

И. М. Гельфанд, Ш. А. Губерман, М. П. Жидков, М. С. Калецкая,
В. И. Кейлис-Борок, Е. Я. Ранцман, И. М. Ротвайн

ПРОГНОЗ МЕСТА ВОЗНИКНОВЕНИЯ СИЛЬНЫХ ЗЕМЛЕТРЯСЕНИЙ КАК ЗАДАЧА РАСПОЗНАВАНИЯ¹

1. **Общие сведения.** Прогноз места землетрясений важен теоретически — для понимания механизма геологического развития Земли. Он имеет также большое практическое значение: в зонах, где возможны разрушительные землетрясения, необходимы специальные мероприятия по повышению устойчивости зданий. О масштабе этих мероприятий можно осудить по тому, что они обязательны на 13% территории СССР (2,9 млн. км² с населением свыше 30 млн. человек), где их стоимость составляет около 100 млн. рублей в год [7].

Рассматриваемая задача возникает из-за того, что сильные землетрясения — явление редкое, а систематически регистрируются они лишь последние 50—70 лет. Поэтому их можно ожидать не только там, где они уже известны, но и в новых местах.

По современным представлениям сильные землетрясения возникают в пределах тех территорий, где верхние слои Земли (земная кора) раздроблены и охвачены интенсивными тектоническими движениями в новейшее геологическое время (последние 25 млн. лет). К таким территориям относятся разнотипные пояса современного горообразования: Кордильеры Америки, западная периферия Тихого океана, Европейско-Азиатский горный пояс, включающий Средиземноморье, юго-восточную Европу, Малую, Центральную и Среднюю Азию и др.

Землетрясения приурочены к поверхностям разрыва земной коры. Следы этих поверхностей на поверхности Земли мы будем называть линеаментами. Они могут прослеживаться не непрерывно и местами — по геологическим данным, а местами — только по контрастным формам рельефа. Те участки линеаментов, которые выражены геологически, называются геологическими разломами. Сами землетрясения связаны с разрывом и быстрым перемещением вещества вдоль некоторого участка поверхности разрыва. Проекция на поверхность Земли точки, в которой начинаются перемещения, называется эпицентром землетрясения.

Чтобы определить места, где возможны землетрясения той или иной силы, собирается и анализируется широкий комплекс геолого-геофизических данных: о строении и геологической истории земной коры, об аномалиях геофизических полей, о слабых землетрясениях (которые гораздо более многочисленны) и др. Однако до сих пор не удавалось точно сформулировать критерии прогноза этих мест, и он делается с большим запасом, чтобы уменьшить риск пропуска.

¹ Полное описание полученных результатов можно найти в [1—6].

2. **Постановка задачи.** Мы рассматривали следующие регионы: Памир и Тянь-Шань, Армянское нагорье с хребтами Малого Кавказа, Анатолию, Балканы и бассейн Эгейского моря. Линеаменты образуют на карте этих регионов сложную сетку. Вокруг пересечения линеаментов развиты дизъюнктивные узлы — специфические структуры, отличающиеся повышенной раздробленностью земной коры и резкой неравномерностью тектонических движений, выраженной в рельефе земной поверхности. Было замечено, что эпицентры землетрясений с энергией от 10^{22} — 10^{23} эрг и выше (магнитуда $M \geq \sim 6,5$) приурочены к дизъюнктивным узлам. Для Тянь-Шаня и Памира это можно проверить непосредственно, поскольку границы узлов примерно известны. Вероятность случайного совпадения оказывается пренебрежимо мала. В дальнейшем мы рассматривали только такие землетрясения; их мы и называем сильными.

Для остальных областей границы узлов не определялись, поскольку это требует специальных полевых исследований. Мы построили распределение расстояний от каждого эпицентра до ближайшего пересечения разломов. Оно согласуется с гипотезой, что сильные землетрясения и в этих областях возникают только в дизъюнктивных узлах (если учесть, что линейные размеры узлов составляют, по меньшей мере, десятки километров). Мы приняли, что и в дальнейшем сильные землетрясения будут происходить только в дизъюнктивных узлах — тех, где они уже известны и, быть может, еще в некоторых других (строго говоря, мы доказали только, что вне узлов вероятность сильных землетрясений гораздо меньше).

Тогда мы имеем следующую задачу. Даны две группы узлов; в первой — В («высокосейсмичные») — землетрясения известны, во второй — нет. Даны характеристики каждого узла. Требуется разделить вторую группу на узлы, похожие (B^*) и непохожие (H) на узлы первой группы. Естественно считать, что сильные землетрясения невозможны в узлах H и возможны, хотя пока неизвестны, в узлах B^* .

Как сказано выше, исходные характеристики узлов можно выбирать из весьма широкого комплекса геолого-геофизических данных. Однако в каждом регионе есть всего по несколько десятков узлов и, как показывает опыт, использовать больше 10—15 характеристик нет смысла. Поэтому мы отобрали для начала небольшую, простейшую часть этого комплекса, соответствующую указанным выше представлениям: простые характеристики рельефа земной поверхности, отражающие интенсивность и контрастность новейших геологических движений (размах высот, максимальная высота и т. п.), и столь же простые характеристики раздробленности земной коры в узле (число пересекающихся разломов, степень различия разделяемых ими областей и т. п.). Каждый узел описывался 11 или 17 характеристиками с общим числом двоичных разрядов соответственно 28 или 41.

3. Методика и результаты. Поставленная задача обладает рядом особенностей, которые могут встретиться при распознавании самых разных, не только геофизических, объектов. В практике распознавания такие задачи пока не рассматривались (и, возможно, избегались). Мы использовали программу Кора-3, которая оказалась достаточно гибкой, чтобы решить эту задачу.

Материал обучения неточно разделен на два класса. Часть объектов заведомо принадлежит $\{B\}$. Но каждый из остальных объектов может быть или B^* или H , причем объекты B^* сходны с B . Наша задача имеет смысл при гипотезе, что в классе $\{B^* + H\}$ большинство узлов относится к H , т. е. что в большинстве высокосейсмичных узлов сильные землетрясения уже успели произойти за период наблюдений, а узлов H достаточно много. Эта особенность должна была учитываться уже при первичной обработке материала — при оценке информативности параметров и выборе порогов для их квантования. Более того, это потребовало модификации критериев отбора признаков. Обычно Кора-3 отбирает «чистые» признаки — конъюнкции трех переменных, встречающиеся только в одном классе и притом достаточно часто [9]. В нашей же задаче признаки узлов B должны встречаться на узлах B^* , если они входят во второй класс. Поэтому мы искали для класса B «нечистые» признаки, которые встречаются не менее K_1 раз среди узлов B и не более \bar{K}_1 раз среди узлов $\{B^* + H\}$.

Для узлов H можно было бы искать чистые признаки, ибо в противоположном классе все узлы принадлежат $\{B\}$. Однако, учитывая, что в силу недостаточной изученности узлов их параметры определены с ошибками, мы и для класса H искали «нечистые» признаки, встречающиеся не менее K_2 раз среди $\{B^* + H\}$ и не более \bar{K}_2 раз среди $\{B\}$; практически принималось $K_2 = 1$.

Эта особенность в значительной мере определяла и выбор решающего правила. Он был основан на разности Δ числа признаков B и H .

Обычное правило состоит в том, что к B^* относят узлы с $\Delta > 0$ («простым большинством голосов»). Однако в нашей задаче это неприемлемо: во-первых, цены ошибок первого и второго рода различны, особенно нежелательны ошибки типа «пропуск цели» (отнесение к H узлов B^*); во-вторых, мы должны идентифицировать B^* по сходству с B , а при использовании нечистых признаков сами узлы B могут иметь $\Delta < 0$. Поэтому решающее правило формулировалось так.

Пусть $\underline{\Delta}$ — минимальное значение Δ для узлов B . К B^* относились узлы с Δ выше или ненамного ниже $\underline{\Delta}$ (с $\Delta = \underline{\Delta}$ для Памира и Тянь-Шаня; с $\Delta = \underline{\Delta} - 2$ для остальных менее изученных регионов). При этом к B^* была отнесена лишь небольшая часть узлов второй группы. При уменьшении указанного порога на 1 или 2 лишь небольшое число узлов второй группы дополнительно переходит в H ; большинство узлов резко разделяется по значениям Δ на две группы, что подтверждает исходную гипотезу.

Регион	Число узлов				Контрольный эксперимент		
	В	Н	В*	\bar{V}^*	T	N	n
Памир и Тянь-Шань	17	19	6	7	1911	7	0
Балканы	13	11	4	4	1905	6	0
Эгейский	12	9	6	8	1944	6	1
Анатолия и Армянское нагорье	15	41	13	15	1938	10	0

Примечание. \bar{V}^* — число узлов, распознанных как V^* при эксперименте, «продолженном в будущее»; начиная с T года удается предсказать все землетрясения, произошедшие с T по 1971 г. Исключение — землетрясение 1954 г. в Эгейском регионе. N — число объектов в материале обучения (т. е. число известных землетрясений до T -го года); n — число ошибок в предсказании землетрясений, происшедших до T -го года.

Проверка результатов путем традиционного экзамена была невозможна, поскольку общее число объектов невелико. Поэтому проводился следующий контрольный эксперимент.

Попытаемся решить рассматриваемую задачу как бы в прошлом — в то время, когда в нескольких узлах V еще не произошли сильные землетрясения. Тогда эти узлы, сохраняя свои характеристики, войдут в материале обучения не в $\{V\}$, а в $\{V^* + N\}$. Применим тот же алгоритм распознавания и посмотрим, удастся ли предсказать высокую сейсмичность этих узлов. Из таблицы видно, что, начиная с довольно далекого прошлого, мы могли бы получать хороший прогноз: заблаговременно выделять как V^* почти все те узлы, в которых в последующие годы произошли крупные землетрясения. Так, на Памире и Тянь-Шане, для которых наши данные наиболее полны, мы, начиная с 1911 г., предсказываем все узлы, в которых за последующие 60 лет произошли сильные землетрясения.

Поскольку сейсмостойкое строительство необходимо ориентировать на десятки лет вперед, следует оценить, не будем ли мы относить к V^* все новые и новые узлы по мере возникновения сильных землетрясений еще в некоторых узлах V^* . Для этого эксперимент был продолжен «в будущее»; все узлы V^* включались в V , как если бы в них уже произошли сильные землетрясения.

Из таблицы видно, что к узлам V добавляется очень мало новых узлов. Устойчивость полученных результатов подтвердили и другие контрольные эксперименты: поочередное исключение узлов из материала обучения, варьирование порогов K_i и \bar{K}_i , варьирование набора исходных характеристик.

Чтобы проверить, отражают ли полученные результаты именно высокую сейсмичность, мы группировали узлы при обучении еще и по-иному: случайно или же по геологическим регионам

(Памир и Тянь-Шань). В обоих случаях узлы нерезко разделялись по значениям Δ , и распознавание оказалось неустойчивым к вариации параметров и материала обучения. Таким образом, полученные результаты выдержали проверку на контрольных экспериментах. Более подробно они изложены в [1—6].

4. Что нам дало применение программы Кора-3?

Изложенные здесь результаты еще не дают материала для прямого обсуждения проблемы организации мышления и того, что в нем можно и стоит формализовать. Однако они позволяют сделать несколько замечаний, косвенно с нею связанных.

Мы будем исходить из того, что эти результаты верны с небольшой долей ошибок типа ложной тревоги. Фактически результаты распознавания пока нельзя доказывать. Их можно только проверять численными экспериментами, пока наши сомнения или терпение не будут исчерпаны. Перечисленные в п. 3 эксперименты представляются достаточно убедительными. Таким образом, программа Кора-3 позволила (по-видимому, впервые) точно сформулировать критерии высокой сейсмичности в виде набора признаков (см. [1—6]). Тогда возникает вопрос — в чем была роль этой программы?

Посмотрим сначала, что нового для сейсмологии в полученных результатах.

Прежде всего, это точная и ограниченная постановка задачи, то, что сильные (с энергией от 10^{22} эрг) землетрясения рассматриваются отдельно и ищутся только в дизъюнктивных узлах. Разрушительными могут быть и гораздо более слабые землетрясения, с энергией от 10^{18} — 10^{19} эрг (максимальная энергия доходит до 10^{25} — 10^{26} эрг). Однако искать общие критерии прогноза места всех разрушительных землетрясений гораздо труднее, если и возможно: при сильном землетрясении охваченный подвижкой участок разлома рассекает значительную часть земной коры до глубин 20 и более километров и тянется на 100 и более километров по горизонтали. При самых же слабых разрушительных землетрясениях подвижки охватывают разве что километры. Естественно, что явления столь разного масштаба необходимо изучать отдельно, они связаны с весьма различными геологическими процессами; кроме того, для их изучения требуются несопоставимые по детальности данные.

Неожиданно также то, что оказались достаточны столь грубые исходные характеристики узлов — небольшая и простейшая часть всего комплекса данных, изучающихся в связи с прогнозом места землетрясений. Это также можно объяснить тем, что сильные землетрясения сами представляют собой очень крупное, грубое явление.

В полученном решении новым представляется его определенность: конкретная формулировка признаков В и Н и указание узлов В*. Наконец, неожиданно, что, признаки, найденные для разных регионов, оказались не только сходными, но и в значительной

мере взаимозаменяемыми. Так, мы провели распознавание узлов Анатолии, Армянского нагорья, Балкан и Эгейского бассейна по тем признакам, которые были найдены для Памира и Тянь-Шаня (исходные характеристики для разных регионов различались только порогами дискретизации) [3]. Из 36 узлов, где по имеющимся данным были сильные землетрясения, 33 были распознаны как В. Это поразительно хороший результат, поскольку рассмотренные регионы имеют существенные различия. Из остальных 83 узлов как В* были распознаны 40, так что этот результат не обусловлен просто избытком ложных тревог. Он указывает на существование какой-то весьма общей закономерности в возникновении сильных землетрясений.

Вернемся к обсуждению роли программы Кора-3. Заметим прежде всего, что полученные результаты хорошо согласуются с установившимися за последние 10—20 лет качественными представлениями об условиях возникновения сильных землетрясений. Постановка задачи согласуется с неоднократно замечавшимся «тяготением» землетрясений к дизъюнктивным узлам. Исходные характеристики узлов отобраны, исходя из известных представлений о связи землетрясений с раздробленностью земной коры и интенсивностью и контрастностью новейших движений.

Найденное решение также согласуется с этими представлениями. Признаки узлов В соответствуют большей раздробленности и более интенсивным и контрастным новейшим движениям, чем признаки узлов Н. Наконец, распознанные по этим признакам узлы В* приурочены к наиболее протяженным и резко выраженным линеаментам, близ которых естественно было считать сейсмичность повышенной.

Таким образом, наши выводы выглядят как точная формулировка некоторых из уже известных качественных представлений. Почему же эти выводы не были получены без помощи программы распознавания?

Попробуем ответить на этот вопрос, исходя из впечатления о ходе предыдущих работ по прогнозу места землетрясений. Ничто ниже не должно пониматься, как недооценка этих работ, на опыт которых мы целиком опираемся. По сути в них также ставилась задача на распознавание — подметить отличительные признаки мест, где возможны сильные землетрясения. Однако не были должным образом разделены придумывание гипотез и их проверка. В утрированном виде такой анализ описан в романе «Война и мир» [10]: Пьер Безухов с помощью различных ухищрений пытается получить из своего имени «звериное» число 666.

Было бы ошибкой считать рассуждения Пьера просто подгонкой: такой подход может привести к подлинному открытию, которое могло быть и не найдено при преждевременной формализации исследования. И вообще, как известно, в создании новых гипотез большую роль играют неформализуемые пока что факторы, например интуиция и здравый смысл. Однако известно также, что,

начиная с какого-то этапа, в разработке новой гипотезы становятся все более важными формализуемые факторы: точная ее формулировка, объективное сравнение с конкурирующими гипотезами и т. п.

Естественно, что и в прогнозе места землетрясения возможность перейти от качественных представлений к конкретному прогнозу зависела от таких формализуемых факторов, как точная формулировка исходных данных и постановки задачи, возможность перебирать и объективно сравнивать множество комбинаций исходных данных, свобода варьировать проверяемые гипотезы и формулировку каждой из них (выбор исходных данных, группирование объектов, выбор параметра и т. п.).

При недостаточной формализации в получении конкретного прогноза возникают большие трудности. Существует тенденция объяснять эти трудности недостатком исходных данных и, следовательно, переходить к сбору новых данных до того, как закончен анализ существующих. Эта тенденция рискованна: она может или приводить к неожиданным открытиям или (в настоящее время, видимо, чаще) запутывать изучаемый вопрос. Так, в постановку задачи о прогнозе места землетрясений она иногда вносила необоснованные осложнения: совместный анализ слишком большого круга данных и одновременный прогноз всех разрушительных землетрясений. В результате решение задачи искалось в излишне сложном виде, так что найденное нами решение выглядит парадоксально простым. Это обстоятельство, видимо, также отражает некоторые общие тенденции человеческого мышления, выраженные известными стихами

...Нельзя в конце не впасть, как в ересь,
В неслыханную простоту.
Но мы пощажены не будем,
Когда ее не утаим:
Она всего нужнее людям.
Но сложное понятней им.

(Б. Пастернак)

Указанные тенденции приводят к тому, что применение ЭВМ чаще всего ассоциируется с формализованной и стандартизованной обработкой массовых данных, тогда как во многих задачах (в том числе в рассмотренных здесь) гораздо важнее другое: что ЭВМ позволяют извлекать больше выводов из ограниченного набора данных. В частности, они позволяют систематизировать и проверять множество гипотез. Эти соображения общие для применения ЭВМ в естественных науках. Однако алгоритм Кора-3, по нашему мнению, обладает исключительной особенностью принимать исходные данные и выдавать результаты в форме, поразительно удобной для обдумывания и проверки; кроме того, выдаваемые результаты кажутся обычно очень естественными. Объяснить эту особенность Кора-3 — важная задача, к которой пока неясно как подойти.

Л и т е р а т у р а

1. И. М. Гельфанд, Ш. А. Губерман, М. Н. Извекова, В. И. Кейлис-Борок, Е. Я. Ранцман. Распознавание мест возможного возникновения сильных землетрясений I. Памир и Тянь-Шань.— Сб. «Вычислительная сейсмология», вып. 6. М., «Наука», 1973.
2. И. М. Гельфанд, Ш. А. Губерман, М. Л. Извекова, В. И. Кейлис-Борок, Е. Я. Ранцман. О критериях высокой сейсмичности.— ДАН СССР, Геофизика, 1972, т. 202, № 6.
3. И. М. Гельфанд, Ш. А. Губерман, М. П. Жидков, М. С. Калецкая, В. И. Кейлис-Борок, Е. Я. Ранцман. Опыт переноса высокой сейсмичности со Средней Азии на Анатолию и смежные регионы.— ДАН СССР, 1973, т. 210, № 2.
4. J. M. Gelfand, Sh. A. Guberman, M. L. Isvekova, V. I. Keilis-Borok, E. Ya. Ranzman. Criteria of high seismicity determined by pattern recognition. Tectonophysics, 1972.
5. И. М. Гельфанд, Ш. А. Губерман, М. П. Жидков, М. С. Калецкая, В. И. Кейлис-Борок, Е. Я. Ранцман, И. М. Ротвайн. Распознавание мест, где возможны сильные землетрясения. II. Четыре региона Малой Азии и Юго-Восточной Европы. Вычислительная сейсмология, вып. 7. М., «Наука», 1974.
6. И. М. Гельфанд, Ш. А. Губерман, М. П. Жидков, В. И. Кейлис-Борок, Е. Я. Ранцман, И. М. Ротвайн. Распознавание мест, где возможны сильные землетрясения. III. Случай, когда границы дизъюнктивных узлов неизвестны. Вычислительная сейсмология, вып. 7. М., «Наука», 1974.
7. С. В. Медведев (ред.) Сейсмическое районирование СССР. «Наука», 1964.
8. М. М. Бонгард. Проблема узнавания. М., «Наука», 1967.
9. М. М. Бонгард, Н. Н. Вайнцвайг, Ш. А. Губерман, М. Л. Извекова, М. С. Смирнов. Использование обучающейся программы для выявления нефтеносных пластов.— Геология и геофизика, 1966, № 6.
10. Л. Н. Толстой. Собр. соч., т. 6, стр. 81, 1951.

М. П. Полякова, М. Н. Вайнцвайг

ОБ ИСПОЛЬЗОВАНИИ МЕТОДА «ГОЛОСОВАНИЯ» ПРИЗНАКОВ В АЛГОРИТМАХ РАСПОЗНАВАНИЯ

При разработке алгоритмов обучения распознаванию, основанных на отборе признаков, соответствующих каждому из данных классов объектов и характеризующих объекты этого класса [1, 2, 3], возникает задача использования отобранных признаков для классификации (узнавания). Применение в этих целях чисто логических методов оказывается, как правило, малоэффективным. Так, например, при построении для каждого из классов дизъюнкции его признаков становится невозможной классификация объектов, характеризующихся одновременно признаками разных классов. Это приводит к тому, что вероятность ошибки при таком правиле будет относительно большой. Одним из возможных методов существенного снижения вероятности ошибки при узнавании служит так называемое «голосование» признаков.