|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| رفع ابهام ضمیر-گروه اسمی | نوع متن ورودی  (جمله پاراگراف) | قبول عبارت و جمله به عنوان آرگومان | تعداد آرگومان | تعیین نوع موجودیت | تعیین جایگاه رابطه  (تعمیم و تخصیص) | محل استخراج رابطه  (فعل-اسم-قید) | سال | Open IE |
| ندارد |  |  |  |  |  |  | 2007 | TextRunner |
| ندارد |  |  |  | دارد |  |  | 2009 | SnowBall |
| ندارد |  |  |  |  |  |  | 2010 | WOE |
| ندارد |  | ندارد | باینری | ندارد | ندارد | فعل | 2011 | ReVerb |
| ندارد |  | ندارد | باینری | ندارد | ندارد | فعل | 2011 | R2A2 |
| ندارد |  | ندارد | باینری | ندارد | ندارد | فعل و اسم | 2012 | OLLIE |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| مشکلات | تعیین نوع موجودیت | محدودیت تعداد رابطه مستخرج | تعیین نوع رابطه | ایده اصلی | سال | Methods/ Proposer Name |
| تهیه دادگان برچسب خورده-  محدود به متن و ژانر خاص-  Doesn’t generalize to different relations |  | دارد |  |  |  | Supervised  {feature based,  Kernel based} |
| Exhaustively iterative to expanding the patterns-  Semantic drift-  Error propagation in the period of time  - Sensitive to original set of seeds- Requires that we have seeds for each relation- Generally have lots of parameters to be tuned- No probabilistic interpretation(Hard to know how confident to be in each result) |  | ندارد |  |  | 1992 | Bootstrapping |
| نویز در داده های آموزشی  محدود به متن و ژانر خاص نسیت | دارد | دارد | دارد | ترکیب روش Bootstraping با Supervised learning | 2009 | Distant Learning |
| Low recall | ندارد | ندارد | ندارد |  |  | Pattern based |
| مشکل در استخراج روابط نهفته در گروههای اسمی-  اطلاعی از نوع آرگومانهای دخیل و ارتباط روابط نمی­دهد |  | ندارد |  |  | 2007 | Unsupervised (Open IE) |
| گیر کردن در مینیمم محلی بدلیل استفاده از الگوریتم EM در ساختارشان- پیچیدگی زیاد مدل و پارامترهای آموزشی |  |  |  |  |  | Multi-instance learning (MIML) |
| Semi-automatic |  |  |  |  |  | ترکیب DS با Active Learning |
| گیر کردن در مینیمم محلی |  |  |  |  |  | Graphical  Models |
| Semi-automatic  Semantic shift  Low precision | دارد | ندارد | دارد | دو هدف اصلی دارند   1. خواندن وب و استخراج اطلاعات از آن 2. هر روز این کار را بهتر از دیروز انجام بدهند   ایده اصلی کارشان Coupled learning | 2010 | NELL  (Coupled semi-supervised learning) |
|  |  |  |  |  |  |  |

ایده: روش های Open IE به تعداد بسیار زیادی رابطه استخراج میکنند. مشکل خروجی این روشها در ندانستن نوع روابط و نوع موجودیت های دخیل در رابطه است. در واقع عملا اطلاعات زیادی را بدون دسته بندی مناسب در اختیار کاربر قرار میدهند.

از طرف دیگر، بقیه روش های غیر Open IE مانند Distant learning, Bootstraping, … اطلاعات کاراتری را در اخیتار ما میگذارند ولی محدود به یکسری رابطه و موجودیت خاص هستند. اصلا نمیتوانند به اندازه روش های Open IE رابطه استخراج کنند.

حالا ایده این است که مزایای هر دو پارادایم را با هم ترکیب کنیم. در مرحله اول از روشهای Open IE برای استخراج رابطه بهره ببریم، سپس به نحوی نوع رابطه و موجودیت های دخیل در آن را تعیین کنیم و در واقع به روابط برچسب بزنیم (نقل قول، کمپانی تکنولوژی، ویژگی های آدم شامل نام و سن و محل ..، رابطه عام-در واقع هرچی رابطه ماند که برچسب نخورد این برچسب را به آن میدهیم- ، برای استخراج روابط بیشتر به Open Calais مراجعه شود ).