



МАТЕМАТИЧЕСКИ МОДЕЛ НА LSTM НЕВРОННИ МРЕЖИ

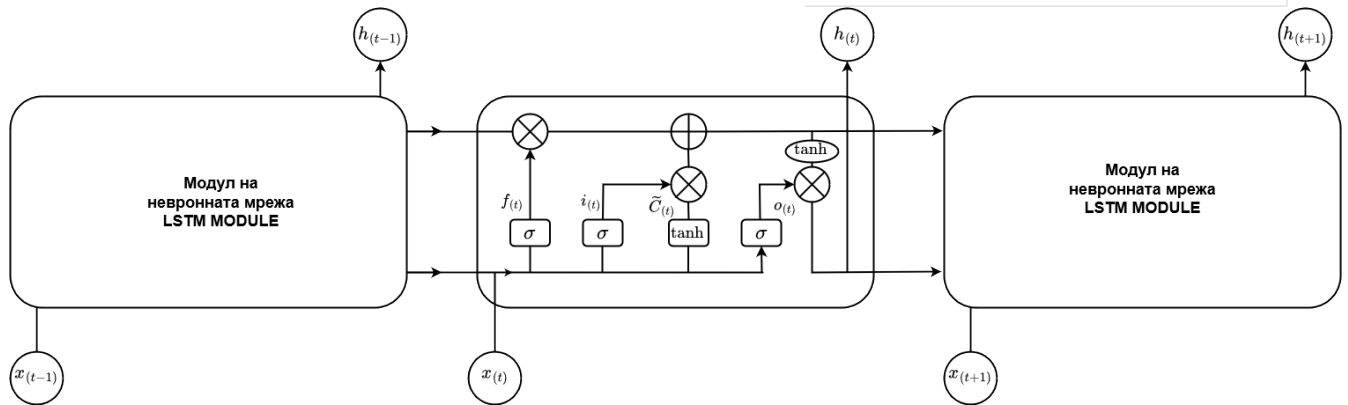
1. LSTM невронни мрежи.

Невронните мрежи с “продължителна краткосрочна памет” (Long Short-Term Memory – LSTM) са специален вид рекурентни невронни мрежи, динамични по своята същност, които могат да научават дългосрочни зависимости между времевите стъпки от последователност от данни. Те могат да се използват за решаване на класификационни и регресионни задачи. LSTM невронните мрежи успяват да научат динамиката в последователността на дадена поредица от обекти. Обикновено се използват за обработка на текст, видео, звук и др.

Рекурентните невронни мрежи имат цикли (loop) за обратни връзки между възлите си, което им позволява да проявяват временно динамично поведение. На входа на рекурентните невронни мрежи се подават дискретни входни величини и информацията започва да се разпространява в мрежата с по една връзка (edge, connection) за дадена единица време t . Този вид мрежи използват вътрешното си състояние (памет) за обработка на последователности от входни величини с променлива дължина. По този начин цикличната връзка на рекурентната мрежа се определя между различните времеви стъпки. След като мрежата се разгърне по оста на времето, рекурентната мрежа става мрежа с право разпространение на информацията (фиг. 1). Като сравнение, конволюционните невронни мрежи работят с вектор с фиксиран размер за входни данни (изображение или видео кадър) и произвеждат вектор с фиксиран размер като изход, отнасяйки го към определен клас, докато рекурентните модели позволяват работа над последователности от вектори във времето.

LSTM невронните мрежи имат топология на линейна верига от повтарящи се модули фигура 2. Класическата архитектурата на мрежата се състои от един входен слой, един скрит слой и един изходен слой. Скрытия слой е напълно свързан и има обратна връзка към себе си. Скрытия слой съдържа клетки с памет и съответни модули на „портата“ (gate), както и скрити единици, осигуряващи достъп до входните единици и клетките с памет.

Повтарящият се модул в стандартна рекурентна мрежа съдържа един \tanh слой, а в LSTM модула си взаимодействат по между си четири слоя – три сигмоидни и един с хиперболична тангента, както е показано на фигури 1 и 2.



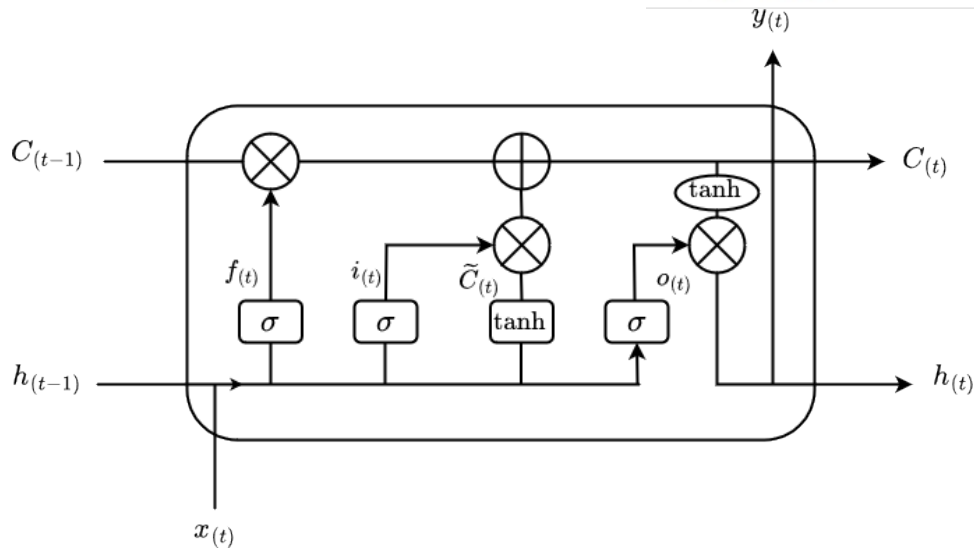
Фиг. 1. Линейна верига от повтарящи се модули в LSTM невронна мрежа.

Ключова при LSTM невронните мрежи е обработката на информационния поток в клетката. Тя има две състояния:

- краткосрочно състояние (Short-term state), което поддържа изхода на текущата времева стъпка;
- дългосрочно състояние (Long-term state), е състоянието на клетката, което има способността да премахва или добавя информация към нея, с помощта на повтарящи се врати, наречени „порти за забравяне“ (forget gates).

Портите са самоадаптивни и определят своите стойности на състояния въз основа на текущото състояние на входа на LSTM единицата. С помощта на портите за забравяне, LSTM мрежата предава две версии на минали събития и по този начин балансира текущия вход и предишни събития. По конкретно притежават три входни порти, през които преминава информацията – входна порта i_t , забравена порта f_t и изходна порта o_t , които са изчислени въз основа на входното и последното скрито състояние, преминавайки през слой от линейни единици последвани от сигмоидна активационна функция.

На фигура 2 е представена диаграма на LSTM клетка.



Фиг. 2. Диаграма на устройството на LSTM клетката.

Горната хоризонтална линия изобразена на фигурата премахва или добавя информация към състоянието на клетка.

Векторът C_t наричан клетка на паметта или селективна памет (memory cell) е комбинация от кандидат характеристика (candidate feature) и предходното състояние на клетката (клетката на последната стъпка), които са претеглени от входната и забравящата порти.

Линията на селективната памет е модифицирана на две места преди да бъде дадена на следващата единица за време, обозначени с времена (\times) и ($+$).

„Спомените“ се премахват в единицата за време (\times) и се добавят в ($+$).

Долната линия е по-скорошната памет h_t . С нея свързваме h линията с предишни h_t и текущото наблюдение x_t , което се подава в линейната единица за „забравяне“ последвана от сигмоидална функция за да произведе f_t сигналът за „забравяне“, който се движи нагоре от лявата страна на фигурата.

Сигмоидната функция σ е ограничена в диапазона (0, 1) и след умножение произвежда резултат, който трябва да бъде намаляване на абсолютната стойност във всяка точка на основната памет, което съответства на „забравянето“.

Изходът на сигмоида се умножава поелементно като $x[i, j]$ -тия елемент от един масив умножен с $u[i, j]$ -тия елемент от друг масив.

Следващата последователност отново идва от долния ляв ъгъл, но този път преминава през два линейни слоя – единият със сигмоидна активационна функция, другия с хиперболична тангенсова функция.



Резултатът от преминалото умножение с \tanh функцията се добавя към състоянието на клетката, в клетката етикетирана като (+), тъй като \tanh функцията може да произвежда както положителни така и отрицателни резултати.

След това линията на паметта на клетката се разделя на две – едно копие тръгва в дясно, а другото се комбинира с линейна трансформация в по-скорошната памет като образуват новия h ред, което е свързано със следващото наблюдение (изображение).

Линията на селективната памет не минава директно през линейни единици. При нея или състоянията са забравени в (\times), или са добавени в (+).

Временните променливи, дефинирани по-горе, могат да се изчислят с помощта на следните уравнения:

$$\begin{aligned}f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \\i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\ \tilde{C}_t &= \tanh(W_c \cdot [h_{t-1}, x_t] + c), \\C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t, \\O_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \\h_t &= O_t * \tanh(C_t),\end{aligned}$$

където с W е означена теглова матрица, а с термина b отклонението (bias).

Проблемът с разпознаването на изображения се представя като класификация с последователност към много „етикети“ (Sequence to Label, Sequential Labelling).

Поредицата от наблюдения $\{x^T \dots, x^T\}$ „научават“ класификатора за генериране хипотези \hat{y} от истинските етикети y .

Индексът t указва стъпките на последователността, а T означава дължината на последователността на всеки пример.

Използваните слоеве на LSTM мрежи са следните (фиг. 2):

- *слой за въвеждане на последователност* (Sequence Input Layer) – входен слой, оразмерен за въвеждане на последователност от 2-D изображения, представени като вектор от три елемента $[h \ w \ c]$. Входният слой прилага и нормализиране на данните, винаги когато данните се разпространяват напред

www.eufunds.bg

Проект BG05M2OP001-2.016-0003 „Модернизация на Национален военен университет "Васил Левски"- гр. Велико Търново и Софийски университет "Св. Климент Охридски" - гр. София, в професионално направление 5.3 Компютърна и комуникационна техника“, финансиран от Оперативна програма „Наука и образование за интелигентен растеж“, съфинансирана от Европейския съюз чрез Европейските структурни и инвестиционни фондове.



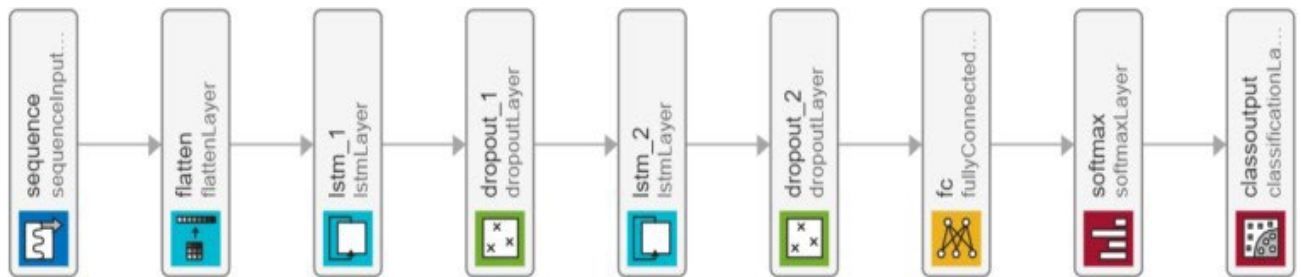
през него.

- „изравняващ“ слой (Flatten Layer) – изравняващият слой свива пространствените размери на входа до измерението на канала.
- *LSTM* слой (LSTM Layer) – научава дългосрочни зависимости между времевите стъпки от данните за последователността;
- „отпадащ“ слой (Dropout Layer) – по време на обучение, слоят за отпадане на случаен принцип настройва входните данни на нула за конкретна маска за отпадане. Тази операция ефективно променя основната мрежова архитектура между итерациите и помага да се предотврати прекомерното напасване на данните (overfitting). По-голямото число пропуска повече елементи по време на обучението.
- *напълно свързан слой* (Fully Connected Layer) – използва се за решаване на класификационна задача като комбинира характеристиките за класифициране на изображенията. Параметърът изходен размер (Output size) е равен на броя на класовете в целевите данни;
- *SoftMax* слой (SoftMax Layer) - използва функция за активиране SoftMax, която нормализира изхода получен от напълно свързания слой. Резултатът от SoftMax слоя се състои от положителни числа, които се сумират до единица, в последствие могат да бъдат използвани като класификационни вероятности от класификационния слой;
- *класификационен слой* (Classification Layer) – изчислява загубата на кръстосана ентропия за класификационни задачи.

На фигура 3 е показана примерна архитектура на LSTM мрежа за разпознаване на изображения.

Първият LSTM слой има входен размер от 16384 неврона, при размер на входното изображение $128 \times 128 \times 1$.

Броят на скритите единици в слоя е зададен на 256, откъдето състоянието на клетката и скритото състояние, са числени вектори с размери 256×1 .



Фиг. 3. Примерна архитектура на LSTM мрежа
с наличие на два LSTM слоя.

LSTM невронните мрежи са лидери в решаване на задачи с дългосрочни зависимости, въпреки изчислителната сложност на моделиране на процесите във всеки модул на мрежата.

Тези мрежи са в състояние да се учат от данни от минали събития и да откриват тенденции в бъдещи събития, което ги прави приложими в случаи на анализиране и прогнозиране на многовариантни времеви редове от данни.

Изборът на LSTM мрежа е обоснован от възможността на този вид невронна архитектура да анализира динамиката в позицията на обекти, проследявайки тяхното движение във времето.