

وزارت علوم، تحقیقات و فناوری
دانشگاه تحصیلات تکمیلی علوم پایه
گاوزنگ، زنجان



پیاده سازی تشخیص دست خط فارسی با استفاده از روش

های ماشین یادگیری

موضوع پروژه درس هوش مصنوعی پیشرفته

رضا خالقی

استاد: دکتر محسن هوشمند

پاییز ۹۸

فصل اول: تعریف مسئله

۱,۱ تعریف پروژه به زبان ساده

مجموعه داده های فراوانی در اینترنت یافت می شود که برای کارهای یادگیری ماشین و روش های دسته بندی و تخمین به کار می رود. یکی از این نمونه های آموزشی مجموعه داده هدی است که شامل چندین هزار تصویر از اعداد فارسی است. هر عدد در تصویری ذخیره شده و دارای برچسب متناظر است که عدد را نشان می دهد. داده به دو مجموعه آموزش (train) و آزمون (test) تقسیم می شود. از روش های یادگیری استفاده کنید که یادگیری از مجموعه آموزش استفاده می شود و نتایج روی مجموعه آزمون امتحان می شود و نتایج در قالب پارامترهایی مشخص می شود.

۲,۱ اهداف پروژه

هدف از انجام این پروژه یادگیری و بکارگیری درست از مباحث تئوری این درس و پیاده سازی آنها در دنیای واقعی می باشد. مسائلی همچون یادگیری توسط الگوریتم های مربوط به یادگیری ماشین با استفاده از پایگاه اطلاعاتی فارسی می باشد.

۳,۱ تکنولوژی های استفاده شده در این پروژه

برای پیاده سازی این پروژه از زبان python نسخه ۳,۷,۱ استفاده شده است. همچنین برای محیط اجرای دستورات نیز از تکنولوژی jupyter استفاده شده که به هنگام مشاهده دستورات و خروجی نتایج نیز این مهم به راحتی صورت گیرد.

فصل دوم: ابزار های بکار رفته

۱,۲ Jupyter Notebook

Jupyter یکی از نرم افزارهای بسیار کاربردی در حوزه برنامه نویسی پایتون در علوم داده و داده کاوی است. توسط این ابزار می‌توانید به راحتی عملیات مختلف تست و آزمایش های خود را در یک محیط ساده همراه با خروجی های در لحظه انجام دهید. این نرم افزار جهت انجام کارهای تحقیق و توسعه در شرکت ها یا دانشگاه ها بسیار کاربرد دارد.

۲,۲ موارد مورد نیاز در پروژه

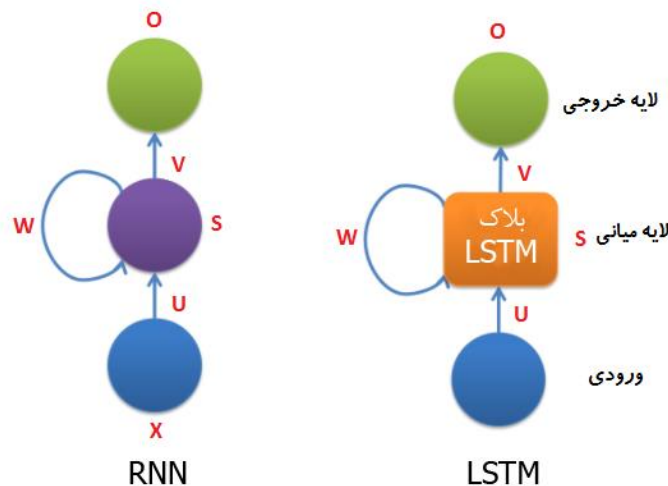
برای پیاده سازی پروژه با استفاده از تکنولوژی های ذکر شده در ابتدا نیاز به زبان پایتون (نسخه ۳,۷ به بالا) و سپس کتابخانه های ذیل مورد نیاز خواهد بود.

- `matplotlib`
- `tensorflow`
- `keras`
- `python-opencv`
- `numpy`
- `pandas`
- `sklearn`

۳,۲ الگوریتم یادگیری دنباله ای

الگوریتم یادگیری دنباله ای (Sequential Learning) جزء شبکه های LSTM می‌باشد که بصورت خلاصه در اینجا به آن پرداخته می‌شود.

LSTM (به معنی حافظه طولانی کوتاه-مدت) مخفف کلمه Long short-term memory یک نوع مدل یا ساختار برای داده های ترتیبی است که در سال ۱۹۹۵ برای توسعه شبکه های عصبی بازگشتی (RNN) ظهور پیدا کرد Sepp Hochreiter. در مقاله LSTM توضیح می‌دهد که عبارت long term memory به وزن های یادگرفته شده و short term memory به حالت های درونی سلول ها اطلاق میشود LSTM. برای حل مشکل پدیده ناپدید شده گرادیان در شبکه های عصبی بازگشتی بوجود آمدند که تغییر عمده آن جایگزین کردن لایه میانی RNN با یک بلاک که بلاک LSTM نام دارد است.



بزرگترین ویژگی LSTM امکان یادگیری وابستگی بلند مدت است که توسط شبکه های عصبی بازگشتی امکان پذیر نبود. برای پیش بینی گام زمانی بعدی نیاز است که مقادیر وزن ها در شبکه بروز رسانی شوند که این کار مستلزم حفظ اطلاعات گام های زمانی ابتدایی است. یک شبکه عصبی بازگشتی فقط می تواند تعداد محدودی از وابستگی های کوتاه مدت را یاد بگیرد ، اما سری های زمانی بلند مدت مثل ۱۰۰۰ گام زمانی قابل یادگیری توسط RNN ها نیستند اما LSTM ها میتوانند این وابستگی های بلند مدت را به درستی یاد بگیرند .همانگونه در جدول ذیر (این جدول از مقاله Sepp Hochreiter استخراج شده است) نمایش داده شده است RNN نتوانست حتی خروجی شبکه با delay 100 را با هر دو متد RTRL, BPTT پیش بینی کند اما LSTM در زمان کوتاه توانست این کار را انجام دهد.

Method	Delay p	Learning rate	# weights	% Successful trials	Success after
RTRL	4	1.0	36	78	1,043,000
RTRL	4	4.0	36	56	892,000
RTRL	4	10.0	36	22	254,000
RTRL	10	1.0-10.0	144	0	> 5,000,000
RTRL	100	1.0-10.0	10404	0	> 5,000,000
BPTT	100	1.0-10.0	10404	0	> 5,000,000
CH	100	1.0	10506	33	32,400
LSTM	100	1.0	10504	100	5,040

داده های سری زمانی

به صورت مختصر اگر بخواهیم تعریف داده های سری زمانی را قبل از ورود به قسمت اصلی را توضیح دهیم، بدینگونه خواهد بود. همانطور که از نام آن پیداست داده های سری زمانی به مجموعه داده هایی گفته میشود که در فواصل زمانی منظمی جمع آوری شده اند:

$$x_1, x_2, x_3, \dots, x_T$$

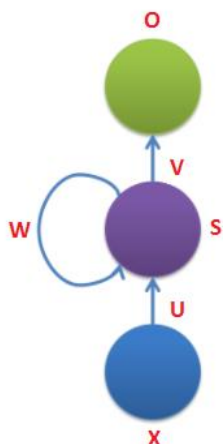
این داده ها می توانند دنباله ای از فریم های یک انیمیشن، کلمات یا اعداد یا حروف یک جمله، امواج صوتی (که با یک بازه زمانی مساوی نمونه گیری شده اند) و ... باشند. نمونه ای از مسایل داده های سری زمانی به شکل زیر هستند:

- تولید جملات و دیابوگ
- باز تشخیص صدا /آواها
- بازتشخیصی ویدئو و ...

اما LSTM چگونه این مسائل را حل می کند؟ به صورت مختصر در مسئله تولید جملات LSTM میتواند کلمه بعدی را با توجه به کلماتی که از طریق ورودی به آن داده شده است پیش بینی کند. در فرآیند آموزش، این شبکه قوانین گرامر را فرا می گیرد.

این شبکه ها در واقع برای پردازش سیگنال های دنباله دار به وجود آمدند. در یک شبکه عصبی معمولی تمام ورودی ها و خروجی ها مستقل از یکدیگر هستند، اما در بسیاری از موارد این ایده میتواند خیلی بد باشد. به عنوان مثال فرض کنید شما در یک جمله به دنبال پیش بینی کلمه بعدی هستید در صورتی که شبکه نتواند روابط بین کلمات را یاد بگیرد مسلماً نمی تواند کلمه بعدی را به درستی پیش بینی کند.

اگر با یک دیدگاه دیگر به این نوع شبکه نگاه کنیم، این شبکه ها دارای یک نوع حافظه هستند که اطلاعاتی تا کنون دیده است را ضبط میکند. در تئوری اینطور به نظر می رسد که این شبکه ها می توانند اطلاعات موجود در یک دنباله طولانی را ضبط و از آنها استفاده کنند اما در عمل اینطور نیست و بسیار محدود هستند، به این صورت که فقط اطلاعات چند گام قبل را ضبط می کنند. در شکل زیر نمونه ای از یک RNN معمولی نمایش داده شده است.



در شکل بالا X ورودی در گام زمانی t است، این ورودی میتواند یک بردار one-hot باشد (one-hot برداری است در یک بعد ۱ و در بقیه ابعاد صفر است).

S حالت پنهان در گام زمانی t است. این قسمت همان جایی است که حافظه شبکه در آن قرار گرفته است S . بر اساس حالت پنهان قبلی و ورودی که در زمان جاری به آن داده میشود تغییر میکند. که به شکل $s(t) = f(Ux(t) + W(s(t-1)))$ است. تابع f معمولاً یک تابع غیر خطی مثل \tanh یا ReLU است.

خروجی در گام زمانی t است. به عنوان مثال در محاسبه کلمه بعدی در یک جمله این خروجی میتواند یک بردار از احتمالات در واژگان ما باشد.

۳,۲ مجموعه ارقام دستنویس هدی

مجموعه ارقام دستنویس هدی که اولین مجموعه ی بزرگ ارقام دستنویس فارسی است، مشتمل بر ۱۰۲۳۵۳ نمونه دست نوشته سیاه سفید است. این مجموعه طی انجام یک پروژه ی کارشناسی ارشد درباره بازشناسی فرمهای دستنویس تهیه شده است. داده های این مجموعه از حدود ۱۲۰۰۰ فرم ثبت نام آزمون سراسری کارشناسی ارشد

سال ۱۳۸۴ و آزمون کاردانی پیوسته‌ی دانشگاه جامع علمی کاربردی سال ۱۳۸۳ استخراج شده است. خصوصیات این مجموعه داده به شرح زیر است:

درجه تفکیک نمونه‌ها: ۲۰۰ نقطه بر اینچ

تعداد کل نمونه‌ها: ۱۰۲۳۵۲ نمونه

تعداد نمونه‌های آموزش: ۶۰۰۰ نمونه از هر کلاس

تعداد نمونه‌های آزمایش: ۲۰۰۰ نمونه از هر کلاس

سایر نمونه‌ها: ۲۲۳۵۲ نمونه

فصل سوم: توضیح کد

۱,۳ موارد پیاده سازی شده

تمامی موارد تعریف شده در پروژه رعایت شده است. همچنین از دیتاست فارسی هدی برای یادگیری استفاده شده است که با دقت بالایی نیز عملیات انجام می شود.

۲,۳ خواندن از پایگاه اطلاعاتی هدی

با استفاده از پایگاه اطلاعاتی هدی که مجموعه دست خط های موجود برای نمونه های آموزشی و آزمون را می خوانیم که این قسمت با استفاده از ارجاعی که در کد آورده شده است پیاده سازی شده است.

۳,۳ بررسی درست خواندن پایگاه اطلاعاتی

یکی از نمونه های آموزشی را خوانده و تصویر آن را نمایش می دهیم که خروجی در کد پیوست شده قابل مشاهده می باشد.

۴,۳ استاندارد سازی متغیر ها برای آغاز یادگیری

در این قسمت دو متغیر آموزش و تست را که هر خانه آن شامل مقادیر ذخیره شده در هر پیکسل (بازه رنگی از صفر تا ۲۵۵) ذخیره شده است را در ابتدا به ۴ بعد تبدیل می کنیم که عدد ۳۲ نیز در اینجا به معنی ابعاد هر تصویر می باشد. در ادامه پس از استاندارد سازی و تبدیل آن به آرایه های `numpy`، بازه مقادیر را از صفر تا ۲۵۵ به صفر تا یک با تقسیم بر ۲۵۵ خواهیم برد.

۵,۳ یادگیری و نمایش نتیجه

پس از آماده سازی اطلاعات، با استفاده از الگوریتم یاد شده در فصل قبل، با اضافه کردن لایه های مشخص شده، یادگیری را آماده شروع کرده و در نهایت با ۱۰ بار تکرار یادگیری (epoch) دقت مدل خود را بالا می بریم.

۶,۳ یادگیری و نمایش نتیجه

برای بررسی اینکه آیا مدل به درستی کار می‌کند یا خیر می‌توان در بخش آخر با مشخص کردن خانه ای از نمونه های آزمون بررسی کنیم که آیا به ازای هر نمونه برچسب متناظر به درستی مشخص شده است. همینطور می‌توان احتمال اینکه چقدر این برچسب با مدل آموزش دیده تطابق دارد را نیز مشاهده کرد.

منابع:

[۱] “بازشناسی ارقام و حروف دستنویس در فرمهای آزمون سراسری”، پایان نامه‌ی کارشناسی ارشد،

حسین خسروی، دانشگاه تربیت مدرس ، ۱۳۸۴

[۲] [شبکه های عصبی LSTM و RNN](#)