ПРОТОКОЛЫ ОЦЕНКИ СИСТЕМ МАШИННОГО ЗРЕНИЯ

И. И. Фролов
Кафедра электронных вычислительных машин, Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: frolov@bsuir.by

Современный этап развития систем машинного зрения позволяет разделить создаваемые решения на несколько категорий: исследовательские работы по созданию отдельных алгоритмов (обработка изображений, машинное обучение), прикладные и учебные системы, построенные из известных алгоритмов, и промышленные системы, которые могут использовать как собственные разработки, так и известные методики, но делающие серьезный акцент на системное решение в целом. Каждый из перечисленных типов систем нуждается в объективном тестировании для оценки качества и возможности применения для решаемой задачи.

Введение

Разработка и внедрение систем машинного зрения содержит в себе этапы получения исходных данных, преобразования и предобработки полученных данных к формату, совместимому с алгоритмами машинного обучения на следующем этапе системы - распознаванию. На практике биометрическое распознавание с помощью систем машинного зрения выполняется для решения задач:

- идентификации;
- верифакации.

Отдельно рассматриваются системы видеонаблюдения и анализа видео-последовательностей, а также системы оценки положения (позы) человека.

I. Оценка алгоритмов машинного обучения

Зачастую при оценке качества алгоритма машинного обучения принимают только отношение правильно распознанных образцов к общему объему коллекции. Тогда как более глубокий подход предполагает использование не только базовых статистических оценок, но и специальных протоколов.

Некоторые сервисы оценки алгоритмов машинного обучения описаны в работе [1], однако приведенные системы оценки направлены на выполнение тестирования задач классификации, при этом в перечисленных [1] системах (MLcomp, TunedIt и Полигон) не достаточно представлена информация о протоколах тестирования, тогда как согласно разным методикам можно получить отличающиеся данные для построения оценок, достаточных для принятия решения по определенному алгоритму машинного обучения (разложение ошибки на смещение и вариацию, Roc-кривая, карта ошибок, распределение отступов, кривая обучения, кривая переобученности, распределение стандартной ошибки, распределение переобученности).

Как правило, большинство протоколов предполагает использование подхода кросс-валидации (cross validation), предполагающего последовательное разбиение выборки на неповторяющиеся тестовые и обучающие наборы, что повышает достоверность статистических оценок. Кроме того, для получения сопоставимых результатов используются идентичные наборы данных – данное правило справедливо не только для тестирования алгоритмов машинного обучения, но и систем машинного зрения в целом.

Наиболее популярными протоколами, определяющими порядок тестирования, вычисление основных характеристик по результатам тестирования, и даже тестовые наборы (для распознавания лиц) являются FERET [2], XM2VTSDB[3], Biometric Testing Best Practices [4], Face Recognition Vendor Test (FRVT) [5] и другие. Основными статистическими характеристиками, вычисляемыми по результатам тестирования являются дисперсия, доверительный интервал, ошибки I и II рода, которые затем используются для построения ROC-кривых и других перечисленных оценок.

Так, например, протокол XM2VTSDB, специально созданный и используемый для оценки производительности фото- и видеосистем аутентификации личности. Протокол определен для задачи верификации, когда человек заявляет права на доступ и свою принадлежность к списку людей, которым разрешен доступ. Система верификации сравнивает признаки распознаваемого человека с биометрическими характеристиками, сохраненными в системе и соответствующими личности, на принадлежность к которой были заявлены права, и вычисляет меру близости (оценку) распознаваемого субъекта к заявленному разрешенному пользователю. В зависимости от рассчитанного значения оценки меры близости система принимает решение о правомерности (истинности) сопоставления (принадлежности) распознаваемого субъекта с разре-

шенным пользователем. Задача аутентификации соответствует сценарию тестирования открытой системы, когда личности, неизвестные системе могут запросить права доступа. Субъекты, чьи признаки хранятся в системе, имеют право доступа и называются клиентами, тогда как люди, неправомерно выдающие себя за клиентов и требующие доступ к объекту, именуются мошенниками.

Оценка работы системы выполняется по расчету ошибок I и II рода. К ошибкам I рода (False Rejection rate, далее – FRR) относятся ошибки типа «отказ в правомерном доступе клиенту», а к ошибкам II рода (False Acceptance rate, далее – FAR) относятся, соответственно, ошибки типа «предоставление доступа мошеннику». Классификация ошибок приведена в соответствии с терминологией протокола [3]. Расчет ошибок FAR и FRR выполняется по следующим формулам:

$$FAR = \frac{EM}{M} * 100$$

$$FRR = \frac{EK}{K} * 100$$

где EM – число предоставлений доступа мошенникам; M – общее число попыток получения доступа мошенниками; EK – число отказов в доступе клиентам; K – общее число попыток получения доступа клиентами. Предполагается, что мошенник пытается получить доступ, используя поочередно все идентификаторы клиентов, т.е. выдавая себя за нового клиента каждую новую попытку. Таким образом, общее число попыток получения доступа мошенниками в рамках тестирования вычисляется по формуле

$$M = M_S * N_M * K_S$$

где M_S – общее количество мошенников; N_M – общее число попыток каждого мошенника выдать себя за одного из клиентов (по одной попытке); K_S – общее количество клиентов в базе. Для клиентов значение K вычисляется по формуле

$$K = K_S * N_K$$

где N_K – общее число попыток каждого клиента получить доступ (по одной попытке).

Таким образом, исследователь или прикладной разработчик должен руководствоваться признанными протоколами оценки алгоритмов машинного обучения для тестирования разработанного или существующего алгоритма для получения общепризнанных оценок.

II. Оценка системы машинного зрения

Для тестирования системы машинного зрения необходимо получить оценку эффективности распознавания, которая зависит не только от выбранного алгоритма, но и от остальных блоков системы. Вопрос оценки конкретной биометрической системы необходимо рассматривать в контексте решаемой задачи с учетом следующих требований:

- производительность (решение допускает временные задержки или имеет жесткие ограничения по времени работы, а также требования к распараллеливанию вычислений и обработке информации);
- точность (важен как процент правильно принятых решений, так и ограничение по допустимому уровню ошибок);
- локализация оборудования (обработка данных и принятие решения выполняются на стороне оборудования захвата входной информации или допускается клиент-серверная архитектура).
- условия эксплуатации системы (отрицательные температуры, вибрации, повышенная влажность, изменяющиеся условия освещенности).
- требованя к обеспечению мобильным контролем и управлением системой (в том числе с помощью мобильных устройств).
- требования по передаче и хранению фото-видеоинформации (в том числе и с использованием облачных технологий). При оценки эффективности распознавания системы могут использоваться приведенные алгоритмы машинного зрения, тогда как остальные параметры должны удовлетворять требованиям разрабатываемой системы.

III. Заключение

Как видно, на данном этапе на передний план в промышленной эксплуатации выходит не только эффективность распознавания, которая может быть сопоставимой между сравниваемыми системами, а именно удобство эксплуатации целостного решения, возможность его поддержки и масштабирования, т.е. именно архитектурная составляющая.

IV. Список литературы

1. Системы тестирования алгоритмов машинного обучения: MLcomp, TunedIt и Полигон / А. В. Лисица [и др.] // Интеллектуализация обработки информации ИОИ-8: материалы международной конференции. – Кипр, г.Пафос. – 17–24 октября 2010. – М.: МАКС Пресс. – 2010. –С. 157–160.
2. FERET face database [Electronic resource] // Mode of access : http://www.nist.gov/itl/iad/ig/colorferet.cfm. – Date of access : 12.09.2015.
3. Messer, K. XM2VTSDB : The extended M2VTS database / K. Messer [etc] // Proceedings of Second International Conference on Audio and Video-based Biometric Person Authentication. – 1999.
4. Mansfield, A. J. Best Practices in Testing and Reporting Performance of Biometric Devices, Version 2.01. Technical Report NPL Report CMSC 14/02 // A. J. Mansfield, J. L. Wayman. – Teddington. – National Physical Laboratory. – Aug., 2002.
5. Face Recognition Vendor Test (FRVT) [Electronic resource] // Mode of access : http://www.nist.gov/itl/iad/ig/frvt-home.cfm. – Date of access : 12.09.2015.