

АЛГОРИТМЫ ПОИСКА ИЗОБРАЖЕНИЙ ПО СОДЕРЖАНИЮ С ИСПОЛЬЗОВАНИЕМ НЕЧЕТКОГО ГИПЕРГРАФА

Н. Н. Воробжанский

Воронежский государственный университет

Поступила в редакцию 20.05.2016 г.

Аннотация. В работе рассматривается новый подход к поиску и индексированию изображений на основе нечеткого гиперграфа. Предложенный трансдуктивный алгоритм существенно ускоряет поиск изображений и способствует более высокой точности результатов. Преимущество нового подхода продемонстрировано на примере его сравнения с уже существующими методами построения алгоритмов обучения для поиска изображений.

Ключевые слова: поиск изображений по содержанию, гиперграф, ранжирование, трансдуктивный алгоритм поиска.

Annotation. In this article a new approach to image retrieval and indexing based on fuzzy hypergraph is taken under consideration. The proposed transductive algorithm sufficiently increases image retrieval velocity and provides better accuracy of the results. Advantages of the new approach are demonstrated in a comparison between simple graph and hypergraph based transductive learning algorithms.

Keywords: content-based image retrieval, hypergraph, fuzzy hypergraph ranking, transductive learning algorithms.

ВВЕДЕНИЕ

Исследования в области поиска изображений по содержанию (CBIR) стали чрезвычайно актуальны в последние десятилетия в связи с увеличением емкости накопителей информации, распространением цифровой фотографии, а также стремительного роста популярности сети интернет. Методики CBIR позволяют искать изображения, основываясь, главным образом, на их внутренних характеристиках, вместо ключевых слов используется визуальная информация. Обычно в CBIR-системах шаблон изображения предоставляется пользователем, а в результате поиска возвращаются наиболее похожие изображения в соответствии с некоторой мерой сходства [1]. С целью улучшения репрезентативности результатов, многие CBIR-фреймворки используют технику онлайн-обучения под названием *релевантная обратная связь* (RF) [1, 2]. Согласно ей, пользователей просят маркировать изображения в возвращаемых

результатах как релевантные или нерелевантные, и затем поисковая процедура повторяется снова с уже новой информацией. Существует достаточно много исследований по релевантной обратной связи [3, 4], большинство из которых нацелены на изучение способов различения релевантных и нерелевантных изображений, как, например, RF методы, основанные на методе опорных векторов (SVM) [5]. Поскольку категоризированные, т.е. имеющие некое признаковое описание, пользователем изображения недостаточно эффективны в применении к методам машинного обучения с учителем в CBIR-системе, последнее время отдается предпочтение трансдуктивным (или полуавтоматическим) методам поиска изображений [6]. Для моделирования отношений между изображениями во многих работах используются графы. В простом графе, категоризированные и некатегоризированные изображения представляются вершинами, два похожих изображения соединяются ребром, а вес ребра вычисляется согласно некоторой мере сходства. Однако для представления отношений между изо-

бражениями простых графов недостаточно, поскольку необходимо брать в расчёт отношения не только между двумя вершинами, но и тремя или более вершинами, содержащими локальную классифицирующую информацию. Эти аспекты возможно учесть с помощью понятия гиперграфа и его обобщений.

В данной статье предложен основанный на гиперграфе трансдуктивный алгоритм поиска изображений по содержанию, который существенно повышает скорость поиска изображений, также предлагается использовать нечеткий гиперграф, позволяющий не только локально группировать информацию, но и определить степень сходства вершин в ребре. В заключении проводится тщательное сравнение трансдуктивных алгоритмов обучения на простом графе, гиперграфе и нечетком гиперграфе.

1. ИСПОЛЬЗОВАНИЕ ГИПЕРГРАФА ДЛЯ ОЦЕНКИ СХОДСТВА ИЗОБРАЖЕНИЙ

Пусть $V = \{v_1, \dots, v_n\}$ – конечное множество элементов. Гиперграфом [7, 8] на множестве V называется пара множеств $G = (V, E)$, где $E = \{e_1, \dots, e_m\}$ – семейство подмножеств множества V такое, что $e_i \neq \emptyset$ для каждого $i = \overline{1, m}$ и $\bigcup_{i=1}^m e_i = V$. Элементы v_1, \dots, v_n множества V называются вершинами, а элементы e_1, \dots, e_m множества E – гиперребрами гиперграфа.

Заметим, что семейство E образует покрытие множества V .

Простой гиперграф – это гиперграф $G = (V, E)$, в котором

$$e_i \subset e_j \Rightarrow i = j.$$

Гиперграф H является обобщением простого графа и представляет собой множество вершин, соединенных гиперребрами, при этом гиперребро e_j является непрерывной кривой, заключающей внутри себя вершины при $|e_j| \geq 3$.

Гиперграф можно задать матрицей инцидентности $A = \{A(i, j)\}_{n \times m}$, в которой строки соответствуют вершинам v_1, \dots, v_n , а столбцы – гиперребрам e_1, \dots, e_m , при этом

$$A(i, j) = \begin{cases} 0, & v_i \notin e_j, \\ 1, & v_i \in e_j. \end{cases} \quad (1)$$

Взвешенным гиперграфом называется гиперграф $H = (V, E, W)$, в котором каждому гиперребру e_j поставлено число $w(e_j) \in W$, называемое его весом.

К основным числовым характеристикам гиперграфа относятся следующие [7, 8]:

- $n(H)$ – число вершин (порядок гиперграфа H);
- $m(H)$ – число гиперребер;
- $|e_j|$ – мощность гиперребра;
- $r(H) = \max_j |e_j|$ – ранг гиперграфа;
- $s(H) = \min_j |e_j|$ – антиранг гиперграфа;
- $d(v)$ – степень вершины v определяется как количество гиперребер, которые ее содержат.

Если $r(H) = s(H)$, то гиперграф называется однородным. Гиперграф, степени вершин которого равны между собой, называется регулярным.

В постановке задачи поиска изображений по содержанию предполагается, что вершины – являются абстракцией изображений, а гиперребра – абстракцией категорий изображений, при этом вершине v_i ставится в соответствие гиперребро e_j с бинарной точностью $A(i, j) \in \{0, 1\}$. Следовательно, относительное сходство между вершинами не учитывается, что приводит к потере информации, и, тем самым, снижает эффективность поисковой процедуры. Для того чтобы решить проблему этого ограничения вводится понятие нечеткого гиперграфа.

Предположим, что с использованием некоторой меры сходства вычисляется матрица сходства $S = \{S(i, j)\}_{n \times n}$ размерности $|V| \times |V|$, элементы которой $S(i, j) \in [0, 1]$ оценивают сходство i -го и j -го изображений, сопоставленных i -ой и j -ой вершинам графа соответственно. Рассматривая каждую вершину графа v_i как центроиду, определим ее k -ближайших соседей и получим соответствующее гиперребро гиперграфа. Таким образом, мощность ребра в конструкции получается равной $(k + 1)$. В этом случае элементы

матрицы инцидентности A нечеткого гиперграфа определяются следующим образом:

$$A(v, e) = \begin{cases} S(u, v), & v \in e, \\ 0, & v \notin e. \end{cases} \quad (2)$$

Согласно этой формуле, вершина v «эластично» сопоставлена гиперребру e с учетом сходства $S(v, u)$ между v и u , где u – центроида гиперребра e .

Вес $w(e)$ гиперребра e вычисляется следующим образом:

$$w(e) = \sum_{v \in e} S(u, v). \quad (3)$$

Заметим, что, согласно определению, чем больше сходство между вершинами, тем больше вес соответствующего гиперребра.

Для вершины $v \in V$, её степень определяется следующим образом:

$$d(v) = \sum_{e \in E} w(e) A(v, e). \quad (4)$$

Для гиперребра $e \in E$ степень определяется следующим образом:

$$\delta(e) = \sum_{v \in V} A(v, e). \quad (5)$$

Через D_v, D_e, W обозначим диагональные матрицы степеней вершин, степеней гиперребер и весов гиперребер соответственно.

2. АЛГОРИТМЫ РАНЖИРОВАНИЯ ГИПЕРГРАФА

Гиперребра могут пересекаться друг с другом, а это означает, что ряд вершин может попасть сразу в несколько разных категорий. Чтобы минимизировать это явление необходимо решить проблему разбиения гиперграфа, т. е. разделить все множество вершин гиперграфа на независимые друг от друга, непересекающиеся подмножества. Существующие подходы к разбиению гиперграфа условно можно разбить на две группы. Первая группа включает алгоритмы, основанные на конструировании простого графа из оригинального гиперграфа, а затем разбиения вершин с помощью спектральных кластерных методик. Вторая группа подходов основана на определении лапласиана гиперграфа, используя аналогии с лапласианом простого графа (лапласиан Болла [9], нормализованный лапласиан Жоу [10] и др.).

Для решения проблемы разбиения гиперграфа можно ввести нормализованную функцию $\Omega(f)$ ошибки обучения, определяемую следующим образом [10]:

$$\Omega(f) = \frac{1}{2} \sum_{e \in Eu, v \in V} \frac{w(e) A(u, e) A(v, e)}{\delta(e)} \times \left(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}} \right)^2, \quad (6)$$

где вектор $f(v)$ – это признаковое описание изображения для обучения.

При минимизации функции ошибки обучения, вершины, делящие между собой множество инцидентных гиперребер, гарантированно получают одни и те же признаковые пометки. Определив $\Theta = D_v^{-\frac{1}{2}} A W D_e^{-1} A^T D_v^{-\frac{1}{2}}$, можно привести выражение (6) к следующему виду:

$$\begin{aligned} \Omega(f) &= \sum_{e \in Eu, v \in V} \frac{w(e) A(u, e) A(v, e)}{\delta(e)} \times \\ &\times \left(\frac{f^2(u)}{d(u)} - \frac{f(u) f(v)}{\sqrt{d(u) d(v)}} \right) = \\ &= \sum_{u \in V} f^2(u) \sum_{e \in E} \frac{w(e) A(u, e)}{d(u)} \sum_{v \in V} \frac{A(v, e)}{\delta(e)} - \\ &- \sum_{e \in Eu, v \in e} \frac{f(u) A(u, e) w(e) A(v, e) f(v)}{\sqrt{d(u) d(v)} \delta(e)} = \\ &= f^T (I - \Theta) f, \end{aligned} \quad (7)$$

где I – единичная матрица.

Заметим, что $\Omega(f) = f^T (I - \Theta) f$ тогда и только тогда, когда $\sum_{v \in V} \frac{A(v, e)}{\delta(e)} = 1$ и $\sum_{e \in E} \frac{w(e) A(u, e)}{d(u)} = 1$, что является верным из определений $\delta(e)$ и $d(u)$. При этом

$$\Delta = I - \Theta \quad (8)$$

– это положительно определенная матрица – лапласиан гиперграфа и тогда $\Omega(f) = f^T \Delta f$.

Данная функция ошибки обучения задается формулой, похожей на формулу нормализованной функции ошибки обычного графа $G_s = (V_s, E_s)$:

$$\begin{aligned} \Omega_s(f) &= \frac{1}{2} \sum_{v_i, v_j \in V_s} A_s(i, j) \left(\frac{f(i)}{\sqrt{D_{ii}}} - \frac{f(j)}{\sqrt{D_{jj}}} \right)^2 = \\ &= f^T \left(I - D^{-\frac{1}{2}} A_s D^{-\frac{1}{2}} \right) f = f^T \Delta_s f, \end{aligned} \quad (9)$$

где D – это диагональная матрица, в которой её (i, i) – элемент равен сумме i -ой строки матрицы сходства A_s . При этом

$$\Delta_s = I - \Theta_s = I - D^{-\frac{1}{2}} A_s D^{-\frac{1}{2}} \quad (10)$$

является лапласианом обычного графа.

В структуре модели обучения без учителя формулы (6) и (3) могут быть оптимизированы с помощью собственного вектора, относящегося к наименьшему ненулевому собственному значению Δ и Δ_s соответственно [10].

В трансдуктивной модели обучения [10] определяется вектор y для описания признаковой категоризирующей информации изображений, поступивших через обратную связь, и присвоения её начальных значений соответствующим элементам y по правилу

$$y(v) = \begin{cases} \frac{1}{|Pos|}, & v \in Pos, \\ \frac{1}{|Neg|}, & v \in Neg. \end{cases} \quad (11)$$

Если v некатегоризировано, то $y(v) = 0$. Для того, чтобы присвоенные признаковые пометки подходили под изначальную признаковую информацию y , вводится регуляризационное выражение, определяемое следующим образом:

$$\|f - y\|^2 = \sum_{u \in V} (f(u) - y(u))^2. \quad (12)$$

После того, как информация обратной связи будет представлена, задача обучения сводится к минимизации следующей функции:

$$\Phi(f) = f^T \Delta f + \mu \|f - y\|^2, \quad (13)$$

где $\mu > 0$ – это параметр управления обучением. Находим минимум $\Phi(f)$:

$$f = (1 - \xi)(I - \xi\Theta)^{-1} y, \quad (14)$$

где $\xi = \frac{1}{1 + \mu}$.

Доказано, что решения этого уравнения совпадают с решениями линейной системы

$$((1 + \mu)I - \Theta) f = \mu y. \quad (15)$$

В случае простого графа можно просто заменить Θ на Θ_s для того, чтобы выполнить трансдуктивное обучение. В [3, 4] простой граф, основанный на трансдуктивной методике, используется для поиска изображений с релевантной обратной связью. Рассмотрим алгоритмы ранжирования нечеткого гиперграфа.

Алгоритм 1. Ранжирование нечеткого гиперграфа

1. Построить матрицу сходства S , основанную на разных дескрипторах свойств, используя формулу (17), где $S(v, u)$ обозначает сходство между вершинами v и u .

2. Построить нечеткий гиперграф H . Для каждой вершины, основываясь на матрице сходства S , выбрать k ближайших соседей и сформировать гиперребро.

3. Построить матрицу инцидентности A гиперграфа по формуле (2). Матрица весов гиперребер вычисляется с помощью формулы (3).

4. Вычислить лапласиан гиперграфа по формуле (8).

5. Для данной вершины и начального признакового описания вектора y , решить линейную систему (15) относительно f . Упорядочить вершины в соответствии с убыванием r (18).

Алгоритм 2 Множественное ранжирование (Manifold Ranking)

1. Построить матрицу сходства S по аналогии с Алгоритмом 1.

2. Построить простой граф G_s . Основываясь на матрице сходства S , каждую вершину соединить с k -ближайшими соседними вершинами.

3. Построить матрицу сходства A_s простого графа такую, что $A_s(i, j) = S(i, j)$, если i и j вершины соединены между собой. Положить $A_s(i, i) = 0$. Построить матрицу D степеней вершин $D_{ii} = \sum_j A_s(i, j)$.

4. Вычислить лапласиан графа G_s по формуле $\Delta_s = I - \Theta_s = I - D^{-\frac{1}{2}} A_s D^{-\frac{1}{2}}$.

5. Решить линейную систему (15) по аналогии с Алгоритмом 1, но Θ заменить на Θ_s .

Упорядочить вершины в соответствии с убыванием r (18).

ДЕСКРИПТОРЫ СВОЙСТВ И МЕРЫ СХОДСТВА

Для определения меры сходства между двумя изображениями было решено использовать следующие дескрипторы: SIFT [11], OpponentSIFT, rgSIFT, C-SIFT, RGB-SIFT [12] и HOG [13]. Известно, что их комбинация даёт наивысшую точность на разных множествах изображений [2]. В соответствии с [10], необходимо извлечь низкочастотные (разреженные) и высокочастотные (плотные) свойства для пяти дескрипторов, для того чтобы улучшить поиск изображений. Для каждого из 11 свойств (5 разреженных + 5 высокочастотных + 1 HOG свойство) используется метод пространственной пирамиды сопоставлений (SPM) для расчёта расстояний между изображениями $\{v, u\}$

$$Dis(v, u) = \sum_{l=0}^L \frac{1}{\alpha^l} \sum_{p=1}^{m(l)} \beta_p^l \chi^2(His_p^l(v), His_p^l(u)). \quad (16)$$

В представленном выше уравнении, $His_p^l(v)$, $His_p^l(u)$ – это две локальные гистограммы изображения на p -ой позиции уровня l , α , β – взвешенные параметры, $\chi^2(\dots)$ – хи-квадратная функция расстояния, используемая для измерения расстояния между двумя гистограммами.

После того как все матрицы расстояний для 11 свойств получены, матрица сходства S между двумя изображениями может быть вычислена следующим образом:

$$S(v, u) = \exp\left(-\frac{1}{11} \sum_{k=1}^{11} \frac{Dis_k(v, u)}{\bar{D}_k}\right), \quad (17)$$

где \bar{D}_k – среднее значение элементов в k -ой матрице.

3. РЕЗУЛЬТАТЫ ВЫЧИСЛИТЕЛЬНОГО ЭКСПЕРИМЕНТА

Возьмем за отправную точку SVM и ранжирование, основанное на сходстве. Метод ранжирования классифицирует найденное изображение v по следующей формуле:

$$r = \frac{1}{|Pos|} \sum_{u \in Pos} S(v, u) - \frac{1}{|Neg|} \sum_{k \in Neg} S(v, k), \quad (18)$$

где r – число ранжирования, а Pos и Neg обозначают релевантные/нерелевантные множества изображений, поступивших по обратной связи, соответственно.

Целью эксперимента является сравнение скорости выполнения ранжирования нечеткого и обычного гиперграфов. В качестве базы данных изображений используется Corel5K [3, 4].

Ранжирования гиперграфа выполняется с помощью Алгоритма 1 с использованием бинарной матрицы инцидентности.

Для оценки производительности алгоритмов используются следующие оценки качества:

- 1) Кривая точности и масштаба (precision-score curve);
- 2) Кривая точности и полноты (precision-recall curve).

Каждое изображение в базе данных используется как тестовый запрос и обе оценки качества усредняются по результатам всех запросов.

Эксперимент тестирования релевантной связи организован следующим образом:

0 цикл: производится поиск изображений по шаблону.

1 цикл: из результатов произвольно отбираются 4 релевантных запросу изображения и 5 нерелевантных с тем, чтобы сформировать обучающую выборку, содержащую 5 релевантных / 5 нерелевантных примеров. Поиск производится повторно.

2 цикл: из результатов поиска с 1 цикла отбираются случайным образом другие 5 релевантных / 5 нерелевантных примеров для обучения. Процедура поиска снова запускается.

3 цикл: из результатов поиска со 2 цикла отбираются случайным образом новые 5 релевантных / 5 нерелевантных примеров для обучения. Процедура поиска запускается в последний раз.

Оставшиеся изображения используются в качестве данных для тестирования. Полученные в результате взаимодействия с релевантной обратной связью, после каждого

цикла, изображения готовы для внесения их в структуру гиперграфа и обработки алгоритмами ранжирования гиперграфов.

Помимо описанной выше техники пассивного обучения также была исследована методика активного обучения на *Corel5K*. Так как каждая из 50 категорий *Corel5K* содержит 100 изображений, кривая точности и масштаба используется в [3, 4] в качестве оценки качества. Именно поэтому эта оценка качества алгоритма была выбрана для анализа, чтобы произвести прямое сравнение с [3, 4].

Сравнение при пассивном обучении.

Сравним методы при пассивном варианте обучения. Как показано на рис. 1, ранжирование нечеткого графа превосходит множественное ранжирование на 4–5 %, а ранжирование традиционного гиперграфа – на 1–3 % после каждого цикла релевантной обратной связи.

Сравнение при активном обучении.

Перейдем к сравнению методик при активном варианте обучения. Как показано на рис. 2, эксперимент начинается с цикла 0, в котором только изображение-запрос используется для поиска. Хотя ранжирование нечеткого гиперграфа достигло приблизительно такой же точности, как и множественное ранжирование, а также ранжирование традиционного гиперграфа в нулевом цикле (т. е. без обратной связи), разница в точности после первого и второго цикла становится ощутима. В конце второго цикла оно превосходит множественное ранжирование на 4–10 %, а традиционное – на 1–2,5 %.

Результаты, полученные от ранжирования нечеткого гиперграфа превосходят результаты в [3, 4] множественного ранжирования приблизительно на 8–20 %.

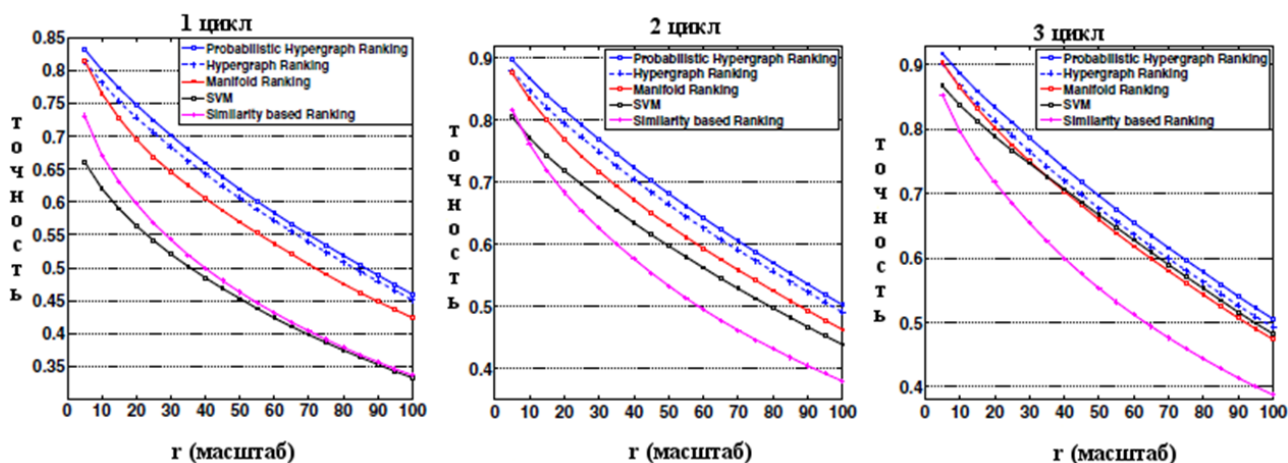


Рис. 1. Кривые точности и масштаба. Пассивное обучение

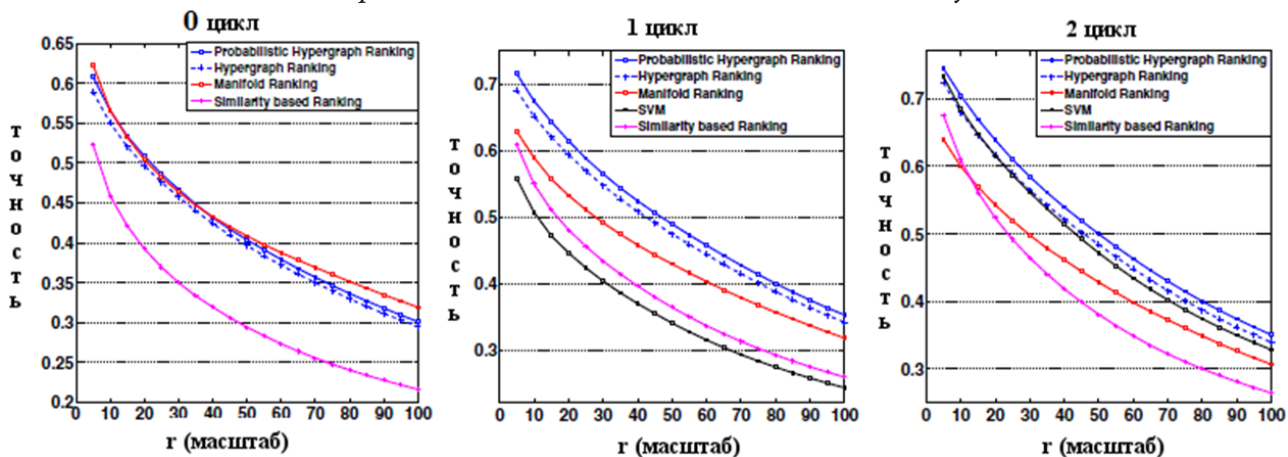


Рис. 2. Кривые точности и масштаба. Активное обучение

ЗАКЛЮЧЕНИЕ

В данной статье был представлен трансдуктивный метод обучения для задачи поиска изображения по содержанию, в котором нечеткий гиперграф используется для описания отношений релевантности между вершинами (изображениями). Каждое изображение рассматривалось как вершина-центроида и, с помощью матрицы сходства, построенной на основе свойств изображений, помещалось вместе с k -ближайшими вершинами внутрь гиперребра. Таким образом, задача поиска изображений с релевантной обратной связью свелась к проблеме трансдуктивного обучения, которая решается алгоритмом ранжирования гиперграфа. Эффективность предложенного метода была продемонстрирована экспериментально на базе данных изображений Corel5K.

СПИСОК ЛИТЕРАТУРЫ

1. He X., Ma W.-Y., King O., Li M., Zhang H. Learning and inferring a semantic space from user's relevance feedback for image retrieval // ACM MULTIMEDIA. – 2002.
2. Rui Y., Huang T. S., Chang S.-F. Image retrieval: Current techniques, promising directions, and open issues // Journal of Visual Communication and Image Representation. – 1999. – vol. 10, no. 1. – pp. 39–62.
3. He J., Li M., Zhang H., Tong H., Zhang C. Generalized manifold-ranking-based image

retrieval // IEEE transaction on Image Processing. – 2006. – vol. 15, no. 10. – pp. 3170–3177.

4. He J., Li M., Zhang H.-J., Tong H., Zhang C. Manifold ranking based image retrieval // ACM MULTIMEDIA. – 2004.

5. Tong S., Chang E. Support vector machine active learning for image retrieval // ACM MULTIMEDIA. – 2001.

6. Hoi S. C. H., Lyu M. R. A semi-supervised active learning framework for image retrieval // CVPR. – 2005.

7. Емеличев В. А., Мельников О. И., Сарванов В. И., Тышкевич Р. И. Глава XI: Гиперграфы // Лекции по теории графов. – М.: Наука, 1990. – С. 298–315. – 384 с.

8. Зыков А. А. Гиперграфы // Успехи математических наук. – 1974. – № 6 (180).

9. Bolla M. Spectra, Euclidean representations and clustering of hypergraphs // Discrete Mathematics. – 1993.

10. Zhou D., Huang J., Schokopf B. Learning with hypergraphs: Clustering, classification, and embedding // NIPS. – 2006.

11. Lowe D. Object recognition from local scale-invariant features // ICCV. – 2009.

12. Van de Sande K. E. A., Gevers T., Snoek C. G. M. Evaluating color descriptors for object and scene recognition // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2010.

13. Dalal N., Triggs B. Histograms of oriented gradients for human detection // CVPR. – 2005.

Воробжанский Н. Н. – аспирант кафедры вычислительной математики и прикладных информационных технологий, факультет прикладной математики, информатики и механики, Воронежский государственный университет.
E-mail: vorobzh@gmail.com

Vorobzhanskiy N. N. – PG student of the Department of Computational Mathematics and Applied Information Technologies, Applied Mathematics, Informatics and Mechanics Faculty, Voronezh State University.
E-mail: vorobzh@gmail.com