 – 2007. – С. 18 - 28.
2. Классификация и кластер/ Под ред. Дж. Вэн-Райзина. – М.: Мир, 1980. – 390 с.
3. *Мандель И. Д.* Кластерный анализ. – М.: Финансы и статистика, 1988. – 176 с.
4. Факторный, дискриминантный и кластерный анализ/ Дж.-О. Ким, Ч. У. Мьюллер, У. Р. Клекка и др. – М.: Финансы и статистика, 1989. – 215 с.
5. *Дубнов П. Ю.* Обработка статистической информации с помощью SPSS – М.: ООО «Издательство АСТ»: Издательство «НТ Пресс», 2004. – 221 с.
6. *Попов Ю. В.* Применения кластерного анализа для группирования угроз информации, зарегистрированной самописцами/ Научный Вестник МГТУ ГА № 99(3) серия «Эксплуатация воздушного транспорта и ремонт авиационной техники. Безопасность полетов» – М.: МГТУ ГА, 2006. – С. 119 - 123.