بسمه تعالی

گزارش کتبی مقاله

Hybrid query expansion model for text and microblog information retrieval

تهیه کننده

مرتضی عیدی پور

۹۸۱۱۶۳۴

استاد : سعید فرضی

فهرست

معرفی

عنوان مقاله :

Hybrid query expansion model for text and microblog information retrieval

شناسه ی یکتا :

https://doi.org/10.1007/s10791-017-9326-6

سال انتشار:

03 February 2018

منتشر کننده :

Springer

خلاصه (۱)

گسترش پرسمان یکی از عملیات های مهم در زمینه ی بازیابی داده برای بهبود پرسمان کاربر به حساب می آید که باعث باز گرداندن مطالب مرتبط میشود.

در این مقاله ما به بررسی روش ترکیبی (HQE) می پردازیم و خواهیم دید که چگونه منابع خارجی میتوانند با قوانین استخراج رابطه ترکیب شده و برای بهبود تولید و انتخاب ترم به کار آیند.

HQE میتواند ساختار های متفاوتی داشته باشد، شروع با استفاده از قوانین ارتباط و ترکیب آن با دانش خارجی.

HQE دو فاز اصلی توسعه ی پرسمان (QE) را که میتوان از آن ها با نام های فاز تولید ترم های کاندید و فاز انتخاب یاد کرد را در برمیگیرد، ما برای فاز اول روش های آماری، معنایی و مفهومی (Conceptual) را برای ساخت ترم های مرتبط با پرسمان استفاده میکنیم، برای بخش دوم یک معیار مشابهت معرفی میکنیم که با نام ESAC شناخته میشود و براساس تحلیل صریح معنایی (ESA) میزان ارتباط بین پرسمان و مجموعه ی ترم های کاندید را محاسبه میکند.

کارایی HQE توسط دو آزمایش مورد بررسی قرار گرفته است:

۱. جستجوی توییت TREC Microblog Track 2011
۲. مجموعه ی موضوعات سخت از TREC Robust 2004

که در مورد دوم با زمینه سازی در مورد توییت ها همراه خواهد بود (INEX 2014)، نتایج حاصل نمایانگر تاثیر گذاری مدل HQE و قوانین استخراج ترکیب شده با منابع خارجی میباشد.

نتیجه گیری (۲)

در روش ترکیبی خود برای گسترش پرسمان های از تو بخش، تولید ترم ها و انتخاب ترم ها استفاده می کردیم و نشان دادیم چگونه می توان با ترکیب منابع خارجی با قوانین استخراج ارتباط کارایی را افزایش داد.

ما برای بخش اول عملیات خود از توابع محلی، کلی و خارجی (local , global , external) برای ساختن ترم های کاندید استفاده کردیم، HQE ترم ها را براساس میزان ارتباط آن ها با پرسمان فیلتر کرده و فقط مرتبط ترین ها را نگه میدارد.

در میان تمام آزمایشات خود (CLEF INEX 2013 , TREC Robust track 2004 , hard topic TREC 2011 and 2014) به این نتیجه رسیدم که فیلتر کردن میتواند برای بهبود نتیجه مناسب باشد وقتی ارتباط خوبی با منابع خارجی داشته باشیم.

برای استخراج، معیار آماری بهترین نتیجه را بدست آورده، درحالی که برای زمینه سازی درمورد تویت ها استفاده از ترکیب همه ی روش های استخراج بهتر میباشد.

کارهای آتی (۳)

در کار های آتی میخواهیم با وزن دادن به ترم های پرسمان اصلی اهمیت آن ها را افزایش داده و از هرگونه دور شدن از موضوع (drift) خودداری کنیم.

این وزن ها میتوانند همان امتیاز شباهت باشند !

ولی همچنین تکرار گسترش بین چندین روش را نیز مدنظر قرار دهیم، به این معنی که وقتی ترم توسعه توسط چندین روش توسعه (مثلا هم آماری و معنایی) تولید میشود احتمالا از اهمیت بیشتر برخوردار خواهد بود.

همچنین بررسی خواهیم کرد چگونه میتوان گسترش پرسمان را توسط بردار های جایگذاری شده (embedding vectors) بهبود بخشید، در این روش باید به دقت ترم های نامناسب را فیلتر کرد زیرا جایگذاری کردن بر بخش های مشترک کلمه اتکا میکند.

کارهای مرتبط (۴)

Almasri, M., Berrut, C., & Chevallet, J. 2016

A Comparison of Deep Learning Based Query Expansion with Pseudo-Relevance Feedback and Mutual Information

خلاصه:

روش های بسیاری برای گسترش پرسمان عرضه شده اند که کارای را افزایش داده اند اما یکی از مشکلات پیداکردن رابطه بین ترم ها برای گسترش پرسمان میباشد در این مقاله میخواهیم با استفاده از روش های یادگیری عمیق این کار را انجام دهیم و به صورت تجربی روش های مختلف را با یکدیگر مقایسه کنیم.

پرسمان های کاربر معمولا بسیار کوتاه هستند تا بتوانند منظور کاربر را به صورت دقیق بیان کنند ، ممکن است کلماتی در آن کم باشند .

برای حل این مشکل از چندین منبع داده برای گسترش پرسمان استفاده میکنیم ، اما انتخاب ترم های مناسب چالش برانگیز خواهد بود.

در روش های استفاده شده (دیپ) یک ترم به صورت یک شیئ ریاضی در ابعاد بالا در می آید و با استفاده از یک معیار میتوان مرتبط بودن را بررسی کرد.

در بخش ترم های گسترش با استفاده از حجم زیادی از داده های متنی بدون ساختار عملیات آموزش را انجام می دهیم ، بردار های خروجی شامل رابطه بین ترم ها میشود به عنوان مثال رابطه ی بین شهر برلین و کشور آلمان ، هر بردار به صورت Vt در نظر گرفته میشود به آن بردار یادگیری عمیق میگوییم .

شباهت بین دو ترم بر اساس معیار کو سینوسی (cos(Vt1 , Vt2)) که در آن اندازه ها نرمال شده اند اندازه گرفته میشود.

برای گسترش پرسمان کاربر آن را به صورت بسته ای از کلمات (Bag of Word) در نظر میگیرم و برای هر ترم آن یک مقدار تکرار به صورت (t,q) داریم و برای گسترش براساس همه ی ترم های موجود در پرسمان ، ترم های مشابه آن ها را بدست می آوریم و در نهایت گسترشی انتخواب میشود که حداکثر شباهت را داشته باشد و تکرار ترم ثابت بماند و در نهایت :

$$\#(t', q') = \alpha \times \#(t, q') \times \widetilde{cos}(v_t, v_{t'})$$

که در آن آلفا یک مقدار بین صفر و یک میباشد و اهمیت ترم گسترش را نشان میدهد.

بخش آموزش خود را براساس CLEF و بر اساس بخش های Image2009 ، Case2012 ، Case2011 انجام دادیم که شامل بیش از ۴۰۰ ملیون کلمه میشدند.

نتایج بدست آمده از روش های نشان دهنده ی بهبود های آماری نسبت به روش های PRF و MI می باشد !

یادگیری عمیق فقط برای یادگیری از داده های آموزشی مناسب نیست بلکه میتوان از آن برای داده های دیگر نیز استفاده کرد و لیل قدرت آن آموزش زیادی است که دیده است.

Belalem, G., Abbache, A., Belkredim, F. Z., & Meziane, F 2016
Arabic Query Expansion Using WordNet and Association Rules

خلاصه:

گسترش پرسمان به معنی اضافه کردن ترم به پرسمان جهت افزایش کارایی میباشد و نشان داده شده است که استفاده از WordNet کمکی به افزایش کارایی نمیکند و چالش گسترش پرسمان در انتخاب ترم مناسب میباشد.

در این مقاله به بررسی WordNet عربی و استفاده از قوانین ارتباط استخراج شده از زبان عربی و تابع انتخاب مناسب پرداخته میشود.

یکی از مشکلات گسترش پرسمان عدم تطابق ترم های پرسمان میباشد که راه حل های بسیاری برای آن عرضه شده است به عنوان مثال AQE (Automatic Query Expansion) که ترم های مرتبط را به دو دسته ی محلی و کلی (local & global) برای تحلیل تقسیم میکند.

- روش های کلی از منابع دانش (داده) خارجی استفاده میکنند (WordNet/ …)
- روش های محلی از متن های باز گشت داده شده از پرسمان استفاده میکنند (بازخورد رابطه ای)

راه های مختلفی برای انتخاب ترم های گسترش وجود دارند که WordNet به عنوان یکی از آن ها شناخته شده است.

زبان عربی یک زبان آوایی میباشد بنابر این نیاز به اضافه کردن علایم آوایی به متن برای تعیین صحیح تلفظ کلمه میباشد و از طرف دیگر کلمه بدون علایم آوایی میتواند معانی متفاوتی داشته باشد از این رو ایجاد ترم های گسترش در عربی میتواند سخت تر باشد درحالی که این مشکلات در زبان های وجود ندارند.

در این مقاله ما در نظر میگیرم که ترم های موجود در نتیجه ی بازگشت داده شده برای پرسمان با پرسمان مرتبط بوده و میتوان از ترم های موجود در آن برای گسترش پرسمان استفاده کرد.

در WordNet مجموعه ای از کلمات عربی وجود دارند در این مجموعه (همانند دیکشنری) برای هر کلمه مجموعه ی هم معنی وجود دارد که به آن sybset میگویند، باید در نظر داشت که WordNet اطلاعات بیشتری نیز فراهم میکند.

در پردازش ما چند قدم را انجام میدهیم ، ابتدا پیش پردازش برای حذف کلمات توقف (به دلیل بار معنایی پایین آن ها) در مرحله ی بعد استخراج و انتخاب هم معنی ها با استفاده از WordNet را انجام می دهیم.

استخراج توکن های کارآمد با استفاده از تحلیل گر (Query Analysis) انجام میشود و این تحلیل گر از عملیات استمینگ (light-stemmin) استفاده میکند.

Bellot, P., Moriceau, V., Mothe, J., SanJuan, E., & Tannier, X. 2016
INEX Tweet Contextualization task: Evaluation, results and lesson learned

خلاصه:

خدمات میکروبلاگ هایی همانند تویتر در حال استفاده ی روز افزون به عنوان ابزار هایی برای بازار یابی و مارکتینگ هستند و این انگیزه ی این مقاله میباشد.

هدف این مقاله آگاه سازی کاربر درمورد تویت با استفاده از یک خلاصه کوتاه (۵۰۰ کلمه) میباشد، این خلاصه باید به صورت خودکار و با استفاده از ویکیپدیا و استخراج متن های مرتبط و تجمیع آن ها ساخته شود.

این مقاله به بررسی عملکرد سیستم های ارایه شده در ۴ سال اخیر و نتایج آن ها می پردازد.

همچنین راه ها و منابع آزاد (open) ساخته شده در این مدت و معیار های ارزشیابی (validation) را نیز بررسی خواهیم کرد.

ما از معیار LogSim برای ارزیابی خلاصه ها استفاده و درنهایت نکاتی را که در این زمینه وجود دارد بررسی میکنیم.

ما هیچ گاه در هنگام نوشتن تمام منظور خود را نمینویسیم!، نویسنده ها بر خوانندگان برای فهمیدن چیز های نوشته نشده حساب میکنند، این موضوع کاملا درست است وقتی درمورد تویت صحبت میکنیم.

دانش مکانی و زمانی و تاریخ در این زمینه (فهمیدن تویت) به ما بسیار کمک میکنند.

سیستم های حساس استخراج داده : سیستم هایی که از دانش های خارجی(منابع خارجی) برای انجام کار خود استفاده میکنند.

مشارکت یکی از روش های ساخت توضیح معنادار برای یک تویت میباشد از آنجایی که یک تویت حداکثر ۱۴۰ کاراکتر میباشد این روش خود در INEX 2011 بررسی شده و به CLEF 2012 تبدیل شد.

شرح نتایج اصلی:

۱. بازگردانی موثر جمله یک چالش است و یک زیر وظیفه برای زمینه سازی در مورد تویت در میان داده های زیاد میباشد.
۲. جملات کوتاه تر از آن هستند که به عنوان متن های اتمیک درنظر گرفته شوند و زمینه سازی برخط (online) بهتر از ایندکس کردن و زمینه سازی غیر برخط (offline) عمل میکند.

۳. بهترین سیستم ها باید متن های بازگشتی را ترکیب کنند، بخش بندی جمله و امتیاز دهی را انجام دهند. (معیار وا گرایی متن ، جابجایی جملات و ...)
۴. مطلوب بودن کلی خلاصه ها میتواند از طریق ترکیب کردن جملات و ان گرام (n-gram) مشخص شود (اگر خوانایی مدنظر باشد)
۵. نیرومندی و پایداری سیستم میتواند افزایش پیدا کند اگر از منابع خارجی علاوه بر ویکیپدیا استفاده کنیم.

Colace, F., Santo, M. D., Greco, L., & Napoletano, P 2015

Improving Relevance Feedback-Based Query Expansion by the Use of a Weighted Word Pairs Approach

خلاصه :

در این مقاله روشی جدیدی برای گسترش پرسمان مورد برسی قرار میگیرد، در این روش پرسمان به صورت زوج کلمات وزن دار (WWP) در می آید، مدل های بولین میتوانند از WWP استفاده کنند.

در یک سیستم بازیابی داده سوال مهم این است که سیستم چگونه میتواند تشخیص دهد که یک متن با پرسمان مرتبط است یا خیر ؟ کدام نتیجه ها مرتبط تر هستند ؟

برای جواب دادن به این سوال چندین مدل بولین، ریاضیاتی و احتمالاتی ارائه شده است.

با وجود اینکه هر روش ویژگی های خود را دارد اما بیشتر روش ها متن را به عنوان بسته ای از کلمات (Bag of Words) در نظر میگیرند، با این روش داده ها در مورد موقعیت یک کلمه درون متن به طور کامل از دست میرود.

نزدیکی یک پرسمان به یک متن را میتوان از طریق محاسبه ی فاصله در فضای برداری بدست آورد.

مشاهدات نشان میدهد که پرسمان هایی که توست کاربران تولید میشود معمولاً طول کمی دارند (۲ تا ۳ کلمه به طور متوسط) به همین دلیل سیستم هایی که براساس تکرار کلمات کار میکنند نمیتوانند عملکرد قابل قبولی ارائه دهند.

هدف گسترش پرسمان ارائه اضافه کردن ترم به پرسمان برای کاهش خطای عدم تطابق متن و پرسمان است.

براساس نظریه ی بازیابی داده فضای برداری روشی موثر برای بازنمایی متن میباشد، در حقیقت متن را میتوان به صورت مجموعه ای از کلمات وزن دار نوشتن (به صورت برداری).

هر وزن نمایانگر میزان ارتباط کلمه با متن میباشد، در معیار tf-idf معمولا رابطه ی مستقیم با میزان تکرار کلمه و رابطه ی عکس با میزان تکرار یک کلمه در متن نسبت به سایز متن دارد.

$$sim(\mathbf{q}, \mathbf{d}) = \sum_{t \in \mathbf{q} \cap \mathbf{d}} w_{t,\mathbf{q}} \cdot w_{t,\mathbf{d}},$$

معیار شباهت به صورت بالا تعریف میشود مه درآن Wd وزن کلمه در متن و Wq وزن کلمه در پرسمان میباشد.

برای افزایش کارایی PRF میتوان موضوعات دیگر را نیز در نظر گرفت، این موضوعات میتوانند توسط کاربر و یا به صورت خودکار انتخاب شوند و کارایی PRF را افزایش دهند.

Al-Shboul and Myaeng 2014

Wikipedia-based query phrase expansion in patent class search

خلاصه :

حل مشکل دور شدن از موضوع اصلی به دلیل نام گذاری نا مناسب عنوان ، فهم نادرست موضوع می تواند مشکلات قانونی نیز ایجاد کند ! (هنگام ثبت اختراع) ، در این مقاله سعی میکنیم ابهام موضوع را با جستجو در مقالات ویکیپدیا حل کنیم و فرض شده که موضوعات به صورت سلسه مراتبی باهم در ارتباط هستند .

پردازش الگو درمتن به دلیل سبک های متفاوت نوشتن معمولا مشکل می باشد ، مشکل به دو دسته تقسیم می شود ، دسته ی اول نویسنده هایی که نوشته های خود را به گونه ای مینویسند که مبهم باشد ، دسته ی دوم کاوشگرانی که میخواهند به همه ی پتنت (اخراعات ثبت شده) مشابه دسترسی پیدا کنند.

گسترش کوری براساس ویکیپدیا (WQPE)

روش گسترش کوری ما هم بر مبنای گسترش کلمات کوری هم گسترش عبارات عمل میکند ، درحالی که گسرش عبارات کلید اصلی میباشد ، از این کار دوهدف را داریم ، افزایش دانش متنی (Contextual information) و کاهش عدم تطابق کلمات پرسمات و نتایج .

برای انجام عملیات صفحات ویکیپدیا پیش پردازش شده اند ، به صورتی که از لحاظ معنایی خلاصه سازی شده باشند.

ما مراحل فیلتر کردن عبارات توقف ، پس برچسب گذار (post tagger) و استفاده از قواعد نوشتاری (RegEx) برای پرسمان انجام می دهیم (این عملیات برای پرسامان های کوتاه بسیار کار آمد می باشد که از مقادیر آماری نمیتوان به خوبی استفاده کرد) .

عبارات پرسمان و کلمات آن به دو دسته ی جداگانه تقسیم میشوند و هرکدام به صورت جداگانه با اسفاده از ویکیپدیا گسترش داده میشوند ، سپس از ترکیب دو روش برای بدست آوردن نتیجه استفاده میشود .

ما بر راه هایی که باضافه کردن ترم های مرتبط به پرسمان کار میکنند تمرکز میکنیم.

در این مقاله ما از گسترش عبارات ، گسترش کامات و ترکیب هردو استفاده کرده ایم و نتایج حاصل را با روش های دیگر مانند RM و WN-Gloss مقایسه کرده ایم.

ما راه حلی را ارایه می دهیم که از ترکیب ویکیپدیا و وردنت (WordNet) برای توسعه ی پرسمان استفاده میکند (IPC-class) راه حل مانشان داده که اضاف کردن کلمه به پرسمان باعث کاهش تاثیر دور شدن از موضوع اصلی (drift away) میشود.

Bouchoucha, A., Liu, X., & Nie, J.-Y 2014
Integrating Multiple Resources for Diversified Query Expansion

خلاصه:

هدف پوشش دادن جنبه ها مختلف یک پرسمان کوتاه و مبهم است.

بیشتر روش ها از یک منبع خارجی مثل ConceptNet استفاده میکنند که پوشش آن توسط متن کم است برای حل این مشکل ما از چندین منبع خارجی استفاده میکنیم و سپس نتایج باز گردانده شده را باهم ترکیب کرده و با استفاده از حاصل پرسمان را گسترش می دهیم سپس با استفاده از حداکثر فاصله ی مرتبط بهترین آن ها را انتخاب میکند.

برای ارزیابی روش خود از داده های TREC استفاده میکنیم.

در روش های مرسوم گسترش پرسمان (مثلا PRF) از نتایج بدست آمده برای پرسمان برای گسترش آن استفاده میشود و نتایج آن ها وابسته به نتایج بازگردانده شده است.

مزیت گسترش پرسمان متنوع شده (DQE) استفاده از چندین منبع خارجی همانند ConceptNet ، Wikipedia و query log برای تولید مجموعه ای متنوع میباشد ، انگیزه ما از استفاده از چندین منبع این میباشد که پرسمان های بسیاری وجود دارند که با استفاده از ConceptNet نمیتوان آن ها را به خوبی گسترش داد (به عنوان مثال TREC 2009).

$$c^* = argmax_{c \in C_{r,Q}} (\lambda_r \cdot sim(c, Q) - (1 - \lambda_r) \cdot max_{c_i \in S_{r,Q}} sim_r(c, c_i))$$

Sim یک تابع برای امتیاز دادن به شباهت می باشد .

لاندا پارامتر ی میباشد که تعادل بین مرتبط بودن و متنوع بودن را برقرار میکند.

برای انجام عملیات خود از حذف کلمات توقف و Krovetz stemmer استفاده میکنیم.

در عمل بااستفاده از Wikipedia در TREC 2009 و TREC 2010 به برتری رسیدیم درحالی که در TREC 2011، ConceptNet عملکرد بهتری داشت !، پس از بررسی نتایج به این نتیجه ریسیدم که پرسمان های موجود در 2010 , 2009 در Wikipedia وجود داشتند به همین دلیل عملکرد خیلی خوبی در این محدوده داشته در حالی که برای TREC 2011 این پرسمان ها موجود نبوده اند.

در پرسمان های سخت احتمالا متن های برگردانده شده نا مرتبط هستند و این متن ها خود دارای خطا (noise) فراوان برای سیستم ما هستند، در نهایت بهترین عملکرد زمانی بدست می آید که همه ی روش ها را باهم ترکیب کنیم.

Liu, C., Qi, R., & Liu, Q 2013
Query Expansion Terms Based on Positive and Negative Association Rules

خلاصه :

از مدل های جدید که شامل قوانین ارتباط مثبت و منفی هستند استفاده میکنیم و با تبدیل متن های دیتابیس به بردار های بولین، الگوریتم ما میتواند ترم های مرتبط و نا مرتبط را براساس ضرب بردار ها تولید کند و قوانین ارتباط مثبت و منفی را بدست آورد.

این روش سریع بوده و تنها یک بار دیتابیس را کاوش میکند و علاوه بر آن حافظه ی کمی مصرف میکند که برای پردازش های بازیابی داده بسیار اهمیت دارد.

گسترش پرسمان میتواند با اضافه کردن ترم های به پرسمان اصلی طول آن را افزایش داده و نارسایی های آن را برطرف نماید، یکی مشکلات گسترش پرسمان انتخاب ترم مناسب می باشد که اخیراً روش های که بر مبنای قوانین ارتباط عمل میکنند مورد توجه بیشتری قرار گرفتهاند و اکثر محقق ها بر قوانین مثبت تمرکز میکنند.

قوانین منفی به دو صورت $T_1 \Rightarrow \neg T_2$ و $\neg T_1 \Rightarrow T_2$ تعریف میشوند.

ماتریس ها بولین به فرمت روبرو نمایش داده میشوند R=(a1,a2,….am) که در آن a یک بردار با m بعد میباشد، ابعاد بردار میتواند با ذخیره فقط بیت مقادیر فشرده سازی شود.

بخش اول الگوریتم: پیدا کردن ترم های پر تکرار و غیر پر تکرار در دیتابیس و باز گرداندن ترم های پر تکرار.

سپس عملیات متن کاوی دیتابیس به عملیات تحلیل بردار های بولی تبدیل میشود و سپس فشرده سازی بردار ها و بعد از آن عملیات مقایسه ی ضرب بردار ها با حد های بالا و پایین برای تعیین پر تکرار و غیر پر تکرار بودن ترم، مهمترین بخش الگوریتم همین بخش میباشد (min_sup).

در مقایسه با روش های مشابه این الگوریتم تنها یک بار دیتابیس را اسکن میکند و دیگر نیازی به انجام مجدد این کار ندارد و الگوریتم ما از چند سطح برای ساخت ترم ها استفاده میکند و همچنین دارا رابطه های منفی میباشد که باعث افزایش کارایی میشود.

در الگوریتم ما اگر وابستگی بین ترم و پرسمان بزرگ تر از یک باشد ترم جدید به پرسمان اضافه میشود و اگر کوچک تر از یک باشد ترم برای حذف شدن از پرسمان کاندید میشود.

در حقیقت الگوریتم ما گسترش پرسمان فقط در اضافه کردن ترم به آن نمیبیند و درصورت لزوم ترم را از آن نیز حذف میکند.

Luo, J., Meng, B., Liu, M., Tu, X., & Zhang, K 2012
Query Expansion using Explicit Semantic Analysis

خلاصه:

گسترش پرسمان روشی در حیطه ی بازیابی داده میباشد که تلاش میکند عدم تطابق پرسمان و متن را کاهش دهد، یک معیار ظهور مجدد ترم میباشد اما میدانیم که این معیار مناسب نیست زیرا برخی ترم ها کلمات پس زمینه ای متن هستند!

به همین منظور ما از تحلیل معنایی صریح (ESA) استفاده میکنیم که از ترکیب که دو وزن را حساب میکند:

۱. وزن ترم های استخراج شده از متن های حاصل از نتیجه ی پرسمان در درجه ی اول
۲. وزن ترم های استخراج شده از جستجوی ترم مورد نظر به عنوان کلید واژه در Google

وزن های بدست آمده برای انتخاب ترم مناسب برای گسترش پرسمان استفاده میشوند.

معمولا افراد منظور خود با کلمات کلیدی در پرسمان بیان میکنند، در بسیاری از موارد متن های بازگردانده شده توسط سیستم های بازیابی داده ارتباط زیادی با پرسمان جستجو شده ندارد و این یک مشکل خود یک مشکل اساسی است.

دو بخش اصلی در همه ی سیستم های توسعه ی پرسمان وجود دارد:

۱. منابع برای انتخاب ترم های کاندید (ساختن ترم ها)
۲. وزن دهی به ترم برای انتخاب ترم مناسب برای توسعه

دور روش کلی و جزئی برای تولید ترم ها وجود دارد (در کل متن ها – در متن های بازگردانده شده) و مشکل روش جزئی این است که اگر متن های بازگردانده شده برای یک پرسمان نا مرتبط باشند نتایج گسترش پرسمان نیز لاجرم نامطلوب خواهند بود.

ESA به صورت صریح معنی هر متن را به صورت بردار وزن داری از مفاهیم (concept) بیان میکند، که توسط انسان ها تعریف شده اند! هنگامی که ماشین به دانش انسانی (هوش انسانی) برای انجام کاری احتیاج دارد طبیعی است که از دانشنامه ی آزاد استفاده کنیم، بدین منظور از Wikipedia استفاده میکنیم

روش ESA متن را با توجه به مفاهیم آن به ابعاد بالاتر میبرد

$$\Phi: T \to \mathbb{R}^{|W|}$$
$$\Phi(t) := \langle v_1, \dots, v_{|W|} \rangle$$

در فرمول بالا قدر مطلق W برابر است با تعداد مقالات موجود در Wikipedia، مقدار vi قدرت بین بردار متن (text) و مقاله ی a را نشان میدهد.

وزن کلی را میتوان با جمع کردن وزن کل وزن ها بدست آورد، یک روش بدست آوردن چنین وزنی در نظر گرفتن متن به بسته ای از کلمات (BOW) است (در اینجا از آن استفاده میشود).

Jabeur 2012

Uprising microblogs: A Bayesian network retrieval model for tweet search

خلاصه :

دارین مقاله به بررسی ، دسترسی برخط به داد با شبکه های بیز برای تویت ها میپردازیم ، مدل ارایه شده به ارتباط بین تویت ها به عنوان احتمال شرطی نگاه میکند .

در حقیقت مدل ارایه شده علاوه بر دقت به شباهت های متنی به جریان داده های میکروبلاگ (تویت) (براساس تگ ها) نیز توجه می کند.

برای بررسی عملکرد مدل خود از TREC Tweets2012 استفاده میکنیم.

سرویس های میکروبلاگ خدماتی برای اطلاع رسانی و ارتباط هستند! ، ویژگی پست ها در این گونه سیستم ها کوتاه بودن پست ها و وابسته بودن آن ها به زمان میباشد. (در این مقاله تویتر به عنوان معروف ترین سرویس میکروبلاگ مورد توجه میباشد) ، مهمترین ویژگی مورد توجه در مورد تویتر قابلیت باز نشر (ریتویت) میباشد که به همین دلیل مورد توجه ما قرار گرفته است.

براساس آمار رسمی تویتر در روز بیش از ۵۰ ملیون تویت منتشر میشود که به همین اساس افراد درگیر اطلاعات بیش از حد میشوند و نمیتوانند به تویت های مورد نظر خود دسترسی پیدا کنند.

برای این منظور یک تابع RSV(q , t, delta) تعریف می شود که ارتباط بین q پرسمان ، t تویت را در زمان دلتا بیان میکند.

هدف جستجوی تویت : پیدا کردن ، کوتاه ، مختصر و برخط داده ها درمورد موضوع خاص یا واقعه ای که اخیراً اتفاق افتاده میباشد .

دارین مقاله کیفیت تویت برای ما اهمیت دارد که آن را از طریق الگوریتم های رتبه بندی صفحه برروی صفحه های اجتماعی و زمان انتشار آن بدست می آوریم.

دلایل استفاده از این روش : به دلیل خواص احتمالاتی بیز در هنگامی که داده ها ناقص باشند با استفاده از احتمال میتوان مقادیر را تخمین زد و در برخی موارد از منابع گوناگون داده استفاده میشود که ممکن برروی یکدیگر تاثیر داشته باشند.

معماری شبکه ی ما به صورت گراف G(X,E) تعریف میشود که در آن X= Q v K v T v U می باشد و E یال های شبکه هستند ، Q پرسمان ، K , T ,U ترم های پرسمان ، ترم های تویت و نود های میکروبلاگ هستند.

ارتباط بین یک تویت و پرسمان در زمان تتا به صورت احتمال توئمان بیان میشود که به صورت زیر قابل محاسبه میباشد.

$$P(q \wedge t_j) = \sum_{\forall \vec{k}} P(q|\vec{k}) P(\vec{k}|t_j) P(t_j)$$

Cao, G., Nie, J., Gao, J., & Robertson, S. 2008
Selecting Good Expansion Terms for Pseudo-Relevance Feedback

خلاصه:

PRF (بازخورد شبه رابطه ای) در نظر میگیرد که پر تکرار ترین ترم ها در متن های بازگشتی با پرسمان رابطه دارند، در این مقاله نشان میدهیم که این فرضیه ی در دنیای واقعی قابل اعتماد نیست و بسیاری از ترم های پیدا شده به روش های سنتی بی ربط به موضوع بوده و کمکی به ما نمیکنند، هم چنین نشان می دهیم که ترم های مناسب قابل تشخیص از ترم های بد در متن های بازگشتی نیستند بلکه باید در کل متن ها بررسی شوند.

آزمایشات ما بر داده های TREC نشان میدهد که کلاس بندی ترم ها میتواند در نتیجه ی کار تاثیر مثبت داشته باشد (باید از آموزش با ناظر استفاده شود به جای آموزش بدون ناظر).

پرسمان های کاربر معمولا آنقدر کوتاه هستند که منظور او را به طور کامل بیان نمیکنند، به همین منظور از روش های گسترش پرسمان استفاده میشود که در میان آن ها PRF کارآمد ترین آن ها بوده است.

مسئله این است چگونه میتوانیم به صورت کارآمد تری ترم های مناسب را با استفاده از PRF انتخاب کنیم؟

از یک روش با ناظر برای انتخاب ترم ها استفاده میکنیم، براساس تاثیر مستقیم ترم ها بر نتایج آن ها را به دو دسته ی مناسب و نامناسب دسته بندی میکنیم.

شیوه ی جدید دارای چند مزیت میباشد :

۱. ترم های انتخاب شده تحت تاثیر توزیع آنها در متن های بازگشتی قرار نمیگیرند.
۲. کلاس بندی میتواند شاخص های مختلفی را به عنوان معیار قرار داده و می توان آن را به یک فرم ورک تبدیل کرد.

- PRF در حوضه های مختلفی پیاده سازی شده است (حوضه ی برداری ، احتمالاتی و اخیراً در حوزه ی زبانی).

$$Score(d,q) = \sum_{w \in V} P(w \mid \theta_q) \log P(w \mid \theta_d)$$

در این رابطه V مجموعه ی تمام لغتنامه، احتمالات باید روان شوند تا از احتمال صفر جلوگیری کنند.

برای انجام عملیات ساده سازی هایی انجام داده ایم، احتمال شرطی ترم های اضافه شده را صفر در نظر گرفته این (ترم های جدید به هم وابستگی ندارند!) با این حال عملکرد سیستم همان خواهد بود و برای وزن دادن به ترم ها از ۰.۰۱ و -۰.۰۱ استفاده میکنیم.

تعریف مسئله (۵)

همان طور که در ابتدا گفته شد گسترش پرسمان یک روش برای افزایش تشابه بین پرسمان کاربر و متن های بازگردانده شده میباشد، برای تعریف دقیق تر آن میتوان گفت به عمل اضافه کردن ترم به پرسمان جهت افزایش کارایی جستجو، گسترش پرسمان (QE) میگویند.

روش های مرسوم برای گسترش پرسمان:

- Local

از ترم های موجود در متن های بازگردانده شده از نتیجه ی جستجوی پرسمان اولیه برای گسترش آن استفاده میکند.

- Global

از ترم هایی که از نظر آماری با پرسمان مرتبط میباشند برای گسترش استفاده میکند.

- External

ترم های توسعه را از منابع خارجی بدست می آوریم (مثلا Wikipedia , Dbpedia).

در این مقاله ما میخواهیم از ترکیب روش های فوق استفاده کنیم.

در توسعه ی پرسمان همواره چالش هایی موجود هستند از جمله این چالش ها در گسترش پرسمان میتوان به عدن تطابق ترم و متن به دلیل کوتاه بودن بیش از حد پرسمان و مشکل گسترش پرسمان براساس رابطه ی معنایی اشاره کرد.

ترم های کاندید برای توسعه ی پرسمان معمولا براساس تکرار آن ها در متن انتخاب میشوند و میدانیم که این معیار مناسبی نیست زیرا بعضی از کلمات پس زمینه ارتباط را نشان میدهند، برای حل این مشکل در وهله ی اول ما از قوانین ارتباط استفاده میکنیم تا به شکل موثر تری ترم های کاندید را انتخاب کنیم و در وهله ی دوم معیار تحلیل صریح معنایی را با مدل خود تطابق داده تا رابطه ی معنایی بین ترم های کاندید و پرسمان داده شده را بررسی کند.

Notation	Description
$\mathcal{C}$	The *whole set* of documents which form the collection
C	A *set* of documents belonging to the collection ($C \subseteq \mathcal{C}$)
d	A *single* document of the collection ($d \in \mathcal{C}$)
V	The *whole set* of *distinct* terms of the collection $\mathcal{C}$
T	A *set* of terms of the collection ($T \subseteq V$)
t	A *single* term of the collection ($t \in V$)
R	An association rule
q	An original query
t_q	A term in a given query q
E_q	A query q extended

تصویر ۱ علائم استفاده شده

در این مقاله ما پرسمان را به صورت مجموعه ای از ترم ها معرفی میکنیم (bag)

$$q = \{t_{q1}, \ldots, t_{qn}\}$$

به صورتی که tqi ترم i ام از پرسمان میباشد.

برای هر مجموعه ی ترم یک معیار ساپورت تعریف میکنیم که برابر است با تعداد متن هایی که تمام اعضای مجموعه در آن حضور دارند و به شکل زیر بیان میشود.

$$Supp(T) = |\{d | d \in \mathcal{C} \wedge \forall t \in T : (d,t) \in I\}|$$

و ساپورت نسبی مجموعه برابر خواهد شد با :

$$\frac{Supp(T)}{|\mathcal{C}|}$$

که در آن C مجموعه ی کل میباشد.

هر رابطه بین ترم ها به صورت زیر نوشته میشود:

$$R\colon T_1 \Rightarrow T_2$$

و به صورتی تعریف میشود که :

احتمال وجود ترم دوم از حدی بالاتر باشد درحالی که ترم اول در متن وجود داشته باشد ، اگر کلمه ی اول در متن وجود داشته باشد احتمالا دومی هم هست .

$$Conf(R) = \frac{Supp(T_1 \cup T_2)}{Supp(T_1)} \qquad Supp(R) = Supp(T_1 \cup T_2)$$

همانطور که در بالا مشاهده میکنید برای یک معیار اعتماد (confidence) تعریف میشود.

- قابل ذکر است که روش های گسترش پرسمان که بر اساس قوانین ارتباط عمل میکنند نیازی به داشتن دانش قبلی یا پردازش زبانی ندارند! (میزان تاثیر آن های در بازیابی داده در سیستم های آی آر قبلا بررسی شده است)

$$\mathcal{R}_C = \{R|\, Conf(R) \geq minconf \;\; and \;\; Supp(R) \geq minsupp\}$$

در نهایت یک رابطه ی قابل اعتماد به شکل بالا بدست میآید (را بط ای که حداقل ساپورت و اطمینان مورد نظر را داشته باشد).

- حداقل ساپورت را طوری در نظر میگیریم که تمام ترم های پر تکرار ساخته شوند.

تحلیل صریح معنایی (ESA) یک معیار مرتبط بودن در فضای برداری میباشد که به صورت گسترده در استخراج داده مورد استفاده قرار میگیرد، در اینجا متن ها براساس تکرار ترم ها در آن ها نمایش داده نمیشوند بلکه براساس مفاهیم مشابه (concepts) که از مقالات Wikipedia بدست می آید بیان میشود.

هر مفهوم Wikipedia به صورت برداری از ترم که در متن ظاهر میشوند نمایش داده میشود، مقادیر های این بردار ها با tf * idf وزن دهی شده اند، این وزن های نمایانگر قدرت بین ترم ها و مفهوم هستند.

برای محاسبه ی شباهت از فرمول زیر استفاده میشود که همان معیار cosine است.

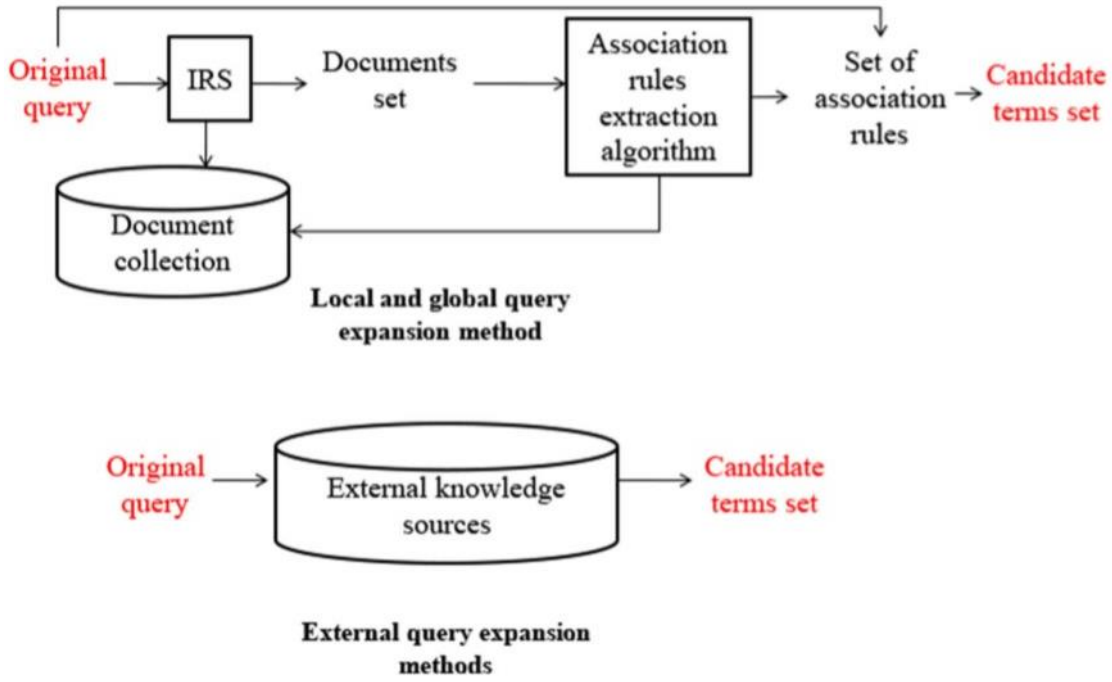
$$ESA(q,t) = \frac{\overrightarrow{q} \times \overrightarrow{t}}{\|\overrightarrow{q}\| \times \|\overrightarrow{t}\|}$$

روش پیشنهادی (۶)

دارین بخش ما دو مشکل گسترش پرسمان، تولید ترم های کاندید و انتخاب آن ها را برطرف میکنیم.

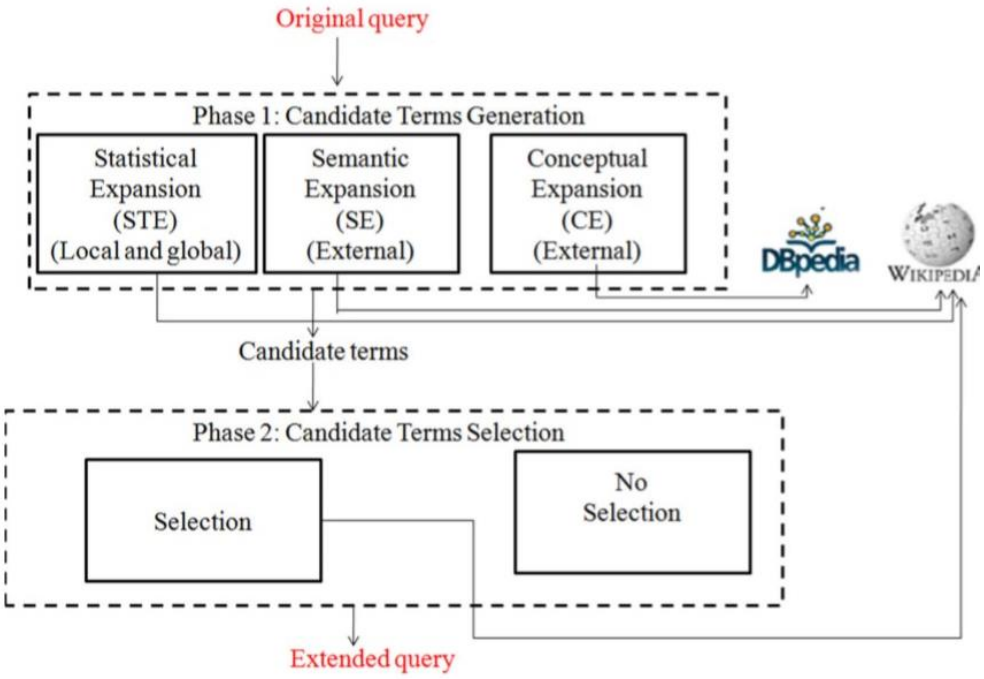
ما از چندین منبع دانش مانند Wikipedia و DBPEDIA علاوه بر مجموعه ی متن ها استفاده میکنیم تا به ترم های گسترش تنوع ببخشیم، به طور دقیق تر مدل ما ابتدا مجموعه ای از ترم های کاندید را براساس هر کدام از منابع دانش تولید کرده و سپس آن ها را توسط معیار ارتباط برای هر کدام از پرسمان های توسعه یافته باهم ترکیب میکند.

هدف ما بررسی تاثیر هر کدام از روش ها به همراه انتخاب به صورت معنایی میباشد.



تصویر ۲ ساخت ترم های کاندید

براساس قوانین کلی و محلی (global , local) قوانین ارتباط را استخراج میکنیم و استخراج ترم ها از منابع خارجی، فرض ما این است که ترم های مرتبط تر بیشتر در متن تکرار میشوند (local).



تصویر ۳ مدل ترکیبی برای گسترش پرسمان

ترم های کاندید پرسمان را به صورت زیر نمایش میدهیم.

$$Candidate\_Set(q) = \{t_1, \ldots, t_p\}$$

ما از سه روش برای تولید ترم های استفاده میکنیم، ترم هایی که از نظر آماری دارای وابستگی بدون در نظر گرفتن معنای آن ها، با استفاده از قوانین ارتباط، ترم هایی که از نظر معنایی با پرسمان رابطه دارند (این ترم ها از تعریف پرسمان بدست می آیند)، ترم هایی که از نظر مفهومی (علم شناختی) با پرسمان در ارتباط هستند.

مجموعه ترم های آماری به صورت زیر بیان میشوند:

$$Candidate\_Set_{STE}(q) = \bigcup_{(T_1 \Rightarrow T_2) \in \mathcal{R}_C \text{ sothat } T_1 \in 2^q} T_2$$

مجموعه ی ترم های معنایی به صورت زیب بیان میشوند:

$$Candidate\_Set_{SE}(q) = \bigcup_{t \in q} Def_{Semantic}(t, RS)$$

مجموعه ی ترم های مفهومی به صورت زیر بیان میشوند:

$$Candidate\_Set_{CE}(q) = \bigcup_{t \in q} Concept(t, O)$$

- توجه هم برای تولید ترم های مفهومی هم معنایی از Wikipedia استفاده خواهد شد! (برای مفهوم از بخش تعاریف استفاده میشود).

برای بخش انتخاب ترم های کاندید معیار های زیر را تعریف میکنیم.

$$relatedness(q,t) = score \in \mathbb{R}$$

معیار مرتبط بودن یک ترم از مجموعه ی ترم های کندید به صورت یک امتیاز

$$E_q = q \cup \{t \in Candidate\_Set(q) \mid relatedness(q,t) = score \geq \mu\}$$

پرسمان گسترش پیدا کرده با استفاده از ترم t به صورتی که امتیاز ترم بیشتر از حداقل مورد نظر باشد.

$$\begin{aligned} relatedness(q,t) &= ESAC(q,t) \\ &= \begin{cases} (\alpha \times ESA(q,t) + (1-\alpha) \times Conf_{max}(R,q,t) & if\ Conf_{max}(R,q,t) \neq 0; \\ ESA(q,t), & otherwise. \end{cases} \end{aligned} \quad (16)$$

از یک ترکیب خطی از تحلیلی صریح معنایی استفاده خواهیم کرد امتیاز دهی نهایی ما بر عهده ی آن خواهد بود، این معیار با ضریب اطمینان بین رابطه، پرسمان و ترم انتخاب شده ترکیب شده است.

$$Conf_{max}(R,q,t) = \max_{t_q \in q, R \in \mathcal{R}_C} Conf(R(t_q,t))$$

و ضریب اطمینان بین رابطه، پرسمان و ترم انتخاب شده به صورت بالا تعریف میشود.

Terms generation	Terms selection	
	With selection	Without selection
STE	$STE_{Selection}$	$STE_{NoSelection}$
SE	$SE_{Selection}$	$SE_{NoSelection}$
CE	$CE_{Selection}$	$CE_{NoSelection}$
ALL	$ALL_{Selection} = STE_{Selection} \cup SE_{Selection} \cup CE_{Selection}$	$ALL_{NoSelection} = STE_{NoSelection} \cup SE_{NoSelection} \cup CE_{NoSelection}$

تصویر ۴ حالت های مختلف HQE

برای مدل خود چندین حالت ممکن (configuration) را در نظر میگیریم که لیست آن ها را در جدول بالا میتوانید ببینید.

نتایج آزمایشات (۷)

برای بررسی مدل ارایه شده توسط ما از دو منبع، TREC 2011 که در آن هم پرسمان ها کوتاه هستند و هم تویت ها و TREC Robust 2004 که در آن پرسمان به دلیل عدم تطابق سخت هستند. ما شیوه ی خود را توسط مقادیر مختلف بررسی کرده و نتایج را با روش سنتی PRF مقایسه کرده ایم.

TREC 2011 Miroblog Track

یک مجموعه ی متنی داده های اجتماعی برخط می باشد که در آن کاریر به دنبال آخرین ولی مرتبط ترین داده (tweet) میگردد. این مجموعه شامل ۱۶ ملیون تویت میباشد که در طی ۲ هفته جمع آوری شده اند و شامل ۵۰ موضوع میباشند که هیچ خلاص و علایمی برای آن ها ارائه نشده است.

اما هدف ما در این بررسی مقایسه صرفاً با مرتبط ترین میباشد و نه وابسته به زمان به همین دلیل با مقیاس های رسمی این مجموعه ی داده مقایسه نمیکنم.

Model	Parameter	Values
BM25	b	0.01; 0.02; 0.03; 0.04; 0.05; 0.06; 0.07; 0.08; 0.09; 0.1; 0.2, 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9
Hiemstra	λ	0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9
Dirichlet	μ_D	50; 300; 500; 1000; 1500; 2000; 2500; 3000

تصویر ۵مقادیر مدل های مقایسه

TREC 2004 Robust retrieval track

این مجموعه شامل چهار مجموعه ی متن میشود که مقالات انگلیسی روزنامه میباشند که در مجموع به بیش از نیم ملیون متن میرسند.
هدف ما مطالعه ی پرسمان های سخت میباشد به همین بروی زیر مجموعه ای از آن که به عنوان موضوعات سخت شناخته میشود تمرکز کرده ایم. این ۵۰ موضوع و از در بازه ی ۳۰۱ تا ۳۵۰ قرار دارند.
ما این مجموعه را انتخاب کرده این زیرا روش های مقایسه نیز از داده های وب برای توسعه ی پرسمان استفاده کرده اند.
تمام نتایج بدست آمده در پلتفرم Terrier 4.0 بدست آمده اند، ما روش های خود را با روش های کلاسیک مثل BM25 ، Dirichlet ، Hiemstra و PRF مقایسه میکنیم.

برای ساخت قوانین ارتباط برای TREC 2011، توییت ها منبع مناسبی نیستند به همین دلیل ما از ۵۰ هزار مقاله ی Wikipediaبرای اینکار استفاده کرده ایم.

Collection	#documents	*Minsupp*	*Minconf*	#Rules
WIKIPEDIA corpus for TREC Microblog 2011 (50 social topics)				
Documents	50,000	15	0.7	402,862
TREC 2004 Robust Track (50 hard topics)				
FBIS	130,471	1,700	0.5	770,359
Federal register 94	55,630	1,500	0.5	211,759
LA times	131,896	2,000	0.5	538,323
Financial times	210,158	2,500	0.5	379,248

تصویر ۶ مقادیر ثابت برای ساخت ترم

Parameter	Values
α	0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9
μ	0.35; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9

تصویر ۷ مقادری ثابت برای انتخاب ترم

نتایج TREC 2011

نتایج این آزمایش نشان میدهد که مسئله ی بازیابی داده از توییت ها فاصله ی زیادی با حل شده بودن دارد!
یک بار دیگر اشاره میکنیم که هدف ما باز گرداندن نتایج با توجه به زمان نیست.
برا نتایج از معیار میانه ی میانگین ها (Mean Average Precision) استفاده نیکنیم.

Run	Configuration	P@5	P@10	P@30	MAP (%Chg.$_{Baseine}$,%Chg.$_{PRF}$)
BM25					
Baseline	–	0.1265	0.1327	0.1238	0.1025
PRF	–	0.1592	0.1551	0.1245	0.1145
–	**STE$_{Selection}$**	**0.4000**	**0.3796**	**0.3197**	**0.3079** $^{\dagger\circ}$ **(200% , 168%)**
–	STE$_{NoSelection}$	0.3551	0.3265	0.2850	0.2804$^{\dagger\circ}$ (173% , 145%)
–	ALL$_{Selection}$	0.3633	0.3429	0.2707	0.2747$^{\dagger\circ}$ (168% , 140%)
–	SE$_{Selection}$	0.3342	0.3184	0.2626	0.2589$^{\dagger\circ}$ (153% , 126%)
–	ALL$_{NoSelection}$	0.3551	0.3388	0.2553	0.2570$^{\dagger\circ}$(151% , 124%)
–	CE$_{Selection}$	0.3224	0.3041	0.2755	0.2505$^{\dagger\circ}$(144% , 118%)
–	CE$_{NoSelection}$	0.2408	0.2227	0.2041	0.2053$^{\dagger\circ}$ (100% , 79%)
–	SE$_{NoSelection}$	0.2367	0.2224	0.2163	0.1676$^{\dagger\circ}$(63% , 46%)
HIEMSTRA					
Baseline	–	0.1429	0.1429	0.1333	0.1148
PRF	–	0.1469	0.1755	0.1374	0.1156
–	**STE$_{Selection}$**	**0.3837**	**0.3673**	**0.3014**	**0.3083** $^{\dagger\circ}$ **(168%, 166%)**
–	STE$_{NoSelection}$	0.3469	0.3347	0.2939	0.2883$^{\dagger\circ}$(151%, 149%)
–	CE$_{Selection}$	0.3306	0.3286	0.2653	0.2690$^{\dagger\circ}$(134%, 132%)
–	SE$_{Selection}$	0.3184	0.3102	0.2605	0.2627$^{\dagger\circ}$(128%, 127%)
–	CE$_{NoSelection}$	0.2857	0.2755	0.2265	0.2439$^{\dagger\circ}$(112% , 110%)
–	ALL$_{Selection}$	0.3102	0.2796	0.2265	0.2177$^{\dagger\circ}$ (89% , 88%)
–	ALL$_{NoSelection}$	0.3020	0.2714	0.2286	0.2075$^{\dagger\circ}$(81%, 79%)
–	SE$_{NoSelection}$	0.2245	0.2286	0.1986	0.1671$^{\dagger\circ}$ (45% , 44%)
DIRICHLET					
Baseline	–	0.1592	0.1571	0.1367	0.1156
PRF	–	0.1184	0.1286	0.1340	0.1177
–	**STE$_{Selection}$**	**0.4000**	**0.3837**	**0.3197**	**0.3152** $^{\dagger\circ}$ **(172% , 167%)**
–	STE$_{NoSelection}$	0.3592	0.3367	0.3061	0.2973$^{\dagger\circ}$(157% , 152%)
–	CE$_{Selection}$	0.3540	0.3286	0.2762	0.2741$^{\dagger\circ}$ (137%, 132%)
–	SE$_{Selection}$	0.3184	0.3122	0.2741	0.2700$^{\dagger\circ}$(133% , 129%)
–	CE$_{NoSelection}$	0.2980	0.2857	0.2510	0.2507$^{\dagger\circ}$ (116%, 112%)
–	ALL$_{Selection}$	0.3347	0.2796	0.2354	0.2377$^{\dagger\circ}$(105%, 101%)
–	ALL$_{NoSelection}$	0.3020	0.2714	0.2286	0.2075$^{\dagger\circ}$(79%, 76%)
–	SE$_{NoSelection}$	0.2163	0.2122	0.2000	0.1739$^{\dagger\circ}$ (50% , 47%)

ت‌صویر ۸ نتایج TREC 2011

برای همه ی انواع حالت های مدل (بدون فیلتر یا با فیلتر) ما عمل کرد بهتری از آن ها داشته ایم.
RPF رو متن های کوتاه به خوبی عمل نمیکند که سیستم ما با استفاده از منابع خارجی این مشکل را حل میکند.

با توجه به نتایج بدست آمده زمانی که تابع انتخاب ترم ما از فیلتر استفاده میکند شاهد عملکرد بهتری هستیم که نشان میدهد عملیات فیلتر کردن در ESAC کارآمد میباشد.

نتایج TREC 2004

Run	Configuration	P@5	P@10	P@30	MAP(%$Chg_{\cdot baseline}$, %$Chg_{\cdot PRF}$)
Title+Description+Narrative					
Baseline	–	0.3760	0.3200	0.2033	0.1339
PRF	–	0.4240	0.3520	0.2633	0.1546
–	$STE_{NoSelection}$	0.4160	0.3640	0.2780	$0.1471^{\dagger}$ (+ 10% , − 4%)
–	$ALL_{NoSelection}$	0.3640	0.3600	0.2680	$0.1453^{\dagger}$ (+ 9% , − 6%)
–	$ALL_{Selection}$	0.3760	0.3480	0.2687	$0.1422^{\dagger}$ (+ 6% , − 8%)
–	$CE_{Selection}$	0.3840	0.3460	0.2587	$0.1418^{\dagger}$ (+ 6% , − 8%)
–	$STE_{Selection}$	0.3800	0.3440	0.2613	$0.1403^{\dagger}$ (+ 4%, − 9%)
–	$SE_{Selection}$	0.3360	0.3060	0.2433	0.1395 (+ 4% , − 9%)
–	$SE_{NoSelection}$	0.3320	0.3060	0.2373	0.1353 (+ 1% , − 12%)
–	$CE_{NoSelection}$	0.3640	0.3240	0.2560	0.1352 (0%, − 12%)
PRF	$STE_{NoSelection}$	0.4120	0.3620	0.2867	$0.1777^{\dagger\circ}$ (+ 33% , + 14%)
Official best (pircRB04td2)		0.4600	0.4020	0.2867	0.1949

تصویر ۹ نتایج TREC 2004

با توجه به نتایج حاصل روش ما فقط درحالی که از انتخاب ترم استفاده نمیکرد (بدون فیلتر) توانست بهتر عمل کند، این نشان میدهد که ترم هایی که توسط روش آماری تولید شده اند خود به اندازه ی کافی خوب بوده اند و فیلتر کردن آن ها باعث کاهش کارایی میشوند، دلیل این امر این گونه توضیح داده میشود که این متن ها خود مقالات روزنامه هستند و این متن بسیار تمیز (دار ساختار استاندارد) هستند به همین دلیل روش آماری به تنهایی عملکرد بهتری خواهد داشت.

برای بیشتر حالات روش آماری و ترکیب همه عملکرد بهتری خواهد داشت.

روش PRF برای ۳۰ متن بسیار کارآمد عمل میکند زیرا برروی مجموعه ی بزرگی از متن ها عمل کرده است.