

СРАВНЕНИЕ РУЧНЫХ И АВТОМАТИЧЕСКИ ВЫДЕЛЕННЫХ ПРИЗНАКОВ В ЗАДАЧЕ РАСПОЗНАВАНИЯ ЭМОЦИЙ

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Жабинский А. В.

Одинец Д.Н. – к.т.н., доцент

В работе рассмотрены автоматические способы выделения визуальных признаков, а также проведено их сравнение с ручными применительно к задаче распознавания эмоций по изображению человеческого лица.

С развитием компьютерного зрения и машинного обучения стало возможным решение новых, более сложных задач. Однако, многие из них всё ещё требуют серьёзного вмешательства со стороны человека. Одной из таких задач является задача распознавания эмоций по изображению человеческого лица. Классический подход к её решению подразумевает использование так называемых моделей активного образа (active appearance models, AAM), требующих трудоёмкого процесса создания обучающей выборки. В этой связи встаёт вопрос поиска менее трудоёмких способов обучения распознающего алгоритма. В данной работе проведено сравнение двух подходов к сбору признаков: ручного и автоматического, основанного на ограниченных машинах Больцмана (restricted Boltzmann machine, RBM).

Для полноценного сравнения прежде всего следует точно определить задачу, входные данные и ожидаемый результат. Очевидно, что для распознавания по изображению лица требуется набор изображений, содержащих лица. Кроме того, по крайней мере часть изображений должна иметь метки тех эмоций, для которых будет проводиться распознавание. Для ручного подхода на основе AAM также разрешается использование отмеченных вручную для каждого изображения ключевых точек (как правило, это набор из 50-80 точек, описывающих контуры основных элементов лица - бровей, глаз, носа и губ). Тогда задачу можно сформулировать следующим образом: имея описанный набор данных, создать классификатор, способный максимально точно определять выражаемые человеком на изображении эмоции. Ниже будет описано, как данная задача была решена с помощью выделенных вручную и автоматически полученных признаков, а также полученные результаты.

Основная цель моделей активного образа ([1, 2]) - сопоставить статистическую модель формы и текстуры (интенсивности пикселей) реальному изображению. В приложении к распознаванию выражений лиц это означает, что вначале по набору размеченных изображений строится модель, описывающая возможные вариации положения ключевых точек и интенсивности пикселей между ними, а затем алгоритм подгонки пытается найти те же точки на новых, неразмеченных изображениях. Распознавание эмоций при этом строится как отдельный шаг поверх моделей активного образа и реализуется как стандартное обучение с учителем. Здесь в качестве обучающих данных используются уже определённые с помощью AAM ключевые точки, а в качестве правильных ответов - метки эмоций. Обзорный материал по теме классификации эмоций на основе AAM можно найти в [3]. В данной же работе использовался один из наиболее распространённых алгоритмов классификации, а именно метод опорных векторов (support vector machine, SVM). Полный эксперимент, а также его результаты, приведены в конце работы.

Ограниченные машины Больцмана ([4]) предлагают другой подход к данной проблеме. Получившие распространение в середине 2000-ых, они чрезвычайно хорошо показали себя в задачах выделения (изучения) визуальных признаков. Например, в задаче оптического распознавания символов они позволили в полностью автоматическом режиме выделить наиболее характерные линии, составляющие символы, и, тем самым, свести задачу к классификации по высокоуровневым признакам вместо классификации по отдельным пикселям. Эффект автоматического изучения устойчивых признаков достигается за счёт использования поражающей вероятностной сети, а также алгоритма, в некоторой степени имитирующего алгоритм Метрополиса ([5]) для марковских цепей. Подробное описание ограниченных машин Больцмана, а также практическое руководство по их обучению можно найти в [6].

В чистом виде RBM стремятся изучить глобальные признаки, что для практических задач с достаточно крупными изображениями не всегда релевантно. Поэтому в данной работе в качестве обучающего материала были использованы не полные изображения, а наборы "вырезанных" случайным образом кусочков (регионов) фиксированного размера, что позволило алгоритму выделить небольшие локальные признаки (дуги, градиенты, группы пикселей и т.д.). По причинам, которые вскоре станут очевидными, назовём полученные признаки фильтрами.

Как и в случае с AAM, в качестве конечного классификатора использовался метод опорных векторов. Однако, вместо координат ключевых точек использовались признаки, полученные путём применения к оригинальным изображениям процедуры свёртки с полученными ранее фильтрами.

В таблице ниже приведены основные шаги двух сравниваемых методов.

AAM+SVM (признаки, выделенные вручную)	RBM+SVM (автоматически полученные признаки)
<ol style="list-style-type: none"> 1. Ручное выделение ключевых точек. 2. Обучение AAM по выделенным точкам. 3. Использование обученной модели для определения координат точек на новых изображениях. 4. Обучение и применением SVM на полученных координатах и метках эмоций. 	<ol style="list-style-type: none"> 1. Выделение случайных регионов изображений. 2. Обучение RBM на этих регионах и получение высокоуровневых признаков (фильтров). 3. Использование свётки начальных изображений полученными фильтрами. 4. Обучение и применение SVM на новых изображениях и метках эмоций.

Для обучения обоих методов был использован расширенный набор данных Cohn-Kanade ([7]), включающий 10708 изображений с 66-ю отмеченными ключевыми точками, 327 из которых также имеют метку одной из 7 базовых эмоций (вместе с нейтральным выражением лица). Размер вырезаемого региона был установлен в 12 пикселей (при размере полного изображения в 256 пикселей по X и 200 пикселей по Y), а количество компонент для RBM - 72. Предобучение RBM проводилось на всех изображениях, в то время как SVM в обоих подходах использовал лишь те изображения, для которых существует метка эмоции.

По результатам перекрёстной проверки (cross-validation) связка AAM+SVM показала в среднем 92% корректно классифицированных результатов, в то время как RBM+SVM - 64%. Это, однако, значительно выше чистого SVM, чья средняя точность всего чуть выше 20%.

Таким образом, несмотря на относительно низкий результат классификации с автоматически выделенными признаками, данный эксперимент явно показывает перспективность поиска альтернативных способов выделения признаков. Вероятно, наилучшим подходом будет некий гибрид RBM и AAM, на что и будет направлена будущая работа.

Список использованных источников:

1. Cootes T. F., Edwards G. J., Taylor C. J. // In Proc. European Conf. on Computer Vision. 1998. Vol. 2, P. 484–498.
2. Matthews I., Baker S. // International Journal of Computer Vision. 2004, Vol. 60 (2) P. 135 - 164.
3. Ratliff, M. S. Active appearance models for affect recognition using facial expressions: thesis by MD in CS. Wilmington, 2010.
4. Hinton, G. E., Salakhutdinov, R. R. Reducing the Dimensionality of Data with Neural Networks // Science, 2006, Vol. 313 (5786), P. 504–507.
5. Metropolis, N., Rosenbluth, A.W. Equations of State Calculations by Fast Computing Machines. Journal of Chemical Physics. 1953, Vol. 21 (6), P. 1087–1092.
6. Hinton, G. E. A Practical Guide to Training Restricted Boltzmann Machines. University of Toronto, 2010.
7. Lucy, P., Cohn, J. F., Prkachin, K. M., Solomon, P., & Matthews, I. Painful data: The UNBC-McMaster Shoulder Pain Expression Archive Database // IEEE International Conference on Automatic Face and Gesture Recognition, 2011.

ГЕНЕРАЦИЯ СУДОКУ И ОЦЕНКА КОЛИЧЕСТВА РЕШАЕМЫХ СУДОКУ

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Митьковец А.А., Савчик П.А.

Борисенко О. Ф. – канд. физ-мат. наук, доцент

Судoku – математическая головоломка с числами. В нашей работе рассматривается классическая версия судoku. Игровое поле представляет собой квадрат размером 9x9 клеток, разделённый на меньшие квадраты со стороной в 3 клетки. Цель игры – заполнить пустые ячейки так, чтобы каждый ряд, столбик и квадрат 3x3 содержали цифры от 1 до 9 по одному разу.

Постановка задачи:

Создать алгоритм (а также программу на его основе), генерирующий незаполненное поле судoku с заданным количеством изначально указанных чисел. Реализовать создание заполненного поля на основании венгерского алгоритма, создание маски для удаления при помощи «банковской последовательности».

Оценить количество всех незаполненных полей судoku, имеющих решение.

Работа состоит из нескольких частей. Краткое описание каждой из них:

1) Генерация заполненного поля судoku.

Для создания судoku по нашему алгоритму необходимо иметь полностью заполненное (по правилам судoku) поле (рис.1, а). Генерация основывается на применении венгерского метода. Поле судoku заполняется построчно. Задача о заполнении строки рассматривается как задача о назначениях, матрица стоимостей изменяется с учётом уже заполненных строк.