

Варламов А.Д., Варламова Е.В.

*Муромский институт (филиал) федерального государственного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264, г. Муром, Владимирская обл., ул. Орловская, 23
E-mail: varlamov_aleks@mail.ru, karetko22@mail.ru*

Использование оценок ассессоров для машинного обучения поиска изображений по визуальному образцу

В основе алгоритмов поиска изображений по визуальному образцу (другому изображению) лежит оценка близости двух изображений. Эта мера рассчитывается формальным путем на основе математических методов. Ее вычисление может быть выполнено автоматически и, при определенных условиях организации поиска и индексирования изображений, достаточно быстро [1]. Ключевой задачей становится достижение как можно более высокого качества поиска на основе этой меры. Поэтому ее следует выбрать из различных математических процедур и оптимизировать, подбирая параметры алгоритма ее вычисления [2-4]. С этой целью следует применять методы машинного обучения.

В задачах машинного обучения часто решающее значение имеет правильное формирование обучающей выборки [5]. В рассматриваемой задаче выборка представляет собой совокупность прецедентов, каждый из которых содержит пару изображений и экспертную оценку их визуального сходства. Сравнение образцов в отобранной коллекции визуальных данных для обучающего набора выполняют ассессоры - специально нанятые для этой работы люди.

Работа ассессора состоит в следующем: ему последовательно предоставляются пары изображений, а он определяют степень подобия в каждой из них. Ассессоры нужны на этапе разработки системы поиска изображений, и не востребованы при ее эксплуатации.

Так как решение человека, визуально воспринимающего информацию, всегда субъективно, может отличаться от мнения большинства людей, а также, учитывая человеческий фактор допущения ошибки, часто вместо одного ассессора к работе приглашают группу людей. При этом могут использоваться следующие три способа сведения оценок:

1. логическое ИЛИ (or) (если хотя бы один из ассессоров признал пару тематически близкой, она признается близкой, иначе не близкой);
2. логическое И (and) (если все ассессоры признали пару тематически близкой, она признается близкой, иначе не близкой);
3. простое большинство (vote) (если большинство ассессоров признало пару тематически близкой, она признается близкой, иначе не близкой).

При таком агрегировании результатов, качество работы ассессоров не проверяется и не поощряется добросовестная работа оценщиков. Кроме того, оценки, выставленные надлежаще выполнившими работу экспертами, имеют такое же влияние, как и оценки их коллег, отработавших не на совесть. Поэтому мы предлагаем дополнить процедуру оценки следующими этапами:

1. вычисление коэффициентов корреляции между итоговой оценкой и оценкой каждого ассессора;
2. пересчет итоговых оценок (по методу vote) с весовыми коэффициентами влияния на результат, равными рассчитанным коэффициентам корреляции.

Таким образом, с уверенностью можно говорить, что предложенный подход повысит ответственность ассессоров в работе и уменьшит влияние на результат человеческого фактора.

Литература

1. Шарапов Р.В., Варламов А.Д. Сравнительный анализ систем поиска графических данных // Современные наукоемкие технологии. 2013. № 1. С. 27-31.
2. Варламов А.Д., Шарапов Р.В. Поиск визуально подобных изображений на основе машинного обучения. // Электронные библиотеки: перспективные методы и технологии,

электронные коллекции: Труды XIV Всероссийской научной конференции RCDL'2012. - г. Переславль-Залесский: изд.-во "Университет города Переславля", 2012, с. 152-159.

3. Varlamov A., Sharapov R. Machine Learning of Visually Similar Images Search // CEUR Workshop Proceedings. 2012. Vol. 934, pp. 113-120.

4. Соченков И.В., Вохминцев А.В., Кузнецов В.В., Хромов Н.А., Григорьева К.С. Поиск изображений по визуальному подобию с применением инвертированных индексов цветовых гистограмм. // Информационные технологии и вычислительные системы. 2015. № 4. С. 86-94.

5. Кафтанников, И.Л. Проблемы формирования обучающей выборки в задачах машинного обучения / И.Л. Кафтанников, А.В. Парасич // Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника». - 2016. - Т. 16, № 3. - С. 15-24. DOI: 10.14529/йсг160302