

1. Spatial pyramid pooling слой, как ясно из названия, это “пирамида” из обобщающих слоёв, в классической ситуации для обобщающего слоя мы задаем:

- 1) Размер окна (W_{win}, H_{win})
- 2) Размер сдвига по горизонтали и вертикали (W_{str}, H_{str}).
- 3) Ширину бордюра добавляемого вокруг входной матрицы (W_{pad}, H_{pad}).
- 4) Обобщающую функцию.

Тогда, подавая на вход такого слоя матрицу особенностей размера (W_{inp}, H_{inp}) на выходе мы получим матрицу с размерами:

$$(W_{out}, H_{out}) = ((W_{inp} - W_{win} + 2 \cdot W_{pad}) / W_{str} + 1, (H_{inp} - H_{win} + 2 \cdot H_{pad}) / H_{str} + 1)$$

Каждый элемент выходной матрицы, равен значению обобщающей функции, вычисленному на соответствующем прямоугольном окне размера (W_{win}, H_{win}) входной матрицы.

Вместо того, чтобы задавать размеры окна и сдвиг, зафиксируем размеры выходной матрицы. А размер окна и размер сдвига будем вычислять исходя из размеров входной матрицы:

$$(W_{win}, H_{win}) = (\lceil W_{out} / W_{inp} \rceil, \lceil H_{out} / H_{inp} \rceil)$$
$$(W_{str}, H_{str}) = (\lceil W_{out} / W_{inp} \rceil, \lceil H_{out} / H_{inp} \rceil)$$

Фактически мы хотим разбить входную матрицу на определенное количество ячеек, вычислить обобщающую функцию на каждой ячейке, и поместить результат в выходную матрицу. Поскольку теперь независимо от размера входного изображения на выходе такого обобщающего слоя будет трехмерная матрица всегда одних и тех же размеров, то и вектор, который мы подадим в классификатор, не будет меняться при изменении размеров изображения.

Чтобы улучшить качество классификации, вместо одного такого слоя можно использовать пирамиду слоёв. Т.е. вместо одной матрицы на выходе будем иметь несколько, которые развернём в несколько векторов, которые объединим и подадим на вход классификатора.

2. LSTM (long short-term memory) — тип рекуррентной нейронной сети, способный обучаться долгосрочным зависимостям. LSTM специально разработаны для устранения проблемы долгосрочной зависимости. Они запоминают информацию в течение длительных периодов времени.

LSTM необходим, когда разрыв между информацией и точкой, где нужно её учесть, становится настолько большим, что РНС уже не справляется, например, рассмотрим попытку предсказать последнее слово в тексте «Я вырос во Франции ... Я свободно говорю на французском». Из предыдущих слов понятно, что следующим словом, вероятно, будет название языка, но если мы хотим назвать правильный язык, нужно учесть упоминание о Франции.

LSTM блок имеет три слоя:

- 1) Слой утраты, который устанавливает, какое значение необходимо выбросить из памяти.

- 2) Слой сохранения, который устанавливает, какую информацию необходимо сохранить в память.
- 3) Далее получаем новое состояние ячейки.

3. Преимущества создания моделей в функциональном виде:

- 1) Возможность разветвления (несколько входов и выходов)
- 2) Возможность именования слоёв