بسم الله الرحمن الرحیم و توکلت علی الله

گزارش دقت سیستم تخصیص کلی عناصر صحنه اولیه در حالات مختلف

معیار ارزیابی: بر اساس عنصر صحنه‌های ~~یادگرفته شده توسط CRF~~ (نه زیرا در این کار هدف اندازه گیری دقت سیستم تخصیص است نه تشخیص عنصر صحنه)، لیبل خورده در کورپس

۱- در حالتی که تنها بر اساس عنصر صحنه‌های لیبل خورده در کورپس، هر عنصر در بخش مربوطه از مدل صحنه قرار میگیرد.

در این حالت ضمایر چون عنصر صحنه ندارند بدرستی تخصیص داده نمیشوند.لذا همه آنها در FN قرار میگیرند و از ریکال می کاهند.

لازم به ذکر است در این حالت 497 تعداد همه کلمات در همه عناصر صحنه (۱۲ عنصر) در کل مدل صحنه‌ها در مدل صحنه طلایی است.

total (not NO and JUNK) word nums: 497

weighted accuracies: 86.27910453828845

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| RoleEmoent evaluation:  TP:0 FN:0 FP:0 TN:1203 | RoleIntent evaluation:  TP:70 FN:0 FP:0 TN:1133  precision:100.0 recall: 100.0 F1\_score: 100.0  total number: 1203 roleIntent Number: 70 | RoleState evaluation:  TP:56 FN:0 FP:0 TN:1147  precision:100.0 recall: 100.0 F1\_score: 100.0  total number: 1203 roleState Number: 56 | RoleAction evaluation:  TP:72 FN:0 FP:0 TN:1131  precision:100.0 recall: 100.0 F1\_score: 72.28915662650601  total number: 1203 roleAction Number: 72 | Role evaluation:  TP:90 FN:69 FP:0 TN:1044  precision:100.0 recall: 56.60377358490566 F1\_score: 72.28915662650601  total number: 1203 role Number: 159 |
|  |  | DynamicObjectStates evaluation:  TP:5 FN:0 FP:0 TN:1198  precision:100.0 recall: 100.0 F1\_score: 100.0  total number: 1203 dynmic\_obejcts\_state Number: 5 | DynamicObjectActions evaluation:  TP:16 FN:0 FP:0 TN:1187  precision:100.0 recall: 100.0 F1\_score: 100.0  total number: 1203 dynamic\_obejct\_Actions Number: 16 | DynamicObjects evaluation:  TP:12 FN:1 FP:0 TN:1190  precision:100.0 recall: 92.3076923076923 F1\_score: 95.99999999999999  total number: 1203 dynamic\_obejcts Number: 13 |
|  |  |  | StaticObjectStates evaluation:  TP:10 FN:0 FP:0 TN:1193  precision:100.0 recall: 100.0 F1\_score: 100.0  total number: 1203 static\_obejct\_states Number: 10 | StaticObjects evaluation:  TP:23 FN:3 FP:0 TN:1177  precision:100.0 recall: 88.46153846153845 F1\_score: 93.87755102040816  total number: 1203 static\_obejcts Number: 26 |
|  |  |  | AlternativeTimes evaluation:  TP:10 FN:0 FP:0 TN:1193  precision:100.0 recall: 100.0 F1\_score: 100.0  total number: 1203 alternative Times Number: 10 | AlternativeLocations evaluation:  TP:56 FN:4 FP:0 TN:1143  precision:100.0 recall: 93.33333333333333 F1\_score: 96.55172413793103  total number: 1203 alternative Locations Number: 60 |

۲- حالتی که مرجع ضمیرها را بصورت دستی مشخص کرده ایم. میزان افزایش دقت سیستم در این مرحله، اهمبت تعیین مرجع ضمیر را مشخص می‌کند.

در این حالت همه پرسیژن و ریکال و اف یک و همه چی ۱۰۰ میشود ☺ بقیه مثل همین رول هستند فقط تعدادشان متفاوت است.

Role evaluation:

TP:159 FN:0 FP:0 TN:1044

precision:100.0 recall: 100.0 F1\_score: 100.0

total number: 1203 role Number: 159

total (not NO and JUNK) word nums: 497

weighted accuracies: 100.0

گزارش دقت سیستم تشخیص کلمات غیر تکراری در یک مدل صحنه

در این سیستم همه کلماتی که در همه جملات یک مدل صحنه به یک عنصر صحنه اشاره دارند، اگر بیش از یک کلمه باشند به عنوان عنصر تکراری شناخته شده و اگر تنها یک کلمه به آن عنصر صحنه اشاره کند، به عنوان کلمه غیر تکراری شناخته میشود.

۱- در حالتی مشابه حالت ۱ بالا (در حالتی که تنها بر اساس عنصر صحنه‌های لیبل خورده در کورپس، هر عنصر در بخش مربوطه از مدل صحنه قرار میگیرد.) دقت ها بشرح زیر است. در این حالت ضمایر چون عنصر صحنه ندارند بدرستی تخصیص داده نمیشوند و در تکراری و غیر تکراری بودن عناصر اختلال ایجاد میکنند و برخی از کلمات غیر تکراری را تکراری تشخیص میدهند و لذا همه در FP قرار میگرند و از پرسیژن می کاهند.)

لازم به ذکر است در این حالت 342 تعداد همه کلمات غیر تکراری در کل مدل صحنه ها برای کل عناصر صحنه (۱۲ عنصر) است.

total (not NO and JUNK) word nums: 342

weighted accuracies: 81.62862271718294

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| RoleEmoent evaluation:  TP:0 FN:0 FP:0 TN:0 | RoleIntent evaluation:  TP:70 FN:0 FP:0 TN:0  precision:100.0 recall: 100.0 F1\_score: 100.0  total number: 70 roleIntent Number: 70 | RoleState evaluation:  TP:56 FN:0 FP:0 TN:0  precision:100.0 recall: 100.0 F1\_score: 100.0  total number: 56 roleState Number: 56 | RoleAction evaluation:  TP:72 FN:0 FP:0 TN:0  precision:100.0 recall: 100.0 F1\_score: 46.35761589403973  total number: 72 roleAction Number: 72 | Role evaluation:  TP:35 FN:0 FP:81 TN:43  precision:30.17241379310345 recall: 100.0 F1\_score: 46.35761589403973  total number: 159 role Number: 35 |
|  |  | DynamicObjectStates evaluation:  TP:5 FN:0 FP:0 TN:0  precision:100.0 recall: 100.0 F1\_score: 100.0  total number: 5 dynmic\_obejcts\_state Number: 5 | DynamicObjectActions evaluation:  TP:16 FN:0 FP:0 TN:0  precision:100.0 recall: 100.0 F1\_score: 100.0  total number: 16 dynamic\_obejct\_Actions Number: 16 | DynamicObjects evaluation:  TP:7 FN:0 FP:2 TN:4  precision:77.77777777777779 recall: 100.0 F1\_score: 87.50000000000001  total number: 13 dynamic\_obejcts Number: 7 |
|  |  |  | StaticObjectStates evaluation:  TP:10 FN:0 FP:0 TN:0  precision:100.0 recall: 100.0 F1\_score: 100.0  total number: 10 static\_obejct\_states Number: 10 | StaticObjects evaluation:  TP:18 FN:0 FP:5 TN:3  precision:78.26086956521739 recall: 100.0 F1\_score: 87.8048780487805  total number: 26 static\_obejcts Number: 18 |
|  |  |  | AlternativeTimes evaluation:  TP:10 FN:0 FP:0 TN:0  precision:100.0 recall: 100.0 F1\_score: 100.0  total number: 10 alternative Times Number: 10 | AlternativeLocations evaluation:  TP:43 FN:0 FP:5 TN:12  precision:89.58333333333334 recall: 100.0 F1\_score: 94.50549450549451  total number: 60 alternative Locations Number: 43 |

۲- حالتی که مرجع ضمیرها را بصورت دستی مشخص کرده ایم. میزان افزایش دقت سیستم در این مرحله، اهمیت تعیین مرجع ضمیر را مشخص می‌کند.

در این حالت همه پرسیژن و ریکال و اف یک و همه چی ۱۰۰ میشود ☺ بقیه مثل همین رول هستند فقط تعدادشان متفاوت است.

Role evaluation:

TP:35 FN:0 FP:0 TN:124

precision:100.0 recall: 100.0 F1\_score: 100.0

total number: 159 role Number: 35

total (not NO and JUNK) word nums: 342

weighted accuracies: 100.0

گزارش دقت سیستم تخصیص دستی (رول بیسد در مقایل دیتا بیسد) رول استیت

در این حالت ابتدا بررسی میشود که اگر رول استیت صفت است و موصوف آن رول است، اگر نه اگر مضاف الیه است و مضافش رول است، اگر نه اگر در گروه اسمی که رول استیت قرار دارد، رولی وجود دارد، این رول استیت به آن رول تحصیص می‌یاید.

۱- حالت اول گروه‌های اسمی را بسیار شکستیم و به ازاء هر فعل در جمله گروه های اسمی جدایی در نظر گرفته و به جمله اصلی افزودیم مثلا در جمله:

1 روبيل N SBJ 4 null null role Arg0 ||| ROLE

2 به PREP VPP 4 null null junk Arg3 ||| JUNK

3 او PR POSDEP 2 يونس§n-23943 نفر§n-13075 role -4||| JUNK

4 گفت V ROOT 0 گفتن§v-7774 رخداد§n-13136 role\_action ||| JUNK

5 که SUBR VCL 4 null null junk Arg1 ||| JUNK

6 اين ADJ NPREMOD 7 null null junk ||| JUNK

7 کار N OBJ 9 کار§n-12781 رخداد§n-13136 role\_intent Arg1 ||| NO

8 را POSTP OBJ 9 null null junk Arg1 ||| JUNK

9 نکن V PRD 5 انجام دادن§v-21737 رخداد§n-13136 role\_action||| ROLE\_ACTION

10 زيراکه SUBR AJUCL 9 null null no ArgM\_CAU ||| JUNK

11 خداوند N SBJ 15 باري تعالي§n-11940 خصوصيت\_رواني§n-12725 no Arg0|||ROLE

12 هلاکت N OBJ 15 کشت§n-12729 رخداد§n-13136 role\_state Arg1 13||| NO

13 آن‌ها PR MOZ 12 عوام§n-13222 گروه§n-12753 role -10||| ROLE

14 را POSTP OBJ 15 null null junk Arg1 ||| JUNK

15 نمي‌پسندد V PRD 10 پسنديدن§v-21834 رخداد§n-13136 no ||| ROLE\_ACTION

16 ، PUNC PUNC 4 null null junk ||| JUNK

----روبيل به او گفت که اين کار را نکن زيراکه خداوند هلاکت آن‌ها را نمي‌پسندد ،

root: گفت

head: روبيل words: روبيل size: 1

head: به words: به او size: 2

head: که words: که size: 1

head: ، words: ، size: 1

sub roots: 9 نکن V PRD 5 انجام دادن§v-21737 رخداد§n-13136 role\_action ||| %%%%

head: کار words: کار اين size: 2

head: را words: را size: 1

head: زيراکه words: زيراکه size: 1

sub roots: 15 نمي‌پسندد V PRD 10 پسنديدن§v-21834 رخداد§n-13136 no ||| %%%%

head: خداوند words: خداوند size: 1

head: هلاکت words: هلاکت آن‌ها size: 2

head: را words: را size: 1

دقت هر یک از این سه روش به شرح زیر است. در این تست کلا ۱۲۰۳ کلمه داشتیم که ۵۶ تا رول استیت در ۳۲ مدل صحنه بود. از اینها ۲۰ مورد تخصیص داده شد و ۳۶ مورد بدون تخصیص ماندند. همه ۳ تخصیص صفتی درست بودند (۱۰۰ ٪). ۶ مورد از ۷ مورد تخصیص مضاف الیه ای درست بود (۷/۸۵ ٪). همه ۴ مورد تخصیص مسندی درست بودند (۱۰۰ ٪). ولی تنها ۲ مورد از ۷ تخصیص گروه اسمی درست بود (۵/۲۸ ٪). در ضمن ۳۶ مورد رول استیت در گروه اسمی (کوچک)ش هیچی رولی نداشت!

یک مورد تخصیص مضافی اش غلط بود و چون در سیستم تست جواب را میدانیم و چک می کنیم که اگر غلط بود با روش گروه اسمی هم پیدا کند، لذا عملاً تعداد تخصیص گروه اسمی یکی اضافی است. آن مورد اضافی هم جوابش غلط است.

all RoleStateNumbers: 56

Adj RoleStateNumbers: 3 ۳ تا درست

Moz RoleStateNumbers: 7 ۶ تا درست

Mosnad RoleStateNumbers: 4 ۴ تا درست

chain RoleStateNumbers: 7 ۲ تا درست

allocated RoleStateNumbers: 21

non-allocated RoleStateNumbers: 36

correct allocations: 15 and fault allocations: 5 from: 21 allocations.

correct alloation in allocated RoleStates: 75.0%

correct alloation in all RoleStates: 26.785714285714285%

۲- در این حالت گروه‌های اسمی را تنها با محوریت فعل اصلی جمله انجام دادیم و بدین ترتیب گروه‌های اسمی بسیار طولانی بدست میآیند. جمله بالا در این حالت تبدیل به گروه‌های اسمی زیر میشود:

----روبيل به او گفت که اين کار را نکن زيراکه خداوند هلاکت آن‌ها را نمي‌پسندد ،

root: گفت

head: روبيل words: روبيل size: 1

head: به words: به او size: 2

head: که words: که نکن کار اين را زيراکه نمي‌پسندد خداوند هلاکت آن‌ها را size: 11

head: ، words: ، size: 1

دقت در این حالت به شرح زیر است:

all RoleStateNumbers: 56

Adj RoleStateNumbers: 3 ۳ تا درست

Moz RoleStateNumbers: 7 ۶ تا درست

Mosnad RoleStateNumbers: 4 ۴ تا درست

chain RoleStateNumbers: 23 ۱۲ تا درست

allocated RoleStateNumbers: 37

non-allocated RoleStateNumbers: 20

correct allocations: 25 and fault allocations: 11 from: 37 allocations.

correct alloation in allocated RoleStates: 69.44444444444444%

correct alloation in all RoleStates: 44.642857142857146%

در این حالت تعداد تخصیص های گروه اسمی بیش از ۳ برابر شده است! از ۳ تخصیص صفتی همه درست (۱۰۰ ٪). از ۷ تخصیص مضافی ۶ مورد درست (۷/۸۵ ٪)،‌ از ۴ تخصیص مسندی همه درست بودند (۱۰۰ ٪) و از ۲۳ تخصیص گروه اسمی ۱۲ مورد درست بودند (۲/۵۲ ٪).

یک مورد تخصیص مضافی اش غلط بود و چون در سیستم تست جواب را میدانیم و چک می کنیم که اگر غلط بود با روش گروه اسمی هم پیدا کند، لذا عملاً تعداد تخصیص گروه اسمی یکی اضافی است. آن مورد اضافی هم جوابش غلط است.

در مجموع در این حالت ۴۴ درصد رول استیت ها درست تخصیص داده شدند در مقایسه با ۲۶ درصد حالت ۱. دقت کل تخصیص‌ها (هر چهار نوع) نیز از ۶۸ به ۶۹ ارتقا پیدا کرد.

گزارش دقت سیستم تخصیص دستی (رول بیسد در مقایل دیتا بیسد) رول اکشن

لازم به ذکر است تعداد رول اکشن ها یکی کم شد. یک رول اکشن اشتباه بود و به خدا برمیگشت که no شد.

در این حالت فاعل جمله ای که رول اکشن در آن قرار دارد را بررسی می‌کنیم و اگر رول بود، به آن تخصیص می‌دهیم.

۱- بررسی فاعل‌ متعلق به جمله

all RoleActionNumbers: 71

Sbj RoleActionNumbers: 49

allocated RoleActionNumbers: 49

non-allocated RoleActionNumbers: 22

correct allocations: 43 and fault allocations: 6 from: 49 allocations.

correct alloation in allocated RoleActions: 87.75510204081633%

correct alloation in all RoleStates: 60.56338028169014%

ملاحظه میشود ۶۹٪ رول اکشن ها در این شیوه تخصیص داده شدند، یعنی ۴۹ تا از ۷۱ای. از بین اینها ۸۷٪ درست تشخیص داده شدند، یعنی ۴۳ مورد از ۴۹ تخصیص. از کل ۷۱ رول اکشن ۶۰٪ درست تشخیص داده میشوند که دقت خوبی است ولی جای ارتقا دارد!

۲- البته بهتر است که فاعل را به خود فعل نسبت داد و نه جمله.

۳- فعلهایی که ضمیر عطفی دارند که فلان کار را کرد و فلان ... فلان. اینها بررسی شوند.