



РАСПОЗНАВАНИЕ ЖЕСТОВ С ПОМОЩЬЮ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ

Кулишова Н.Е., профессор, к.т.н., кафедра МСТ, ХНУРЭ
Казакова А.Д., студент, кафедра ИИ, ХНУРЭ

Основой функционирования общества является взаимодействие между людьми, которое в свою очередь осуществляется с помощью различных видов коммуникаций. Однако существуют люди с ограниченными возможностями, которые не могут стать полноценными участниками общества из-за того, что не могут наладить общение с окружающими. К этой категории людей относятся глухонемые, глухие или плохо слышащие люди. Язык жестов, который они используют для общения, известен очень ограниченному кругу лиц. Поэтому люди с нарушением слуха испытывают трудности даже при выполнении таких повседневных задач как поход за покупками. К тому же для них ограничен выбор потенциальных рабочих мест, которые бы они могли занимать, и круг людей, с которыми они могли бы общаться. Это является проблемой, но не такой неразрешимой, как может показаться на первый взгляд.

Мы хотели бы рассмотреть, каким образом компьютерные технологии потенциально могли бы помочь преодолеть языковой барьер, который ограждает людей с упомянутыми выше ограниченными возможностями от всего остального мира. Современные исследования в области искусственного интеллекта, разнообразные разработки и инструменты уже сейчас предоставляют нам различные способы решения этой проблемы. С помощью машинного зрения можно дислоцировать и распознать жесты языка глухонемых на фотографии или видеозаписи и сопоставить их с буквами соответствующего алфавита. Таким образом можно создать своеобразный переводчик с языка глухонемых на письменную/устную речь, рассматриваемого языка. В данном случае будет взят за основу английский язык как язык международного общения и соответствующие ему язык жестов – ASL (American Sign Language).

Уже осуществлялись попытки создать подобное приложение. Однако тема не изучена до конца и пока что не найдено полноценного решения, которое могло бы быть с легкостью использовано любым человеком в повседневной жизни. К основным разработкам в этой сфере относятся проекты Microsoft Kinect Sign Language Translator и Motionsavvy. Первый из них находится на стадии исследований и экспериментов и уже сейчас достиг высоких показателей точности распознавания жестов. Кроме того, он позволяет осуществить перевод как с языка жестов в устную/письменную речь, так и наоборот. Однако для этого необходимо использовать специальный прибор – Microsoft Kinect, что делает это решение непортативным и неудобным для повседневного использования. Motionsavvy в свою очередь тоже способен осуществлять двунаправленный перевод. При



этом используется портативное устройство MotionLeap, которое напоминает обычный планшет. Он более удобен в использовании, однако качество распознавания жестов у него ниже, чем у Microsoft Kinect. Это связано с тем, что он не позволяет перенести жесты в электронную форму с достаточной для распознавания точностью в режиме реального времени.

Жесты не поддаются точному шаблону. Один и тот же жест будет демонстрироваться различными людьми по-разному. Запечатленная в электронном формате версия жеста практически уникальна – достаточно сложно сделать две одинаковые фотографии. Однако на основании общих черт, которые отличают один жест от другого, можно разбить их на соответствующие классы. В любом случае в этой задаче есть некоторая доля неопределенности и следует использовать интеллектуальные методы анализа, которые на основании заданных примеров, смогут выделить необходимые характеристики для верной классификации примеров. Среди всего разнообразия методов анализа данных следует выделить глубокие нейронные сети, а именно сверточную нейронную сеть. Она широко используется для обработки изображений, так как использует принцип скользящего окна, который очень часто применяется при выполнении таких манипуляций как размытие, увеличение контраста или нанесение эффектов на изображение. С помощью ядра свертки рассматриваемая НС позволяет выявить и зафиксировать наличие определенных признаков в той или иной части изображения, например, линий или дуг с учетом их угла наклона и направления. Далее формируются карты признаков – наборы основных характеристик, описывающих рассматриваемое изображение. На основании их будет осуществляться дальнейшая классификация примеров. При подготовке сети используется стратегия обучения с учителем, которая заключается в том, что после каждой эпохи классификации входных примеров, НС сообщаются классы, которые на самом деле имеют поступившие на вход примеры, и на основании этой информации итеративно настраиваются соответствующие веса всех слоев сети.

В качестве входных данных, которые понадобятся нам для обучения и проверки качества классификации, была использована общедоступная база изображений Sign Language MIST. Каждый отдельный пример характеризуется значением яркости всех пикселей изображения, представленного в градации серого цвета. Картинка имеет расширение 28x28, то есть характеризуется 784 пикселями. На основании этих данных нейронная сеть будет принимать решение, к какому классу принадлежит тот или иной пример. Всего имеется 24 класса, которые соответствуют буквам английского алфавита (следует учесть, что буквы J и Z были исключены из рассмотрения, так как их обозначающие жесты требуют движения, что нельзя изобразить на картинке). Выборка хранится в CSV-файле и содержит 27.455 различных примеров.



Данные были разбиты на 2 блока соотношении 70 на 30, где 70 – это обучающая выборка, а 30 – тестовая. Был выбран пакетный режим обучения: в каждом пакете содержалось по 128 примеров. Количество эпох обучения нейронной сети составляло 50, а размер ядра был выбран 3x3. После каждой эпохи обучения была замерена точность предсказаний НС, чтобы отследить процесс обучения, который достигается путем настройки весов сети. Результаты представлены на рисунке 1.

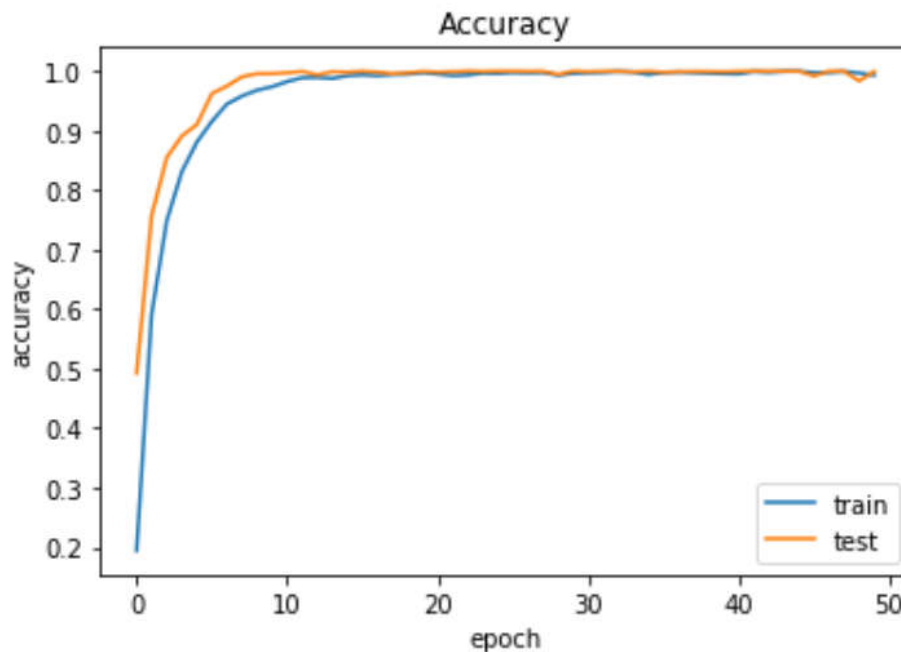


Рисунок 1 – Изменение точности предсказаний сети

Далее было проверено качество обучения на тестовой выборке. Как следует из графика результаты коррелируют с теми, что были получены при работе с обучающей выборкой. Точность увеличивается экспоненциально с увеличением количества эпох обучения, однако достигнув определенного значения больше не изменяется и в этот момент следует остановить обучение, чтобы достигнуть баланса между потраченными ресурсами (в первую очередь временными) и необходимым качеством классификации примеров с помощью нейронной сети.

Список литературы

1. Хайкин С. Нейронные сети: полный курс. 2-е издание. М.: Вильямс, 2006. 1104 с.
2. Chollet F. Deep learning with Python. М.: Manning Publications Co., 2018. 386 p.
3. TensorFlow. URL: <https://www.tensorflow.org>.