АНАЛИЗ РАЗНОТИПНЫХ ДАННЫХ
(Вычислительные системы)
1983 год Выпуск 99

УДК 519.95:681.3.06

ПРИМЕНЕНИЕ ZET-МЕТОДА В ЭКСПЕРТНЫХ СИСТЕМАХ

В.Н.Ёлкина, Н.Г.Загоруйко

Под "экспертной" мы понимаем машинную систему, способную накапливать разнородные данные о какой-либо прикладной области, обнаруживать на этих данных эмпирические закономерности, углублять и уточнять их по мере поступления новых данных и в итоге выходить на уровень высококвалифицированного эксперта в изучаемой прикладной области. Сформированная (обученная) экспертная система используется для получения справок, выработки прогнозов и рекомендаций в процессе принятия решений.

В процессе создания и развития экспертных систем типичной является ситуация, при которой нужно сопоставлять имеющиеся знания с вновь поступающими фактами. Знания могут быть представлены в виде теории (формализованной гипотезы h_0) предметной области или в виде набора таблиц (протоколов pr) типа "вход (X)-выход (У)", описывающих поведение изучаемых объектов. Могут встретиться два различных случая: 1) новые факты (pr_0) не противоречат старым представлениям, тогда теория h_0 получает дополнительное подтверждение, ее значение на новом протоколе истинно, что можно пометить такой записью: $h_0(pr_0)=1$; 2) новые факты (pr_0') противоречат старым взглядам, при этом $h_0(pr_0')=0$. В том и другом случае новые факты можно (и нужно) использовать для углубления, уточнения имеющейся теории h_0. В работах [1,2] рассмотрен механизм усиления теории для первого случая. Рассмотрены свойства алгоритма F усиления эмпирических гипотез, который по непротиворечивой паре $\langle h_0, pr_0 \rangle$ вырабатывает новую гипотезу h_1:

$$F\langle h_0, pr_0 \rangle \to h_1 .$$

Гипотеза h_1 сильнее h_0 по ее способности делать более конкретные предсказания новых (будущих) фактов. Если гипотезы (модели

мира) h есть правила предсказания "выхода" У по заданному значению "входа" Х, то алгоритм F есть правило изменения (усиления) таких правил.

Во втором случае нужно использовать некоторый другой алгоритм ($\mathcal{F}$), который должен моделировать процесс $\mathcal{F}\langle h_0, pr_0' \rangle \to h_1^*$ критического пересмотра старых гипотез h_i и замены их новыми гипотезами h_1^*, которые не являются логическим продолжением (усилением) старых, а могут отличаться от них коренным образом.

В данной работе мы не будем обсуждать сами эти правила $\mathcal{F}$ перестройки гипотез h. Содержательная суть $\mathcal{F}$-процесса хорошо описана в методологической литературе [3,4], а попытки формального изучения свойств $\mathcal{F}$-алгоритмов, аналогичного изучению свойств F -алгоритмов, нам не известны.

Здесь мы опишем один из возможных способов лишь установления того, согласуется ли новый факт со старыми представлениями или противоречит им.

На ранней стадии создания экспертных систем знания могут быть представлены набором фактов, еще не описанных соответствующей теорией h. Чтобы в этих условиях обнаружить несоответствие нового факта закономерностям, скрытым в старых фактах, можно использовать алгоритм ZET-M, являющийся современной модификацией алгоритма ZET [5-7].

Алгоритм ZET-M разработан для заполнения пробелов в эмпирических таблицах, содержащих данные преимущественно в сильных шкалах. Суть его состоит в следующем.

Пусть ранее накопленные факты сведены в таблицу "объект-свойство" размером m x n, в которой строки имеют номера 1,2,... ..., i ,..., m, а столбцы - 1,2,..., j ,..., n. Каждая строка таблицы соответствует объекту, а каждый столбец - признаку, характеризующему объекты. Клетка a_{ij} таблицы указывает значение j-й характеристики i-го объекта.

Теперь представим себе, что значение некоторого элемента a_{ij} таблицы неизвестно. Можно ли предсказать это значение? Оказывается, можно, если использовать имеющуюся в таблице избыточность: в реальных таблицах многие столбцы связаны друг с другом определенной зависимостью, есть в таблице и строки (объекты), похожие друг на друга по своим характеристикам. Эти связи и похожести можно использовать для заполнения пробелов в таблице с помощью программы, реализующей алгоритм ZET-M.

В алгоритме ZET-M используется линейная зависимость между строками и(или) между столбцами таблицы. Элемент a_{ij} предсказывается по группе столбцов, похожих на j-й столбец, и (или) по группе строк, похожих на i-ю строку. "Подсказки" ($\hat{a}_{ij}$) от этих столбцов и строк вычисляются с использованием уравнений линейной регрессии и усредняются с весом, пропорциональным степени "компетентности" этих столбцов и строк. Компетентность вычисляется как функция сходства и взаимной заполненности столбцов (строк).

Программа ZET-M входит в состав пакета прикладных программ ОТЭКС [8], разработанного Институтом математики СО АН СССР и Новосибирским университетом. Описание алгоритма и программы ZET-M приведено в приложении.

Программа ZET-M может работать в режиме "редактирования", при котором машина по очереди закрывает все известные элементы таблицы и предсказывает их. Если обнаруживается большое различие между тем, что было в таблице и что предсказано, то машина сообщает об этом. Чаще всего оказывается, что в таблице содержалась ошибка, но иногда несовпадение a_{ij} и $\hat{a}_{ij}$ выявляет аномалии, т.е. говорит о том, что элемент a_{ij} "выпадает" из закономерности, характерной для данной таблицы. Если описанным способом контролировать новые факты (строки), поступающие в экспертную систему, то по большим ошибкам редактирования можно обнаружить несоответствие этих фактов закономерностям, скрытым в ранее накопленных данных.

Этот же прием можно использовать и для сжатия табличных данных, уже хранящихся в памяти экспертных систем. Нужно поочередно устранять из таблицы строки и редактировать остальную часть таблицы. Если при удалении строки a_i качество не ухудшается, то это значит, что закономерности хорошо проявляются и без строки a_i и ее можно из таблицы убрать. Аналогично можно сокращать и число столбцов.

Если каждая строка a_i, $i = \overline{1,m}$, таблицы отражает состояние некоторого динамического процесса в моменты времени $t_1, t_2, \ldots, t_i, \ldots, t_m$ и предполагается, что закономерности этого процесса меняются со временем, то ZET-метод можно использовать для устранения устаревших данных. Для этой цели следует редактировать самую "молодую" строку a_m и наблюдать, какие строки при этом выбираются в число наиболее компетентных и каково качество редактирования. Если ошибка редактирования меньше заданного порога, то строки, которые "старше самой старой" из компетентных для a_m,

Таблица I

Результаты редактирования данных

m	n					Ошибка редактирования	№ ближайших строк
	I	2	3	4	5	6	7
I	IO/9,5	2/2,4	9/9,6	4/4	I/2,I	5,2%	2,3,4,6
2	IO/9,6	I/2,2	IO/IO,2	3/3,I	3/2,5	4,4%	I,3,4,6
3	9/IO,4	3/I,7	IO/IO,I	I/3,2	2/2	9,8%	2,I,4,6
4	IO/8,4	4/4,I	8/8,6	5/4,4	4/4	5,8%	I,5,2,6
5	8/7,2	5/4,7	8/6,7	8/5,8	5/4,6	IO,0%	4,6,7,3
6	6/8,I	3/3,4	9/7,8	6/5,2	3/3,2	9,4%	4,5,I,3
7	5/4,I	6/6,3	5/4,I	IO/9,2	6/6,3	6,4%	8,5,IO,6
8	I/2,4	7/7,8	2/2,6	IO/IO,4	7/7,6	7,8%	IO,7,9,5
9	2/2,5	9/8,8	I/2,6	8/I2,3	IO/8,5	I6,2%	IO,8,7,5
IO	3/3	IO/8,3	2/2,4	IO/IO,5	9/8,8	5,6%	9,8, 7,5
Ошибка редактирования	II	9,6%	6,7%	7,5%	II,9%	4,8%	
№ ближайших столбцов	I2	4,3	5,3	4,2	I,3	2,3	

можно из таблицы изъять, как "устаревшие". По-видимому, закономерности, скрытые в начале таблицы, уже существенно отличаются от за-

кономерностей, которые достаточно хорошо проявляются в более поздних данных.

Табл.I иллюстрирует сказанное выше. Здесь количество объектов (или моментов наблюдения) $m = 10$, число характеристик $n = 5$. Исходные значения элементов таблицы приведены в числителе, а предсказанные алгоритмом ZET - в знаменателе. В столбце 6 показана ошибка редактирования элементов строки a_i по четырем самым компетентным строчкам, номера которых в порядке компетентности указаны в столбце 7. Аналогично, в строке II указаны средние ошибки редактирования столбцов по 2-м компетентным столбцам, номера которых приведены в строке I2.

Как видно из этой таблицы, для редактирования строк 7,8,9 и I0 использовались строки с номерами ≥ 5. Можно считать, что к моменту t_{10} уже достаточно хорошо сформировались закономерности, на которые не влияют события, наблюдавшиеся до момента t_5.

Судя по величине ошибок редактирования, общему содержанию табл.I меньше других соответствуют строки 9 и 5. Если это возможно, то их нужно перепроверить, причем можно надеяться, что ошибка может быть обнаружена в первую очередь в 4-м столбце.

В ряде случаев программа ZET-M может быть использована для прогноза новых данных, для продолжения временного ряда. Рассмотрим некоторые из этих случаев.

Если в качестве исходных данных есть сведения о значениях n параметров для одного объекта за m временных циклов, например, значения показателей для одного и того же хозяйства за последовательные m лет, то может быть выполнен прогноз значений этих показателей в $(m+1)$-м цикле. Для этого входная информация должна быть организована следующим образом. Обозначим символом P_i n-мерный вектор значений параметров для одного объекта в i-м цикле (за i-й год). На основе ряда $P_1, P_2, \ldots, P_i, \ldots, P_m$ данных за m лет образуем матрицу "объект-свойство" размерности $M \times N$ ($M = m - K + 2$, $N = K \cdot n$, $K \geq 2$) вида:

$$\begin{array}{cccc} P_1 & P_2 & \ldots & P_{K-1}\ P_K \\ P_2 & P_3 & \ldots & P_K\ \ P_{K+1} \\ \ldots & \ldots & & \ldots \\ P_{m-K+1} & & \ldots & P_{m-1}\ P_m \\ P_{m-K+2} & & \ldots & P_m\ \ P_{m+1} \end{array}$$

Таблица 2

Матрица исходных данных

	1	2	3
1	121,0	136,0	165,0
2	144,0	162,0	208,0
3	161,0	180,0	216,0
4	142,0	146,0	189,0
5	128,0	140,0	187,0
6	139,0	148,0	192,0
7	142,0	163,0	205,0
8	133,0	142,0	191,0
9	139,0	152,0	199,0
10	137,0	150,0	202,0
11	133,0	147,0	193,0
12	133,0	143,0	193,0
13	133,0	145,0	200,0
14	140,0	150,0	207,0
15	128,0	141,0	193,0
16	140,0	147,0	206,0
17	121,0	128,0	175,0
18	141,7	147,0	188,9

Другими словами,строками такой матрицы могут быть данные за 2,3 и более последовательных временных цикла (К лет) со сдвигом начала каждой строки относительно предыдущей на один цикл. В этом случае программа ZET-M будет строить прогноз на основе сходства последовательностей К циклов. На месте значений вектора P_{m+1} в этой матрице должны быть поставлены символы пробелов. Программа ZET-M в режиме заполнения пробелов вычислит значения параметров для (m+1)-го цикла временного ряда.

Например, пусть есть исходная матрица I8x3 данных о результатах работы одного хозяйства по трем показателям за последовательные I8 лет (табл.2). Программой ZET-M можно спрогнозировать значения аналогичных показателей в I9-м году, организовав имеющиеся данные так, как показано в табл.3 и 4.

Таблица 3

Организация данных при К = 2

	№ показателя					
	I	2	3	I	2	3
№ строки исходной матрицы,которая должна быть вписана в таблицу	I 2 3 : I7 I8			2 3 4 : I8		

Таблица 4

Организация данных при К = 4

	Номер показателя											
	I	2	3	I	2	3	I	2	3	I	2	3
№ строки исходной матрицы, которая должна быть вписана в таблицу	I 2 3 : I5 I6			2 3 4 : I6 I7			3 4 5 : I7 I8			4 5 6 : I8		

Если известны данные для t объектов по n показателям за m лет, а за (m+1)-й год для тех же объектов значения показателей неизвестны или известна только часть из них, то для прогноза неизвестных значений исходные данные могут быть организованы в таблицу "объект-свойство", аналогичную предыдущему варианту, только в ней символу P_i будет соответствовать матрица размерности t x n

данных за i-й год для t объектов, а размерность будет M x N, где M = (m-K+2)t, N = K·n . Строка такой матрицы - K-летние данные для одного объекта. Так как одновременно рассматриваются t объектов, то в сформированной матрице будет t строк с данными для этих t объектов за один и тот же период времени, t строк за следующий период (K-летие) и т.д. Последние t строк матрицы будут содержать по n пробелов на местах, отведенных для данных прогнозируемого цикла (года).

После выполнения прогноза полезно для контроля и корректировки выполнить по той же программе ZET-M редактирование полученных элементов, переформировав входные данные (например, задавая K = 1,2,3).

Пусть есть однопараметрический временной ряд с длиной цикла n - $P_1, P_2, \ldots, P_i, \ldots, P_m$, где P_i - вектор n значений одного параметра в i цикле (например, среднемесячные удои за каждый из 12 месяцев i-го года, данные о ежемесячном или поквартальном выполнении плана и т.п.), m - количество циклов. Для этого варианта данных матрица может быть организована так же, как и в предыдущих случаях, и заполнение пробелов, т.е. прогнозирование неизвестных значений, может выполняться по программе ZET-M .

Для этого же варианта исходных данных вычисление новых членов ряда может быть выполнено также по программе ZET-MC (ZET-M циклический), разработанной на основе программы ZET-M и существенно использующей особенности цикличности однопараметрического ряда. Для программы ZET-MC исходные данные должны быть представлены таблицей "объект-свойство" вида

$$\begin{matrix} P_1 & P_2 & \ldots & P_K \\ \ldots & \ldots & & \\ P_{m-K+1} & \ldots & P_m & \end{matrix} \quad \text{, где} \quad K \geq 1.$$

Программа ZET-MC в качестве одной из основных подпрограмм использует программу ZET-M, поэтому все сказанное о применимости программы ZET-M справедливо и для ZET-MC . В программе ZET-MC предусмотрена возможность по исходной матрице размерности m x n организовывать матрицу с заданным количеством K циклов в строке. Программа ZET-MC дает возможность прогнозировать значения заданного количества элементов временного ряда и выполнять "редактирование" известных элементов таблицы, т.е. сравнивать вычисленные и имеющиеся в таблице значения и обнаруживать те из них, где полу -

чено расхождение больше заданного порога. В отличие от программы ZET-M , в программе ZET-MC редактирование и прогнозирование данного элемента выполняется с использованием только предыдущей информации.

Поскольку в программе ZET-M для формирования "предсказывающих" подматриц вначале отбираются наиболее "похожие" строки, а затем на этих строках - похожие столбцы (признаки), то формирование входных матриц - количество циклов в строке - определяет прогноз значений по сходным годам , двухлетиям ,..., K -летиям . Для каждой конкретной задачи имеет смысл попробовать выполнить предсказание (редактирование) известных элементов таблицы при разной разумной длине строки и отобрать для прогноза неизвестных элементов этой таблицы лучший вариант. В каждой задаче может быть свой оптимальный цикл, поэтому универсального значения K , пригодного одновременно для всех задач, естественно,нет и быть не может.

Так как результаты прогноза могут до некоторой степени зависеть от количества данных, по которым делается этот прогноз, т.е. от размера "предсказывающих" подматриц,то неизбежно возникает вопрос о том, как в каждом конкретном случае определять число строк и столбцов, формирующих эти подматрицы. Этот вопрос связан как с проблемой сокращения времени вычислений,так и с проблемой получения наиболее качественного прогноза. Мы постоянно сталкиваемся с тем, что чрезмерное увеличение количества данных не приводит к увеличению точности результата. Если закономерность четко проявлена на многих элементах таблицы, то и использование части информации для предсказания дает хороший результат. Если же в таблице представлено только малое количество объектов (строк), подчиняющихся используемой закономерности, то опора прогноза на большой объем информации только ухудшит результат. Априорных критериев достаточности информации для каждого конкретного случая нет.Поэтому в ал-

Таблица 5

Результаты редактирования по программе ZET-M

№ столбца	% отклонения при подматрицах			
	3x3	5x5	7x7	10x10
1	2,88	3,79	3,67	3,99
2	20,28	21,16	23,76	24,17
3	4,91	4,37	4,05	3,98
4	3,26	5,75	5,46	4,5
5	14,9	10,49	9,89	13,27
6	4,26	3,50	3,87	4,36

Таблица 6

Результаты многошаговой процедуры заполнения пробелов для таблицы I86x6 элементов

Допустимая ошибка	Размер подматриц	Заполнилось пробелов	Осталось незаполненным
5%	3x3	27	195
		12	183
		6	177
		4	173
		2	171
		0	171
	5x5	36	135
		14	121
		12	109
		6	103
		8	95
		9	86
		2	84
		0	84
	7x7	1	83
		0	83
	9x9	0	83
I0%	5x5	13	70
		9	61
		7	54
		0	54
	7x7	7	47
		13	34
		2	32
		3	29
		0	29
	9x9	2	27
		6	21
		0	21
	5x5	4	17
		1	16
		0	16
	7x7	0	16
	9x9	0	16

горитме ZET качество прогноза косвенно оценивается по качеству предсказания известных элементов таблицы, связанных с интересующим нас пробелом (стоящих с ним в одной строке или одном столбце). Если известные элементы таблицы (строки, столбца) предсказываются достаточно уверенно или, наоборот, плохо, то из этого можно сделать и заключение о том, стоит или нет по этим имеющимся данным прогнозировать неизвестные элементы. Для того чтобы решить вопрос об оптимальном размере предсказывающих подматриц, следует провести предварительное редактирование известных элементов таблицы с разными размерами подматриц и использовать для прогноза тот вариант, при котором была достигнута лучшая точность. Заметим, что использовать размерности выше I0xI0 заведомо нецелесообразно. Для иллюстрации приведем достаточно типичный результат, полученный при полном редактировании реальной матрицы геофизических исходных данных, содержащей I86 строк, описанных 6 параметрами. В исходных данных первоначально содержалось 222 пробела. Было выполнено предварительное редактирование всех известных элементов при размерах прогнозирующих подматриц 3x3, 5x5, 7x7, I0xI0. Результаты приведены в табл.5.

Очень часто первичный материал имеет в таблицах данных пробелы, возникшие по разным причинам - данные не сняты, утеряны, сомнительны. К сожалению, не всегда в таблицах бывает достаточно информации, чтобы сразу по ним с хорошей (приемлемой) точностью предсказать все неизвестные значения. Но можно организовать многошаговую процедуру, до некоторой степени моделирующую процесс принятия решения человеком. Задавая допустимую (ожидаемую) точность, можно спрогнозировать в первую очередь "наиболее надежные" элементы, затем, поставив их в таблицу и считая уже известными, предсказать с той же ожидаемой точностью и при том же размере подматриц новую порцию данных и т.д., пока не прекратятся заполнения при этих параметрах. Тогда следует постепенно наращивать размер предсказывающих матриц, сохраняя ту же требуемую точность предсказания. Если не удалось с первоначальным качеством заполнить пробелы на подматрицах до размера IOxIO, то это означает, что для строки (столбца), содержащих рассматриваемый пробел, нет достаточной информации в таблице, не удается найти хорошую референтную группу.

Если условия задачи позволяют, то для заполнения оставшихся пробелов можно повторить цикл заполнения при менее жестких ограничениях на ожидаемую точность предсказания. Достаточно наглядно описанную процедуру иллюстрирует процесс заполнения пробелов на той же матрице данных, для которой в табл.5 приводились результаты редактирования. Если допустимая ошибка прогноза должна находиться в пределах IO%, то из табл.5 видно, что большую часть пробелов можно с этой точностью заполнить. Процесс заполнения можно начать с минимальных подматриц 3x3. Результаты многошаговой процедуры заполнения приведены в табл.6. Напомним, что матрица, состоящая из I86 строк и 6 столбцов, содержала 222 пробела.

Косвенная оценка точности прогноза последних I6 пробелов превышает IO%, и поэтому эти пробелы остались незаполненными.

Литература

I. САМОХВАЛОВ К.Ф. О теории эмпирических предсказаний. - В кн.: Вычислительные системы. Вып.55. Новосибирск, 1973, с.3-35.

2. ЗАГОРУЙКО Н.Г. Эмпирическое предсказание. - Новосибирск: Наука, Сиб.отд., 1979.

3. LAKATOS J. History of science and its reconstructions. - In: Boston Studies in the philosophy of Science. Vol.111. Dortrecth, 1972, p.91-136.

4. POPPER K. The Logic of Scientific Discovery. - London, 1959.- 282 p.

5. ЗАГОРУЙКО Н.Г., ЁЛКИНА В.Н., ТИМЕРКАЕВ В.С. Алгоритм ZET-75 заполнения пробелов в эмпирических таблицах и его применение. - В кн.: Машинные методы обнаружения закономерностей. Новосибирск,Наука,1976, с.57-63.

6. YOLKINA V.N., ZAGORUIKO N.G. Some classification algorithms developed at Novosibirsk.- R.A.I.R.O.Informatique/Computer Science,1978,vol.12, N 1, Paris, Trance.

7. ZAGORUIKO N.G., YOLKINA V.N. Inference and Data Tables with Missing Values.- Handbook of Statistics.Vol.2, 1982. North-Holland Publishing Company, Amsterdam-New York-Oxford.

8. ОТЭКС. Пакет прикладных программ для обработки таблиц экспериментальных данных. Новосибирск, НГУ, 1977.

Поступила в ред.-изд. отд.
17 ноября 1983 года

Приложение

ОПИСАНИЕ АЛГОРИТМА И ПРОГРАММЫ ZET-M

Загоруйко Н.Г., Ёлкина В.Н., Киприянова Т.П.

Назначение.

Программа предназначена для заполнения пропущенных элементов в эмпирических таблицах, содержащих данные, замеренные в сильных шкалах, и для редактирования (проверки) всей таблицы или ее части.

Основные этапы работы алгоритма ZET-M.

I. Производится нормировка столбцов таблицы исходных данных по дисперсиям.

2. Выбирается пробел a_{ij}, находящийся на пересечении i-й строки и j-го столбца.

3. Вычисляется евклидово расстояние от i-й строки до всех строк исходной матрицы.

4. Выбирается заданное число N2 строк, ближайших к i-й строке.

5. Формируется подматрица из N2 строк и исходного количества N столбцов.

6. Столбцы полученной подматрицы нормируются к интервалу [0, I] (без учета строки, содержащей рассматриваемый пробел).

7. Вычисляется евклидово расстояние от j-го столбца до всех столбцов подматрицы.

8. Выбирается заданное число N1 столбцов, ближайших к j-му.

9. Формируется подматрица из N1 столбцов и N2 строк.

I0. Столбцы полученной подматрицы нормируются к интервалу [0, I] с учетом элементов строки, содержащей рассматриваемый пробел.

II. Из уравнений линейной регрессии для элемента a_{ij} вычисляются "подсказки" a^{x}_{ij} от ближайших строк и столбцов.

I2. Вычисляется коэффициент α, определяющий степень учета взаимного сходства столбцов (строк) подматрицы при вычислении значения прогнозируемого элемента a_{ij}. Для этого при различных значениях коэффициента α (из заданного диапазона) отыскивается минимум функционала $\delta(\alpha)$ ошибки предсказания известных элементов подматрицы, находящихся в i-й строке или (и) в j-м столбце. Если найденное δ_{min} не будет превосходить заданное ограничение на оши-

даемую точность предсказания, то при α , соответствующем этому значению δ_{min}, предсказывается пропущенный элемент $\hat{a}_{ij}$.

I3. Пункты 2-I2 повторяются для каждого пропущенного элемента.

I4. Все значения, вычисленные в режимах заполнения I,2,3 или 4 (см. "Режимы работы программы", с. 86), ставятся в таблицу и повторяются пп. I-I3. Количество итераций задается параметром IT.

Обращение к программе:

```
CALL ZETM (N,M,NG,NK,IT,GAP,BEG,FIN,STEP,REST,N1,N2,KR1,KT1,T,
                IW,MW,C1,RP) .
```

Описание параметров.

N - количество признаков (столбцов таблицы);

M - количество объектов (строк таблицы);

NG - количество пропусков, либо количество редактируемых элементов для режимов 6,8,9; если заранее количество пропусков неизвестно, то NG полагают равным максимальному предполагаемому числу;

NK - константа, задающая номер столбца или строки для режимов редактирования 7,8;

IT - количество итераций для режимов заполнения I,2,3,4;

GAP - обозначение пропуска в таблице, кодируется положительным числом, заведомо превышающим числа, имеющиеся в таблице;

BEG - начальное значение коэффициента α;

FIN - конечное значение коэффициента α;

STEP - шаг изменения коэффициента α;

REST - ограничение на ожидаемый (допустимый) процент ошибок предсказания;

N1 - требуемое количество столбцов для формирования подматрицы;

N2 - требуемое количество строк для формирования подматрицы;

KR1 - константа, указывающая номер режима работы программы (см. "Режимы работы программы");

KT1 - константа режима печати: KT1 = 0 - программа работает без выдачи результатов на печать; KT1 = 1 - печатается информация о заполняемых или редактируемых элементах: (в строчку) - номер по порядку, номер столбца, номер строки, коэффициент α , средний процент ожидаемой ошибки, чему отдано предпочтение при предсказании данного элемента - строке (0) или столбцу (I), предсказанное

значение элемента. В режиме редактирования дополнительно печатается исходное значение и % отклонения вычисленного значения от исходного. В режиме прогнозирования на печать выдается только информация о предсказанных элементах, в режиме редактирования - только об элементах, для которых получено расхождение, превышающее заданный параметром REST порог; KT1 = 2 - дополнительно к информации, выдаваемой при KT1 = 1, на печать выводится матрица исходных данных и матрица результатов; KT1 = 3 - во всех режимах, кроме 2,4,5 и 7, дополнительно к печати при KT1 = 1 выдаются номера строк и столбцов, отобранных для формирования подматрицы; KT1 = 4 - одновременно выдается на печать все, что предусмотрено при KT1 = 1,=2, =3.

T - массив исходных данных размерности M * N,

IW - входной вектор длины N, задающий номера столбцов, которые можно использовать при отборе ближайших строк: если IW(I) = = 1, то I-й столбец можно использовать, если IW(I) = 0, то нельзя,

MW - целый входной массив размерности 2* NG, в котором содержатся номера столбцов и строк, на пересечении которых находятся прогнозируемые элементы (задается только для режимов 3,4,6,9),

C1 - рабочий массив длины не менее, чем

$$9 \times NG + 10 \times N3 + N \times M + N \times N2 + 2 \times M5 + N1 + N2,$$

где M5 = max(M,N), N3 = max(N,N2),

RP - выходной вектор: для режимов 5,7 длины M, для режимов 6,8 длины NG ; в нем хранятся отредактированные элементы.

KR1	Режимы работы программы
	Режимы прогнозирования
I	На каждой итерации все элементы, отмеченные символом GAP, прогнозируются независимо друг от друга; при последующей итерации учитываются значения, полученные на предыдущей.
2	Предсказываются все элементы, отмеченные символом GAP, заполнение каждого пробела ведется с учетом всех предсказанных к этому моменту элементов.
3	Прогнозируются только те элементы, для которых в массиве MW указаны координаты; прогнозирование выполняется аналогично режиму I.
4	Прогнозируются только те элементы, для которых в массиве MW указаны координаты; прогнозирование выполняется аналогично режиму 2.

	Режимы редактирования (без исправлений)
5	Редактирование всех элементов таблицы.
6	Редактирование заданных массивом MW элементов.
7	Редактирование элементов столбца с номером NK.
8	Редактирование элементов строки с номером NK .
9	Редактирование либо заполнение заданных элементов таблицы с выдачей на печать дополнительной информации - "вклада" каждой строки (столбца) в предсказанное значение (в относительных единицах от 0 до 1000).

Режим подбора значений коэффициента α можно задавать следующим образом:

BEG > 0 ,FIN > 0 - подбор производится по строкам и столбцам одновременно,

BEG < 0 ,FIN > 0 - подбор производится только по столбцам,

BEG > 0 ,FIN < 0 - подбор производится только по строкам.

Ограничения.

Для таблиц, содержащих шкалы наименований и порядка, программа неприменима. Это связано с тем, что используемая мера похожести (евклидово расстояние) применима только для сильных шкал. Для таблиц со шкалами наименований и порядка необходимо программиро - вать другие меры похожести. Требуемый объем оперативной памяти:

$$V \geq 2*N*M + 9*NG + 10*N3 + N*N2 + 2*M5 + N1 + N2 + M^{*)} + NG^{**)} + 2*NG^{***)}.$$

Точность метода.

Точность метода зависит от числа пропусков и от того, выполняются ли гипотезы "линейной избыточности" и "компетентности".

Критерий оценки результатов вычислений.

Критерием служит точность предсказания известных элементов таблицы - "ожидаемая ошибка предсказания".

---

*) Для режимов 5,7.

**) Для режимов 6,8.

***) Для режимов 3,4,6,8,9.