



دانشگاه صنعتی امیرکبیر  
( پلی تکنیک تهران )

## گزارش پروژه ۴

محمد توکلی ۹۷۳۱۰۱۴

امیر محمد راعی ۹۷۳۱۰۸۳

جدول مربوط به مدل زبانی bigram

تاثیر حذف کلمات کم تکرار	تاثیر حذف کلمات پر تکرار	لاندا ۰	لاندا ۱	لاندا ۲	اپسیلون	دقت
+	-	0.1	0.2	0.7	0.84	0.65625
-	+	0.1	0.2	0.7	0.84	0.60375
+	+	0.1	0.2	0.7	0.84	0.610625
-	-	0.1	0.2	0.7	0.84	0.62625
-	-	0.84	0.04	0.12	0.84	0.611875
-	-	0.3	0.6	0.1	0.84	0.619375
-	-	0.1	0.2	0.7	0.4	0.646875

طبق بررسی ای که بالا شده است حالات مختلف را در نظر گرفته ایم و طبق ستون دقت، حالتی که هیچ یک از کلمات پرتکرار و کم تکرار را حذف نکنیم دقت بالاتری دارد.

حال برای بررسی تاثیر لاندا ها، در زمانی که هیچ یک از کلمات کم تکرار و پرتکرار را حذف نکنیم، هر سری مقدار یک لاندا را تغییر میدهیم که به این نتیجه میرسیم که اگر لاندا ۲ را تاثیر بیشتری بدهیم دقت بالاتری میگیریم و علت آن این است که مدل bigram برای ما اهمیت بیشتری دارد.

جدول مربوط به مدل زبانی unigram

تأثیر حذف کلمات کم تکرار	تأثیر حذف کلمات پر تکرار	لاندا ۰	لاندا ۱	اپسیلون	دقت
+	-	0.3	0.7	0.84	0.501875
-	+	0.3	0.7	0.84	0.500625
+	+	0.3	0.7	0.84	0.500625
-	-	0.3	0.7	0.84	0.5025
-	-	0.88	0.12	0.84	0.49755
-	-	0.3	0.7	0.4	0.51450

مطابق جدول بالا حالتی کلمات کم تکرار و پرتکرار هردو را حذف کنیم دقت بالاتری میدهد.  
و به این دلیل است که در unigram با کلمات کار داریم و کلمات پرتکرار و کم تکرار میتوانند مارا به اشتباه بندازند.