

به نام خدا



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر



درس پردازش زبان های طبیعی

سجاد پاکدامن ساوجی

۸۱۰۱۹۵۵۱۷

فروردین ۱۴۰۰

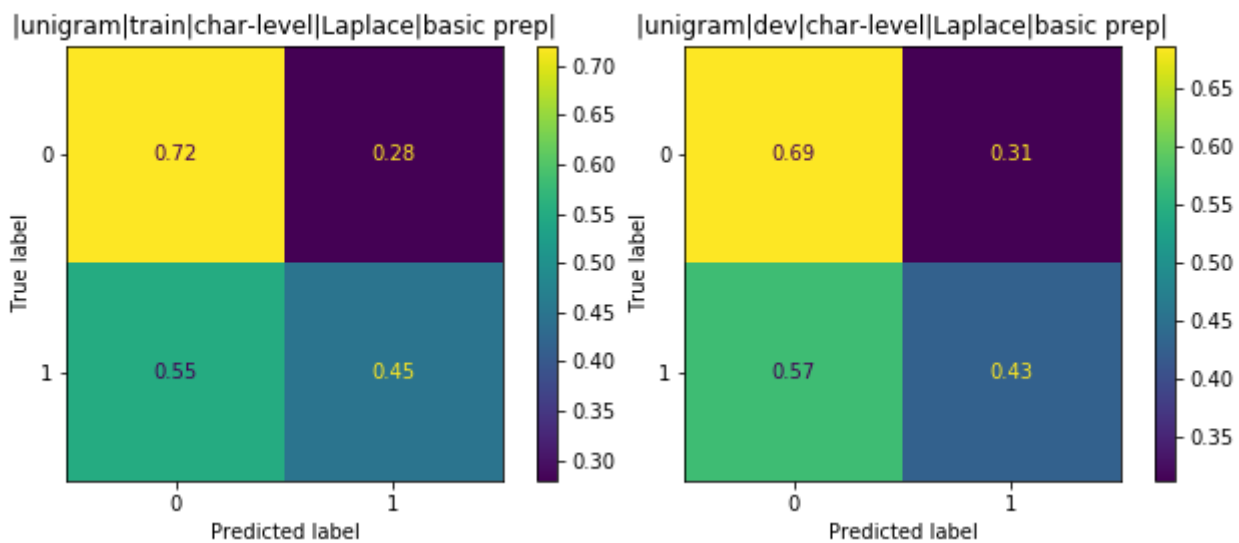
فهرست مطالب

3	سوال ۱ - پیش پردازش
3	سوال ۲ - ایجاد مدل های زبانی
4	سوال ۳ - معیار سرگشتگی
8	سوال ۴ - دادگان تست

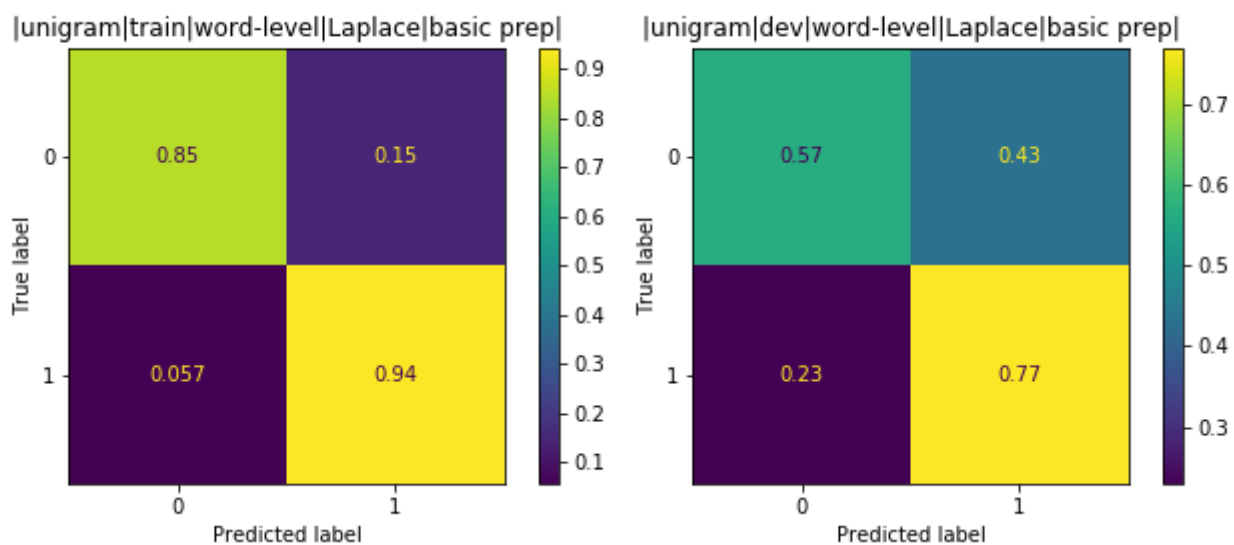
سوال ۳ - معیار سرگشتگی

الف) در این قسمت از method موجود در Language Model به نام perplexity استفاده شد.

ب) عملکرد مدل ها با استفاده از ماتریس confusion در شکل های ۳ و ۴ نشان داده شده است.

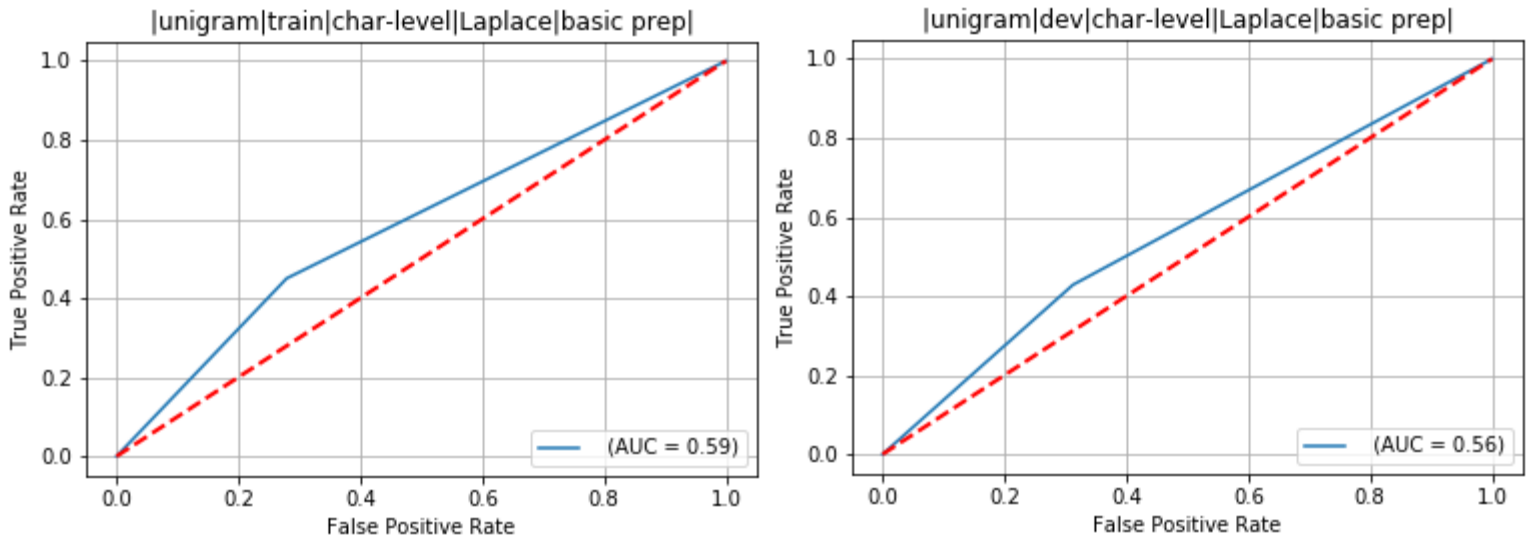


شکل ۳. ماتریس های پراکنده برای char level

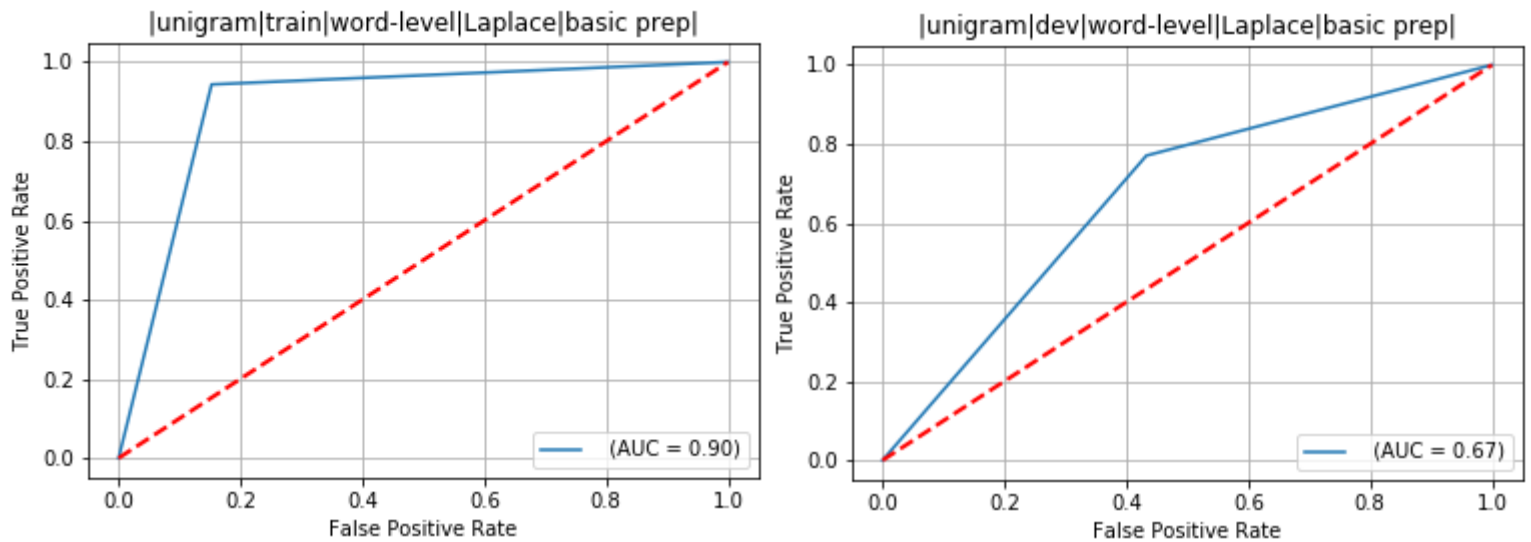


شکل ۴. ماتریس های پراکنده برای word level

همچنین برای مقایسه بهتر مدل ها نمودار های ROC نیز در شکل های ۵ و ۶ نشان داده شده است.



شکل ۵. نمودار های ROC برای Char level



شکل ۶. نمودار های ROC برای word level

ج) برای ارزیابی مدل علاوه بر confusion matrix از متریک های دیگری نیز استفاده شده است که این معیار ها را می توان در جدول ۱ و جدول ۲ مشاهده نمود.

model	F1-macro	F1-micro	acc	prec	recall
unigram word	0.878189419	0.878888889	0.878888888	0.7969128996	0.95532699
bigram word	0.980595760	0.98100000	0.98100000	0.9609494640	0.99524187
unigram char	0.557689596	0.59033333	0.59033333	0.5173160171	0.37906423
bigram char	0.679534197	0.68566666	0.68566666	0.6200906344	0.65107057

جدول ۱. عملکرد مدل روی داده های آموزش

model	F1-macro	F1-micro	acc	prec	recall
unigram word	0.6429588680	0.6440000000	0.6440000000	0.5566188197	0.8173302107
bigram word	0.5393494507	0.5670000000	0.5670000000	0.4963325183	0.9508196721
unigram char	0.5309847198	0.5580000000	0.5580000000	0.4774774774	0.3723653395
bigram char	0.6405463695	0.6480000000	0.6480000000	0.5874125874	0.5901639344

جدول ۲. عملکرد مدل روی داده های dev

علت تفاوت معیار های F1-micro و F1-macro در تعریف آنها است. برای مثال F1-macro برای هر یک از کلاس ها F1 را محاسبه میکند و روی تعداد کلاس ها میانگین گیری میکند. در حالی که F1-micro کلاس ها در نظر گرفته نمی شوند و مقادیر FN، TP و FP بدون در نظر گرفتن کلاس ها محاسبه می شوند.

د) همانطور که در متن پروژه اشاره شده است، داده های آموزش نامتوازن هستند. در شکل های ۷ و ۸ این عدم توازن نشان داده شده است. برای متوازن سازی داده ها، از داده های off_train استفاده شده است و به صورت تصادفی تعدادی از داده ها که لیبیل آن ها ۱ بوده است را به داده های آموزش اضافه میکنیم. در ادامه برای مقایسه عملکرد مدل های آموزش داده از متریک هایی که در قسمت قبل معرفی شد، استفاده خواهیم کرد. این اعداد در جدول ۳ و جدول ۴ گزارش شده است.

model	F1-macro	F1-micro	acc	prec	recall
unigram word	0.8951963615	0.8954379911	0.8954379911	0.8607904861	0.9434540923
bigram word	0.9832245790	0.9832279087	0.9832279087	0.9699850857	0.9973164654
unigram char	0.5776370987	0.5853939045	0.5853939045	0.6171443597	0.4498754073
bigram char	0.6882800772	0.6884224650	0.6884224650	0.6806985294	0.7097949012

جدول ۳. عملکرد مدل روی داده های آموزش پس از توازن

model	F1-macro	F1-micro	acc	prec	recall
unigram word	0.6539944639	0.6540000000	0.6540000000	0.5701906412	0.7704918032
bigram word	0.6211982355	0.6280000000	0.6280000000	0.5388967468	0.8922716627
unigram char	0.5572901131	0.5770000000	0.5770000000	0.5055248618	0.4285714285
bigram char	0.6319521961	0.6350000000	0.6350000000	0.5643153526	0.6370023419

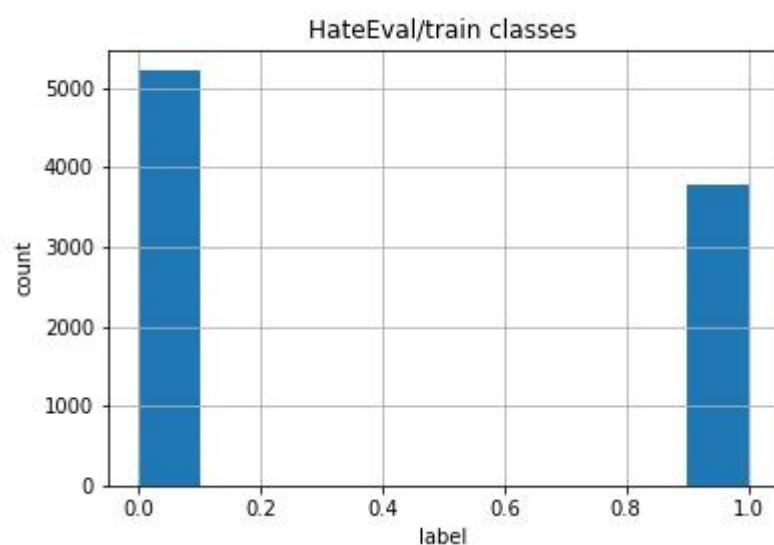
جدول ۴. عملکرد مدل روی داده های dev پس از توازن

در مقایسه جداول ۲، ۱ و ۴، ۳ این نتیجه بدست می آید که متوازن سازی داده های آموزش با استفاده از داده های جدید، عملکرد مدل را هم روی داده های Train و هم روی داده های Dev بالا خواهد برد.

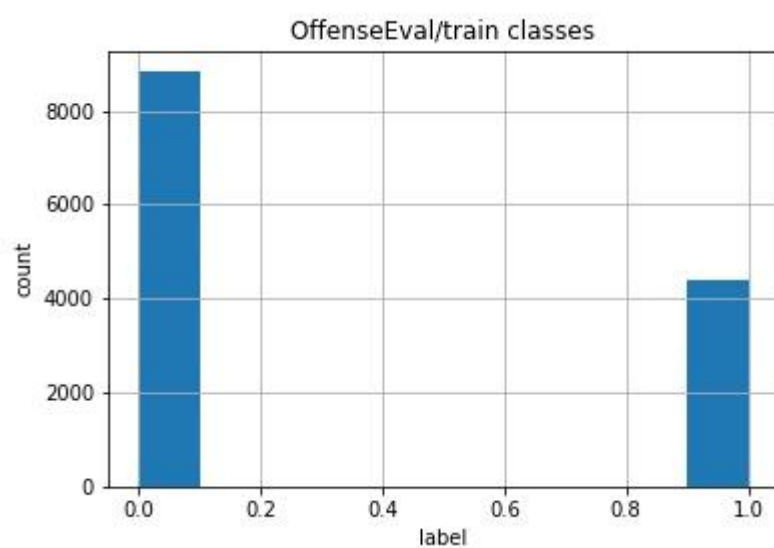
ه) طراحی مدل در Char-Level نسبت به طراحی مدل در Word-Level این مزیت را دارد که داده کمتر sparse خواهد بود. همچنین احتمال زیاد نیازی به smooth کردن جداول نباشد، زیرا که در حالت Char-Level همه حرف ها تکرار خواهند شد. همچنین بر اساس داده های جدول ۳، ۲، ۱ و ۴ می توان به این نکته رسید که مدل world-level بر روی داده ها overfit می شود زیرا که اختلاف عملکرد آن مدل بر روی داده های Train و Dev زیاد است در حالی که این اختلاف برای Char-Level بسیار کم است.

شیوه دیگری که این دو مدل را میتوان مقایسه کرد، در موضوع bias-variance trade off است. مدل Char level از آن جایی که پیچیدگی کمتری ایجاد میکند، مقدار خطای بایاس بیشتری دارد در حالی که مدل word-level مقدار خطای variance بیشتری دارد.

در قسمت پیش پردازش نیز، در حالت world level تشخیص کلمات و جملات کار دشوار تری است. به همین دلیل آموزش مدل در لایه حرف، نیاز به پیش پردازش کمتری دارد.



شکل ۷. عدم توازن در داده های Hate Train



شکل ۸. عدم توازن در داده های Off Train

سوال ۴ - دادگان تست

دادگان تست طبقه بندی شده اند و مطابق خواسته سوال در فایل Result.csv قرار گرفته است.