# ПРИМЕНЕНИЕ ТЕОРИИ НЕЧЕТКИХ МНОЖЕСТВ К ЗАДАЧЕ ГЕНЕАЛОГИЧЕСКОЙ КЛАССИФИКАЦИИ В ТЕКСТОЛОГИЧЕСКОМ ИССЛЕДОВАНИИ

# APPLICATION OF FUZZY SET THEORY FOR GENEALOGICAL CLASSIFICATION IN TEXTOLOGY

**Шпирко Сергей Валерьевич**
Кандидат физико-математических наук, старший научный сотрудник лаборатории структурных методов анализа данных в предсказательном моделировании (ПреМоЛаб) факультета управления и прикладной математики (ФУПМ) Московского физико-технического института (МФТИ), магистрант I курса (специалитет) исторического факультета Московского государственного университета имени М. В. Ломоносова
E-mail: shpirko@yahoo.com

**Sergey V. Shpirko**

Предлагается новый подход к построению генеалогической классификации текстов списков (вариантов текста произведения). Вводится формальным образом отношение предпочтения для каждой пары текстов. Введенное отношение предпочтения позволяет применить для данной задачи классификации аппарат теории нечетких множеств. Описывается алгоритм нечеткой классификации текстов. Работа предложенного алгоритма демонстрируется на двух конкретных примерах.

*Ключевые слова*: текстология, генеалогическая классификация, нечеткие множества, отношение порядка, нечеткая классификация.

In this paper we present a new approach for genealogical classification in textology. We formally introduce a preference relation between pairs of texts. This allows apply Fuzzy Set Theory for genealogical classification. We propose an algorithm of fuzzy classification of texts. We illustrate the algorithm by two examples.

*Key words*: textology, genealogical classification, fuzzy sets, preference relation, fuzzy classification

## ВВЕДЕНИЕ

Теория нечетких множеств является обобщением многих важнейших разделов классической математики. В основе данной теории лежит осознание нечеткости как универсальной реальности человеческого существования. Так, при качественном анализе сложных систем аппарат теории нечетких отношений позволяет учитывать силу связи между объектами. Отсюда понятен интерес к данной теории со стороны исследователей гуманитарно-исторических направлений. Так, теория нечетких множеств уже применялась для анализа структуры текстов[1]. В данной статье предпринята попытка применения теории нечетких множеств к задачам генеалогической классификации в текстологии, что представляется, на наш взгляд, весьма актуальной целью.

Сделаем неформальное введение в рассматриваемую задачу. Предположим, что текстолог имеет в своем распоряжении некоторое количество сохранившихся списков некоторого произведения (архетипа). Каждый из списков имеет порой в качестве своего источника не один, а несколько списков (протографов). Каждый из списков создавался в свое время и несет на себе следы как бессознательных изменений (извод), так и сознательных (редакция). В общем, каждый из списков слагается из разнородных и разновременных частей. В данной статье анализируются только тексты списков. В дальнейшем вместо слова «текст списка» будем употреблять просто «список». Текстологическая реконструкция архетипа является архисложной, часто невыполнимой задачей[2]. Поэтому на первый план выступает задача выявления взаимосвязей, классификация списков, построение генеалогического древа (стеммы) списков.

Когда текстолог говорит, например, о том, что какие-либо списки принадлежат одному изводу, то он исходит из своих представлений о «похожести» или «непохожести» списков друг на друга, из детального анализа изучаемого текста. Возникает вопрос — возможно ли формализовать в какой-то степени подобный процесс, носящий, вне сомнения, творческий характер?

Так, для задачи хронологического упорядочивания списков можно ввести неким формальным образом отношение предпочтения между двумя элементами, а далее переходить к задаче отыскания максимального (минимального) элемента на множестве альтернатив с заданным отношением предпочтения, как это делается в классической теории принятия решений. Однако во многих реальных ситуациях и, в частности, текстологических задачах можно определить лишь нечеткое отношение предпочтения. Например, можно говорить, что один список предшествует другому со степенью достоверности не ниже заданного порога. В таком случае задача определения максимального элемента становится неопределенной, поскольку неясно, что такое максимальный (минимальный) элемент для нечеткого отношения предпочтения. Как будет показано далее, теория нечетких множеств дает положительный ответ на поставленный вопрос и позволяет строить нечеткую классификацию списков. Необходимо отметить, что проводимое упорядочивание списков невозможно без наличия эксперта, который и ориентирует стемму, задает направление времени, «стрелу времени». Также стоит сказать, что развиваемый в данной статье формальный подход носит предварительный, вспомогательный характер для дальнейшего текстологического исследования.

Задача построения генеалогической стеммы списков решалась с помощью метода групп[3], который не только позволил выяснить взаимосвязи между существующими списками, но и реконструировать предполагаемые списки. На основе полученных результатов была построена стемма из списков «Закона Судного людем», как сохранившихся, так и реконструированных.

Напомним, что генеалогическая стемма есть ориентированный граф. Вершинами данного графа являются списки, а дугами соединены протографы со своими ближайшими списками. Стемму, построенную методом групп, можно представить в виде стройного дерева, разрастающегося от самой макушки книзу, причем каждая вершина непосредственно соединяется только с одной верхней вершиной.

Применение метода групп опиралось на безусловное соблюдение трех важных условий: каждый список имеет только один протограф; все изменения (ошибки) в протографе наследуются во всех его последующих списках (потомках); у списков, относящихся к различным протографам, не могут возникать одинаковые изменения (ошибки).

Предложенный в данной статье подход по сравнению с методом групп обладает следующими особенностями. Во-первых, в модель вводится понятие нечеткости, что позволяет учитывать силу связи между списками (на заданном уровне достоверности). Во-вторых, предложенный подход учитывает возможность наличия нескольких протографов у любого из списков. В-третьих, на основе данного подхода приводится формальный алгоритм нечеткой классификации. Нечеткая классификация разбивает множество реальных списков на классы и хронологически их упорядочивает. Списки, оказавшиеся в одном классе, являются эквивалентными (схожими). Можно сказать, что предложенный алгоритм минимизирует внутриклассовую «неоднородность». При этом важно подчеркнуть, что и эквивалентность, и упорядоченность рассматриваются исключительно с позиции нечеткости. Причем уровень нечеткости регулируется задаваемым порогом (например, равным 0,5).

В то же время существенным недостатком предложенного подхода, по крайней мере на данный момент, является невозможность реконструкции предполагаемых списков. В оправдание можно заметить, что если предполагаемые списки все же реконструированы, то дальнейшее построение генеалогической стеммы можно проводить с помощью предложенного подхода. При этом еще раз подчеркнем, что данный подход включает экспертную оценку, ориентирующую стемму определенным образом.

Структура статьи следующая.

В первой части введены необходимые понятия и формулировки утверждений из теории нечетких множеств. Это в первую очередь понятия нечеткого отношения и его свойств, эквивалентности нечетких множеств, отношения квазипорядка и порядка. При первом прочтении данную часть можно пропустить. Во второй части статьи описан собственно сам алгоритм нечеткой классификации. Его работа будет проиллюстрирована на простом примере.

В третьей части работа алгоритма будет продемонстрирована на реальном материале, приводимом Я. С. Лурье[4].

Было бы весьма интересно оценить работу предложенного алгоритма на качественно ином (большом) материале, например, «Повести временных лет»[5].

## 1. ЭЛЕМЕНТЫ ТЕОРИИ НЕЧЕТКИХ МНОЖЕСТВ

### 1.1. Нечеткое множество

Обстоятельное введение в теорию нечетких множеств можно найти, например, в работе Л. Заде[6]. Мы можем говорить о нечетком множестве, когда не можем четко указать его границы. В обычной практике можно с полной определенностью сказать — входит тот или иной элемент в заданное множество или не входит. С точки зрения нечетких множеств можно говорить лишь о степени достоверности такого суждения. Таким образом, мы приходим к понятию «функция принадлежности», которое показывает, насколько можно считать отдельный элемент входящим во множество. Возможен и крайний вариант, когда функция принадлежности принимает на элементе единичное или нулевое значение. Мы оказываемся в рамках классической теории четких множеств тогда, когда элемент либо принадлежит множеству, либо не принадлежит.

Далее, если это особо не оговаривается, мы следуем в своем изложении теории нечетких множеств согласно работе А. Н. Мелихова (с соавторами)[7].

Пусть $X$ — произвольное непустое множество.

*Определение 1.* Нечетким подмножеством $\tilde{A}$ называется множество пар $\tilde{A} = \{< \mu_A(x)/x >\}$, где $x \in X$.

Функция $\mu_A(x): X \to [0,1]$ называется функцией принадлежности нечеткого множества $\tilde{A}$, а $X$ — базовым множеством или базовой шкалой.

### 1.2. Операции над нечеткими высказываниями

Функцию принадлежности можно трактовать как логическую формулу (высказывание), заданную на множестве элементов. Но, в отличие от традиционной, бинарной логики, данная функция принимает значения из всего единичного отрезка.

*Определение 2.* Отрицанием нечеткого высказывания $\mu_A(x)$ называется нечеткое высказывание, обозначаемое $\overline{\mu}_A(x)$, степень истинности которого определяется выражением

$$\overline{\mu}_A(x) = 1 - \mu_A(x).$$

*Определение 3.* Конъюнкцией нечетких высказываний $\mu_A(x)$ и $\mu_B(x)$ называется нечеткое высказывание, обозначаемое $\mu_A(x) \,\&\, \mu_B(x)$, степень истинности которого определяется выражением

$$\mu_A(x) \,\&\, \mu_B(x) = \min(\mu_A(x), \mu_B(x)).$$

*Определение 4.* Дизъюнкцией нечетких высказываний $\mu_A(x)$ и $\mu_B(x)$ называется нечеткое высказывание, обозначаемое $\mu_A(x) \vee \mu_B(x)$, степень истинности которого определяется выражением

$$\mu_A(x) \vee \mu_B(x) = \max(\mu_A(x), \mu_B(x)).$$

*Определение 5.* Импликацией нечетких высказываний $\mu_A(x)$ и $\mu_B(x)$ называется нечеткое высказывание, обозначаемое $\mu_A(x) \to \mu_B(x)$ степень истинности которого определяется выражением

$$\mu_A(x) \to \mu_B(x) = \max(1 - \mu_A(x), \mu_B(x)).$$

Данные определения позволяют ввести операции над нечеткими множествами.

### 1.3. Операции над нечеткими множествами

Пусть заданы нечеткие подмножества $\tilde{A}$ и $\tilde{B}$ множества $X$.

*Определение 6.* Множество $\tilde{A}$ нечетко включается во множество $\tilde{B}$, если

$$(\mu_A(x) \to \mu_B(x)) \geq 0{,}5 \;\; \forall x \in X.$$

Например, определенная археологическая культура может нечетко включаться в более крупную археологическую культуру (культурно-историческую общность).

*Определение 7.* Будем полагать, что множество $\tilde{A}$ нечетко эквивалентно (равно) множеству $\tilde{B}$ и обозначать $\tilde{A} \approx \tilde{B}$, если $\tilde{A}$ нечетко включается в $\tilde{B}$, а $\tilde{B}$ нечетко включается в $\tilde{A}$.

Например, близкородственные культуры на множестве всех археологических культур являются эквивалентными.

Пусть

$$\tilde{A} = \{< \mu_A(x)/x >\}, \tilde{B} = \{< \mu_B(x)/x >\}, x \in X.$$

*Определение 8.* Объединением множеств $\tilde{A}$ и называется нечеткое множество, обозначаемое $\tilde{A} \cup \tilde{B}$ и определяемое как

$$\tilde{A} \cup \tilde{B} = \{< \mu_{A \cup B}(x)/x >\}, x \in X,$$

где $\mu_{A \cup B}(x) = \mu_A(x) \vee \mu_B(x)$.

*Определение 9.* Пересечением множеств $\tilde{A}$ и $\tilde{B}$ называется нечеткое множество, обозначаемое $\tilde{A} \cap \tilde{B}$ и определяемое как

$$\tilde{A} \cap \tilde{B} = \{< \mu_{A \cap B}(x)/x >\}, x \in X,$$

где $\mu_{A \cap B}(x) = \mu_A(x) \,\&\, \mu_B(x)$.

Наряду с нечетким множеством, центральную роль в данной теории играет понятие нечеткого отношения на множестве.

## 1.4. Нечеткое отношение на множестве

*Определение 10*. Нечетким отношением на $X$ называется и через $\tilde{\eta} = (X,V)$ обозначается пара множеств, в котором $V$ является нечетким подмножеством $X \times X$.

Перечислим множество $X = \{x_i\}, i \in I$ , и пусть

$$V = \{< \mu_V < x_i, x_j > / < x_i, x_j > \in X \times X\}, \mu_V < x_i, x_j > \in [0,1].$$

Тогда нечеткое отношение представимо в виде ориентированного графа. Каждой вершине данного графа соответствует свой элемент из $X$. А ненулевое значение функции $\mu_V < x_i, x_j >$ задает дугу, соединяющую соответствующие вершины $x_i$ и $x_j$.

## 1.5. Основные свойства нечетких отношений

*Определение 11*. Отношение $\tilde{\eta} = (X,V)$ называется нечетко рефлексивным, если

$$\mu_V < x, x > \geq 0{,}5 \; \forall x \in X.$$

*Определение 12*. Отношение $\tilde{\eta} = (X,V)$ называется нечетко антирефлексивным, если

$$\overline{\mu}_V < x, x > \geq 0{,}5 \; \forall x \in X.$$

*Определение 13*. Отношение $\tilde{\eta} = (X,V)$ называется нечетко симметричным, если

$$(\mu_V < x, y > \to \mu_V < y, x >) \geq 0{,}5 \; \forall x, y \in X, x \neq y.$$

*Определение 14*. Отношение $\tilde{\eta} = (X,V)$ называется нечетко антисимметричным, если

$$\overline{(\mu_V < x, y > \& \mu_V < y, x >)} \geq 0{,}5 \; \forall x, y \in X, x \neq y.$$

*Определение 15*. Отношение $\tilde{\eta} = (X,V)$ называется нечетко транзитивным, если

$$(\bigvee_{y \in X} (\mu_V < x, y > \& \mu_V < y, z >) \to \mu_V < x, z >) \geq$$
$$\geq 0{,}5 \; \forall x, y, z \in X, x \neq y, y \neq z, x \neq z.$$

*Определение 16*. Отношение $\tilde{\eta} = (X,V)$ называется нечетко связанным, если

$$(\mu_V < x, y > \vee \mu_V < y, x >) \geq 0{,}5 \; \forall x, y \in X, x \neq y.$$

Все введенные определения отношений имеют наглядную графическую интерпретацию. Так, каждая вершина графа нечеткого рефлексивного отношения содержит петлю со значением, большим порогового (≥ 0,5).

В графе нечеткого симметричного отношения, если какая-либо пара вершин соединена дугой, то между ними обязательно имеется дуга с противоположной ориентацией. Причем данные дуги имеют оба значения: либо больше, либо меньше порогового.

В графе нечеткого связанного отношения между любыми двумя вершинами обязательно найдется дуга со значением, большим порогового.

В графе нечеткого транзитивного отношения, если пара вершин соединена дугой со значением большим порогового, то обязательно найдется такая цепочка промежуточных вершин, что по ней можно «пройти» от начальной вершины к конечной. Причем все соответствующие дуги имеют значения больше порогового.

Комбинируя приведенные основные свойства, можно описать любое нечеткое отношение. Нам понадобятся следующие типы нечетких отношений.

## 1.6. Нечеткое отношение эквивалентности

*Определение 17*. Отношение $\tilde{\eta} = (X,V)$ называется отношением нечеткой эквивалентности (сходства), если оно нечетко рефлексивно, нечетко симметрично и нечетко транзитивно.

Примерами отношений эквивалентности являются отношения «быть примерно одного цвета», «иметь схожие интересы».

Любое множество $X$ с заданным на нем отношением нечеткой эквивалентности $\tilde{\eta} = (X,V)$ можно разбить на семейство эквивалентных классов. Каждый класс данного семейства есть непустое нечеткое подмножество $X$. Все они попарно нечетко не пересекаются и в сумме составляют (нечетко) $X$.

По определению два нечетких множества нечетко не пересекаются, если функция принадлежности их пересечения не превосходит порогового значения. Понятно, что таких разбиений на эквивалентные классы можно привести бесконечно много.

Среди всех разбиений нас будет интересовать оптимальное в том смысле, что функция принадлежности пересечений любых его классов принимает минимальные значения. Приведем процедуру построения такого разбиения. В дальнейшем будем называть такое разбиение каноническим.

*Процедура 1. Построение канонического разбиения*

1. Для любого $x \in X$ построим нечеткое множество $\tilde{\mathrm{A}}(\mathrm{x}) = \{< \mu_{A(x)}(\mathrm{y}) / \mathrm{y} >\}, \mathrm{y} \in \mathrm{X}$, где

$$\mu_{A(x)}(\mathrm{y}) = \mu_V < x, y >.$$

2. Если $\tilde{A}(x) \approx \tilde{A}(\mathrm{y})$, то элементы $x$ и $y$ входят в один класс $\tilde{A}(x) \cup \tilde{A}(\mathrm{y})$.

## 1.7. Нечеткие отношения порядка

*Определение 18*. Отношение $\tilde{\eta}=(X,V)$ называется отношением нечеткого порядка, если оно нечетко антисимметрично и нечетко транзитивно.

Примерами отношений порядка является отношение «быть примерно старше возрастом» на множестве людей.

*Определение 19*. Отношение $\tilde{\eta}=(X,V)$ называется отношением нечеткого строгого порядка, если оно нечетко антисимметрично, нечетко транзитивно и нечетко антирефлексивно.

*Определение 20*. Отношение $\tilde{\eta}=(X,V)$ называется отношением нечеткого совершенного строгого порядка, если оно является нечетким строгим порядком и является связанным.

Нетрудно видеть, что в графе совершенного порядка все вершины соединены друг с другом.

Если $\mu_V<x,y>\geq 0{,}5$, то будем говорить, что элемент $x$ нечетко предшествует элементу $y$. Интуитивно понятно, что введя такую меру предшествования, можно линейно упорядочить все элементы базового множества.

Для такого упорядочивания необходимо прежде всего ввести понятие нечетко наименьшего элемента. Ниже приводится конструктивный способ определения наименьшего элемента.

Для этого выберем произвольный элемент $y_1 \in X$ и проверим выполнение неравенств $\mu_V<y_1,y>\geq 0{,}5\,\forall y\in X, y\neq y_1$. Если все неравенства выполнены, то элемент $y_1$ назовем наименьшим. В противном случае, так как порядок нечеткий совершенный, в $X$ найдется элемент $y_2 \neq y_1$, что будет верно $\mu_V<y_2,y_1>\geq 0{,}5$, т. е. элемент $y_2$ предшествует $y_1$. Тогда либо будет наименьшим, либо найдется такой элемент $y_3 \neq y_2$, что будет верно $\mu_V<y_3,y_2>\geq 0{,}5$. Продолжая аналогичные рассуждения, на k-м шаге приходим к цепочке элементов $y_k,y_{k-1},..,y_2,y_1$. В данной цепочке каждые соседние элементы связаны неравенством $\mu_V<y_{i+1},y_i>\geq 0{,}5, i=1,..,k-1$. А поскольку отношение транзитивное, то среди данных элементов $y_k$ будет наименьшим: $\mu_V<y_k,y_i>\geq 0{,}5, i=1,..,k-1$. Понятно, что за конечное число шагов мы придем к нечетко наименьшему элементу всего множества $X$.

*Процедура 2. Построение линейного упорядочивания $\tilde{P}(X)$ базового множества X*

Пусть $x_1$ — наименьший элемент множества $X$. Припишем элементу $x_1$ номер один и образуем пару $<x_1,\beta(x_1)>$, где $\beta(x_1)=\underset{x\in X}{\&}\mu_V<x_1,x>(\geq 0{,}5)$. Рассмотрим множество $X_1=X\setminus\{x_1\}$ и найдем на нем наименьший элемент. Получим элемент $x_2$ и пару $<x_2,\beta(x_2)>$, где $\beta(x_2)=\underset{x\in X_1}{\&}\mu_V<x_2,x>(\geq 0{,}5)$. Продолжая аналогичным образом, приходим к нечеткой линейной последовательности $\tilde{P}(X)=<<x_1,\beta(x_1)>,<x_2,\beta(x_2)>,..,<x_n,\beta(x_n)>>$. Полученная последовательность в определенном смысле упорядочивает элементы $X$ оптимальным образом. Будем называть такое упорядочивание *каноническим.*

Совершенный порядок — это идеальный случай. Обычно для графа отношения трудно добиться того, чтобы все элементы были бы друг с другом связаны.

Рассмотрим отношение $\tilde{\eta}=(X,V)$ нечеткого строгого порядка, которое не является совершенным. В этом случае от рассмотрения наименьшего элемента переходят к рассмотрению понятия минимального элемента.

*Определение 21*. Элемент $a \notin X$ называют нечетко минимальным, если не существует элемента $x \in X$, что $\mu_V<x, a>>0{,}5$.

Заметим, что в случае совершенного порядка понятия наименьшего и минимального элемента совпадают.

Понятно, что в случае порядка, не являющегося совершенным, на множестве $X$ найдутся несравнимые элементы (не соединенные дугами), поэтому его нельзя линейно упорядочить. Однако можно выделить такое семейство нечетких подмножеств множества $X$, что в сумме они дают все множество $X$ (покрывают его), а каждое из них линейно упорядочивается.

*Определение 22*. Подмножество $Y \subseteq X$ называют максимальным, если отношение $\tilde{\eta}$ задает на $Y$ нечеткий совершенный строгий порядок и не существует $Z \subseteq X$ такого, что $Y \subseteq Z$ и отношение $\tilde{\eta}$ на $Z$ также является нечетким совершенным строгим порядком.

*Процедура 3. Построение нечеткого покрытия из линейно упорядоченных подмножеств*

1. Выделим все максимальные совершенные подмножества множества $X$;
2. Для каждого максимального совершенного подмножества $Y$ построим линейную последовательность $\tilde{P}(\mathrm{Y})$ (см. процедуру 2);
3. Для каждого $\tilde{P}(\mathrm{Y})$ сопоставим нечеткое множество $\tilde{P}_y(\mathrm{Y})$, состоящее из тех же элементов, что и $\tilde{P}(\mathrm{Y})$, но порядок которых несуществен.
4. Совокупность всех таких $\tilde{P}_y(\mathrm{Y})$ будет являться искомым покрытием.

Заметим, что требование антирефлексивности в Определении 19 не является для нас критичным. Действительно, пусть отношение нечеткого порядка является рефлексивным (нестрогий порядок). Тогда, удалив из графика данного отношения все петли, приходим к антирефлексивному порядку. Для полученного отношения остается применить описанную Процедуру 3 и получить соответствующую структуру множеств.

## 1.8. Нечеткие отношения предпорядка

*Определение 23*. Отношение $\tilde{\psi}=(X,F)$ называется отношением квазипорядка (предпорядка), если оно нечетко рефлексивно и нечетко транзитивно.

Примером квазипорядка является отношение «быть примерно не старше курсом» на множестве студентов.

Из Определения 23 нетрудно заключить, что и отношение эквивалентности, и отношение нестрогого порядка являются частными случаями предпорядка.

Справедливо следующее важное утверждение: нечеткий квазипорядок на некотором непустом множестве $X$ порождает как нечеткую эквивалентность (с нечетким разбиением множества $X$ на классы), так и некоторый нечеткий нестрогий порядок, построенный на данных классах.

Для построения отношения эквивалентности нам потребуются понятия инверсии отношения и пересечения отношений

*Определение 24*. Инверсией отношение $\tilde{\psi}=(X,\mathrm{F})$ называется и через $\tilde{\psi}^{-1}=(X,F^{-1})$ обозначается нечеткое отношение, у которого график имеет те же вершины, что и у графика $F$, а дуги имеют противоположную направленность по сравнению с $F$. Для любых $x_i$, $x_j \in X$ справедливо $\mu_{F^{-1}}<x_i,x_j>=\mu_F<x_j,x_i>$.

*Определение 25*. Пересечением отношений $\tilde{\psi}=(X,F)$ и $\tilde{\varphi}=(X,\mathrm{P})$ называется нечеткое отношение $\tilde{\pi}=(X,\mathrm{S})$, обозначаемое $\tilde{\pi}=\tilde{\psi}\cap\tilde{\varphi}$, если $S=F\cap P$. Иными словами,

$$\mu_S<x_i,x_j>=\mu_F<x_i,x_j>\&\,\mu_P<x_i,x_j>.$$

*Процедура 4. Построение нечеткого отношения эквивалентности*

Пусть $\tilde{\psi}=(X,F)$ — нечеткий квазипорядок.

1. Построим инверсию $\tilde{\psi}^{-1}=(X,F^{-1})$;
2. Построим пересечение $\tilde{\varphi}=\tilde{\psi}\cap\tilde{\psi}^{-1}$;

Полученное отношение $\tilde{\varphi}=(X,\mathrm{P})$ является отношением эквивалентности, и на нем можно построить каноническое разбиение множества $X$ на классы (см. Процедуру 1). Ниже приведена процедура построения на множестве таких классов $\Im$ нечеткого нестрогого порядка, индуцируемого исходным квазипорядком $\tilde{\psi}$ [8].

*Процедура 5. Построение нестрогого порядка на множестве классов канонического разбиения*

Пусть задано каноническое разбиение множества $X: X=\bigcup\limits_{A\in\Im}A$.

1. Для каждого нечеткого множества $A$ построим четкое множество $A^*$ следующим образом:

$$\mathrm{A}^*=\{<\mu_{A^*}(x)/x>\},\ \text{где}\ \mu_{A^*}(x)=\begin{cases}1, & \mu_A\geq 0{,}5\\ 0, & \mu_A<0{,}5.\end{cases}$$

2. Зададим нечеткое отношение $\tilde{\rho}=(\Im,T)$ следующим образом. Для любых двух классов $A_i, A_j\in\Im$ положим $\mu_T<A_i,A_j>=\mu_F<x_i,x_j>$, где $x_i\in A_i^*, x_j\in A_j^*$ и $x_i\neq x_j$.

Заметим, что такие пары элементов $x_i$ и $x_j$ (эталоны нечетких множеств) обязательно найдутся. Иначе соответствующие им классы $A_i$ и $A_j$ были бы эквивалентны, что противоречит определению канонического разбиения.

Построенное нечеткое отношение $\tilde{\rho}$ является нечетким нестрогим порядком.

## 1.9. Транзитивное замыкание

Все рассмотренные нами отношения (эквивалентности, порядка и предпорядка) обладают одним общим свойством — транзитивностью (см. Определение 17). Данное свойство определяет некоторую «правильную» структуру базового множества, обеспечивает возможность его разбиения на непересекающиеся классы сходства, возможность естественного упорядочения элементов. Но что делать в случае, когда изучаемое нечеткое отношение не обладает этим свойством? Оказывается, существует процедура, позволяющая перейти от исходного нетранзитивного отношения к транзитивному.

*Определение 26*. Композицией нечетких отношений $\tilde{\psi}=(X,\mathrm{F})$ и $\tilde{\varphi}=(X,\mathrm{P})$ будем называть нечеткое отношение $\tilde{\eta}=(X,V)$, обозначаемое $\tilde{\eta}=F\circ P$, где

$$\mu_V<x,z>=\bigvee_{y\in X}\mu_F<x,y>\&\,\mu_P<y,z>, x,z\in X.$$

*Определение 27*. Транзитивным замыканием отношения $\tilde{\eta}$ называется отношение $\hat{\eta}=\eta^1\cup\eta^2\cup..\cup\eta^k\cup..$, где отношения $\eta^k$ определяются рекуррентным образом:

$$\eta^1=\tilde{\eta}, \eta^k=\eta^{k-1}\circ\tilde{\eta}, k=2,3,4...$$

Можно доказать, что транзитивное замыкание $\hat{\eta}$ транзитивно и является наименьшим транзитивным отношением, включающим $\tilde{\eta}$.

*Процедура 6. Построение транзитивного замыкания нечеткого отношения*

1. Пусть задано нечеткое отношение $\tilde{\eta}$. Если $\tilde{\eta}$ транзитивно, то завершаем выполнение процедуры. Иначе положим $V=\tilde{\eta}$ и перейдем к пункту 2.

2. На очередном $k$-м шаге. Построим $\eta^k=\eta^{k-1}\circ\tilde{\eta}$. Положим $V=V\bigcup\eta^k$. Если $\eta^k=\eta^{k-1}$, то положим $\hat{\eta}=V$ и завершаем выполнение процедуры. Иначе положим $k:=k+1$ и переходим к пункту 2.

Можно доказать, что описанная процедура остановится за конечное число шагов[9].

## 2. ОПИСАНИЕ АЛГОРИТМА КЛАССИФИКАЦИИ

Будем считать, что задача уже формализована, т. е. мы имеем некоторое нечеткое отношение предпорядка $\tilde{\psi} = (X, \mathrm{F})$.

*Алгоритм нечеткой классификации*

1. Если $\tilde{\psi}$ не транзитивно, то осуществляем процедуру 6;

2. Если $\tilde{\psi}$ не рефлексивно, то добавляем в граф *F* петли с пороговым значением больше 0,5;

В результате шагов 1,2 имеем отношение квазипорядка $\tilde{\psi} = (X, \mathrm{F})$.

3. Процедура 4. В результате имеем отношение эквивалентности $\tilde{\varphi} = \tilde{\psi} \cap \tilde{\psi}^{-1} = (X, P)$;

4. Процедура 1. В результате получаем каноническое разбиение $\Im$ множества *X*;

Процедура 5. В результате имеем нечеткий нестрогий порядок $\tilde{\rho} = (\Im, T)$;

6. Построение по $\tilde{\rho}$ нечеткого строгого порядка $\tilde{\delta} = (\Im, R)$;

7. Если $\tilde{\delta}$ — совершенный порядок, то Процедура 2, иначе Процедура 3 (включающая нахождение всех максимальных совершенных подмножеств множества $\Im$).

## 3. ПРИМЕНЕНИЕ АЛГОРИТМА НЕЧЕТКОЙ КЛАССИФИКАЦИИ

Пусть *X* есть множество сохранившихся списков изучаемого нами произведения. Рассмотрим все списки и выделим из них все имеющиеся разночтения. Пусть число всех разночтений есть *N*.

Пронумеруем все места в текстах списков, в которых выявлены разночтения. Для произвольного $x \in X$ будем обозначать через $x[i]$ чтение *i*-го места в списке $x (i = 1,\dots, N)$.

Наша задача состоит в построении некоторым формальным образом предпорядка $\tilde{\psi} = (X, \mathrm{F})$ на множестве *X*. Будем проводить данную формализацию в два этапа.

На первом этапе мы рассмотрим все разночтения списков и неформально зададим совершенный порядок на множестве чтений $x[i]$ каждого *i*-го места.

На втором этапе введем формальным образом предпорядок уже на самом множестве списков *X*.

*Определение 28*. Для произвольного $i = 1,..,N$ будем говорить, что чтение $x[i]$ предшествует чтению $y[i]$, и обозначать через $x[i] \prec \mathrm{y[i]}$, если $x[i]$ старше $y[i]$.

*Предположения модели:*

1) введенное в Определении 28 отношение предшествования является совершенным. Следовательно, используя экспертную оценку, любые чтения какого-либо места можно сравнить и расположить их в хронологической последовательности. В частности, отсюда следует, что для любого места текста можно среди всех чтений выявить древнейшее, а по отношению к нему все остальные чтения являются ошибками;

2) в процессе переписывания происходит как перенос существующих ошибок из протографов, так и появление собственных ошибок. Отсюда следует, что чем чаще в одном списке присутствуют ошибки другого, тем достовернее выглядит гипотеза о том, что второй генетически предшествует первому, и наоборот;

3) чем больше у двух списков несовпадающих ошибок, тем более они независимы друг от друга. При этом абсолютные показатели ничего не значат. Необходимо рассматривать относительные показатели (наподобие того, как это делалось Л. И. Бородкиным (с соавторами) при оценке близости синтаксических инвариантов текстов)[10];

4) ошибки не различаемы с точки зрения сознательного или бессознательного внесения в текст.

С учетом приведенных выше предположений мы предлагаем использовать следующую процедуру для вычисления отношения предпочтения.

*Процедура 7. Вычисление функции отношения*

Выберем произвольные $x, y \in X$. Обозначим через $S(x)$ число ошибок в *x*.

1. Положим $S(x, y) := 0$ и $i := 1$;

2. Если $i > N$, то перейдем к пункту 5. Иначе сравним $x[i]$ и $y[i]$;

3. Если $x[i] \prec \mathrm{y[i]}$, то $S(x, y) := S(x, y) + 1$;

4. Положим $i := i + 1$ и перейдем к пункту 2.

5. Положим $V(x, y) := S(x, y) / S(x)$.

Грубо говоря, в качестве отношения предпочтения между двумя списками берем арифметическое отношение числа совпадающих ошибок к их общему числу в первом списке.

**Пример 1**. В качестве простейшего примера рассмотрим четыре списка некоторого произведения. Пусть выявлено 15 мест, в которых обнаружены разночтения. Припишем каждому месту свою букву алфавита. При этом все ошибочные чтения данного места будем обозначать той же буквой, но с натуральным индексом. Причем, чем старше чтение, тем меньше у него индекс. Пусть *x, y, z, w*∈*X*, и *x* = {$А_1$, Б, В, Г, Д, Е, Ж, З, И, К, Л, М, Н, О, П}, *y* = {А, Б, В, $Г_1$, $Д_1$, Е, Ж, З, И, К, Л, М, Н, О, П}, *z* = {$А_1$, $Б_1$, $В_1$, $Г_1$, $Д_1$, $Е_1$, $Ж_1$, $З_1$, $И_1$, $К_1$, Л, М, Н, О, П}, *w* = {$А_2$, $Б_2$, $В_1$, $Г_1$, $Д_1$, $Е_1$, $Ж_1$, $З_1$, $И_1$, $К_1$, $Л_1$, $М_1$, $Н_1$, $О_1$, $П_1$}.

Следовательно, список *z* является протографом для *w*. Списки *x, y* не зависят друг от друга, и оба являются протографами для *z*. В списке *x* всего одна ошибка (по сравнению с оригиналом), и стоит она на первом месте. В $y$ две ошибки, и находятся они на четвертом и пятом местах. Список $z$ унаследовал ошибки от обоих своих протографов, а также у него появились еще семь своих ошибок. Его непосредственный список $w$ унаследовал все ошибки от $z$ и добавил свои (стоящие на местах с одиннадцатого по пятнадцатое). Данная генеалогическая схема изображена на рисунке 1.

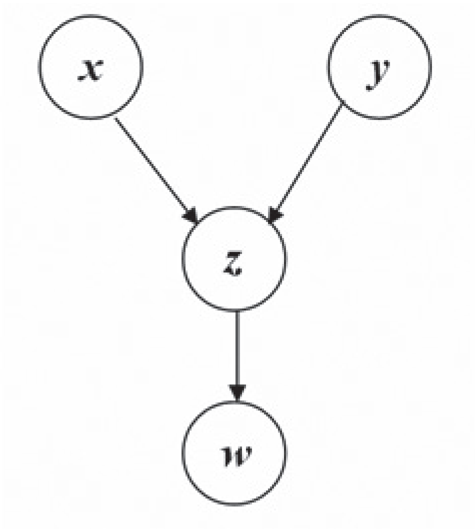


Рис.1. Генеалогическая стемма списков (пример 1)

Подсчитаем, например, $V(x, y)$. Общих ошибок у этих списков нет. А всего в списке $x$ ошибка одна. Следовательно, $V(x,y) = 0/1 = 0$. Подсчитаем $V(w, y)$. Общих ошибок две ($Г_1$, $Д_1$). Всего в $w$ 15 ошибок, т. е. $V(w, y) = 2/15$. Результаты всех подсчетов удобно свести в так называемую матрицу отношения (см. табл. 1).

*Таблица 1*

**Таблица отношения** $\tilde{\eta} = (X, V)$

| | x | y | z | w |
|---|---|---|---|---|
| x | 1 | 0 | 1 | 1 |
| y | 0 | 1 | 1 | 1 |
| z | 1/10 | 2/10 | 1 | 1 |
| w | 0 | 2/15 | 8/15 | 1 |

По строкам и столбцам данной матрицы перечислены списки, а в клетках стоят соответствующие значения функции отношения. Нетрудно увидеть, что отношение $\tilde{\eta}$ рефлексивное (на диагоналях матрицы отношения стоят единицы).

Преобразуем $\tilde{\eta}$ к транзитивному отношению с помощью процедуры 6. Сначала необходимо вычислить композицию $\tilde{\eta}^2 = \tilde{\eta} \circ \tilde{\eta}$ (см. Определение 26). Результаты вычислений приведены в таблице 2.

*Таблица 2*

**Таблица отношения** $\tilde{\eta}^2$

| | x | y | z | w |
|---|---|---|---|---|
| x | 1 | 2/10 | 1 | 1 |
| y | 1/10 | 1 | 1 | 1 |
| z | 1/10 | 2/10 | 1 | 1 |
| w | 1/10 | 2/10 | 8/15 | 1 |

Объединение двух отношений $\tilde{\eta}$ и $\tilde{\eta}^2$ представлено в таблице 3.

*Таблица 3*

**Таблица отношения** $\tilde{\eta} \cup \tilde{\eta}^2$

| | x | y | z | w |
|---|---|---|---|---|
| x | 1 | 2/10 | 1 | 1 |
| y | 1/10 | 1 | 1 | 1 |
| z | 1/10 | 2/10 | 1 | 1 |
| w | 1/10 | 2/10 | 8/15 | 1 |

Вычислим еще раз композицию двух отношений и убедимся, что $\tilde{\eta}^3 = \tilde{\eta}^2$. Таким образом, мы конструктивно показали, что полученное отношение $\hat{\eta} = \tilde{\eta} \cup \tilde{\eta}^2$ является транзитивным замыканием ($\tilde{\psi} = (X, \mathrm{F}) = \hat{\eta}$) исходного отношения (см. Определение 27).

Теперь, согласно алгоритму нечеткой классификации, переходим к построению отношения эквивалентности $\tilde{\varphi} = \hat{\eta} \cap \hat{\eta}^{-1}$ (см. Процедуру 4).

*Таблица 4*

**Матрица отношения** $\tilde{\varphi} = (X, \mathrm{P})$

| | x | y | z | w |
|---|---|---|---|---|
| x | 1 | 2/10 | 1/10 | 1/10 |
| y | 1/10 | 1 | 2/10 | 2/10 |
| z | 1/10 | 2/10 | 1 | 8/15 |
| w | 1/10 | 2/10 | 8/15 | 1 |

Перейдем к построению канонического разбиения $\Im$ множества списков. Согласно Процедуре 1

$$\tilde{A}(x) = \{<\mu_F<x,x>/x>, <\mu_F<x,y>/y>, <\mu_F<x,z>/z>, <\mu_F<x,w>/w>\} = \{<1/x>, <2/10\ /y>, <1/10\ /z>, <1/10\ /w>\},$$

$$\tilde{A}(y) = \{<1/10\ /x>, <1/y>, <2/10\ /z>, <2/10\ /w>\},$$

$$\tilde{A}(z) = \{<1/10\ /x>, <2/10\ /y>, <1/z>, <8/15\ /w>\},$$

$$\tilde{A}(w) = \{<1/10\ /x>, <2/10\ /y>, <8/15\ /z>, <1/w>\}.$$

Заметим, что $\tilde{A}(z) \approx \tilde{A}(w)$ (см. Определение 7). Объединяя оба нечетких множества, приходим к следующему каноническому разбиению из трех классов: $\Im = \{A_1, A_2, A_3\}$, где

$$A_1 = \tilde{A}(x), A_2 = \tilde{A}(y), A_3 = \tilde{A}(z) \cup \tilde{A}(w) =$$
$$\{<1/10 \,/\, x>, <2/10 \,/\, y>, <8/15 \,/\, z>, <1/w>\}.$$

Введем теперь нечеткий нестрогий порядок на $\Im$ индуцируемый исходным квазипорядком $\tilde{\psi} = (X, F)$. Вспомним, что Процедура 5 проводится в два этапа.

На первом этапе с каждым классом $A_i$ связывается свое четкое множество $A_i^*$. В нашем случае два таких множества состоят из одного элемента, а одно — из двух элементов:

$$A_1^* = \{x\}, A_2^* = \{y\}, A_3^* = \{z, w\}.$$

На втором этапе собственно и строится нестрогий порядок $\tilde{\rho} = (\Im, T)$. Напомним, что отношение $\tilde{\rho}$ задается следующим образом:

$$\mu_T < A_i, A_j > = \mu_F < x_i, x_j >,$$

где $x_i \in A_i^*$, $x_j \in A_j^*$ и $x_i \neq x_j$. В нашем случае это отношение можно задать двумя способами. При первом способе в расчет берется элемент $z$ из $A_3^*$, а при втором — элемент $w$. Матрица отношения $\tilde{\rho}$, рассчитанная первым способом, приведена в таблице 5.

*Таблица 5*

**Матрица нестрогого порядка $\tilde{\rho} = (\Im, T)$**

| | $A_1$ | $A_2$ | $A_3$ |
|---|---|---|---|
| $A_1$ | 1 | 2/10 | 1 |
| $A_2$ | 1/10 | 1 | 1 |
| $A_3$ | 1/10 | 2/10 | 1 |

Обнулив главную диагональ данной матрицы (что эквивалентно удалению всех петель на соответствующем графе), приходим к нечеткому строгому порядку $\tilde{\delta}$.

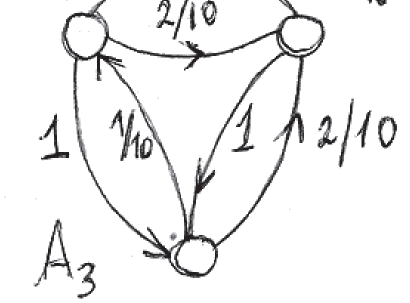


Рис. 2. Граф отношения строгого порядка $\tilde{\rho} = (\Im, R)$

Нетрудно увидеть, что порядок не является связанным. Подмножества $Y_1 = \{A_1, A_3\}$ и $Y_2 = \{A_2, A_3\}$ являются максимальными совершенными подмножествами. Применяя Процедуру 3, получим две нечеткие линейные последовательности, порожденные данными подмножествами:

$$\tilde{P}(Y_1) = << 1/A_1 >, <1/A_3 >>$$
$$\text{и } \tilde{P}(Y_2) = << 1/A_2 >, <1/A_3 >>.$$

В свою очередь данные линейные последовательности порождают нечеткие множества

$$\tilde{P}_{y_1} = \{<1/A_1>, <1/A_3>\},$$
$$\tilde{P}_{y_2} = << 1/A_2 >, <1/A_3 >>,$$

образующие покрытия базового множества $X$.

Итак, в результате работы алгоритма нечеткой классификации была построена стемма, отражающая взаимосвязи исходных четырех списков *x, y, z, w*. Со степенью правдоподобия больше порога 0,5 исходные списки были разбиты на три эквивалентных класса, между ними было проведено генеалогическое упорядочение. Отметим, что оба списка *z* и *w* попали в один класс эквивалентности. Это можно объяснить большим числом совпадающих ошибок в обоих списках. Перейдем теперь к рассмотрению реального примера.

**Пример 2.** В работе «Новые списки Царева государева послания во все его Российское царство» Я. С. Лурье вводит в научный оборот два найденных списка «Царева государева послания во все его Российское царство»: Библиотеки Академии наук СССР, № 230 (А), Государственной Публичной библиотеки, собрания Погодина № 1311 (Б). Выясняя вопрос о первоначальной редакции Послания Ивана IV и ее назначения, Я. С. Лурье проводит текстологическое сличение списков А и Б с уже имевшимися (на тот момент) списками Государственной Публичной библиотеки, собрания Погодина, № 1567 (П), Археографической комиссии, № 41 (К).

Проведем нечеткую классификацию данных списков и сравним полученные результаты с выводами, сделанными Я. С. Лурье.

Обозначим для удобства списки *А, Б, П, К* как x, y, z, w. Используя материал, приведенный в Приложении 2, выявим и пронумеруем все места с разночтениями. Все древнейшие чтения будем обозначать заглавными буквами, а все ошибочные — теми же буквами, но с натуральным индексом внизу. Чем «моложе» ошибки, тем соответствующий индекс больше. Результаты нашего анализа приведены в таблице 6[11].

*Таблица 6*

**Распределение ошибок по спискам (пример 2)**

| № | Древнейшее чтение | x | y | z | w |
|---|---|---|---|---|---|
| 1 | В ряду предков указано имя Дмитрия (Донского) | $А_1$ | А | $А_1$ | А |
| 2 | «Писание же твое вразумлено внятельно» | $Б_1$ | Б | $Б_1$ | Б |
| 3 | «Поправшим вам с своими советники» | $В_2$ | В | $В_2$ | $В_1$ |
| 4 | Приведено отчество у князя Федора — «Ростиславич» | $Г_1$ | Г | $Г_1$ | Г |
| 5 | «Ипаты» | $Д_1$ | Д | $Д_1$ | Д |
| 6 | «Рим со всею Италиею от греческого царства отторжеся» | $Е_1$ | Е | $Е_1$ | Е |
| 7 | «Почто есмя сильных во Израиле побили» | $Ж_1$ | Ж | $Ж_1$ | Ж |
| 8 | «…не станет казнить подвласных имуще разум» | $З_1$ | З | $З_1$ | З |
| 9 | «Не по времени» | $И_1$ | И | $И_1$ | И |
| 10 | Корректно указаны все три сына императора Константина | $К_2$ | К | $К_2$ | $К_1$ |
| 11 | **«Его благословение отпустившим»** | Л | $Л_1$ | $Л_1$ | $Л_1$ |
| 12 | **«Янушу»** (Заполи) | М | М | $М_1$ | $М_1$ |
| 13 | **«Евтропию скопцу»** | Н | Н | $Н_1$ | $Н_1$ |
| 14 | **Приведено имя Дионисия** (Ареопагита) | О | О | $О_1$ | $О_1$ |
| 15 | **«…понеже он ему дядя»** | П | П | $П_1$ | $П_1$ |
| 16 | «…понеже и утешен не…» | Р | Р | $Р_2$ | $Р_1$ |
| 17 | «Кроновы» жертвы | $С_1$ | $С_1$ | С | $С_1$ |

Построим отношение предпорядка на данном множестве списков, применяя Процедуру 7. Из таблицы 6 нетрудно подсчитать общее число ошибок, попавших из одного списка в другой.

*Таблица 7*

**Общее число ошибок, попавших из одного списка в другой**

| $S(.,.)$ | x | y | z | w |
|---|---|---|---|---|
| x | 11 | 1 | 10 | 1 |
| y | 1 | 2 | 1 | 2 |
| z | 10 | 1 | 16 | 5 |
| w | 3 | 2 | 8 | 9 |

Заметим, что общее число ошибок в списках $S(x) = 11$, $S(y) = 2$, $S(z) = 16$, $S(w) = 9$.

Нормируя полученную функцию $S(.,.)$ согласно п. 5 Процедуры 7, приходим к отношению $\tilde{\eta} = (X,V)$ , представленному в таблице 8.

*Таблица 8*

**Таблица отношения $\tilde{\eta} = (X,V)$**

| $V(.,.)$ | x | y | z | w |
|---|---|---|---|---|
| x | 1 | 1/11 | 10/11 | 1/11 |
| y | 1/2 | 1 | 1/2 | 1 |
| z | 10/16 | 1/16 | 1 | 5/16 |
| w | 3/9 | 2/9 | 8/9 | 1 |

Применим теперь к отношению $\tilde{\eta}$ алгоритм нечеткой классификации. Ниже приведены результаты последовательной работы данного алгоритма.

*Нечеткая генеалогическая классификация*

1. Процедура 6 (транзитивное замыкание)

*Таблица 9*

**Таблица транзитивного замыкания $\tilde{\psi} = (X,\mathrm{F})$**

| $F(.,.)$ | x | y | z | w |
|---|---|---|---|---|
| x | 1 | 1/11 | 10/11 | 5/16 |
| y | 1/2 | 1 | 8/9 | 1 |
| z | 10/16 | 2/9 | 1 | 5/16 |
| w | 10/16 | 2/9 | 8/9 | 1 |

2. Рефлексивность

Отношение $\tilde{\psi}$ является рефлексивным (на главной диагонали матрицы стоят единицы).

3. Процедура 4 (построение отношения эквивалентности)

*Таблица 10*

**Матрица отношения $\tilde{\varphi} = \tilde{\psi} \cap \tilde{\psi}^{-1} = (X,P)$**

| $P(.,.)$ | x | y | z | w |
|---|---|---|---|---|
| x | 1 | 1/11 | 10/16 | 5/16 |
| y | 1/11 | 1 | 2/9 | 2/9 |
| z | 10/16 | 2/9 | 1 | 5/16 |
| w | 5/16 | 2/9 | 5/16 | 1 |

4. Процедура 1 (каноническое разбиение)

$$\tilde{\mathrm{A}}(\mathrm{x}) = \{<1/x>, <1/11\ /y>, <10/16\ /z>, <5/16\ /w>\},$$

$$\tilde{\mathrm{A}}(\mathrm{y}) = \{<1/11\ /x>, <1/y>, <2/9\ /z>, <2/9\ /w>\},$$

$$\tilde{\mathrm{A}}(\mathrm{z}) = \{<10/16\ /x>, <2/9\ /y>, <1/z>, <5/16\ /w>\},$$

$$\tilde{A}(w)=\{<5/16\ /x>,<2/9\ /y>,$$
$$<5/16\ /z>,<1/w>\}.$$

Заметим, что $\tilde{A}(x)\approx\tilde{A}(z)$ (см. Определение 7). Тогда искомое каноническое разбиение $\Im=\{A_1,A_2,A_3\}$, где

$$A_1=\tilde{A}(x)\cup\tilde{A}(z)=\{<1/x>,<2/9\ /y>,$$
$$<1/z>,<5/16\ /w>\},$$
$$A_2=\tilde{A}(y),\ A_3=\tilde{A}(w).$$

5. Процедура 5 (нестрогий порядок на каноническом разбиении)

Возьмем $A_1^*=\{x,z\},\ A_2^*=\{y\},\ A_3^*=\{w\}$. Нестрогий порядок $\tilde{\rho}=(\Im,T)$, индуцированный исходным предпорядком $\tilde{\psi}$, представлен в следующей таблице 11.

*Таблица 11*

**Матрица отношения $\tilde{\rho}=(\Im,T)$**

| $T(.,.)$ | $A_1$ | $A_2$ | $A_3$ |
|---|---|---|---|
| $A_1$ | 1 | 1/11 | 5/16 |
| $A_2$ | 1/2 | 1 | 1 |
| $A_3$ | 10/16 | 2/9 | 1 |

Здесь во всех вычислениях с $A_1$ участвует $x$:

$$\mu_T<A_1,A_3>=\mu_F<x,w>;$$

$$\mu_T<A_3,A_1>=\mu_F<w,x>;$$

$$\mu_T<A_2,A_1>=\mu_F<y,x>;$$

$$\mu_T<A_1,A_2>=\mu_F<x,y>.$$

6. Построение строгого порядка

*Таблица 12*

**Матрица отношения $\tilde{\delta}=(\Im,R)$**

| $R(\cdot,\cdot)$ | $A_1$ | $A_2$ | $A_3$ |
|---|---|---|---|
| $A_1$ | 0 | 1/11 | 5/16 |
| $A_2$ | 1/2 | 0 | 1 |
| $A_3$ | 10/16 | 2/9 | 0 |

7. Процедура 3 (построение покрытия множества $X$)

Видим, что отношение $\tilde{\delta}$ является совершенным (см. Определение 20). Элемент $A_2$ является наименьшим. Последовательность, нечетко линейно упорядочивающая $\Im$, выглядит следующим образом:

$$\tilde{P}(X)=<<A_2,1/2>,<A_3,10/16>,<A_1,1>>.$$

**Вывод.** Заметим, что а через x, y, z, w были обозначены списки *А, Б, П, К* соответственно. Результаты применения алгоритма нечеткой классификации представимы в виде следующей стеммы:

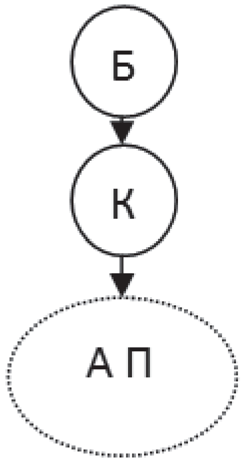


Рис. 3. Граф, отражающий генеалогические связи списков «Послания»

Таким образом, нами установлены следующие факты:

— списки *А* и *П* нечетко эквивалентны, т. е. создавались примерно в одно и то же время и входят в одну и ту же редакцию или извод;
— список *К* нечетко предшествует спискам *А* и *П*;
— список *Б* нечетко предшествует списку *К*.

Для сравнения приведем выдержки из выводов Я. С. Лурье при работе над теми же списками: «…сопоставляя между собой уже известные списки (*П, К. — С. Ш.*) и вновь найденные списки (*А, Б. — С. Ш.*) — мы обнаруживаем прежде всего, что новый список *А* отличается особой близостью к известному уже списку *П*… Следует отметить, что совпадение между *К* и *Б* здесь объясняется только тем, что *ПА* в этих случаях дают явно неверное чтение; никакой специальной близости списков *Б* и *К* мы не обнаруживаем… Таким образом, списки *П* и *А* имеют общий протограф; в свою очередь этот общий протограф, список *К* и список *Б* независимо друг от друга восходят к единому архетипу» (см. текст статьи Я. С. Лурье, приведенный в приложении 2).

Таким образом, предложенный алгоритм выделяет два близких списка *П* и *А*. Данные списки хронологически следуют за списком *Б* и *К*, что совпадает с выводами Я. С. Лурье. Но Я. С. Лурье указывает для *П, А* существование некоего промежуточного списка (протографа), который совместно со списками К и Б восходит к единому архетипу.

В заключение хочу выразить горячую благодарность профессору Л. И. Бородкину, доценту

Т. А. Кругловой за все то внимание и ценные советы, оказываемые ими автору в процессе написания данной статьи.

## Приложение 1

### Доказательство корректности Процедуры 5

**Транзитивность**

Как следует из Определения 15, отношение $\tilde{\rho}=(\Im,T)$ является транзитивным, если $\mu_T<A_i,A_j>\geq 0{,}5$ и $\mu_T<A_j,A_l>\geq 0{,}5$, то

$$\mu_T<A_i,A_l>\geq 0{,}5\ \ \forall A_i,A_j,A_l\in\Im.$$

Неравенства $\mu_T<A_i, A_j>\geq 0{,}5$ и $\mu_T<A_j, A_l>\geq 0{,}5$ означают, что найдутся такие $x\notin A_i^*$, $y$, $y'\notin A_j^*$, $z\notin A_l^*$, $x\neq y$, $y'\neq z$, что $\mu_F<x, y>\geq 0{,}5$ и $\mu_F<y', z>\geq 0{,}5$.

Поскольку $y$, $y'\notin A_j^*$, то $\mu_P<y, y'>\geq 0{,}5$. А поскольку $\tilde{\varphi}=\tilde{\psi}\cap\tilde{\psi}^{-1}$, то и $\mu_F<y, y'>\geq 0{,}5$ (см. Определение 25).

С учетом транзитивности $\tilde{\psi}$ из цепочки неравенств $\mu_F<x,y>\geq 0{,}5$, $\mu_F<y,y'>\geq 0{,}5$, $\mu_F<y',z>\geq 0{,}5$ следует, что $\mu_F<x,z>\geq 0{,}5$, т. е. $\mu_T<A_i,A_l>\geq 0{,}5$.

**Рефлексивность**

Как следует из Определения 11, отношение $\tilde{\rho}=(\Im,T)$ является рефлексивным, если $\mu_T<A,A>\geq 0{,}5\ \forall A\in\Im$. Данное неравенство означает, что найдется $x\in A^*:\mu_F<x,x>\geq 0{,}5$. А последнее неравенство справедливо в силу рефлексивности $\tilde{\psi}=(X,F)$.

**Антисимметричность**

Как следует из Определения 14, отношение $\tilde{\rho}=(\Im,T)$ является антисимметричным, если $\mu_T<A_i, A_j>\geq 0{,}5$ и $\mu_T<A_j, A_i>\geq 0{,}5$, то $A_i=A_j\ \forall\ A_i$, $A_j\in\Im$. Неравенства $\mu_T<A_i, A_j>\geq 0{,}5$ и $\mu_T<A_j, A_i>\geq 0{,}5$ означают, что найдутся такие $x$, $x'\notin A_i^*$, $y$, $y'\notin A_j^*$, $x\neq y$, $x'\neq y'$, что $\mu_F<x, y>\geq 0{,}5$ и $\mu_F<y', x'>\geq 0{,}5$. Поскольку $x$, $x'$ нечетко входят в один класс эквивалентности $A_i$, то $\mu_F<x', x>\geq 0{,}5$. Аналогично имеем $\mu_F<y, y'>\geq 0{,}5$.

Вспомним, что $\tilde{\psi}=(X,F)$ транзитивно. Следовательно, из $\mu_F<x, y>\geq 0{,}5$ и $\mu_F<y, y'>\geq 0{,}5$ получаем $\mu_F<x, y'>\geq 0{,}5$. Аналогично из $\mu_F<y', x'>\geq 0{,}5$ и $\mu_F<x', x>\geq 0{,}5$ следует $\mu_F<y', x>\geq 0{,}5$, что эквивалентно $\mu_{F-1}<x, y'>\geq 0{,}5$ (см. Определение 24).

Поскольку $\tilde{\varphi}=\tilde{\psi}\cap\tilde{\psi}^{-1}=(X,P)$, то из $\mu_F<x, y'>\geq 0{,}5$ и $\mu_{F-1}<x, y'>\geq 0{,}5$ следует $\mu_P<x, y'>\geq 0{,}5$. Таким образом, $x$ и $y'$ принадлежат одному классу эквивалентности. Следовательно, $A_i=A_j$ (поскольку $\Im$ — каноническое разбиение).

## Приложение 2

### Лурье Я. С. Новые списки Царева государева послания во все его Российское царство // Труды отдела древнерусской литературы. М. ; Л., 1954. С. 308–309 (фрагмент из публикации)

…сопоставляя между собой списки — ГПБ, Погодинский № 1567 (П) и Археографической комиссии № 41 (К) — и вновь найденные списки — БАН, № 230 (А) и ГПБ, Погодинский № 1311 (Б), — мы обнаруживаем прежде всего, что новый список А отличается особой близостью к уже известному списку П. Сходен состав этих списков: в обоих сборниках тексту «Послания» царя предшествуют «Послания» Курбского Ивану Грозному, Тетерина и Сарыхозина — Морозову, Полубенского (в списке А нет только «Послания» Курбского в Псковско-Печерский монастырь). Текстологическое сличение всех четырех списков обнаруживает ряд совпадений между П и А — даже в явных описках; в списках К и Б в соответствующих местах текст читается иначе. Так, например, в ПА в начале «Послания» (стр. 9)[12] при перечислении предков Ивана IV пропущено имя Дмитрия (Донского) — в КБ оно читается; в ПА читается: «Писание же твое … вразумлено и внятельно» (стр. 13) — в КБ правильнее: «вразумлено внятельно», в ПА читается непонятное обвинение Курбского, что он «вправду с вами злобесовскими советники» (стр. 14) — в К яснее: «Поправшу с вами» (т. е. «поправшу» заветы христианства), в Б еще лучше: «поправшим вам с своими.. советники»; в ПА пропущено отчество смоленского князя Федора «Ростиславич» (стр. 19) — в КБ оно читается; в ПА при перечислении правителей Византии называются непонятные «пилаты» (стр. 24) — в КБ правильно: «ипаты»; В ПА читается: «Рим со всею Италиею от греческого царства оттого же вся» (стр. 26) — в КБ: «отторжеся»; в ПА в процитированном царем вопросе Курбского: «почто есмя сильных во Израиле побили» (стр. 29) — пропущено слово «сильных» — в КБ оно читается; в ПА царь доказывает, что он не станет казнить «подвласных, имущих разум» (стр. 31) — в КБ правильно: «имуще разум»; в ПА: «по не времени» (стр. 33) — в КБ: «не по времени» и т. д. Следует отметить, что совпадение между К и Б здесь объясняется только тем, что ПА в этих случаях дают явно неверное чтение, никакой специальной близости списков Б и К мы не обнаруживаем: список Б нигде не повторяет многочисленных ошибок и погрешностей списка К. Так, список Б правильно называет имена трех сыновей императора Константина — «Константин, Констанций и Конста» (стр. 24), в то время как список К одного из них пропускает, а двух других сливает в единого «Константина»; в списке К об отставке Сильвестра говорилось: «мы же его благословив, не отпустившу» (стр. 41) — создавалось впечатление, что царь отказал Сильвестру в отставке[13]; в А читается: «его благословение отпустившим»; вместо ошибочного «Янущу» (Заполе) в списке К (стр. 42), в АБ читается: «Янушу»; вместо «Евтропию скупцу» — «Евтропию скопцу»; в АБ читается имя Дионисия (Ареопагита) как автора рассказа, приводимого Грозным; в К оно пропущено (стр. 57), и т. д. Таким образом, списки П и А имеют общий протограф[14]; в свою очередь этот общий протограф, список К и список Б независимо друг от друга восходят к единому архетипу.

Новые списки А и Б дают возможность правильно восстановить несколько мест «Послания», до сих пор непонятных вследствие искажений. В особенности это относится ко второй части «Послания», которая была известна только по одному списку первой редакции — списку К, богатому описками и бессмыслицами. Вот как читаются, например, в списках АБ слова Ивана IV об отце Курбского: «Се вси видят, каков же пред ним ты и колко у отца твоего начальников поселяном, колико же у тебя; отец твой был князя Михаила Кубенского боярин, понеже он ему дядя» (в списке К читалось бессмысленное «дадя»; ср. стр. 54)[15].

Гораздо лучше читаются по новым спискам и заключительное предостережение царя Курбскому по поводу его надежд на утешение в Польше, искаженное во всех других редакциях и с трудом восстанавливающееся по списку К: «Понеже и утешен не[16] можеши быти, понеже там кождо о своем попечение имея — кто убо может избавити тя от насильных рук и от обидящего восхитити тя возможет, иже сиру и вдовицу суду не внемлюще, их же вы, желающие на христианство злая, составляете» (ср. стр. 62).

Но главное значение новонайденных списков в том, что они окончательно помогают установить текст первой редакции «Послания». Пока в нашем распоряжении было только два списка, притом дефектных, нельзя было (несмотря на сходство их совпадающих частей) полностью быть уверенным в том, что перед нами единый текст. Теперь в нашем распоряжении четыре текста, несомненно принадлежащих к одной редакции (отсутствие перестановок текста, имеющихся во всех списках остальных редакций, наличие пропущенных там мест), причем один из них (А) содержит весь текст с начало до конца, без пропусков, а в другом (Б) отсутствуют только несколько страниц в конце.

С несомненностью восстанавливается и название «Послания»; если раньше оно читалось только в одном списке полной редакции, то теперь мы читаем его в трех списках. При этом особое значение приобретают установленные нами взаимоотношения списков ПА, К и Б. Если общность названия в ПА можно было бы объяснить особой близостью между ними, то наличие того же (несущественными вариациями) названия в Б свидетельствует о том, что это название органически присуще тексту первой редакции…».

## ПРИМЕЧАНИЯ


1 Москин Н. Д., Петров А. М. Использование нечетких теоретико-графовых моделей для анализа синтаксической структуры музыкально-стихотворного фольклора (на материале эпических духовных стихов) // Методы и принципы современных гуманитарных исследований: сборник научных работ аспирантов и молодых ученых. Петрозаводск, 2011. С. 13–23.

2 Лихачев Д. С. Текстология. Л., 1983. С. 640.

3 Бородкин Л. И., Милов Л. В. О некоторых аспектах автоматизации текстологического исследования (Закон Судный людем) // Математические методы в историко-экономических и историко-культурных исследованиях : сборник статей. М., 1977. С. 235–280.

4 Лурье Я. С. Новые списки Царева государева послания во все его Российское царство // Труды отдела древнерусской литературы. М.; Л., 1954. С. 305–310.

5 Михеев С. М. Кто писал «Повесть временных лет?» М., 2011. С. 280.

6 Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений. М., 1976. С. 166.

7 Мелихов А. Н., Бернштейн Л. С., Коровин С. Я. Ситуационные советующие системы с нечеткой логикой. М., 1990, С. 272.

[8] А. Н. Мелихов с соавторами предлагает построить такой алгоритм в качестве упражнения. Поскольку найти решение предложенного упражнения не удалось, автор данной статьи приводит свой собственный алгоритм. Ввиду его важности доказательство его корректности приведено в приложении 1.

[9] Кофман А. Введение в теорию нечетких множеств. М., 1982. С. 432.

[10] Бородкин Л. И., Милов Л. В., Морозова Л. Е. К вопросу о формальном анализе авторских особенностей стиля в произведениях Древней Руси // Математические методы в историко-экономических и историко-культурных исследованиях: сборник статей. М., 1977. С. 298–326.

[11] К сожалению, не все чтения приведены для всех списков. В таком случае автор считает их ошибочными и выделяет соответствующие места жирным шрифтом.

[12] В дальнейшем в скобках указываются страницы «Послания» в издании 1951 г.

[13] Ср. Послания Ивана Грозного, комментарии, прим. 33 (с. 599).

[14] Характерным отличием списка А от списка П является наличие в нем киноварных заголовков на полях. Заголовки эти, несомненно, принадлежат уже переписчику XVII в. и являются как бы его комментарием к тексту. К месту об Адашеве сделан, например, следующий заголовок: «О Алексее Адашеве, что назвал его царь батожником». Иногда автор заголовков не вполне понимал текст; например, место о «Кроновых жертвах» писец озаглавил: «О Иродовых жертвах».

[15] Ср.: Послание Ивана Грозного, комментарии, прим. 46 (с. 606).

[16] Так в Б; в А нет «не».