

Answer Validation in Question-Answering System using Support Vector Machine

Yaghoob Ghanbari¹, Shahram Golzari^{2*} and Shyamala Doraisamy³

1- Department of Electrical and Computer Engineering, University of Hormozgan, Bandar Abbas, Iran.

2*- Department of Electrical and Computer Engineering, University of Hormozgan, Bandar Abbas, Iran – Deep Learning Research Group, University of Hormozgan, Bandar Abbas, Iran.

3- Department of Multimedia, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Malaysia.

¹ghanbari.yaghoob@gmail.com, ^{2*}golzari@hormozgan.ac.ir, and ³shyamala@fsktm.upm.edu.my

Corresponding author's address: Shahram Golzari, Department of Electrical and Computer Engineering, University of Hormozgan, Bandar Abbas, Iran, PO Box: 3995.

Abstract- Question-answering system is a special type of information retrieval system. Considering a set of documents (such as Web or local set of documents), the system should answer the questions asked in the natural language form. Question-answering systems generally consist of question processing and analysis, keyword generation, information retrieval, answer extraction, and answer validation. In such systems, selecting proper answers for user question is an important task which influences performance of the whole system. An appropriate answer validation technique can increase the performance of the question-answering system. In this study, Support Vector Regression (SVR) is employed in order to perform answer validation. SVR eliminates the risk of getting stuck in local minima by reducing the operational risk. The proposed system is applied on some of the TREC and Wikipedia questions repositories. In order to evaluate performance of the proposed system, following measures are calculated: Mean Reciprocal Ranking (MRR) and F-measure. Based on experimental results, the proposed system achieved 81% MRR and 49.7% F-measure which shows higher performance than systems with no answer validation and also systems using neural network-based answer validation.

Keywords- Question-answering system, information retrieval, answers validation, Support Vector Machine.

اعتبارسنجی پاسخ‌ها در سیستم پرسش و پاسخ با استفاده از رگرسیون ماشین بردار پشتیبان

یعقوب قنبری^۱، شهرام گلزاری^{۲*}، Shyamala Doraisamy^۳

۱- کارشناس ارشد هوش مصنوعی، گروه مهندسی برق و کامپیوتر، دانشکده فنی و مهندسی، دانشگاه هرمزگان، بندرعباس، ایران.

۲* - گروه مهندسی برق و کامپیوتر، دانشکده فنی و مهندسی، دانشگاه هرمزگان، بندرعباس، ایران - هسته پژوهشی یادگیری عمیق، دانشگاه هرمزگان، بندرعباس، ایران.

۳- Department of Multimedia, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Malaysia

¹ ghanbari.yaghoob@gmail.com, ^{2*} golzari@hormozgan.ac.ir, ³ shyamala@fsktm.upm.edu.my

* نشانی نویسنده مسئول: شهرام گلزاری، بندرعباس، دانشگاه هرمزگان، دانشکده فنی و مهندسی، گروه مهندسی برق و کامپیوتر، صندوق پستی ۳۹۹۵.

چکیده - سیستم پرسش و پاسخ شکل خاصی از بازیابی اطلاعات است. با در نظر گرفتن مجموعه‌ای از اسناد (مانند شبکه جهانی وب یا یک مجموعه محلی) سیستم باید بتواند پاسخ پرسش‌های مطرح شده به زبان طبیعی را بازیابی نماید. سیستم‌های پرسش و پاسخ به صورت متداول از بخش‌های پردازش و تجزیه پرسش، تولید عبارت جست‌وجو، بازیابی اطلاعات، استخراج پاسخ و رتبه‌بندی پاسخ تشکیل شده است. یکی از علل کاهش کارایی سیستم‌های پرسش و پاسخ، پایین بودن دقت انتخاب پاسخ‌های مناسب به پرسش کاربران است. به وسیله اعتبارسنجی مناسب پاسخ‌ها می‌توان این مشکل را حل نموده و کارایی سیستم پرسش و پاسخ را ارتقا داد. در این پژوهش از رگرسیون ماشین بردار پشتیبان (SVR) برای اعتبارسنجی و رتبه‌بندی پاسخ‌ها استفاده شده است. الگوریتم رگرسیون ماشین بردار پشتیبان با کمینه نمودن ریسک عملیاتی، خطر افتادن در کمینه‌های محلی را بر طرف می‌کند. برای ارزیابی کارایی سیستم پیشنهادی از مجموعه پرسش‌های TREC و پرسش‌های ویکی‌پدیا استفاده گردیده است. در این پژوهش از معیارهای میانگین رتبه‌بندی معکوس و معیار F برای ارزیابی سیستم پیشنهادی استفاده گردیده است. بر اساس نتایج حاصل از آزمایش‌ها، مقادیر به دست آمده برای این معیارها توسط سیستم پیشنهادی به ترتیب ۸۱٪ و ۴۹.۷٪ بوده است. سیستم پیشنهادی در مقایسه با سیستم بدون رتبه‌بندی پاسخ و سیستم مبتنی بر رتبه‌بندی پاسخ با شبکه عصبی از کارایی مناسب‌تری برخوردار است. همچنین سیستم پیشنهادی در مقایسه با پژوهش‌های پیشین دارای میانگین رتبه‌بندی معکوس بهتری می‌باشد.

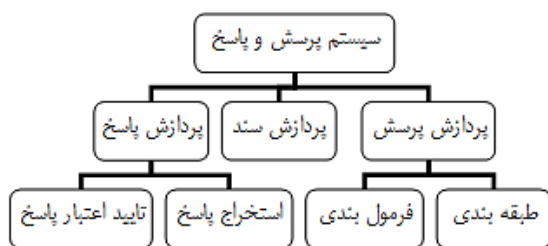
واژه‌های کلیدی: سیستم پرسش و پاسخ، بازیابی اطلاعات، اعتبارسنجی پاسخ، رگرسیون ماشین بردار پشتیبان (SVR).

۱- مقدمه

منبعی مانند متن، تصویر، صوت و ویدیو باشد. این اطلاعات معمولاً از ساختار مشخصی پیروی نکرده و عموماً دارای معانی تعریف شده و مشخصی نیستند. هدف بازیابی اطلاعات در چنین شرایطی، کمک به کاربر برای یافتن اطلاعات مورد نظر در انبوهی از اطلاعات ساختار نیافته است. هنگامی که کاربر برای کسب اطلاعات، جست و جو می‌نماید، گستردگی نتایج حاصل از موتورهای جست و جو، وی را

سیستم پرسش و پاسخ در بین محققان پردازش زبان طبیعی بسیار مورد توجه قرار گرفته است. با افزایش روز افزون حجم اطلاعات قابل دسترس در منابع اطلاعاتی بزرگ مانند شبکه جهانی وب و زیرمجموعه‌های آن، فرآیند بازیابی و استخراج اطلاعات اهمیت ویژه‌ای یافته است. اطلاعات مورد نظر ممکن است شامل هر نوع

در فرم خاصی دریافت می‌کند و از میان منابع موجود، اسناد مرتبط را استخراج می‌کند. پردازش پاسخ شامل دو بخش اصلی، استخراج پاسخ و تایید اعتبار پاسخ می‌باشد. در بخش استخراج پاسخ، پاسخ‌های کاندید شده توسط الگوریتم‌های استخراج پاسخ، از اسنادی که توسط موتور جست‌وجو در واحد استخراج پاسخ، استخراج می‌شوند، بازیابی شده‌اند. بعد از اینکه پاسخ‌های نامزد در بخش استخراج پاسخ بازیابی شدند، پاسخ‌های کاندید شده توسط فیلترگذاری و رتبه‌بندی که دو گام اساسی در تایید اعتبار پاسخ هستند، اعتباردهی می‌شوند [۹-۵]. در شکل ۱ بخش‌های یک سیستم پرسش و پاسخ نشان داده شده است.



شکل ۱: بخش‌های سیستم پرسش و پاسخ

یکی از مراحل مهم و اساسی سیستم پرسش و پاسخ انتخاب پاسخ از میان پاسخ‌های موجود می‌باشد که اگر این مرحله به درستی صورت گیرد باعث بهبود کارایی سیستم می‌گردد. استفاده از الگوریتم‌های یادگیری ماشین مانند ماشین بردار پشتیبان به دلیل متکی نبودن به قوانین برای انتخاب پاسخ، موجب بهبود کارایی سیستم می‌شود. سیستم‌هایی که بر اساس قانون، پاسخ‌های موردنظر را انتخاب می‌کنند نیازمند قوانین متعددی برای یافتن پاسخ هستند و اگر قوانین جامع و کامل نباشد سیستم با مشکل مواجه خواهد شد که این مشکل در الگوریتم‌های یادگیری ماشین رفع شده است.

از آنجایی که سیستم پرسش و پاسخ یک سیستم سلسه مراتبی است، هر مرحله می‌تواند نقش بسزایی در افزایش کارایی سیستم پرسش و پاسخ داشته باشد. یکی از مراحل مهم در افزایش دقت سیستم‌های پرسش و پاسخ، مرحله‌ی انتخاب پاسخ بوده که اگر این مرحله به درستی صورت گیرد، کارایی سیستم نیز افزایش پیدا می‌کند. هدف از این پژوهش استفاده از رگرسیون ماشین بردار پشتیبان^۱ (SVR) جهت اعتبارسنجی پاسخ‌ها در سیستم پرسش و پاسخ به منظور افزایش کارایی سیستم پرسش و پاسخ می‌باشد.

سازمان‌دهی ساختار مقاله به این صورت است که در بخش ۲، پیشینه پژوهش بیان می‌گردد. در بخش ۳، رگرسیون ماشین بردار پشتیبان شرح داده خواهد شد. در بخش ۴، الگوریتم پیشنهادی به طور مفصل توضیح داده می‌شود و مثالی از رتبه‌بندی توسط سیستم

دچار سردرگمی می‌کند؛ زیرا موتورهای جست‌وجو مانند Google و Yahoo لیست طولانی از لینک‌هایی که حاوی قطعه‌هایی از درخواست کاربر است را برمی‌گردانند؛ بررسی محتوی هر یک از این یافته‌ها برای استخراج پاسخ مورد نیاز کار عظیمی می‌طلبد. زمینه پژوهشی سیستم‌های پرسش و پاسخ به عنوان تلاشی برای مقابله با این مشکل فراوانی اطلاعات، پدید آمده است [۱].

در سیستم پرسش و پاسخ، پرسش‌هایی که از طرف کاربر پرسیده می‌شوند به زبان طبیعی است. از همین رو سیستم باید بتواند پاسخ‌هایی متناسب با پرسش کاربر عرضه کند. پرسش‌ها به دو دسته حقیقت و غیر حقیقت گروه‌بندی می‌شوند. پرسش‌های حقیقت به پرسش‌هایی گفته می‌شود که پاسخ آن‌ها واقعیت‌های ساده‌ای است که از یک سند واحد بازیابی می‌گردد. در زبان انگلیسی پرسش‌های حقیقت با استفاده از کلمات پرسشی wh بیان می‌شوند. به طور مثال می‌توان به پرسش حقیقت Where was Julius Caesar born? اشاره کرد. پرسش‌های غیر حقیقت معمولاً دربرگیرنده‌ی چندین بخش از اطلاعات است که از یک یا چند سند بازیابی می‌گردد و معمولاً پاسخ‌گویی به آن‌ها دشوارتر است. به طور مثال می‌تواند به پرسش غیر حقیقت What were the consequences of World War II? اشاره کرد. قسمت اصلی این سیستم‌ها، پردازش پرسش‌ها و به دست آوردن نوع پرسش و پاسخ موردنظر است [۲].

سیستم‌های پرسش و پاسخ از لحاظ دامنه تحت پوشش به دو دسته دامنه باز و دامنه بسته تقسیم بندی می‌شود. سیستم‌های با دامنه باز یا نامحدود، قادر به پاسخ‌گویی پرسش‌های کاربران در زمینه‌های مختلف می‌باشند. سیستم‌های با دامنه بسته، قادر به پاسخ‌گویی پرسش‌ها در یک زمینه خاص مانند پزشکی هستند و غالباً بروی یک سایت یا یک کتاب خاص، کار می‌کنند. عملکرد این سیستم‌ها در مقایسه با سیستم‌های با دامنه باز، از پیچیدگی کمتری برخوردار است [۳ و ۴].

هدف سیستم پرسش و پاسخ، پیدا کردن پاسخ دقیق و صحیح برای پرسش‌های کاربر است. تجزیه و تحلیل پرسش، بازیابی اطلاعات و انتخاب پاسخ، سه مورد مهم در یک سیستم پرسش و پاسخ هستند. این سیستم‌ها برای برقراری ارتباط با کاربر شامل حداقل سه بخش پردازش پرسش، پردازش سند و پردازش پاسخ می‌باشد. اجزای مهم پردازش پرسش، طبقه‌بندی پرسش و فرمول‌بندی است. برای استخراج پاسخ در یک مجموعه‌ی بزرگ از اسناد و متون، در ابتدا سیستم باید بداند که چه چیزی را جست‌وجو می‌کند. در این مورد پرسش‌ها بایستی با توجه به نوعشان طبقه‌بندی شوند. فرمول-بندی پرسش برای شناسایی روش‌های مختلف ارائه پاسخ به پرسش زبان طبیعی، صورت می‌گیرد. در بخش بازیابی اطلاعات، اسناد مرتبط با پرسش کاربر بازیابی می‌گردد. این بخش پرسش کاربر را

و پاسخ، روش‌های نمایش منطقی یا نحوی، نمونه‌هایی از روش‌های بدون ناظر می‌باشند. آن‌ها در پژوهش خود برای تولید نمونه آموزش از ترکیب هسته‌های مختلف بر روی هر یک از اعضای پرسش و پاسخ استفاده کرده‌اند. همچنین برای کاهش قابل توجه میزان حاشیه‌نویسی دستی^۳، با به کار بردن هسته‌های ساختاری^۴ بر روی ساختارهای معنایی/نحوی، نمونه‌های آموزشی (یعنی همان زوج پرسش و پاسخ) با استفاده از روش‌های عمومی‌سازی^۵، نمایش داده می‌شوند. این روش توانست صحت روش کیسه کلمات^۶ (BoW) را بر روی مجموعه داده TREC به میزان ۶۳ درصد بهبود ببخشد [۱۴].

کو و همکاران از یک روش ترکیبی بر اساس احتمالات و شواهد جهت استخراج پاسخ بهره برده‌اند. آن‌ها یک چارچوب احتمالاتی با ترکیب شواهد به دست آمده از پرسش، جهت رتبه‌بندی پاسخ‌ها در سیستم پرسش و پاسخ ارائه داده‌اند. این سیستم از شواهد مختلفی جهت شناسایی پاسخ مرتبط استفاده می‌کند. ادغام پاسخ‌های منتخب، توسط تکنیک‌های مختلف صورت می‌گیرد. این سیستم به پرسش‌های لیستی همانند پرسش‌های حقیقی با دقت بالایی پاسخ می‌دهد [۱۵].

ممیزی و همکاران سیستم پرسش و پاسخی طراحی کرده‌اند که هدف آن تولید الگو از وب بر اساس معنای واژگان و محدودیت‌های نحوی است. محدودیت‌ها در این سیستم برای ارزیابی و رتبه‌بندی پاسخ‌های کاندید، استفاده شده است. مجموعه پرسش‌های TREC-8، TREC-9، TREC-10 به‌عنوان مجموعه آموزشی استفاده شده است. سیستم پیشنهادی پاسخ‌ها را به طور خودکار از مجموعه TREC بازیابی می‌کند. در این سیستم از شبکه واژگان برای کمک به رابطه معنایی و برچسب نحوی بین پرسش و پاسخ استفاده شده است. در پایان، وزن‌دهی به هر پاسخ کاندید بر اساس طول پاسخ، سطح رابطه معنایی بین پرسش و پاسخ و فاصله بین کلمات کلیدی در نظر گرفته شده است [۱۶].

متایاس و همکاران یک سیستم پرسش و پاسخ مبتنی بر روش‌های آماری ارائه کرده‌اند. هدف از به کارگیری روش‌های آماری، ایجاد سیستم پرسش و پاسخی است که بتواند بدون نظر گرفتن ماژول‌های زبانی، برای اکثر زبان‌ها قابل استفاده باشد. برای بازیابی پاسخ و طبقه‌بندی پرسش، از مدل‌های ریاضی استفاده می‌شود. در سیستم پیشنهادی برای یافتن پاسخ به جای استفاده از داده‌های وب از داده‌های مجموعه‌ی AQUANIT استفاده گردیده است. هرچند در داده‌های وب افزونگی ذاتی اطلاعات نسبت به یک مجموعه ثابت از اطلاعات بیشتر به چشم می‌خورد که این افزونگی ذاتی اطلاعات باعث می‌شود که برای هر پرسش چندین جواب بازیابی گردد. برای

مورد نظر نیز در این بخش ذکر می‌گردد. ارزیابی و نتایج آزمایش‌ها در بخش ۵ آورده شده است. در نهایت نیز در بخش ۶ نتیجه‌گیری صورت می‌گیرد.

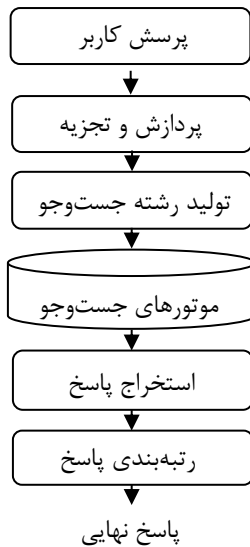
۲- پیشینه پژوهش

نخستین سیستم پرسش و پاسخ در سال ۱۹۶۱ درزمینه‌ی ورزش بیس‌بال مورداستفاده قرار گرفت که به سیستم بیس‌بال معروف گردید. سیستم بیس‌بال باهدف پاسخ‌گویی به پرسش‌ها، در مورد بازی‌های بیس‌بال صورت گرفته در لیگ آمریکا، طراحی شده بود. این سیستم یک برنامه کامپیوتری بود که به پرسش‌های انگلیسی مرتبط با بازی بیس‌بال از قبیل ماه، روز، مکان برگزاری، گروه‌ها و امتیازات کسب‌شده در بازی‌ها پاسخ می‌داد. منابعی که در این نوع سیستم‌ها و سیستم‌های مشابه مورداستفاده قرار می‌گیرد با توجه به اهداف سیستم می‌تواند متفاوت باشد. به‌طور مثال سیستم می‌تواند اسناد از پیش جمع‌آوری شده را استفاده کند و یا اینکه از مجموعه‌ای بزرگ مانند اینترنت بهره بگیرد [۲].

اولین سیستم دستی توسط هرمی ژاکوب در سال ۲۰۰۱ ارائه شد. ایشان برای شناسایی نوع پرسش مورد نظر، از قوانین hand-crafted استفاده کرد. با وجود اینکه این قوانین از دقت بالایی برخوردارند اما نمی‌توان از مشکلات این نوع سیستم‌ها از قبیل افزایش زمان محاسبات، خسته کننده بودن قوانین و غیر انعطاف‌پذیر بودن قوانین نسبت به پرسش‌های جدید اجتناب کرد. به همین دلیل طبقه‌بندی‌های خودکار اهمیت بیشتری پیدا کردند. سیستم‌های طبقه‌بند خودکار در مواجهه با پرسش‌های جدید و تعیین نوع پرسش موردنظر دارای انعطاف بیشتری می‌باشند [۱۰ و ۱۱]. با توسعه سیستم‌های طبقه‌بند خودکار، روش‌ها و الگوریتم‌های جدیدی نیز برای این نوع طبقه‌بندی ابداع گردید که می‌توان به روش‌های یادگیری ماشین و مدل سازی زبان اشاره کرد. هاسی‌اوگلو و وارد برای طبقه‌بندی پرسش، از ماشین بردار پشتیبان استفاده کردند [۱۲].

موردا و همکاران در پژوهش خود برای استخراج پاسخ از روش‌های معنایی، نقش‌های معنایی و شبکه واژگان استفاده کرده‌اند که سیستم پیشنهادی یک نمونه از سیستم‌های پرسش و پاسخ دامنه نامحدود است. هدف از این پژوهش تعیین عملکرد سیستم و میزان تحت تأثیر قرار گرفتن کارایی سیستم با به کارگیری اطلاعات معنایی بوده که در آن، مقایسه‌ای با سیستم‌های پرسش و پاسخ مبتنی بر شناسایی نهاد اسمی^۲ صورت گرفته است [۱۳].

موشتی و کوارترونی در پژوهش خود از مدل‌های با ناظر با به کارگیری نمونه‌هایی از پرسش و پاسخ‌ها برای انتخاب پاسخ، استفاده کرده‌اند. روش‌هایی همانند اندازه‌گیری مشابهت بین پرسش



شکل ۳: فلوجارت الگوریتم پیشنهادی

۴-۱- پردازش و تجزیه پرسش

قبل از پردازش و تجزیه پرسش، علائم نقطه گذاری با استفاده از عبارات منظم مانند علامت پرسش، کاما از پرسش حذف می شوند و پرسش برای تعیین نوع پرسش به بخش طبقه بند پرسش فرستاده می شود. بخش طبقه بند پرسش یک گام مهم و اساسی در سیستم های پرسش و پاسخ می باشد که نوع پاسخ مورد انتظار را تعیین می کند. اگر این بخش با دقت بالایی انجام گیرد، دقت سیستم پرسش و پاسخ در یافتن پاسخ های مرتبط با پرسش کاربر نیز افزایش می یابد. می توان دلیل این امر را وابستگی نوع پرسش و پاسخ مورد نظر دانست. ویژگی هایی که برای طبقه بندی پرسش در این پژوهش استفاده گردیده است ویژگی های لغوی و نحوی می باشد. ویژگی های لغوی یک پرسش بر اساس کلمات محتوای آن پرسش استخراج می گردد [۱۹-۲۱]. ویژگی نحوی به شکل کلمات درون پرسش گفته می شود. شکل کلمه به خواص ظاهری و بخش سخن تکواژه ها اشاره دارد. در دستور زبان، بخش های سخن، طبقه بندی های زبانی از کلمات هستند که رفتار نحوی یک قسمت از جمله را بیان می کنند. به طور عموم، تمامی زبان ها دارای دو بخش سخن، اسم و فعل می باشند و مابقی بخش های سخن در زبان های مختلف، متفاوت است. برای مثال، اسم، ضمیر، صفت، قید و حرف اضافه، از جمله مهم ترین بخش های سخن در زبان فارسی هستند.

ارزیابی سیستم پیشنهادی از پرسش های TREC 2006 استفاده گردیده است [۱۷].

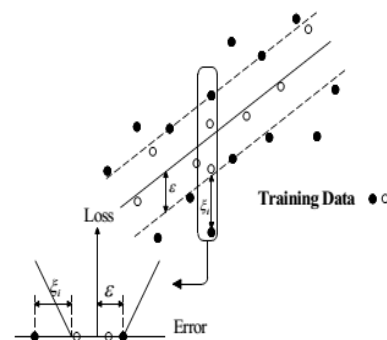
۳- رگرسیون ماشین بردار پشتیبان

ماشین بردار پشتیبان مبتنی بر کمینه کردن ساختاری ریسک می باشد که از تئوری آموزش آماری گرفته شده است. وپنیک برای ساختن شکل رگرسیون ماشین بردار پشتیبان از تابع خطای جدیدی استفاده نمود که به نام تابع خطای ϵ -insensitive خوانده می شود و به صورت رابطه ۱ بیان می گردد.

(۱)

$$L(y, f(x, \alpha)) = |y - f(x, \alpha)|_{\epsilon} = \begin{cases} 0 & \text{if } |y - f(x, \alpha)| \leq \epsilon \\ |y - f(x, \alpha)| - \epsilon & \text{otherwise} \end{cases}$$

با توجه به رابطه ۱، می توان دریافت که خطاهایی با مقادیر کمتر از ϵ در نظر گرفته نشده اند که در شکل ۲ نشان داده شده است. به عبارت دیگر در این تابع، خطاهایی که در محدوده کمتر از ϵ قرار دارند، متحمل جریمه نمی شوند. از این محدوده به نام لوله ϵ یاد می شود و در مسائل چندبعدی ناحیه ϵ -insensitive شکلی نظیر دال خواهد داشت و یا به طور کلی این محدوده بین دو آبرصفحه موازی قرار گرفته است.



شکل ۲: تابع خطای ϵ -insensitive [18]

۴- الگوریتم پیشنهادی

سیستم طراحی شده در این پژوهش یک سیستم پرسش و پاسخ می باشد که از بخش های پردازش و تجزیه پرسش، تولید عبارت جست و جو، موتورهای جست و جو، استخراج پاسخ و رتبه بندی پاسخ تشکیل شده است. برای رتبه بندی پاسخ ها از رگرسیون ماشین بردار پشتیبان استفاده گردیده است. فلوجارت روش پیشنهادی در شکل ۳ نشان داده شده است.

در ادامه بخش های الگوریتم پیشنهادی را بصورت جداگانه شرح

می دهیم.

۴-۲- تولید عبارت جست‌وجو

در این بخش با وارد کردن پرسش از طریق کاربر، درخت تجزیه پرسش مربوطه، ساخته شده و با استفاده از بخش سخن، نوع کلمات (کلمه پرسشی، اسم، فعل، صفت، قید) مشخص می‌گردد. حال با توجه به بخش سخن شناسایی شده از پرسش، کلمات را طی فرآیند زیر در دو دسته‌ی کلمات اصلی و کمکی قرار می‌دهیم.

✓ به عبارت اسمی با بخش سخن (NN|NNP|NNS|NNPS) و بیش‌ترین طول و دارای نهاد اسمی (اسم مکان، شخص) با توجه به این که احتمال داشتن صفحه مجزا برای آن‌ها در فضای وب بیشتر می‌باشد، اولویت بیشتری داده شده و در دسته‌ی کلمات اصلی قرار می‌گیرند.

✓ بعد از آن عبارات اسمی با بیش‌ترین طول و فاقد برچسب نهاد اسمی، در اولویت بعدی در دسته کلمات اصلی قرار می‌گیرند و اولویت‌دهی به عبارات اسمی به همین ترتیب بر اساس بیش‌ترین طول و داشتن برچسب نهاد اسمی یا فاقد آن ادامه می‌یابد.

✓ سایر عبارات استخراج شده از پرسش در صورتی که جزء افعال یا عبارات اسمی غیر اصلی نباشند در دسته‌ی کلمات کمکی قرار می‌گیرند

۴-۳- موتورهای جست‌وجو

بعد از شناسایی کلمات اصلی و کمکی، در این بخش با استفاده از کلمات اصلی، در صفحات ویکی‌پدیا جست و جو کرده و صفحه مرتبط با کلمات اصلی بازگردانده می‌شود. اگر با استفاده از کلمات کلیدی صفحه‌ای در ویکی‌پدیا یافت نشد، از کلمات کمکی برای جست‌وجو استفاده می‌شود و جست‌وجو بر اساس کلمات کمکی صورت می‌گیرد.

۴-۴- استخراج پاسخ

در بخش استخراج پاسخ سعی بر آن شده که از میان پاراگراف‌های استخراج شده در بخش موتورهای جست‌وجو، مرتبط‌ترین پاراگراف‌ها برای بخش انتخاب پاسخ فرستاده شود. روش کار به شرح زیر می‌باشد:

اگر کلمه کمکی تنها یک کلمه باشد، بخش سخن کلمه را به دست آورده و ریشه کلمه کمکی در صورت وجود به دست آورده می‌شود. اگر کلمه کمکی فعل باشد، زمان گذشته، گذشته کامل، حال و حال استمراری آن به دست می‌آید. اگر کلمه کمکی اسم باشد، جمع آن به دست می‌آید. اگر کلمه کمکی صفت باشد فقط ریشه آن ذخیره می‌گردد.

حال برای ذخیره پاراگراف‌ها شرایط زیر را در نظر می‌گیریم:

✓ اگر بخش سخن کلمه کمکی صفت نباشد، پاراگراف مورد نظر را جست‌وجو کرده و اگر شامل خود کلمه کمکی یا ریشه آن باشد، آن پاراگراف ذخیره می‌شود.

✓ اگر بخش سخن کلمه کمکی صفت و یک کلمه‌ای باشد، پاراگراف مورد نظر را جست و جو کرده و اگر شامل خود کلمه کمکی یا ریشه آن باشد، آن پاراگراف ذخیره می‌شود.

✓ اگر در پاراگراف مورد نظر هیچ‌کدام از موارد فوق یافت نشد، کلمه کمکی به شبکه واژگان فرستاده می‌شود و هم‌معنی‌های کلمه کمکی از شبکه واژگان استخراج می‌گردد، اگر هم‌معنی‌های کلمه کمکی بیشتر از پنج مورد باشد، پنج هم‌معنی اول کلمه کمکی مرتبط‌تر هستند و احتمال حضور آن‌ها در پاراگراف بیشتر است. در صورت وجود یکی از هم‌معنی‌ها در پاراگراف، آن پاراگراف ذخیره می‌شود.

۴-۵- رتبه‌بندی پاسخ

آخرین و مهم‌ترین مرحله در سیستم پرسش و پاسخ، بخش انتخاب پاسخ می‌باشد [۲]. در این بخش، سیستم پاراگراف‌ها را امتیازدهی کرده و مناسب‌ترین پاراگراف را انتخاب می‌کند. مراحل رتبه‌بندی پاراگراف به صورت زیر می‌باشد:

برای هر پاراگراف منتخبی که از بخش استخراج پاسخ به دست آمده است، ویژگی‌های معنایی، ساختاری و TF-IDF استخراج می‌شود و برای دادن امتیاز به پاراگراف، به رگرسیون ماشین بردار پشتیبان داده می‌شود. رگرسیون ماشین بردار پشتیبان بر اساس اطلاعاتی که قبلاً آموزش داده شده است به پاراگراف‌ها امتیاز می‌دهد و در نهایت پاراگراف‌ها بر اساس امتیاز مرتب و بالاترین امتیاز به‌عنوان پاراگراف مورد نظر انتخاب می‌شود. در شکل ۴ فلوچارت روش پیشنهادی بخش رتبه‌بندی پاسخ نشان داده شده است. ویژگی معنایی بر اساس ویژگی‌های معنایی قابل شناسایی از پاسخ به دست می‌آید و متکی به نوع پرسش تعیین شده در بخش طبقه‌بند پرسش است. برای مثال اگر نوع پرسش مورد نظر در جست و جو اسم مکان بود، در پاراگراف‌های به دست آمده در بخش استخراج پاسخ، این امر در نظر گرفته می‌شود که پاراگراف‌هایی با برچسب مکان در جملات به دست آمده، برای پاسخ موجود باشد. پاراگرافی که چنین تطابقی را ایجاد کند امتیاز یک و پاراگرافی که چنین تطابقی ایجاد نکند امتیاز صفر می‌گیرد.

کلیدی نقش فاعلی یا مفعولی داشته باشد امتیاز ۰.۵ به این ویژگی داده می‌شود.

- ✓ فقط در صورت حضور عبارت کلیدی بدون رخ دادن حالات فوق، به ویژگی امتیاز ۰.۲۵ می‌دهیم.
- ✓ اگر هیچ کدام از شرایط بالا در پاراگراف بازگشتی اتفاق نیفتاد به ویژگی مورد نظر امتیاز صفر داده می‌شود.
- ✓ در نهایت بیشترین امتیاز یافته شده در کل پاراگراف مورد نظر به عنوان امتیاز این پاراگراف در نظر گرفته می‌شود.

یکی از پرکاربردترین روابط در حوزه بازیابی اطلاعات ویژگی TF-IDF می‌باشد، این ویژگی برابر با حاصل ضرب فراوانی کلمه در فراوانی معکوس سند است. TF-IDF یک روش مبتنی بر چند سند می‌باشد، که در آن منظور از فراوانی کلمه، فقط تعداد تکرار کلمه در یک سند خاص است. همچنین منظور از فراوانی معکوس سند، تعداد اسنادی است که این کلمه در آن اسناد ظاهر شده است. دلیل مقبولیت این ویژگی نسبت به سایر روش‌ها را می‌توان سهولت استفاده، محاسبات کم و نتایج قابل قبول دانست. برای محاسبه این ویژگی از رابطه‌های زیر استفاده می‌شود:

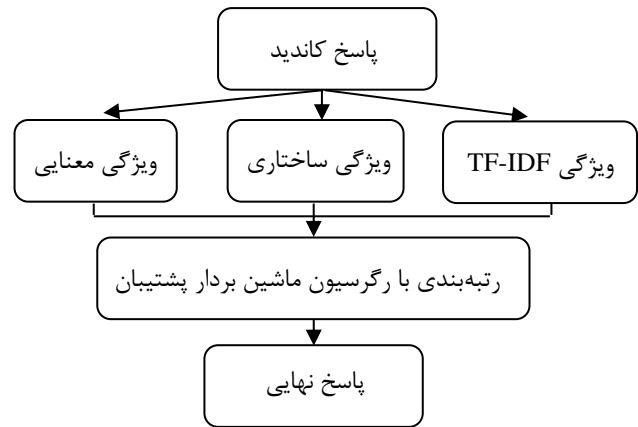
$$TF_{i,j} = N(t_i, d_j) \quad (2)$$

$$IDF_i = \log \frac{N}{N(t_i)} \quad (3)$$

$$W_{i,j} = TF_{i,j} \times IDF_i \quad (4)$$

که در روابط بالا $N(t_i, d_j)$ بیانگر تعداد رخ داد کلمه t_i در پاراگراف d_j ، N تعداد کل پاراگراف‌ها و $N(t_i)$ بیانگر تعداد پاراگراف‌هایی که کلمه t_i در آن‌ها وجود دارد، می‌باشد. مقدار $W_{i,j}$ که همان ویژگی TF-IDF است، حاصل ضرب فراوانی کلمه در فراوانی معکوس سند بوده و به عنوان ویژگی سوم در نظر گرفته می‌شود.

همانطور که در شکل ۴ نیز مشخص است، برای اعتبارسنجی پاسخ‌ها در این پژوهش از رگرسیون ماشین بردار پشتیبان استفاده گردیده است. در مدل رگرسیون ماشین بردار پشتیبان باید سه پارامتر C ، ϵ و پارامتر هسته γ تعیین گردند که نقش مهمی در رتبه‌بندی توسط رگرسیون ماشین بردار پشتیبان دارند. پارامتر C درجه ریسک تجربی رگرسیون ماشین بردار پشتیبان می‌باشد. ϵ عرض بازه خطا و γ عرض تابع گوسی از تابع هسته می‌باشد. مقادیر استفاده شده برای این پارامترها به ترتیب برابر ۰.۳، ۰.۰۰۳ و ۰.۰۱ می‌باشد. همچنین در این بخش، از ۲۰۰ داده‌ی آموزشی برای آموزش ماشین بردار پشتیبان استفاده شده است [۱۸ و ۲۲].



شکل ۴: فلوجارت الگوریتم پیشنهادی در بخش رتبه بندی پاسخ

ویژگی ساختاری، خصوصیات نحوی یک پاراگراف را برمی‌گرداند. بعد از تجزیه پاراگراف به اجزای سازنده‌اش، وابستگی‌های مربوط به هر جمله را به دست می‌آوریم. جهت یافتن ویژگی مورد نظر در هر پاراگراف در ابتدا در صورت وجود کلمات کلیدی (کلمات اصلی و کمکی استخراج شده در بخش تولید عبارت جست‌وجو) در پاراگراف، وابسته‌های پیشین و پسین آن را می‌یابیم. وابسته پیشین می‌تواند صفت اشاره، پرسشی، مبهم، تعجبی، شمارشی، عالی و شاخص باشد. وابسته پسین نیز می‌تواند ی نکره، نشانه‌های جمع، صفت شمارشی ترتیبی، صفت تفضیلی، مضاف الیه و یا صفت بیانی باشد. برای مثال، در عبارت این کتاب خواندنی، کلمه این وابسته پیشین و خواندنی وابسته پسین می‌باشد. پس از آن به صورت زیر عمل می‌شود:

✓ در صورتی که کلمه‌ی کلیدی استخراج شده در بخش تولید عبارت جست‌وجو، فعل باشد و در پاراگراف یافته شده برای پاسخ، کلمه‌ی فوق حضور داشته باشد، ویژگی ساختاری مورد نظر به این صورت امتیازدهی می‌شود که اگر عبارت کلیدی یافته شده در بخش تولید عبارت جست‌وجو یا وابسته‌های پیشین و پسین آن، نقش فاعلی یا مفعولی در رابطه با ریشه داشته باشند امتیاز یک به این ویژگی نسبت داده می‌شود.

✓ در صورتی که کلمه کلیدی استخراج شده در بخش تولید عبارت جست‌وجو اسم یا صفت باشد و در پاراگراف بازگشتی، کلمه فوق حضور داشته باشد، جمله‌ای که در پاراگراف شامل کلمه‌ی کلیدی است، مورد بررسی قرار می‌گیرد. اگر عبارت کلیدی یا وابسته‌های آن در جمله مورد نظر، در رابطه با اجزای جمله نقش فاعلی یا مفعولی داشته باشد برای ویژگی فوق امتیاز ۰.۷۵ در نظر می‌گیریم.

✓ اگر دو حالت فوق در پاراگراف بازگشتی رخ ندهد، در صورت وجود پاراگرافی که در آن جمله‌ای وجود داشته باشد که عبارت

۴-۶- تشریح روش پیشنهادی با طرح یک مثال

در این بخش مراحل روش پیشنهادی براساس یک مثال آورده می‌شود. برای مثال فرض کنید بخواهیم جواب سوال "When did Alfred Hitchcock die?" را بیابیم. در زیر دو پاراگراف و شیوه امتیازدهی به آن‌ها توسط روش پیشنهادی آورده شده است:

Paragraph 1: "Hitchcock died in his Bel Air home of renal failure at 9:17-am on 29 April 1980. While biographer Spoto wrote that Hitchcock "rejected suggestions that he allow a priest . . . to come for a visit, or celebrate a quiet, informal ritual at the house for his comfort," Jesuit priest Mark Henninger wrote that he and fellow priest Tom Sullivan performed Mass and heard Confession at the filmmaker's home."

Paragraph 2: "When the real Hitchcock died, the fictional Hitchcock in the Three Investigators books was replaced by a retired detective named Hector Sebastian."

در ادامه امتیازهای داده‌شده به پاراگراف‌ها توسط سیستم پیشنهادی آورده شده است:

ویژگی ساختاری: براساس تجزیه پاراگراف به اجزای سازنده‌اش، وابستگی پسین و پیشین کلیدی استخراج می‌گردد. از آنجایی که کلمه‌ی کلیدی در پاراگراف ۱ نقش فاعلی دارد، امتیاز ۰.۷۵ کسب می‌کند. به همین صورت، چون در پاراگراف ۲ نیز کلمه کلیدی نقش فاعلی دارد امتیاز ۰.۷۵ کسب می‌کند.

ویژگی معنایی: ویژگی معنایی دلالت بر نوع پرسش دارد و از آنجایی که نوع پرسش مربوط به زمان بوده و در پاراگراف ۱ زمان قید گردیده، بنابراین امتیاز یک کسب می‌کند. اما پاراگراف ۲ حاوی اسم زمان نبوده و در نتیجه امتیاز صفر به آن تعلق می‌گیرد.

ویژگی TF-IDF: این ویژگی از حاصلضرب فراوانی کلمه در فراوانی معکوس سند بدست می‌آید. امتیاز این ویژگی براساس رابطه‌ی چهار برای هر دو پاراگراف برابر ۰.۱۸ می‌باشد.

۵- نتایج آزمایش‌ها

در این بخش ابتدا داده‌ها و محیط پیاده‌سازی بیان شده و سپس پارامترهای ارزیابی و آزمایش‌ها شرح داده می‌شوند.

۵-۱- داده‌های مورد استفاده

یکی از مشکلات پژوهشی در حوزه‌ی سیستم‌های پرسش و پاسخ، کمبود مجموعه داده‌های استاندارد می‌باشد. جهت ارزیابی سیستم‌های پرسش و پاسخ، سالانه کنفرانس بازیابی متن برگزار می‌شود. در هر دوره، پرسش‌هایی مربوط به اتفاقات مهم سال گذشته به همراه اسنادی که جهت بازیابی اطلاعات می‌توان از آن‌ها

استفاده نمود به شرکت‌کنندگان ارائه می‌گردد. از آنجایی که داده‌های موجود در این کنفرانس وابسته به حوزه‌ی خصوصی بوده و برای عموم در دسترس نیست، جهت ارزیابی سیستم پیشنهادی، از داده‌های یکی از دوره‌های پیشین این کنفرانس (یعنی داده‌های کنفرانس بازیابی متن در سال ۲۰۰۷ که قابل دسترسی عمومی است) استفاده شده است [۱].

سیستم مورد نظر توسط ۴۰۰ پرسش مورد آموزش و ارزیابی قرار گرفته است که از این میان ۲۰۰ پرسش برای آموزش و ۲۰۰ پرسش برای ارزیابی سیستم استفاده گردیده است. ۱۴۴ پرسش از مجموعه پرسش‌های TREC-2007 و ۲۵۶ پرسش از مجموعه پرسش‌های ویکی‌پدیا انتخاب شده است.

۵-۲- محیط پیاده‌سازی

جهت پیاده‌سازی روش پیشنهادی از زبان برنامه‌نویسی جاوا در محیط توسعه NetBeans 11 استفاده گردیده است. جهت اجرای شبیه‌سازی از سیستمی با امکانات زیر استفاده گردیده است:

Operating System: Microsoft Windows 7
RAM: 8 G DDR3
CPU: Intel Core i7 5500 (2.4 GHz)

۵-۳- پارامترهای ارزیابی

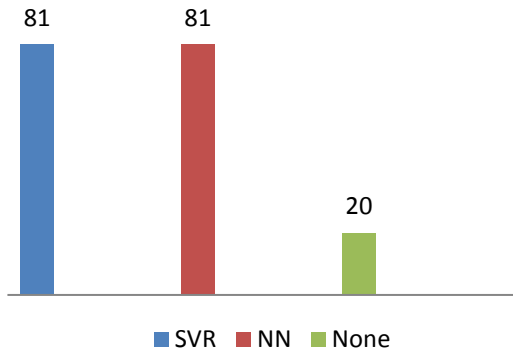
در این پژوهش جهت ارزیابی کارایی سیستم، از معیارهای صحت، میانگین رتبه‌بندی معکوس، دقت، فراخوانی و معیار F استفاده شده که صحت سیستم از رابطه ۵ به دست می‌آید:

$$Accuracy = \frac{m}{n} \times 100 \quad (5)$$

در اینجا m تعداد پرسش‌هایی است که برای آن‌ها حداقل یک پاسخ صحیح در میان پاسخ‌های بازیابی شده از سیستم پیشنهادی وجود داشته باشد و n تعداد کل پرسش‌های پرسیده شده است [۲]. پارامتر ارزیابی دیگر، میانگین رتبه‌بندی معکوس می‌باشد که طبق رابطه ۶ محاسبه می‌گردد [۱].

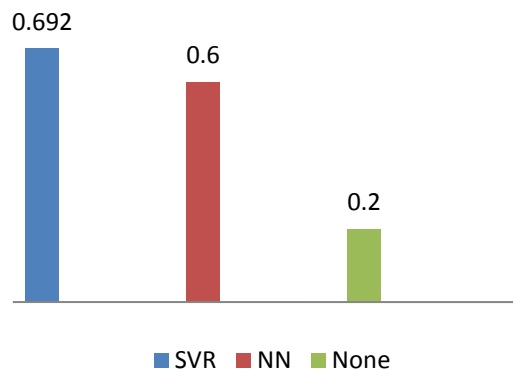
$$MRR = \frac{1}{N} \sum_{i=1}^N \frac{1}{Rank(Q_i)} \quad (6)$$

که N تعداد پرسش‌ها و $Rank(Q_i)$ رتبه صحیح‌ترین پاسخ به پرسش Q_i است. میانگین رتبه‌بندی معکوس، نشان‌دهنده‌ی دقت امتیازدهی به پاسخ‌هاست. هرچه مقدار میانگین رتبه‌بندی معکوس به یک نزدیک‌تر باشد نشان از برتری روش امتیازدهی است.



شکل ۵: مقایسه روشها براساس معیار صحت

همانطور که در شکل ۵ نشان داده شده است، صحت الگوریتمهای رتبه‌بندی با شبکه عصبی و رتبه‌بندی با رگرسیون ماشین بردار پشتیبان برابر می‌باشد. دلیل این برابری، یکسان بودن بخش‌های طبقه‌بندی پرسش، بازیابی اطلاعات و استخراج پاسخ می‌باشد. از همین رو پاسخ‌های استخراج شده به عنوان پاسخ‌های کاندید، در دو سیستم یکسان می‌باشد. دلیل برتری سیستم‌های رتبه‌بندی نسبت به سیستم‌های بدون رتبه‌بندی در پاراگراف‌های استخراج شده است، زیرا در سیستم‌های بدون رتبه‌بندی، اولین پاراگراف به عنوان جواب سیستم انتخاب می‌گردد در حالی که در سیستم‌های رتبه‌بندی، پاسخ‌های کاندید رتبه‌بندی شده و بهترین پاسخ که دارای بالاترین رتبه می‌باشد به عنوان جواب به کاربر نمایش داده می‌شود.



شکل ۶: مقایسه روشها براساس معیار میانگین رتبه‌بندی معکوس

همانطور که در شکل ۶ نشان داده شده است بر اساس معیار میانگین رتبه‌بندی معکوس، رتبه‌بندی با استفاده از ماشین بردار پشتیبان نسبت به سیستم‌های دیگر دارای برتری می‌باشد. دلیل این برتری گیر نیفتادن رگرسیون ماشین بردار پشتیبان در کمینه محلی، ایجاد حاشیه امن بزرگ و امتیاز دهی مناسب تر این روش

همچنین در این پژوهش از معیارهای دقت، فراخوانی و معیار F نیز برای ارزیابی سیستم استفاده شده است. این معیارها به ترتیب بر اساس روابط ۸، ۷ و ۹ محاسبه می‌گردند.

$$Precision = \frac{Questions\ Correct}{Question\ Answered} \quad (7)$$

$$Recall = \frac{Questions\ Correct}{Question\ Total} \quad (8)$$

$$F = \frac{2 * Precision * Recall}{Precision + Recall} \quad (9)$$

که در معادلات فوق $Questions\ Correct$ برابر تعداد پرسش‌های است که در آن پاسخ‌های صحیح بالاترین رتبه را دارند. $Questions\ Answered$ برابر تعداد پرسش‌هایی است که در پاسخ‌های کاندید به آن پرسش، حداقل یک پاسخ صحیح وجود داشته باشد. $Question\ Total$ برابر تعداد کل پرسش‌ها می‌باشد.

در جدول ۱ نتایج به دست آمده از ۲۰۰ پرسش برای سیستم‌های بدون رتبه‌بندی، رتبه‌بندی با شبکه عصبی و رتبه‌بندی با رگرسیون ماشین بردار پشتیبان آورده شده است.

جدول ۱: تعداد پاسخ‌های صحیح و بالاترین امتیاز

الگوریتم	تعداد پاسخ‌های صحیح	تعداد پاسخ‌های صحیح با بالاترین امتیاز
بدون رتبه‌بندی	۴۰	۴۰
رتبه‌بندی با شبکه عصبی	۱۶۲	۶۵
رتبه‌بندی با رگرسیون ماشین بردار پشتیبان	۱۶۲	۹۰

در ادامه برای بررسی رفتار الگوریتم پیشنهادی و ارزیابی سیستم، سیستم پیشنهادی را با دو سیستم بدون رتبه‌بندی (None) و رتبه‌بندی با شبکه عصبی (NN) براساس معیارهای معرفی گردیده مورد مقایسه قرار می‌دهیم. در این پژوهش از پارامترهای موجود در جدول ۲ برای شبکه‌ی عصبی استفاده شده است.

جدول ۲: پارامترهای شبکه عصبی

تعداد بکار	تعداد انتقال	تعداد آزمون ریزش	تعداد دوره‌های یادگیری	تعداد لایه‌ها
۱۰۰۰	سیگموئید	۰.۴	۷	شبکه عصبی FeedForward

همان‌طور که در شکل ۹ نشان داده شده است، معیار F سیستم پیشنهادی نسبت به سیستم‌های دیگر بالاتر بوده که نشان دهنده برتری سیستم پیشنهادی به عنوان یک سیستم پرسش و پاسخ می‌باشد.

در جدول ۳، مقایسه‌ای بر اساس معیار میانگین رتبه‌بندی معکوس بین سیستم پیشنهادی با دیگر پژوهش‌های انجام شده در حوزه سیستم‌های پرسش و پاسخ صورت گرفته است. لازم به ذکر است که از مجموعه داده TREC برای انجام این مقایسه استفاده شده است. علاوه بر این، در جدول ۴ نیز مقایسه‌ای بر اساس معیار صحت بین سیستم پیشنهادی با دیگر پژوهش‌های انجام شده در حوزه سیستم‌های پرسش و پاسخ صورت گرفته است. بر اساس نتایج این مقایسه‌ها، سیستم پیشنهادی عملکرد مناسبی از خود نشان داده است.

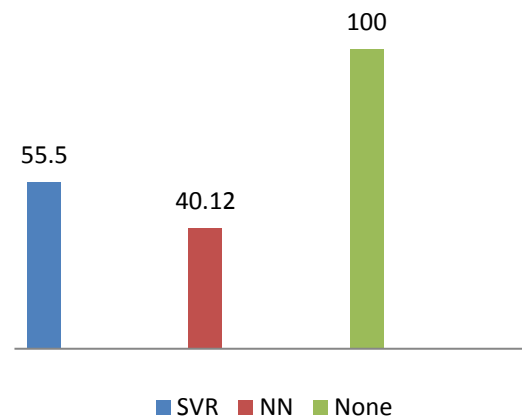
جدول ۳: مقایسه میانگین رتبه‌بندی معکوس سیستم پیشنهادی

سیستم مورد نظر	میانگین رتبه‌بندی معکوس
سیستم پیشنهادی	۰.۶۹۲
شو جون یی و همکاران (۲۰۱۳) [۲۳]	۰.۵۶۳
متایاس و همکاران (۲۰۱۲) [۱۷]	۰.۵۷
هیو جانگ و همکاران (۲۰۱۱) [۲۴]	۰.۷۱۵

جدول ۴: مقایسه صحت سیستم پیشنهادی

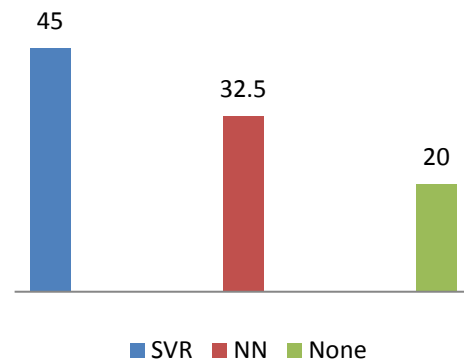
سیستم مورد نظر	مجموعه داده استفاده شده	تعداد پرسش‌های صحیح از کل پرسش‌ها	صحت سیستم (درصد)
سیستم پیشنهادی	مجموعه پرسشی ویکی‌پدیا و TREC - 2007	۱۶۲/۲۰۰	۸۱
قبادی (۱۳۹۲) [۲]	مجموعه پرسش ویکی‌پدیا و TREC - 2007	۱۴۱/۲۰۰	۷۰.۵
طهری و همکاران (۲۰۱۳) [۲۵]	TREC - 10	۷۲/۲۰۰	۳۶
کو و همکاران (۲۰۱۰) [۱۵]	TREC - 13-15	۵۲۴/۹۹۸	۵۲.۵
همکاران (۲۰۱۲) [۱۶]	TREC - 8-11	۲۵۳/۴۹۲	۵۱.۳

در مقایسه با شبکه عصبی می‌باشد.



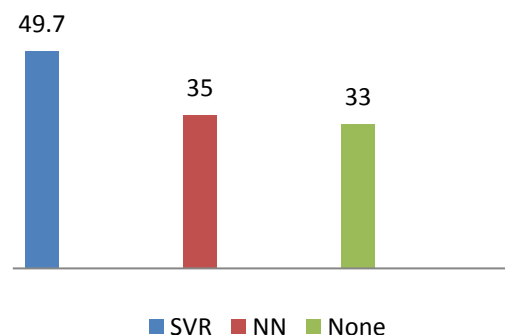
شکل ۷: مقایسه روش‌ها بر اساس معیار دقت

همان‌طور که در شکل ۷ نشان داده شده است، سیستم بدون رتبه‌بندی از لحاظ معیار دقت نسبت به دیگر سیستم‌ها دارای برتری می‌باشد زیرا به پاسخ‌های صحیح انتخابی در سیستم‌های بدون رتبه‌بند بالاترین امتیاز، که امتیاز یک می‌باشد، اختصاص داده می‌شود.



شکل ۸: مقایسه روش‌ها بر اساس معیار فراخوانی

همچنین نتایج شکل ۸ نشان از برتری سیستم پیشنهادی نسبت به سیستم‌های دیگر در معیار فراخوانی دارد.



شکل ۹: مقایسه روش‌ها بر اساس معیار F

۶- نتیجه‌گیری

در این پژوهش از رگرسیون ماشین بردار پشتیبان برای اعتبار-سنجی و رتبه‌بندی پاسخ‌ها در سیستم پرسش و پاسخ استفاده شد. سیستم پیشنهادی با استفاده از ۲۰۰ پرسش از مجموعه پرسش‌های TREC و پرسش‌های ویکی‌پدیا ارزیابی گردیده است. همچنین در این پژوهش سیستم پیشنهادی با استفاده از معیارهای صحت، میانگین رتبه‌بندی معکوس، دقت، فراخوانی و معیار F مورد بررسی و ارزیابی قرار گرفته است. نتایج به دست آمده از این معیارها نشان‌دهنده کارایی قابل قبول سیستم پیشنهادی در مقایسه با سیستم‌های بدون رتبه‌بندی و رتبه‌بندی با شبکه‌های عصبی بوده است.

سپاسگزاری

نویسندگان بر خود لازم می‌دانند از جناب آقای دکتر خلیلی که نظرات ارزشمند ایشان موجب بهبود کیفیت پژوهش و مقاله حاضر گردید، تشکر و قدردانی نمایند.

مراجع

- [1] سمیرا بختیار قلعه تکی، میر حسین دزفولیان و محرم منصوری زاده، "ارائه روشی ترکیبی جهت امتیاز دهی پاسخ در سیستم‌های پرسش و پاسخ"، کارشناسی ارشد، دانشگاه بوعلی سینا، همدان، صفحه‌های ۵-۲۵، بهمن ۹۲.
- [2] راحیل قبادی، میر حسین دزفولیان و محرم منصوری زاده، "سامانه هوشمند سیستم پرسش و پاسخ چند سطحی در محیط‌های یادگیری الکترونیکی"، کارشناسی ارشد، دانشگاه بوعلی سینا، همدان، صفحه‌های ۱۰-۴۵، بهمن ۹۲.
- [3] M. R. Kangavari, S. Ghandchi, M. Golpour, "A New Model for Question Answering Systems." Proceedings of World Academy of Science: Engineering & Technolig, pp. 3-5, 2008.
- [4] O. Kolomiyets and M. F. Moens, "A survey on question answering technology from an information retrieval perspective", Information Sciences., vol.181, pp. 5412-5434, 2011.
- [5] M. A. GreenWood, "Open-Domain Question- Answering", Ph.D. Thesis, University of Sheffield, UK, pp. 23-156, 2005.
- [6] P. Gupta and V. Gupta, "A survey of text question answering techniques", International Journal of Computer Applications., vol. 53, no. 4, pp. 1-8, 2012.
- [7] C. D. Manning, P. Raghavan and H. Schütze, "Introduction to information retrieval", Cambridge University press, 2008.
- [8] S. Tellex, B. Katz, J. Lin, A. Fernandes and G. Marton, "Quantitative evaluation of passage retrieval algorithms for question answering", In Proceedings of the 26th Annual International ACM SIGIR Conference, pp. 41-47, 2003.
- [9] G. Lee, J. Seo, S. Lee and H. Jung, "SiteQ: Engineering High Performance QA System Using Lexico-Semantic Pattern Matching and Shallow NLP", in TREC, 2001.
- [10] U. Hermjakop, "Parsing and question classification for question answering", In Proceeding of the workshop on Open domain question answering; Vol. 12, pp. 1-6; 2001.
- [11] S. K. Ray, S. Shailendra and P. Bhagwati, "A Semantic Approach for Question Classification using WordNet and Wikipedia." Pattern Recognition Letters, vol. 31, pp. 1935-1943, 2010.
- [12] K. Hacioglu and W. Ward, "Question Classification With Support Vector Machine and Error Correcting Codes", In Proceeding of the 2003 conference of the North American Chapter of Association for Computational Linguistics on Human Language Technology: comparing volume of the Proceeding of HLT-NAACL, pp. 28-30, 2003.
- [13] P. Moreda, H. Lioners, E. Saquete and M. Palomar, "Combining Semantic Information in Question Answering System", Information Processing & Management, vol.47, pp. 870-885, 2011.
- [14] A. Moschitti and S. Quarteroni, "Linguistic Kernel for Answer Ranking in Question Answering"; Information Processing & Management, vol.47, pp. 825-842, 2011.
- [15] J. Ko, L. Si and E. Nyberg, "Combining Evidence With a Probabilistic Framework for Answer Ranking and Answer Merging in Question Answering", Information Processing & Management, vol.46, pp. 541-554, 2010.
- [16] M. Umamehaswari, M. Ramprasath and H. Hariharan, "Improved Question Answering System by Semantic Reformulation", In Advanced Computing (ICOAC), 40th International Conference on, pp. 1-4, 2012.
- [17] H.H. Matthias, W.D.W. Edward and F. Sadaoki, "Question Answering Using Statistical Language Modeling", Computer Speech and Language, vol.26, pp. 193-209, 2012.
- [18] W. Wang, C. Men and W. Lu, "Online Prediction Model Based on Support Vector Machine", Neurocomputing, vol.71, pp. 550-558, 2008.
- [19] B. Loni, M. long and D. MJ. Tax, "Enhanced Question Classification with Optimal Combination of Features", Master of Science Thesis, Department of Media and Knowledge Engineering, Delft university of Technology, 2011.
- [20] Z. Huang, M. Thint and Z. Qin, "Question Classification Using Headwords and their Hypernyms", In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 927-936, 2008.
- [21] P. Blunsom, K. Kocik and J.R. Curran, "Question Classification With Loglinear Models"; In Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 615-616, 2006..
- [22] V. Cherkassky and Y. Ma, "Practical Selection of SVM Parameters and Noise Estimation for SVM Regression", Neural Network, vol.17, pp. 113-126, 2004.
- [23] S. J. Yen, Y. C. Wu, J. C. Yang, Y. S. Lee, C. J. Lee and J.J. Liu, "A Support Vector Machine Based Context Ranking Model for Question Answering." Information Sciences, vol.224, pp. 77-87, 2013.
- [24] H. J. Jang, K.Y. Sung, M.G. Jang and S.H. Myaeng, "Compositional question answering: A divide and conquer approach", Information Processing & Management, vol.47, pp. 808-824, 2011.
- [25] A. Tahri, and O. Tibermercine "DBPEDIA Based Factoid Question Answering System", International Journal of Web & Semantic Technology, vol.4, pp. 1-16, 2013.

پاورقی‌ها:

⁴ Structural kernels
⁵ Generalization methods
⁶ Bag of Words (BoW)

¹ Support Vector Regression (SVR)
² Named Entity Recognition (NER)
³ Manual annotation