УДК 004.932, 004.855

ИСПОЛЬЗОВАНИЕ МНОЖЕСТВЕННЫХ ПРЕДСТАВЛЕНИЙ ВИДЕОИНФОРМАЦИИ В СИСТЕМАХ АВТОМАТИЧЕСКОГО АНАЛИЗА ИЗОБРАЖЕНИЙ

А. С. Потапов, доктор техн. наук

Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург

E-mail: pas.aicv@gmail.com

Проведен анализ необходимости одновременного использования многих представлений информации в системах обработки и анализа изображений. Исследовано различие между критериями Колмогоровской сложности и алгоритмической вероятности при решении задач индукции и принятия решений. Показано, что принятие оптимальных решений (например, в задачах распознавания или прогнозирования) требует использования многих представлений информации, в рамках которых конструируются альтернативные описания изображений. Выведен критерий репрезентационной алгоритмической вероятности для определения оптимального набора представлений по заданной выборке изображений.

Ключевые слова: *анализ изображений, представление информации, теория информации, алгоритмическая сложность.*

Коды OCIS: 150.1135

Поступила в редакцию 29.05.2012

Введение

Понятие представления является центральным для области обработки изображений. В явном виде это понятие введено Д. Марром. Под представлением он понимал формальную систему, содержащую алгоритмы, позволяющие получить описание объектов заданного класса [1]. Несмотря на такое четкое определение, до недавнего времени понятие представления оставалось не полностью формализованным. Как результат, возникала проблема оценки качества представлений, обоснования выбора того или иного представления, их оптимизация.

Проблема оценки качества представлений была теоретически решена в рамках принципа репрезентационной минимальной длины описания (РМДО) [2]. Данный принцип является расширением на случай массовых проблем (проблем, требующих нахождения единого алгоритма решения всех задач некоторого класса) принципа минимальной длины описания (МДО), дающего общее решение индивидуальных задач индуктивного вывода. Принцип РМДО дает общее теоретическое решение оценки качества представлений и выбора наилучшей модели в рамках используемого представления, однако при условии, что может использоваться только одно представление данных.

Хорошо известно [3], что принятие промежуточных решений (в частности, в задачах распознавания изображений) делает невозможным принятие оптимального конечного решения. Например, выбор наиболее вероятной гипотезы при асимметричной матрице потерь может не согласовываться с принятием решения, минимизирующего математическое ожидание потерь. Если пропуск цели несет более значительные потери, чем ложная тревога, то мы предпочтем чаще ошибаться, поднимая ложную тревогу, нежели минимизировать число ошибок (правильность распознавания того, является ли объект целью).

В этой связи задачей индуктивного вывода можно считать не выбор одной лучшей гипотезы о содержании данных, но назначение вероятностей многим гипотезам, чтобы на основе этих вероятностей далее проводился анализ или принимались решения, которые минимизируют риски, исходя из прагматического критерия конечной задачи. Такое понимание задачи индукции (в частности, построения описаний изображений) достаточно распространено. Однако принцип РМДО показывает, что выбор наилучшего представления также сводится к задаче индукции, в связи с чем возникает вопрос, насколько оправданно выбирать одно лучшее представление, или необходимо использовать многие представления, но учитывать их с разными весами. В данной работе дается теоретическое решение этой проблемы.

Принцип РМДО

Пусть имеется строка исходных данных f, в качестве которых здесь рассматривается изображение, и необходимо определить наилучшую модель источника этих данных. Если никакой другой информации не дано, то универсальное решение такой задачи индукции заключается в том, что множество моделей определяется как множество алгоритмов (заданных, например, в форме программ p для универсальной машины Тьюринга (УМТ) U, порождающих f: $U(p) = f$, а лучшей считается модель с наименьшей длиной [4]. Длина такой наикратчайшей программы называется алгоритмической (Колмогоровской) сложностью строки f

$$K(f) = \min_p \left[l(p) \middle| U(p) = f \right].$$

Разделение программы p на регулярную и случайную составляющие приводит к принципу минимальной длины описания. Нас здесь, однако, структура программы p интересовать не будет. Важнее то, что задачи анализа изображений являются массовыми проблемами, то есть для разных изображений некоторой предметной области необходимо независимо одним и тем же алгоритмом осуществлять анализ. При этом в разных изображениях содержится взаимная информация, поэтому верно неравенство $K(f_1 f_2 \dots f_n) \ll K(f_1) + \dots + K(f_n)$, т. е. независимое описание изображений оказывается неэффективным без использования априорной информации. При этом можно показать [2], что

$$\min_S \left[K(S) + \sum_{i=1}^{n} K(f_i|S) \right] \approx K(f_1 f_2 \dots f_n), \quad K(f|S) = \min_\mu \left[l(\mu) \middle| U(S\mu) = f \right]. \tag{1}$$

Следовательно, для данной выборки изображений можно найти такую программу S для УМТ, что для каждого изображения f_i будет существовать его описание μ_i, по которому программа S воспроизведет это изображение (очевидно, такая программа будет удовлетворять определению представления по Марру), причем сумма длин описаний (с учетом сложности самой программы, или представления S) будет близка к Колмогоровской сложности конкатенации всех изображений f_i. Иными словами, представление тем лучше, чем меньше в его рамках суммарная длина описаний изображений, что и составляет принцип репрезентационной МДО.

Колмогоровская сложность и алгоритмическая вероятность

В рамках алгоритмической теории информации дается критерий выбора наилучшей модели. Однако, как отмечалось, при принятии решений необходимо учитывать разные модели с соответствующими им вероятностями. Поскольку Колмогоровская сложность играет роль количества информации (выражаемого в битах, если программы p кодируются бинарными строками), а количество информации равно минус логарифму по основанию два от вероятности, можно, обратив эту зависимость, получить следующее выражение для априорной вероятности некоторой строки f

$$P_K(f) = 2^{-K(f)} = 2^{-\min\limits_{p:U(p)=f}(l(p))} = \max_{p:U(p)=f} 2^{-l(p)}.$$

Однако строка f может быть порождена разными программами, с учетом которых корректнее использовать понятие алгоритмической вероятности [4]

$$P_{ALP}(f) = \sum_{p:U(p)=f} 2^{-l(p)}. \tag{2}$$

В частности, именно алгоритмическая вероятность дает теоретически оптимальное решение задачи предсказания (то есть предсказание должно выполняться путем взвешивания предсказаний по всем моделям с учетом их вероятностей) [4]. В случае, когда многие модели имеют близкие значения Колмогоровской сложности, использование единственной модели, выбранной по принципу МДО, оказывается не вполне адекватным [5, 6].

Очевидно, выполняется соотношение $P_K(f) \le P_{ALP}(f)$, откуда следует интересный факт: $-\log_2 P_K(f) = K(f) \ge -\log_2 P_{ALP}(f)$, то есть Колмогоровская сложность превосходит количество информации в строке f, определенное по алгоритмической вероятности. В этом смысле, хотя одновременное описание некоторых данных f множеством альтернативных моделей выглядит громоздким, оно формально по критерию количества информации лучше, чем использование одной простейшей модели.

Расширение принципа РМДО на основе алгоритмической вероятности

Поскольку понятие представления и критерий его оптимальности возникают при попытке факторизации Колмогоровской сложности конкатенации выборки изображений (или других информационных объектов), а сама Колмогоровская сложность может рассматриваться лишь как аппроксимация алгоритмической вероятности, то необходимо пересмотреть и уточнить принцип РМДО. Зададимся вопросом, можно ли алгоритмическую вероятность конкатенации нескольких изображений представить в факторизованном виде, аналогичном (1)?

Для этого каждую программу p, порождающую весь данный набор изображений $f_1...f_n$, неким произвольным образом разобьем на две части $p = S\mu$. Пока смысл этих частей не имеет существенного значения, но подразумевается, что S – это представление изображений, а μ – их совместное описание в рамках этого представления. Тогда выражение для алгоритмической вероятности можно преобразовать следующим образом

$$P_{ALP}(f_1\ldots f_n)=\sum_{p:U(p)=f_1\ldots f_n}2^{-l(p)}=$$
$$=\sum_{S\mu:U(S\mu)=f_1\ldots f_n}2^{-l(S)-l(\mu)}=$$
$$=\sum_S 2^{-l(S)}\sum_{\mu:U(S\mu)=f_1\ldots f_n}2^{-l(\mu)}=$$
$$=\sum_S 2^{-l(S)}P_{ALP}(f_1\ldots f_n|S).$$

При этом следует иметь в виду, что все пары $S\mu$ должны соответствовать уникальным строкам, чтобы при расчете алгоритмической вероятности эти строки не учитывались несколько раз.

К примеру, при пустом представлении S мы уже получим равенство вероятностей $2^{-l(S)}$ $P_{ALP}(f_1...f_n|S) = P_{ALP}(f_1...f_n)$, поэтому учет какого-либо дополнительного представления, помимо пустого, даст противоречие (алгоритмическая вероятность не равна сама себе). Чтобы избежать противоречия, проще проводить суммирование по некоторому набору представлений (программ для УМТ), среди которых ни одно не является префиксом другого и $\sum_S 2^{-l(S)}=1$ (либо при вычислении вероятностей $P_{ALP}(f_1...f_n|S)$ избегать того, чтобы разные $S\mu$ оказались идентичны).

Нас интересуют лишь такие представления, в рамках которых условная алгоритмическая вероятность может быть факторизована. Таким образом, можно сделать два вывода. Во-первых, использование многих представлений может быть вполне оправданным. Во-вторых, даже при использовании наиболее полной версии алгоритмической вероятности суммирование ведется не по всем возможным представлениям, но может быть поставлена задача определения набора представлений $\{S\}$, наилучшим образом соответствующих данной выборке (наилучшим образом обеспечивающих факторизацию)

$$\max_{\{S\}}\sum_S 2^{-l(S)}\prod_{i=1}^n P_{ALP}(f_i|S)\approx$$
$$\approx\sum_S 2^{-l(S)}P_{ALP}(f_1\ldots f_n|S)=P_{ALP}(f_1\ldots f_n). \qquad (3)$$

Этот результат можно сформулировать как *критерий репрезентационной алгоритмической вероятности*: наилучшей совокупностью представлений, в рамках которых осуществляется описание изображений данной предметной области, можно полагать такое множество, для которого максимальной оказывается оценка факторизованной алгоритмической вероятности (3).

Обсуждение

На практике использование слишком большого числа представлений проблематично из соображений вычислительной эффективности. Естественно, если в рамках какого-то представления S не описываются закономерности, содержащиеся в изображениях данной предметной области, то для него будет верно

$$\prod_{i=1}^n P_{ALP}(f_i|S)\approx\prod_{i=1}^n P_{ALP}(f_i)\ll P_{ALP}(f_1\ldots f_n).$$

Такие представления будут вносить очень малый вклад в суммарную алгоритмическую вероятность и их можно игнорировать.

Но можно ли ограничиться одним лучшим представлением

$$S^* = \arg\max_S \left[2^{-l(S)} \prod_{i=1}^{n} \sum_{p:U(Sp)=f_i} 2^{-l(p)} \right] =$$

$$= \arg\min_S \left[l(S) - \sum_{i=1}^{n} \log_2 P_{ALP}(f_i|S) \right]?$$

Это бы дало формулировку, очень близкую к обычному РМДО, за исключением того, что в рамках этого представления следовало бы рассматривать многие модели.

Мы уже ссылались на то, что выбор одной лучшей модели может быть неадекватным для принятия последующих решений. Аналогичный вывод можно сделать и относительно представлений. Улучшение качества принятия решений (решения каких-либо задач, опирающихся на результаты анализа изображений) может достигаться путем построения и использования множества моделей в рамках различных представлений. Вероятно, даже неполные наборы, состоящие просто из нескольких представлений, могут обладать некоторым преимуществом по сравнению с одиночными представлениями.

Этот вывод может вызвать некоторые сомнения. Ведь до сих пор на практике достаточно успешно использовали именно одиночные представления. Однако при увеличении степени априорной неопределенности анализируемых изображений традиционные методы оказываются недостаточными. В этой связи не столь удивительным является то, что в последнее время стали приобретать все большую популярность различные методы "голосования экспертов", где каждый "эксперт" выражает, по сути, свой независимый способ описания анализируемых объектов.

Подобные методы наиболее хорошо проработаны в области распознавания образов. Это разнообразные композиции классификаторов и методы бустинга [7], которые привлекают одновременно различные представления для описания распределений образов по классам, и результат классификации определяется путем взвешенного голосования или суммирования "мнений разных экспертов".

Ранее мы предлагали синтетические методы распознавания [8], в которых задается несколько семейств представлений (набор классификаторов), и в процессе обучения выбирается лучшее семейство и лучшее представление в рамках этого семейства. Иными словами, среди разных методов распознавания в процессе обучения на основе информационного критерия выбирается лучший. Такой подход позволяет обоснованным образом без перекрестной проверки выбирать лучший (для данной выборки образов) классификатор.

Однако в отличие от известных композиций классификаторов в синтетических методах осуществляется не голосование и не взвешенное суммирование классификаторов, а делается однозначный выбор одного из них. В то же время именно композиции классификаторов доказали свою полезность. Но при этом остается загадкой, почему они не склонны к переобучению. Действительно, композиции классификаторов оказываются чрезмерно сложными. Такое усложнение обычных классификаторов всегда ведет к эффекту переобучения. Имеются попытки дать ответ на вопрос, почему этого не наблюдается для композиций [9], но их нельзя признать общезначимыми без обращения к теории универсальной индукции в рамках алгоритмической теории информации.

Критерий репрезентационной алгоритмической вероятности дает полное решение этой проблемы. Если взвешенное "голосование" представлений ("экспертов") будет осуществляться в соответствии с данным критерием, то композиция классификаторов не будет склонна к переобучению. Если же совокупность представлений совместно используется для построения описаний, то это будет приводить к переобучению, поскольку

$$2^{-l(S_1 \dots S_N)} \prod_{i=1}^{n} P_{ALP}(f_i|S_1 \dots S_N) \ll P_{ALP}(f_1 \dots f_n)$$

в силу того, что априорные вероятности представлений $2^{-l(S)}$ в этом случае будут не суммироваться, а перемножаться.

К примеру, чем композиция нормальных распределений будет отличаться от гауссовой смеси? Смесь как целое пытается подстроиться под распределение образов (параметры ее компонентов оптимизируются совместно). В композиции же рассматриваются альтернативные нормальные плотности, каждая из которых в отдельности пытается описать распределение (параметры каждой плотности оцениваются независимо) всех образов.

Таким образом, структура критерия репрезентационной алгоритмической сложности подсказывает, как осуществлять совместное использование многих представлений так, чтобы вместо эффекта переобучения получить эффект бустинга.

Феномен полезности различных алгоритмов бустинга, смесей экспертов, взвешенных голосований, композиций классификаторов и т. д. дает мощное (хотя и косвенное) эмпирическое подтверждение тех методологических следствий, которые вытекают из рассмотрения различий между алгоритмической вероятностью и Колмогоровской сложностью.

Следовательно, методы интерпретации изображений должны использовать одновременно различные способы описания изображений (множественные представления), в рамках которых строятся различные описания (модели), в совокупности повышающие качество принятия решений, например при распознавании. Этот вывод наиболее значим в контексте теоретико-информационного подхода, который доказал свою чрезвычайную полезность в области анализа изображений и машинного обучения, но который традиционно используется для выбора одной лучшей модели данных или одного представления информации.

Заключение

Проведен анализ необходимости при решении задач индукции формировать множество гипотез о содержании данных с оценкой их вероятностей. Показано, что принцип репрезентационной минимальной длины описания, основывающийся на Колмогоровской сложности, которая определяется по одной наилучшей модели, требует уточнения с помощью алгоритмической вероятности, учитывающей сложности всех возможных моделей. В результате такого уточнения показана необходимость использования при анализе изображений наборов взаимоисключающих представлений, причем лучший набор должен обеспечивать наилучшую факторизацию алгоритмической вероятности для заданной выборки изображений.

Установлена связь между использованием множественных представлений в соответствии с критерием алгоритмической вероятности и широко применяемыми на практике методами распознавания образов, основанными на композициях классификаторов и процедурах голосования. Это подтверждает корректность сделанных методологических выводов и необходимость их применения в области анализа изображений. Дальнейшее исследование в данном направлении подразумевает создание конкретных методик использования множественных представлений на практике.

Работа выполнена при поддержке Министерства образования и науки Российской Федерации.

* * * * *

ЛИТЕРАТУРА

1. *Марр Д.* Зрение. Информационный подход к изучению представления и обработки зрительных образов. М.: Радио и связь. 1987. 400 с.
2. *Потапов А.С.* Выбор представлений изображений на основе минимизации репрезентационной длины их описания // Изв. вузов. Приборостроение. 2008. Т. 51. № 7. С. 3–7.
3. *Ковалевский В.А.* Локальные и глобальные решения в распознавании изображений // ТИИЭР. 1979. Т. 67. № 5. С. 50–58.
4. *Solomonoff R.J.* Does Algorithmic Probability Solve the Problem of Induction? // Oxbridge Research, P.O.B. 391887, Cambridge, Mass. 02139. 1997.
5. *Solomonoff R.J.* Algorithmic Probability, Heuristic Programming and AGI // Proc. 3rd Conf. on Artificial General Intelligence (AGI-2010), Lugano, Switzerland. March 5–8, 2010. P. 151–157.
6. *Poland J., Hutter M.* MDL convergence speed for Bernoulli sequences // Statistics and Computing. 2006. V. 16. P. 161–175.
7. *Bauer E., Kohavi R.* An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants // Machine Learning. 1999. V. 36. P. 105–139.
8. *Potapov A.S.* Synthetic pattern recognition methods based on the representational minimum description length principle // Proc. OSAV'2008, The 2nd Int. Topical Meeting on Optical Sensing and Artificial Vision, St. Petersburg, Russia. 12–15 may, 2008. P. 354–362.
9. *Schapire R.E., Freund Y., Bartlett P., Lee W.S.* Boosting the margin: a new explanation for the effectiveness of voting methods // Annals of Statistics. V. 26. № 5. P. 1651–1686.