С. В. Знаменский

Модель и аксиомы метрик сходства

Аннотация. В современных приложения метрики сходства обычно комбинируются с учётом сложности алгоритмов, особенностей восприятия человека, ресурсов и выборок данных. Для оптимизации требуется унифицированное формальное описание основных показателей подобия. Для оптимизации требуется выделить формально и строго описанное абстрактное понимание сходства между объектами.

Расширена система аксиом метрики сходства и для неё построена универсальная модель, обобщающая известные модели сходства, не сводящиеся к евклидовой метрике. Модель базируется на взвешенном частично упорядоченном множестве.

Ключевые слова и фразы: сходство строк, выравнивание последовательностей, аксиомы сходства, LCS, метрика Левенштейна.

Различные метрики и меры сходства широко используются в психологии, химии, биоинформатике и при обработке текстов. Иногда в этом качестве используются метрики расстояния, представленные в математике понятиями метрического пространства, псевдометрики [1], квазиметрики [1, 2], псевдометрики [1], праметрики [3], семи метрики [4] и частичной метрики [5]. Общеизвестно [6], что метрик расстояния недостаточно для описания метрик и мер сходства. Известные формализации метрик [7–9] и мер [10] сходства прояснили важные аспекты сходства. Прикладные исследования [8, 11–13] убедительно показали необходимость комбинирования различных метрик.

Поиск оптимальных комбинаций метрик нуждается в прозрачной общей картине используемых понятий сходства.

Множество объектов, сопоставимых по сходству, обозначим X . Буквами s, t, u, x, y, z будем обозначать произвольные его элементы. Будем опускать все внешние кванторы $\forall_{s \in X} \dots \forall_{z \in X}$ во всех формулах, акцентируя в ближайшем контексте редко присутствующие внешние переменные.

© С. В. Знаменский, 2017

© Институт программных систем имени А. К. Айламазяна РАН, 2017

© Программные системы: теория и приложения, 2017

DOI: 10.25209/2079-3316-2017-8-4-347-357

Таким образом, любая метрика сходства s предстаёт функцией на $X \times X$ со значениями в $V \subset \mathbb{R}$. Обычно для метрики сходства $V = [0, \infty) \stackrel{\text{def}}{=} \{v \in \mathbb{R} : v \geq 0\}$, а для меры сходства — единичный отрезок $V = [0, 1]$, к которому рассмотрение как-нибудь сводится. Зачастую 0 считается минимально возможным значением, показывающим отсутствие какого бы то ни было сходства, но иногда 0 лежит внутри V и означает отсутствие выразительного сходства или различия.

Различная семантика нуля сигнализирует о путанице.

Чтобы достичь прозрачности и строгости, вспомним о практике нормализации метрик. Исходные метрики часто приводятся к требуемому диапазону значений различными преобразованиями. Возникает задача разделить качества метрик, которые могут быть получены различными преобразованиями при нормализации и те качества, которые должны характеризовать фундаментальное понятие сходства.

Требуется не ограничивая приложений предельно упростить базовую модель, основываясь на разнообразии преобразований нормализации, каждое из которых влияет сразу на несколько качеств иногда по-разному в разных практических ситуациях.

Для решения этой задачи воспользуемся системой аксиом [14], органично совмещающая различные понимания сходства и расстояния. Проанализируем её основные аксиомы и возможные ограничения на область значений V .

1. Система аксиом сходства

1.1. Аксиомы направленности и отделимости

В соответствии с [14] будем идентифицировать аксиомы либо первой буквой названия, исключая «a», «d», «s» и «t», либо второй буквой, если идентификация первой не удаётся. *Аксиома направленности* (*direction*) сходства

$$(i) \quad \forall_{x,y \in U} \quad s(x, y) \leq \min(s(x, x), s(y, y)).$$

отличается от родственной аксиомы для частичной метрики противоположным знаком неравенства. В случае $V = \mathbb{R}$ функции из $U \times U$ со значениями в V образуют линейное пространство, в котором (i) выделяет выпуклый конус направленных на сходство функций. Этот конус пересекается с центрально-симметричным ему конусом направленных на расстояние функций по линейному пространству, состоящему из постоянных функций.

Направленность на сходство эквивалентна условию, что любой непустой замкнутый шар

$$\overline{B}_l(a, r) = \{x \in U : s(a, x) - s(a, a) \geq r\},$$

$$\overline{B}_r(a, r) = \{x \in U : s(x, a) - s(a, a) \geq r\}$$

имеет неотрицательный радиус и содержит свой центр a . и эквивалентна наличию системы вложенных открытых шаров $B_l(a, r) = \{x \in U : s(a, x) - s(a, a) > r\}$ и $B_r(a, r) = \{x \in U : s(x, a) - s(a, a) > r\}$ с центром в произвольной точке a — её окрестностей.

Иными словами, направленная на сходство функция задаёт топологию. Сразу возникает вопрос об отделимости. Классическая аксиома T0 отделимости топологических пространств для направленной (на сходство или на расстояние) функции s означает, что для любых разных точек есть окрестность одной из них, не содержащая другую

$$(0) \quad s(x, y) = s(x, x) = s(y, y) \implies y = x,$$

Аксиома T1 сильнее: она означает, что любая точка не пересекается с подходящей окрестностью любой другой

$$(1) \quad ((s(x, y) = s(x, x)) \vee (s(x, y) = s(y, y))) \implies y = x,$$

Эта аксиома существенно ограничивает общность: многие модели сходства дают полезную возможность судить о вложенности по превращению неравенства в равенство, а T1 (как и её усиление T2, означающая существование непересекающихся окрестностей у любых двух точек) такую возможность принципиально исключает.

1.2. Неравенство треугольника

Хорошо известно [6], что в классической форме из определения метрических пространств *неравенство треугольника* (*triangle*) неприменимо к сходству. Однако оно необходимо для алгоритмов, работающих с метриками расстояния, к которым метрика сходства сводится нормализацией, и для метрики сходства имеет несколько иной вид [7]:

$$(r) \quad s(x, y) + s(y, z) \leq s(x, z) + s(y, y),$$

Нелинейные нормализации могут только усложнить вид неравенства треугольника.

Важная для биоинформатики возможность наглядного представления иерархии близости деревом *phylogenetic tree* or *evolutionary tree* [15] обеспечивается заменой неравенства треугольника на более сильное, чем неравенство треугольника, *аддитивное* (*additive*) *неравенство* (аксиому *четырёх точек*)

$$(d) \quad s(x, y) + s(u, v) \geq \min(s(x, u) + s(y, v), s(x, v) + s(u, y)),$$

либо на ещё более сильное *неравенство ультраметрики* (*ultrametric*)

$$(u) \quad s(x, z) \geq \min(s(x, y), s(y, z)).$$

однако эти варианты существенно ограничивают общность.

1.3. Аксиома симметрии

Хорошо известно [6], что аксиома *симметрии* (*symmetry*)

$$(y) \quad s(x, y) = s(y, x)$$

не всегда применима к метрикам сходства. При этом практически используемые ассиметричные варианты сводятся к использованию ассиметричной однородной нормализации

$$(1) \quad \frac{s(x, y)}{\alpha s(x, x) + \beta s(y, y) + (1 - \alpha - \beta)s(x, y)}$$

симметричной функции сходства $s(x, y)$ или других ассиметричных нормализаций этой же функции.

Будем это интерпретировать как фундаментальность аксиомы симметрии.

1.4. Значения сходства и избранные аксиомы

Поскольку аксиома треугольника использует сложение, то множество V в абстрактной общей формулировке удобно считать замкнутым относительно сложения, а, значит, неограниченным. Ограниченность его снизу однако не противоречит ни одному из известных подходов и приложений. Инвариантность неравенства треугольника γ относительно добавления к s константы, то есть сдвига, позволяет в абстрактной общей формулировке обойтись без отрицательных чисел, корректируя при необходимости смещение сдвигом.

Описанные соображения приводят к однозначному выбору в качестве V замкнутого луча $[0, \infty)$ вещественной оси и в итоге к предложенной в [7] системе аксиом, обозначенных выше (0,i,r,y).

Если исключить в отдельную группу метрики сходства, полученные нормализацией евклидовой метрики и метрики, основанные на эвристических алгоритмах и не имеющие строгих математических оснований, то оставшиеся метрики сходства как правило удовлетворяют аксиоме *выравнивания* (*alignment*) о существовании (суб)оптимальной *общей части*

$$(l) \quad s(x, y) = \sup\{s(z, z) : s(z, z) = s(x, z) = s(z, y), z \in X\},$$

которая в случае если V не имеет конечных предельных точек, например, при $V = \mathbb{N}$, приобретает вид

$$(2) \quad \forall_{x,y \in X} \exists_{z \in X} \quad s(x, y) = s(z, z) = s(x, z) = s(z, y).$$

Рассмотрим аксиому *исчерпания* (*exhausting*):

$$(e) \quad \left. \begin{array}{l} \sup_k s(t_k, t_k) = s(x, x) \\ \forall_k s(t_k, t_k) = s(t_k, x) = s(t_k, y) \end{array} \right\} \implies (s(x, x) = s(x, y)),$$

смысл которой несложен: Если x изнутри приближается сходящимися по s частями, входящими в y , то x входит в y .

Эта аксиома заведомо выполняется если V не имеет конечных предельных точек, например, при $V = \mathbb{N}$. Не выявлено моделей, в которых добавление этой аксиомы ограничивало бы практическую применимость.

2. Содержательная общность

Метрики редактирования и меры сходства строятся на основе содержательной общности (общей информации, общей подпоследовательности, общего подграфа и т.д.) Рассмотрим абстрактную модель такого построения.

ОПРЕДЕЛЕНИЕ 1. *Пространством содержательности* назовём тройку $(X, \preceq, w)$, состоящую из произвольного частично упорядоченного отношением *содержательного вхождения* $\preceq$ множества X , с *весом содержательности* $w : X \rightarrow [0, \infty)$, для которых выполнены следующие условия:

- (1) *монотонность* веса: $x \preceq z \implies w(x) \leq w(z)$,
- (2) существование *минимального элемента* с весом $w(0_x) = 0$.

- (3) *супермодулярность* веса: при $x \preceq z$ и $y \preceq z$ выполняется

$$(3) \quad w(x) + w(y) \leq w(z) + \sup\{w(t) : (t \preceq x) \& (t \preceq y)\},$$

- (4) *исчерпания*:

$$(4) \quad \left. \begin{array}{l} \sup_k w(x_k) = w(x) \\ x_k \preceq x \\ x_k \preceq y \end{array} \right\} \implies (x \preceq y);$$

Смысл исчерпания здесь прежний: если x изнутри приближается сходящимися по w частями, входящими в y , то x входит в y .

Если неравенство в (3) заменить равенством, то супермодулярность превращается в модулярность, а пространство содержательности становится *аддитивным*.

2.1. Содержательностное сходство

Мера содержательностного сходства на пространстве содержательности X это неотрицательная функция $s : (X \times X) \rightarrow [0, \infty)$, определённая равенством

$$s(x, y) \stackrel{\text{def}}{=} \sup\{w(t) : (t \preccurlyeq x) \& (t \preccurlyeq y)\}.$$

ЛЕММА 1. *Отношение и вес содержательности восстанавливаются из содержательного сходства формулами $w(x) = s(x, x)$ и $x \preccurlyeq y \iff s(x, x) = s(x, y)$*

ДОКАЗАТЕЛЬСТВО. Первая формула и импликация вправо во второй непосредственно следуют из монотонности меры w . Обратно, пусть $s(x, x) = s(x, y)$. Тогда для любого $k \in \mathbb{N}$ существует $(t \preccurlyeq x) \& (t \preccurlyeq y)$ для которого $w(t) > w(x) - \frac{1}{k}$. Применяя условие исчерпаемости (4), получаем $x \preccurlyeq y$. $\square$

ТЕОРЕМА 1. *Мера содержательностного сходства $s(x, y)$ при всех $x, y, z \in X$ неотрицательно и удовлетворяет условиям (0, i, 1, r, y, e), то есть является метрикой сходства, удовлетворяющей аксиомам выравнивания и исчерпания (1, e).*

ДОКАЗАТЕЛЬСТВО. Условия неотрицательности и симметричности непосредственно вытекают из неотрицательности w и определения s . Отделимость вытекает из антисимметричности отношения порядка. Направленность (i) вытекает из вложения $\{w(z) : (z \preccurlyeq y) \& (z \preccurlyeq x)\} \subset \{w(z) : (z \preccurlyeq x)\}$, а аксиома исчерпания получается подстановкой $s(x, x) = w(x)$.

Докажем неравенство треугольника (r). Зафиксируем произвольное $\varepsilon > 0$. По определению $s(x, y)$ существует такое $y_x \preccurlyeq x$, что $y_x \preccurlyeq y$ и выполнено $w(y_x) \leq s(x, y) < w(y_x) + \varepsilon$. По определению $s(y, z)$ существует такое $y_z \preccurlyeq z$, что $y_z \preccurlyeq y$ и выполнено $w(y_x) \leq s(y, z) < w(y_z) + \varepsilon$. Применяя к $y_x \preccurlyeq y$ и $y_z \preccurlyeq y$ условие супермодулярности и используя монотонность s получим $s(x, y) + s(y, z) - 2\varepsilon \leq w(y_x) + w(y_z) \leq w(y) + s(y_x, y_z) \leq w(y) + s(x, z)$. $\square$

2.2. Примеры пространств содержательности

Важнейшие подходы, ведущие к пространствам содержательности перечислены в таблице 1. Ключевым объектом в ней назван тот, на котором достигается верхняя грань в определении меры содержательностного сходства. В доступных описаниях метрик сходства не видно прозрачных моделей (кроме базирующихся на евклидовом расстоянии), которые не являлись бы мерами содержательностного сходства.

Таблица 1. Подходы, ведущие к пространствам содержательности

Содержательность (объекта) w , отношение $\preceq$	ключевой объект
аддитивная мера (множества), $\subset$	пересечение множеств
вероятность (события), включается/входит	произведение событий
мощность (множества признаков), $\subset$	множество общих признаков
длина (последовательности символов), подпоследовательность	длиннейшая общая подпоследовательность
длина (символьной строки), подстрока	длиннейшая общая подстрока
колмогоровская сложность (информационного объекта), часть/целое	общая часть
размер (графа), подграф	максимальный общий подграф

Возможные меры содержательностного сходства характеризуются системой аксиом:

ТЕОРЕМА 2. Пусть на множестве X задана метрика сходства s , удовлетворяющая аксиомам выравнивания и исчерпания (1, e) и элемент $0_x \in X$ удовлетворяет $s(0_x, x) = 0$. Тогда вес $w(x) = s(x, x)$ и отношение $x \preceq y \stackrel{\text{def}}{\iff} w(x) = s(x, y)$ превращают X в пространство содержательности, в котором s оказывается мерой содержательного сходства.

ДОКАЗАТЕЛЬСТВО. Рефлексивность и антисимметричность отношения $\preceq$ вытекают из определения $\preceq$ и аксиомы отделимости.

Покажем транзитивность. Пусть $x \preceq y$ и $y \preceq z$. Вычитая из неравенства треугольника $s(x, y) + s(y, z) \leq s(x, z) + s(y, y)$ равенство $s(y, y) = s(y, z)$, получаем, $s(x, z) \leq s(x, y)$, откуда используя $s(x, x) = s(x, y)$ и $s(x, x) \leq s(x, z)$, получаем искомое $s(x, x) = s(x, z)$.

Монотонность следует из определения веса w , минимальный элемент дан в условии. Супермодулярность $s(x, x) + s(y, y) \leq s(z, z) + s(x, y)$ получается из неравенства треугольника $s(x, z) + s(y, z) \leq s(x, y) + s(z, z)$ применением равенств, вытекающих из $x \preceq z$ и $y \preceq z$. Исчерпание вытекает из аксиомы (e). $\square$

2.3. Метрические нормализации пространства содержательности

Мера содержательного сходства $s(x, y)$ может быть по-разному представлена нормализующими предметриками. Среди них наиболее

характерны:

Размер дефицита $d_{\text{dfc}}^{\mu}(x, y) \stackrel{\text{def}}{=} w(x) - s(x, y)$ — псевдометрика, показывающая насколько x далёк от вхождения в y .

Различие в весах $d_{\delta}^{\mu}(x, y) \stackrel{\text{def}}{=} |w(x) - w(y)|$ — псевдометрика, полностью игнорирующая предпорядок.

Расстояние редактирования $d_{\text{edt}}^{\mu}(x, y) \stackrel{\text{def}}{=} w(x) + w(y) - 2s(x, y)$ — метрика расстояния, подсчитывающая вставки и удаления (пример — метрика Левенштейна).

В качестве предметрик (Джаккарда, Дайса, Тверски (1) и других) часто используются различные функции от $a = s(x, y)$, $b = w(x) - a$ и $c = w(y) - a$. Подробнее об этом можно прочитать в [16, 17].

Список литературы

- [1] J. C. Kelly. «Bitopological spaces», *Proc. London Math. Soc.*, **13**:3 (1963), с. 71–89. ³⁴⁷
- [2] W. A. Wilson. «On quasi-metric spaces», *Am. J. Math.*, **53**:3 (1931), с. 675–684. ³⁴⁷
- [3] A. V. Arkhangel'skii, L. S. Pontryagin (eds.), *General Topology I: Basic Concepts and Constructions Dimension Theory*, Encyclopaedia of Mathematical Sciences, т. **17**, Springer, Berlin–Heidelberg, 1990, vii+202 с. ³⁴⁷
- [4] C. Alexander. «Semi-developable space and quotient images of metric spaces», *Pacific J. Math.*, **37**:2 (1971), с. 277–293. ³⁴⁷
- [5] S. G. Matthews. «Partial metric topology», Proc. 8th Summer Conference on General Topology and Applications (Queens College, City University of New York, USA, June 18–22, 1992), *Ann. New York Acad. Sci.*, **728** (1994), с. 183–197. ³⁴⁷
- [6] A. Tversky. «Features of similarity», *Psychol. Rev.*, **84**:4 (1977), с. 327–352. ^{347, 349, 350}
- [7] S. Chen, B. Ma, K. Zhang. «On the similarity metric and the distance metric», *Theoretical Computer Science*, **410**:24–25 (2009), с. 2365–2376. ^{347, 349, 350}
- [8] R. Kunimoto, M. Vogt, J. Bajorath. «Maximum common substructure-based Tversky index: an asymmetric hybrid similarity measure», *J. Comput. Aided Mol. Des.*, **30** (2016), с. 523–531. ³⁴⁷
- [9] E. Amigo, J. Gonzalo, F. Giner, F. Verdejo. «An Axiomatic Account of Similarity», SIGIR'17 Workshop on Axiomatic Thinking for Information Retrieval and Related Tasks (ATIR) (August 2017, Tokyo Japan), 3, 10 с., URL: https://www.eecis.udel.edu/~hfang/atir_proceedings/atir17-paper3.pdf ³⁴⁷

- [10] D. Guessoum, M. Miraoui, Ch. Tadj. «Survey of semantic similarity measures in pervasive computing», *International journal on smart sensing and intelligent systems*, **8:1** (2015), с. 125–158. ↑³⁴⁷
- [11] Th. Slimani. «Description and Evaluation of Semantic Similarity Measures Approaches», *International Journal of Computer Applications*, **80:10** (2013), с. 25–33. ↑³⁴⁷
- [12] F. Giner, E. Amigó. «General Representation Model for Text Similarity», *Future and Emerging Trends in Language Technology. Machine Learning and Big Data (FETLT 2016)* (Seville, Spain, November 30 –December 2, 2016), 2017, с. 158–169 (Revised Selected Papers). ↑³⁴⁷
- [13] A. Gupta, A. Kumar, J. Gautam. «A Survey on Semantic Similarity Measures», *IJIRST-International Journal for Innovative Research in Science & Technology*, **3:12** (2017), URL: <http://www.ijirst.org/articles/IJIRSTV3I12083.pdf> ↑³⁴⁷
- [14] S. V. Znamenskij. «From Similarity to Distance: Axiom Set, Monotonic Transformations and Metric Determinacy», *Journal of Siberian Federal University. Mathematics & Physics*, 2018 (в печати). ↑³⁴⁸
- [15] A. J. Dobson. «Unrooted Trees for Numerical Taxonomy», *Journal of Applied Probability*, **11:1** (1974), с. 32–42. ↑³⁴⁹
- [16] S. Jimenez, F.A. Gonzalez, A. Gelbukh. «Soft Cardinality in Semantic Text Processing: Experience of the SemEval International Competitions», *Polibits*, **51** (2015), с. 63–72, URL: <http://www.scielo.org.mx/pdf/polibits/n51/n51a10.pdf> ↑³⁵⁴
- [17] N.J.P. van Eck, L. Waltman. *How to Normalize Co-Occurrence Data? An Analysis of Some Well-Known Similarity Measures*, ERIM report series research in management Erasmus Research Institute of Management, № ERS-2009-001-LIS, Erasmus Research Institute of Management, 2009, URL: <http://hdl.handle.net/1765/14528> ↑³⁵⁴

Рекомендовал к публикации

д.ф.-м.н. Н. Н. Непейвода

Пример ссылки на эту публикацию:

С. В. Знаменский. «Модель и аксиомы метрик сходства», *Программные системы: теория и приложения*, 2017, **8:4(35)**, с. 347–357.

URL: http://psta.psiras.ru/read/psta2017_4_347-357.pdf

Об авторе:

Сергей Витальевич Знаменский

Научные интересы мигрировали от функционального анализа и комплексных аналогов выпуклости к основаниям разработки коллаборативного ПО и метрикам сходства

e-mail:

svz@latex.pereslavl.ru

Sergej Znamenskij. *Model and axioms for similarity metrics.*

ABSTRACT. Modern applications usually combine different similarity metrics taking into account the algorithms complexity, the peculiarities of human perception, data resources and samples. The optimization requires a unified formal description of the basic similarity metrics.

The system of the similarity metric axioms is enhanced and its universal model is constructed which generalizes known models of similarity that do not reduce to the Euclidean metric. The model is based on a weighted partially ordered set. (In Russian).

Key words and phrases: similarity of strings, sequence alignment, edit distance, LCS, Levenshtein metric.

2010 *Mathematics Subject Classification:* 97P20; 91C05, 91C20.

References

- [1] Kelly J. C.. “Bitopological spaces”, *Proc. London Math. Soc.*, **13**:3 (1963), pp. 71–89.
- [2] Wilson W. A.. “On quasi-metric spaces”, *Am. J. Math.*, **53**:3 (1931), pp. 675–684.
- [3] Arkhangel'skii A. V., Pontryagin (eds.) L. S., *General Topology I: Basic Concepts and Constructions Dimension Theory*, Encyclopaedia of Mathematical Sciences, vol. **17**, Springer, Berlin–Heidelberg, 1990, vii+202 p.
- [4] C. Alexander. “Semi-developable space and quotient images of metric spaces”, *Pacific J. Math.*, **37**:2 (1971), pp. 277–293.
- [5] S. G. Matthews. “Partial metric topology”, Proc. 8th Summer Conference on General Topology and Applications (Queens College, City University of New York, USA, June 18–22, 1992), *Ann. New York Acad. Sci.*, **728** (1994), pp. 183–197.
- [6] A. Tversky. “Features of similarity”, *Psychol. Rev.*, **84**:4 (1977), pp. 327–352.
- [7] Chen S., Ma B., Zhang K.. “On the similarity metric and the distance metric”, *Theoretical Computer Science*, **410**:24–25 (2009), pp. 2365–2376.

- [8] R. Kunitomo, M. Vogt, J. Bajorath.. “Maximum common substructure-based Tversky index: an asymmetric hybrid similarity measure”, *J. Comput. Aided Mol. Des.*, **30** (2016), pp. 523–531.
- [9] E. Amigo, J. Gonzalo, F. Giner, F. Verdejo. “An Axiomatic Account of Similarity”, SIGIR’17 Workshop on Axiomatic Thinking for Information Retrieval and Related Tasks (ATIR) (August 2017, Tokyo Japan), 3, 10 p., URL: https://www.eecis.udel.edu/~hfang/atir_proceedings/atir17-paper3.pdf
- [10] D. Guessoum, M. Miraoui, Ch. Tadj. “Survey of semantic similarity measures in pervasive computing”, *International journal on smart sensing and intelligent systems*, **8**:1 (2015), pp. 125–158.
- [11] Th. Slimani. “Description and Evaluation of Semantic Similarity Measures Approaches”, *International Journal of Computer Applications*, **80**:10 (2013), pp. 25–33.
- [12] F. Giner, E. Amigó. “General Representation Model for Text Similarity”, Future and Emerging Trends in Language Technology. Machine Learning and Big Data (FETLT 2016) (Seville, Spain, November 30 –December 2, 2016), 2017,, pp. 158–169 (Revised Selected Papers).
- [13] A. Gupta A. Kumar J. Gautam. “A Survey on Semantic Similarity Measures”, *IJIRST –International Journal for Innovative Research in Science & Technology*, **3**:12 (2017), URL: <http://www.ijirst.org/articles/IJIRSTV3I12083.pdf>
- [14] S. V. Znamenskij. “From Similarity to Distance: Axiom Set, Monotonic Transformations and Metric Determinacy”, *Journal of Siberian Federal University. Mathematics & Physics*, 2018 (to appear).
- [15] Dobson A. J.. “Unrooted Trees for Numerical Taxonomy”, *Journal of Applied Probability*, **11**:1 (1974), pp. 32–42.
- [16] S. Jimenez, F.A. Gonzalez, A. Gelbukh. Soft Cardinality in Semantic Text Processing: Experience of the SemEval International Competitions. Polibits [online]. 2015, n.51, pp.63–72. ISSN 1870-9044. <http://dx.doi.org/10.17562/PB-51-9.>, URL: <http://www.scielo.org.mx/pdf/poli/n51/n51a10.pdf>
- [17] N.J.P. van Eck, L. Waltman. How to Normalize Co-Occurrence Data? An Analysis of Some Well-Known Similarity Measures (No. ERS-2009-001-LIS). ERIM report series research in management Erasmus Research Institute of Management. Erasmus Research Institute of Management. (2009), URL: <http://hdl.handle.net/1765/14528>

Sample citation of this publication:

Sergej Znamenskij. “Model and axioms for similarity metrics”, *Program systems: Theory and applications*, 2017, **8**:4(35), pp. 347–357. (In Russian).

URL: http://psta.psiras.ru/read/psta2017_4_347-357.pdf