

УДК 911:001.18

И.В. Малыгин¹

МЕТОДИКА ПРОГНОЗА ОБРАЗОВАНИЯ ЛЕДОВЫХ ЗАТОРОВ НА РЕКАХ НА ОСНОВЕ ТЕОРИИ РАСПОЗНАВАНИЯ ОБРАЗОВ

Предложен способ решения задачи прогноза наличия ледовых заторов с использованием методов теории распознавания образов. Описаны возможная структура и содержание базы исходных данных. Приводятся схема прогнозного алгоритма и результат его работы на примере участка реки Северная Двина от г. Котлас до г. Великий Устюг.

Ключевые слова: прогноз, прогнозирование заторообразования, прогнозный алгоритм, задача классификации, комбинаторно-логический подход.

Введение. Прогноз ледовой обстановки актуален для ряда районов европейской территории России и создается, как правило, по данным гидрометеорологических наблюдений. При этом некоторые районы обеспечены наблюдениями за длительный период, например, для р. Нева есть структурированные данные больше чем за 200 лет, однако есть районы, наблюдения в которых охватывают период 15—30 лет. Это имеет существенное значение при выборе методики прогноза, причем как в методическом смысле, так и при оценке практической достоверности. В гидрологии разработаны разные подходы к прогнозу ледового режима рек [2]. Среди методов исследования наиболее часто применяются моделирование и вероятностно-статистический подход, которые хорошо “работают” для статистически значимых рядов наблюдений. Для оперативного слежения за ледовой обстановкой хорошо подходят методы космического мониторинга с использованием данных дистанционного зондирования [6]. В статье предложен способ, основанный на методах теории распознавания образов, который способствует выявлению и прогнозированию ледового заторообразования в случае, когда многолетние и одинаково структурированные полные данные труднодоступны либо отсутствуют. При этом под прогнозом понимается отнесение прогнозируемого года к одному из выделенных сценариев развития ледовой обстановки. В статье представлена схема прогнозного алгоритма и результаты его применения для прогноза возникновения ледовых заторов на примере участка р. Северная Двина между городами Котлас и Великий Устюг.

Постановка проблемы. Затопы льда на реках относятся к опасным гидрологическим явлениям. Они приводят к наводнениям, нарушению условий эксплуатации водозаборов, ледовых переправ, железных и автомобильных дорог, водного транспорта, повреждению гидротехнических сооружений и др. [3].

Прогнозирование заторообразования — сложная и актуальная задача, имеющая большое значение для разработки методов и средств защиты хозяйственной деятельности от неблагоприятных гидрологических явлений.

В связи с изменениями климата необходимо изучать ледовый режим рек на основе современных гидрометеорологических данных с применением современных математических и компьютерных алгоритмов обработки данных и прогнозирования [3].

Общую постановку задачи классификации сформулируем следующим образом. Пусть M — множество наблюдений, оно может быть разбито на непересекающиеся подмножества — классы K_1, \dots, K_l . Целиком само разбиение неизвестно, однако в каждом классе есть подмножество элементов, о которых полностью известны их принадлежность и описание (характеристики). Совокупность таких подмножеств всех классов образует *обучающую выборку*: $T_1, \dots, T_l, T_i \subset K_i, i = 1, \dots, l$. Элементы обучающей выборки называются *эталомами*. Для элементов множества M не из обучающей выборки принадлежность к классу неизвестна. Для распознавания (классификации) предъявляется элемент множества M , не входящий в обучающую выборку. Требуется классифицировать этот элемент, т.е. отнести его к одному из существующих классов, представленных обучающей выборкой.

Материалы и методы исследований. В теории распознавания образов для классификации объектов и явлений применяются вероятностный и комбинаторно-логический подходы [4, 5]. Первый подход предназначен для исследования таких явлений, для которых имеются статистически значимые выборки, второй подход может быть использован при отсутствии таковых.

Для решения задачи прогнозирования ледовых заторов на участке р. Северная Двина выбран комбинаторно-логический подход. В этой задаче элемент множества M представлен известным набором гидрологических и метеорологических данных по отдельным

¹ Московский государственный университет имени М.В. Ломоносова, географический факультет, кафедра картографии и геоинформатики, аспирант; e-mail: malygin.iv@gmail.com

гидрологическим постам бассейна Северной Двины за каждый год, входящий в обучающую выборку. Каждый элемент множества M характеризуется набором признаков, каждый признак принимает либо числовое значение, либо набор числовых значений. Классы — результат экспертной классификации объектов наблюдения (лет) по критерию наличия или отсутствия заторов; обучающую выборку можно рассматривать как классифицированные данные за прошедшие годы [1, 7]. Распознаваемый элемент — прогнозируемый год — будет классифицирован построенным алгоритмом, т.е. отнесен к году с затором или без него, что и является прогнозом. Для решения поставленной задачи предложена следующая схема прогнозного алгоритма.

Исходные данные для прогнозирования. Для исследования в качестве признаков заторообразования могут быть использованы гидрологические и метеорологические показатели, которые приведены в табл. 1.

В районе наблюдения использованы данные со следующих гидрологических постов: д. Каликино, г. Великий Устюг, д. Медведки, г. Котлас, д. Абрамково, пос. Подосиновец. Схема расположения этих постов представлена на рисунке. Выбор гидрологических постов обусловлен тем, что места слияния рек обладают довольно большой повторяемостью заторообразования.

Исходные данные образуют базу данных, которую упрощенно можно представить в виде набора таблиц для каждого года за период наблюдений 1991—2010 гг.

В каждой таблице содержатся значения всех признаков по каждому посту за один год. Допускается отсутствие данных в отдельных ячейках таких таблиц, пример представлен в табл. 2.

Признаки разделены на две группы — гидрологические и метеорологические. Числовые значения признаков первой группы непосредственно зависят от местоположения гидрологического поста. Для второй группы признаков использованы фактические значения по данным метеостанции г. Великий Устюг.

Выделяются два сценария ледохода:

1) наличие заторов с разными мощностью и продолжительностью на участке г. Великий Устюг — г. Котлас;

2) отсутствие заторов на участке г. Великий Устюг — г. Котлас (в этот класс попадают и ситуации, когда затор произошел выше или ниже по течению, чем исследуемый участок). Указанные сценарии ледохода определяют классы K_1 , K_2 за период наблюдений (табл. 3).

В случае большей обучающей выборки возможно разделение на большее число классов, соответствующих более подробной классификации исследуемого явления. Например, становится возможным классифицировать заторы по мощности: средний, сильный, катастрофический и т.д.

Процедура сравнения однородных признаков за разные годы. Для работы прогнозного алгоритма необходимо уметь “сравнивать” числовые значения однородных признаков за разные годы. Если различие в числовых значениях находится в определенном ин-

Таблица 1

Список гидрологических и метеорологических показателей

Номер признака	Название признака	Характеристика признака	Единицы измерения
1	Предлежавший уровень воды	Гидрологический признак	см
2	Продолжительность осеннего ледохода	Гидрологический признак	сут
3	Наличие заборов	Гидрологический признак	есть (1) — нет (0)
4	Особенности температурного режима в период замерзания	Метеорологический признак, переход температуры воздуха через 0 °С	число суток с 1 сентября
5	Сумма отрицательных значений температуры воздуха за холодный период	Метеорологический признак	°С
6	Сумма положительных значений температуры воздуха за холодный период	Метеорологический признак	°С
7	Число дней с положительными температурами воздуха за холодный период	Метеорологический признак	сут
8	Суммарное количество твердых осадков	Метеорологический признак	мм
9	Особенности температурного режима в период вскрытия	Метеорологический признак, переход температуры воздуха через 0 °С	число суток с 1 февраля
10	Толщина льда перед вскрытием	Гидрологический признак	см
11	Интенсивность увеличения уровней и расходов воды в период подвижек льда	Гидрологический признак	см/сут

Таблица 2

Пример исходных данных прогнозирования за 1991 г.

1991 г., пост	Номер признака										
	1	2	3	4	5	6	7	8	9	10	11
	м	сут	да/нет	сут	°С	°С	сут	мм	сут	см	см/сут
Д. Каликино	337	13	0	59	-1449	3,9	6	192,44	34	41	76
Г. Великий Устюг	121	10	0								
Д. Медведки	135	5	0								
Г. Котлас	214	22	0								
Д. Абрамково	112	23	0								
П. Подосиновец	94	1	0								

Таблица 3

Сравнение результатов работы алгоритма с фактическими результатами

Год	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Номер сценария (факт)	1	1	2	2	2	1	2	1	2	2	1	1	2	1	1	2	2	1	1	1
Результат проверки алгоритма	1	1	1	2	2	1	2	1	2	2	1	1	2	1	1	1	2	2	1	1

тервале, то полагается, что два сравниваемых года по этому признаку одинаковы, в противном случае различны. Для определения различия для гидрологических признаков в алгоритме вычисляется сумма изменений значений по всем шести постам

$$H_k(\Gamma_n, \Gamma_m) = \frac{1}{6} \sum_{j=1}^6 |p_{knj} - p_{kmj}|,$$

где $H_k(\Gamma_n, \Gamma_m)$ — величина различия n -го и m -го годов по k -му признаку, p_{knj} и p_{kmj} — числовые значения k -го признака на j -м посту в n -й и m -й годы.

Для метеорологических признаков вычисляется изменение

$$H_i(\Gamma_n, \Gamma_m) = |q_m - q_{im}|,$$

где $H_i(\Gamma_n, \Gamma_m)$ — величина различия n -го и m -го годов по i -му признаку, q_m и q_{im} — числовое значение i -го признака в n -й и m -й годы.

Если эта величина превышает пороговое эвристическое значение, то полагается, что различие по этому признаку в этой паре лет есть, в противном случае оно отсутствует:

$$H_l(\Gamma_n, \Gamma_m) > \delta_l \text{ — различие есть,}$$

$$H_l(\Gamma_n, \Gamma_m) < \delta_l \text{ — различия нет,}$$

где δ_l — пороговое эвристическое значение по l -му признаку.

Принятая в настоящее время процедура сравнения, описанная выше, является естественной, она использует интегральные характеристики для гидрометеорологических признаков. Пороговые эвристические

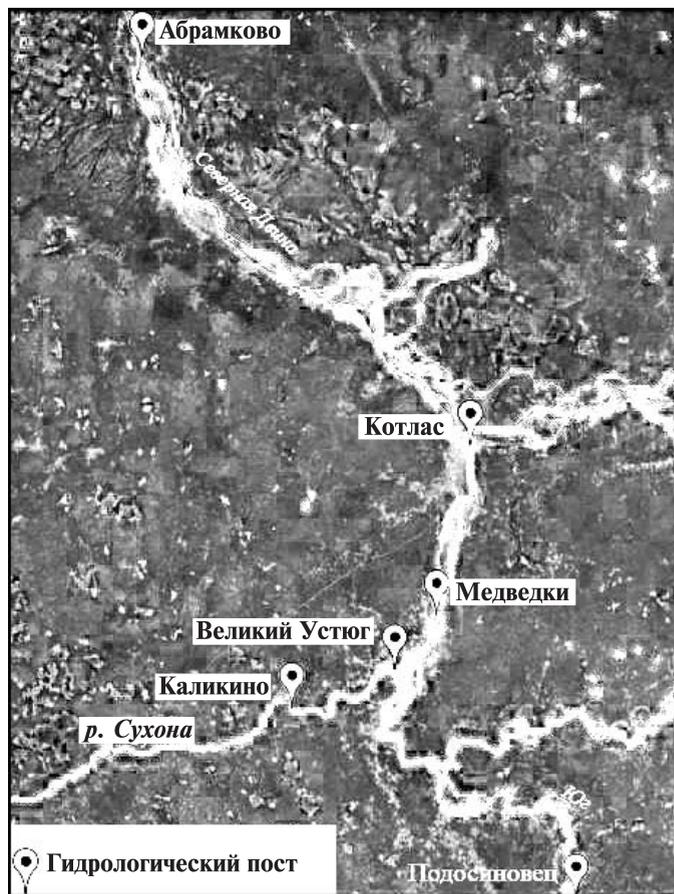


Схема расположения гидрологических постов на р. Северная Двина и ее притоках

константы δ_i устанавливаются либо экспертами исходя из природы признаков и особенностей конкретной территории исследования, либо с использованием автоматизированного алгоритма поиска таких чисел, критерием в котором выступает максимум правильно распознанных элементов обучающей выборки.

Процедура сравнения имеет модульный характер, что позволяет корректировать ее и адаптировать под конкретные условия.

Формирование таблицы сравнения классов. Для каждой пары лет из разных классов с использованием процедуры сравнения однородных признаков определяется обобщенный вектор различия этой пары лет по всем признакам. Вектор формируется следующим образом: если в результате работы процедуры сравнения установлено различие в паре лет по i -му признаку, то в качестве координаты с номером i обобщенного вектора принимается 1, в противном случае 0, следовательно, это булевский вектор:

$$V(\Gamma_n, \Gamma_m) = (v_1, \dots, v_{11}),$$

$$v_i = 1, \text{ если } H_i(\Gamma_n, \Gamma_m) > \delta_i,$$

$$v_i = 0, \text{ если } H_i(\Gamma_n, \Gamma_m) < \delta_i,$$

$$i = 1, \dots, 11.$$

Этот вектор не нулевой, так как годы берутся из разных классов. Равенство нулю этого вектора означает, что выбранная процедура сравнения является грубой моделью реальной ситуации. Совокупность всех таких векторов составляет таблицу сравнения классов. Исследуемый период 1991—2010 гг. разбит на два класса: в первом классе — 11 лет, во втором — 9, следовательно, число пар лет из разных классов равно $11 \cdot 9 = 99$, т.е. в таблице сравнения содержится 99 булевских векторов размерности 11.

Построение множества тестов таблицы сравнения классов. Цель этого этапа — формирование всех наборов признаков (координат булевских векторов), по которым различаются все пары лет из разных классов. *Тестом* называется такой набор признаков, когда для любой пары лет из разных классов есть различие между исследуемыми годами хотя бы по одному признаку из этого набора. По этому набору формируется булевский вектор, т.е. если признак принадлежит набору, то в соответствующую координату ставится 1; если признак не принадлежит набору, то в соответствующую координату ставится 0.

Векторы этого множества хранят информацию о том, насколько отличается один год из одного класса от другого года из другого класса. На этом этапе очень хорошо выявляется не только корреляция между признаками, но и корреляция между группами признаков.

Данные по прогнозируемому году. Для построения прогноза, т.е. отнесения прогнозируемого года к одному из выделенных классов, необходимо сформировать набор X , аналогичный исходным данным прогнозирования (табл. 2). Данные для гидрологических признаков предоставляются постами наблюдений, данные по метеорологическим признакам — по метеопрогнозу на текущую весну.

Построение прогноза. На этом этапе происходит принятие решения о том, по какому сценарию предполагается развитие ледовой обстановки в прогнозируемом году.

По входным данным X прогнозируемого года, исходным данным за период наблюдений, принятой процедуре сравнения признаков и построенному множеству тестов проводится *голосование* за принадлежность прогнозируемого года одному из классов.

Голосование представляется следующей логической процедурой. Фиксируется тест t и год Γ_n из периода наблюдения. Единичные значения координат теста определяют набор признаков, по которым сравнивается текущий год и год периода наблюдения. Если в каждой единичной координате теста значение признака текущего года “совпало” (т.е. они равны в смысле процедуры сравнения) с признаками года из периода наблюдения, то предполагается присвоить один голос тому классу, к которому принадлежит год из периода наблюдения:

1) $g(X, \Gamma_n, t) = 1$, если для каждого i -го признака такого, что i -я координата t равна 1, i -й признак распознаваемого года X “совпал” с i -м признаком года Γ_n ;

2) $g(X, \Gamma_n, t) = 0$, если хотя бы для одного i -го признака, такого, что i -я координата t равна 1, а i -й признак распознаваемого года X “не совпал” с i -м признаком года Γ_n .

Затем проводится суммирование голосов по всем тестам и всем годам из класса $K_j, j = 1, 2$:

$$G_j = \frac{1}{K_j} \sum_t \sum_{\Gamma_n \in K_j} g(X, \Gamma_n, t).$$

Решающее правило F прогнозного алгоритма заключается в отнесении прогнозируемого года к тому классу, за который подано больше голосов:

$$F(X) = j_0, \text{ где } G_{j_0} = \max G_j.$$

Логическая процедура голосования имеет разные модификации по решающему правилу от “жесткого” (идеальное соответствие прогнозируемого года эталону из периода наблюдения) до “мягкого” (предполагает неполное соответствие).

На выходе формируется прогноз вида “ситуация в прогнозируемом году будет развиваться по сценарию j_0 ”. Достоверность этого результата оценивается в процентах.

Результаты исследований и их обсуждение. В табл. 3 представлены результаты автоматизированной обработки данных за период наблюдения построенным выше прогнозированным алгоритмом. Из периода наблюдения последовательно удалялся каждый год, оставшиеся годы рассматривались в качестве обучающей выборки, удаленный год подавался на вход прогнозируемому алгоритму в качестве распознаваемого, строился прогноз. Как видно из данных табл. 3, число правильно классифицированных лет составило 85%. Допущенные алгоритмом прогнозирования ошибки, вероятно,

связаны с пробелами в исходных данных и принятым решающим правилом для поиска пороговых значений признаков (максимум верно распознанных элементов обучающей выборки).

Использование теоретико-вероятностного подхода при решении задач прогнозирования предполагает наличие статистически значимых обучающих выборок. В исследуемой проблеме эти выборки могут не удовлетворять такому требованию, если информативно полные и однородные данные возможно структурировать за небольшой период наблюдения (15–20 лет). Для малообъемных выборок разработан комбинаторно-логический подход. В его рамках тестовые процедуры голосования решают задачи распознавания для параметров разных природы, структуры, описания, происхождения [5].

Применение этого подхода в исследуемой задаче потребовало ряда доработок. Во-первых, в отличие от общепринятого в теории распознавания образов число-

вого описания признаков, имеет место пространственный характер группы признаков, поэтому алгоритмы были модифицированы с учетом этой особенности. Во-вторых, предложены формальные процедуры сравнения гидрологических и метеорологических данных в конкретной задаче с целью выявления их значимого различия.

Выводы:

— представленный алгоритм решает задачу прогнозирования возникновения заторов льда для района с коротким периодом наблюдения за ледовым режимом рек (~20 лет). Результат работы описанного выше метода — отнесение прогнозируемого года к одному из возможных случаев развития обстановки;

— использование прогноза наличия заторов позволяет учитывать этот фактор при составлении прогноза наводнений, что в свою очередь может сократить ущерб для экономической и хозяйственной деятельности на исследуемом участке реки.

СПИСОК ЛИТЕРАТУРЫ

1. Агафонова С.А. Опасные ледовые явления на реках России: классификация, возможность прогнозирования // Проблемы прогнозирования чрезвычайных ситуаций: Мат-лы IV науч.-практ. конф. М., 2004. С. 3–4.

2. Агафонова С.А., Фролова Н.Л. Особенности ледового режима рек бассейна Северной Двины // Водные ресурсы. 2007. Т. 34, № 2. С. 123–131.

3. Агафонова С.А., Фролова Н.Л. Влияние ледового режима рек севера европейской территории России на гидроэкологическую безопасность в условиях изменения климата // Вестн. Моск. ун-та. Сер. 5. География. 2009. № 4. С. 55–61.

4. Вапник В.Н., Червоненкис А.Я. Теория распознавания образов. М.: Наука, 1980. 416 с.

5. Константинов Р.М., Королева З.Е., Кудрявцев В.Б. Комбинаторно-логический подход к задачам прогноза рудоносности // Проблемы кибернетики. Вып. 31. М.: Наука, 1976. С. 5–38.

6. Лупачев Ю.В., Скрипник Е.Н., Кучейко А.А. Опыт космического мониторинга развития весеннего половодья на р. Северная Двина в 2010 г. // Земля из космоса. 2010. № 6. С. 57–68.

7. Симонов Ю.Г., Невяжский И.И. Экспертные оценки при географическом прогнозировании // Вестн. Моск. ун-та. Сер. 5. География. 1978. № 4. С. 25–31.

Поступила в редакцию
12.09.2013

I.V. Malygin

FORECASTING OF ICE CLOGGING IN RIVERS USING THE THEORY OF IMAGE RECOGNITION

A technique of solving the task of ice clogging forecasting using methods of the theory of image recognition is suggested. Possible database structure and content are described. The scheme of forecasting algorithm is presented, as well as the results of its application for the Severnaya Dvina River segment from the town of Kotlas to the town of Veliky Ustyug.

Key words: forecast, forecasting of ice clogging, forecasting algorithm, classification task, combinatorial-logical approach.