

یک مدل آماری جهت ارزیابی سامانه های پرسش و پاسخ تعاملی با استفاده از رگرسیون

محمد مهدی حسینی^{۱*}, مرتضی زاهدی^۲ و حمید حسن پور^۳

^۱ دانشکده مهندسی برق و کامپیوتر، دانشگاه آزاد اسلامی، واحد شهرود، شهرود، ایران

^{۲,۳} دانشکده مهندسی کامپیوتر و فن آوری اطلاعات، دانشگاه صنعتی شهرود، شهرود، ایران

چکیده

همانند بسیاری از زمینه های دیگر زبان شناسی محاسباتی، ارزیابی نقش مهمی در سامانه های پرسش و پاسخ تعاملی ایفا می کند. با این وجود، در زمینه ارزیابی سامانه های پرسش و پاسخ تعاملی به طور تقریبی هیچ روش خاصی وجود ندارد که به ارزیابی کلی این سامانه ها پرداخته و همواره انسان باید در فرآیند ارزیابی مشارکت داشته باشد. ارائه مدلی که بتواند جایگزین انسان در فرآیند ارزیابی شود، یکی از موضوعات مورد توجه در این حوزه است. در این مقاله، یک مدل آماری مناسب برای ارزیابی سامانه های پرسش و پاسخ تعاملی جهت جایگزین کردن به جای انسان توسط مجموعه ای از ویژگی های جدید و رگرسیون ارائه شده است. با استفاده از چهار سامانه تعاملی موجود پایگاه داده ای مناسب ایجاد شد. تعداد ۵۴۰ نمونه به عنوان داده مناسب در نظر گرفته شد تا مجموعه آزمون و آموزش بر اساس آن تشکیل شود. ابتدا پیش پردازش بر روی مکالمات صورت پذیرفت و بر اساس روابط تعریف شده، ویژگی های آماری از متن مکالمه ها استخراج و بر اساس آن ماتریس ویژگی تشکیل و سپس با استفاده از انواع رگرسیون سعی شد تا بهترین مدل استخراج شود که در نهایت رگرسیون غیرخطی سری توانی با RMSE به میزان ۰/۱۳ بهترین مدل را ارائه کرد.

واژگان کلیدی: ارزیابی، سامانه پرسش و پاسخ تعاملی، رگرسیون غیرخطی، استخراج ویژگی

A New Statistical Model for Evaluation Interactive Question Answering Systems Using Regression

Mohammad Mehdi Hosseini^{*1}, Morteza Zahedi² & Hamid Hassanpour³

¹Department of Computer Engineering, Islamic Azad University, Shahrood branch,
Shahrood, Iran

^{2,3}Faculty of Computer Engineering, Shahrood University of Technology, Shahrood, Iran

Abstract

The development of computer systems and extensive use of information technology in the everyday life of people have just made it more and more important for them to make quick access to information that has received great importance. Increasing the volume of information makes it difficult to manage or control. Thus, some instruments need to be provided to use this information. The QA system is an automated system for obtaining the correct answers to questions posed by the human in the natural language. In these systems, if the response is found, and if it is not the user's expected response or if it needs more information, there is no possibility of exchanging information between the system and the user to ask more questions and get answers related to it. To solve this problem, interactive Question answering (IQA) systems were created. Interactive question answering (IQA) systems are associated with linguistic ambiguous structures, so these systems are more accurate than QA systems. Regarding the probability of ambiguity (ambiguity in the user question or ambiguity in the answer provided by the system), the repetition is possible in these systems to obtain the clarity. No standard methods have been developed on IQA systems evaluation, and the existing evaluation

* Corresponding author

نویسنده عهده دار مکاتبات

methods have been developed based on the methods used in QA and dialogue systems. In evaluating IQA systems, in addition to quantitative evaluation, a qualitative evaluation is used. It requires users' participation in the evaluation process to determine the success level of interaction between the system and the user. Evaluation plays an important role in the IQA systems. In the context of evaluating IQA systems, there is partially no specific methodology for evaluating these systems in general. The main problem with designing an assessment method for IQA systems lies in the rare possibility to predict the interaction part. To this end, human needs to be involved in the evaluation process. In this paper, an appropriate model is presented by introducing a set of built-in features for evaluating IQA systems. To conduct the evaluation process, four IQA systems were considered based on the conversation exchanged between users and systems. Moreover, 540 samples were considered as suitable data to create a test and training set. The statistical characteristics of each conversation were extracted after performing the preprocessing on them. Then a feature matrix was formed based on the obtained characteristics. Finally, using linear and nonlinear regression, human thinking was predicted. As a result, the nonlinear power regression with 0.13 Root Mean Square Error (RMSE) was the best model.

Keywords: Evaluation, Interactive Question Answering Systems, Nonlinear Regression, Feature Extraction.

ارائه نکرده‌اند؛ بنابراین نبود تعامل دوطرفه بین سامانه و کاربر یکی از مهم‌ترین مشکلات سامانه‌های QA محسوب می‌شود [7]. با اضافه‌شدن سطح تعامل در سامانه‌های پرسش و پاسخ تعاملی (IQA⁴) این مشکل رفع شده است. سامانه‌های موجود در زمینه IQA می‌توانند با توجه به شرایط و کاربردهایشان در سه گروه مختلف شامل مدیریت محدودیت، QA ارتقا‌یافته و سؤالات متوالی قرار گیرند [5]. وجود یک سامانه ارزیابی استاندارد نقش بسیار مهمی در ارتقای سامانه‌های IQA ایفا می‌کند. با این وجود به طور تقریبی هیچ روش استانداری در زمینه ارزیابی این سامانه‌ها وجود ندارد و روش‌های ارزیابی فعلی بر مبنای روش‌های مورد استفاده در QA و سامانه‌های دیالوگ بنا شده‌اند. یک سامانه IQA از دو موجودیت سامانه و کاربر تشکیل شده و ممکن است، کاربر تحت تأثیر عوامل بسیاری با سامانه کار کند؛ لذا کار ارزیابی بسیار سخت و پیچیده است. اگرچه روش‌های استانداردی وجود دارند که می‌توانند اطلاعات مربوط به عملکرد سامانه از قبیل زمان، دقیقت و یا ارزیابی را با استفاده از آنها به دست آورد، اما هنوز، نیاز به شناسایی سهم سامانه و کاربر در عملکرد مطلوب یک سامانه است؛ بنابراین بیشتر سامانه‌های ارزیابی موجود، از ارزیابی انسانی بهره می‌گیرند که در نتیجه، عملکرد یک سامانه از کاربری به کاربر دیگر متفاوت خواهد بود [4]. بر اساس مطالعات صورت گرفته در زمینه ارزیابی سامانه‌های IQA توسط انسان، پارامترهای مختلفی جهت ارزیابی مورد توجه قرار می‌گیرد؛ بنابراین با توجه به گسترش روزافزون سامانه‌های IQA و کم کردن خطای ارزیابی انسانی، نیاز به جایگزینی یک مدل خودکار به جای ارزیاب انسانی است؛ لذا ابتدا باید پارامترهای مؤثر در ارزیابی مشخص و روشنی برای اندازه‌گیری خودکار آنها ارائه شود که این مسئله خود یکی از

⁴ Interactive Question Answering system (IQA)

۱- مقدمه

سامانه‌های پرسش و پاسخ¹ (QAS) به عنوان سامانه‌ای با پتانسیل بالا شناخته می‌شود که کاربران را قادر می‌سازد تا به منابع علمی با استفاده از زبان طبیعی (از طریق پرسش) دسترسی داشته باشند و یک پاسخ مرتبط، مناسب و مختصراً دریافت کنند. با این حال، همچنان مشکلات چالش برانگیز فراوانی جهت مرتفع کردن در این سامانه‌ها موجود است. سامانه‌های QA شکل پیچیده‌تر سامانه‌های بازیابی اطلاعات هستند که در این سامانه‌ها به جای ارائه کل سند، تنها بخش‌های خاصی از اطلاعات سند به عنوان پاسخ بازگردانده می‌شود؛ بنابراین پاسخ ارائه شده ممکن است یک واژه، یک جمله یا یک پاراگراف باشد. یک سامانه QA از سه بخش پردازش پرسش، بازیابی اطلاعات و پردازش پاسخ تشکیل می‌شود. از سال ۱۹۹۹ هر ساله پژوهش درباره سامانه‌های پاسخ دامنه‌باز² که از منابع اطلاعات غیرساختاری بهره می‌برند، توسط کمپین ارزیابی TREC³ به طور مرتب در حال انجام است [4, 7, 12]. در کمپین‌های بعدی TREC با توجه به افزایش تعداد و پیچیدگی درخواست‌ها، اسناد مورد استفاده و پیچیدگی سؤالات روش‌های ارزیابی پاسخ نیز پیشرفت‌تر شدند. با گسترش صفحات وب، استفاده از این مجموعه اطلاعات برای سامانه‌های QA مورد توجه قرار گرفت و چندین سامانه QA بر مبنای وب توسعه یافتند [5-8]. سامانه‌های QA مبتنی بر وب را می‌توان به QAS دامنه‌باز و QAS دامنه‌بسته طبقه‌بندی کرد [11].

سامانه‌های QA راهکاری برای رفع ابهام در زمانی که پرسش کاربر دارای ابهام بوده، یا اینکه پاسخ سامانه مطلوب کاربر نبوده یا کاربر نیازمند دریافت اطلاعات بیشتری باشد،

¹ Question Answering System

² Open Domain System

³ Text Retrieval Evaluation conference

بر اساس خروجی تولیدشده از سؤال و پرسش‌های ردوبل شده بین کاربر و سامانه (که در این مقاله مکالمه نامیده می‌شود) با استفاده از مجموعه ویژگی‌های آماری جدید و رگرسیون، مدلی ارائه شده است تا بتوانیم گام مهمی در این راستا برداریم.

ساختار مقاله بدین صورت است که در بخش نخست مروری بر کارهای انجامشده در زمینه ارزیابی سامانه‌های QA، IQA صورت پذیرفته است. در بخش دوم مجموعه ویژگی‌های پیشنهادشده، سامانه IQA پایه تولیدشده و مدل دسته‌بندی تشریح شده است. بخش سوم نتایج به دست آمده را نشان می‌دهد و در بخش آخر به نتیجه‌گیری و پیشنهادها پرداخته شده است.

۲- کارهای مرتبط

ارزیابی سامانه‌های QA بسته به ارزیابی سؤالات پیچیده یا ساده متفاوت است. یکی از روش‌های ارزیابی مورد استفاده در سامانه‌های QA استفاده از مجموعه‌ای از سؤالات و پاسخ‌ها به نام «مجموعه استاندارد طلایی» است [7]. در این روش توانایی یک سامانه بر اساس میزان منطبق‌بودن سامانه با این مجموعه استاندارد طلایی، مورد سنجش قرار می‌گیرد. البته این روش برای سؤالات پیچیده و مبهم هنوز تقویت نشده است. در ارزیابی سامانه‌های QA با استفاده از کاربران واقعی پژوهش‌های قابل توجهی وجود دارد. بیشتر ارزیابی‌های صورت‌پذیرفته در این حوزه، بیشتر در زمینه ارزیابی استخراج پاسخ، نحوه تعامل و استفاده از آن انجام شده است. بیشتر روش‌های پیاده‌سازی شده در زمینه ارزیابی سامانه‌های CWS¹, MRR², K1 C@1 استفاده کردند که هر کدام از این روش‌ها خود دارای نقاط ضعف بوده و قابلیت تعمیم به همه سامانه‌های مختلف می‌باشد. می‌دانند که در این حوزه تقویت نشده است. در این روش مثال معیار MRR زمانی به کار گرفته می‌شود که سامانه برای پاسخ به سؤال مطرح شده، چندین جواب را ارائه می‌کرد؛ اما در سامانه‌هایی که در مجموعه داده‌های خود تنها یک پاسخ برای هر سؤال ارائه می‌کردند، از روش ارزیابی C@1 استفاده می‌شد. بنابراین، این یکی از معضلات استفاده از این روش‌ها در سامانه‌های IQA بود و از طرفی این معیارها بیشتر در جهت انتخاب پاسخ به کار گرفته می‌شدند و توانایی سامانه کلی کاربر از سامانه بود.

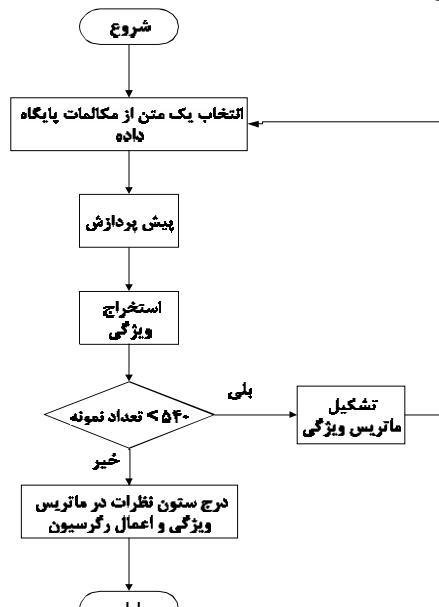
³ Cross Evaluation

¹ Confidence Weighted Score

² Mean Reciprocal Rank

۳- روش پیشنهادی

در فرآیند ارزیابی سامانه‌های IQA افراد متخصص و خبره نقش دارند، حدس این که این افراد در پس زمینه ذهن خود از چه تابع ارزیابی برای نمره‌دهی به یک سامانه استفاده می‌کنند، یکی از چالش‌های موجود در زمینه ارزیابی سامانه‌های IQA است؛ بنابراین تعریف ویژگی‌هایی که بتواند در مدل‌سازی به کار گرفته شوند، تا مدل به دست آمده کمترین خطأ را نسبت به نظرات موجود داشته باشد، امری ضروری است. از آنجایی که ویژگی‌های متعددی در ارزیابی یک سامانه IQA دخالت دارند و اندازه‌گیری خودکار آنها برای ایجاد یک مدل دارای اهمیت است، در این مقاله مجموعه‌ای از ویژگی‌های جدید برای ارزیابی خودکار این سامانه‌ها پیشنهاد شده است. اساس روش پیشنهادی بر روی معرفی مجموعه‌ای از این ویژگی‌های جدید و تأثیر هر ویژگی در مدل‌سازی است. هدف از این کار، یافتن ویژگی‌هایی بود که، خروجی مدل بتواند با کمترین خطأ، خروجی نزدیک به نظر داده شده توسط ارزیاب تولید کند. شکل (۱) روند نمایی روش پیشنهادی را نمایش می‌دهد.



(شکل-۱): روند کلی روش پیشنهادی
(Figure-1): Flow chart of Proposed algorithm

۱-۳- سامانه تعاملی پایه

از سامانه تعاملی پایه طراحی شده در آزمایشگاه پژوهشی فناوری زبان‌های طبیعی^۱ دانشگاه صنعتی شهرورد جهت انجام فرآیند مدل‌سازی ارزیابی استفاده شد. از فناوری‌های آماری جهت پاسخ به سوالات کاربران در این سامانه بهره گرفته شده

(جدول-۱): یک نمونه از پایگاه داده جهت آموزش سامانه پایه
(Table-1): A sample of database to learn base IQA

<Query>
<Question> غیبی که دانشجوی کارشناسی ارشد مجازی در
هر درس مجاز است چقدر است</Question>
<General-answer> تأیید میزان حضور دانشجو در فضای
الکترونیکی بر اساس سامانه آموزشی موسسه و عدم مشاهده به
منزله غیبت نلقي می‌گردد. میزان غیبت مجاز در مشاهده دروس
و سایر فعالیت‌های یادگیری هر درس حداقل معادل سه شانزدهم
است. نحوه محاسبه حضور و غیاب و ضوابط مرتبط با آن از سوی
شورای آموزشی دانشگاه تعیین می‌شود.</General-answer>
<acceptable-answer> میزان غیبت مجاز در مشاهده دروس و
سایر فعالیت‌های یادگیری هر درس حداقل معادل سه شانزدهم
است.</acceptable-answer>

² <http://IILtech.ir/fa/index/category/53>

¹ www.IILtech.ir



متون فارسی یکسان‌سازی حروف (مثل حروف "ی" و "ک") صورت گرفت.

تمامی این کارها به صورت خودکار انجام و جواب نهایی توسط ناظر انسانی کنترل شد.

۳-۳- استخراج ویژگی

یکی از مهم‌ترین مراحل مربوط به هر سامانه، تشخیص یا مدل‌سازی استخراج ویژگی است. در این مرحله، تعدادی ویژگی آماری بر اساس ۱۱-گرم‌ها و بزرگ‌ترین رشته مشترک (LCS^۴) تعریف شد. در ادامه هر کدام از این ویژگی‌ها تشریح شده است. ایده برخی از فرمول‌های تعریف شده برای این کار، از روابط تعریف شده در زمینه ارزیابی خودکار متن‌های خلاصه‌سازی شده اقتباس شده‌اند [۱۰]؛ اما مناسب با کار زمینه ارزیابی سامانه‌های پرسش و پاسخ تعاملی این روابط بازنویسی و بهروزسازی شده‌اند. از آنجایی که خروجی هر مکالمه بین کاربر و سامانه به صورت مجموعه‌ای از سؤال‌ها و پاسخ‌ها است، بعضی از ویژگی‌های تعریف شده، علاوه‌بر این که برای مجموعه سؤال-جواب مورده استفاده قرار گرفت، به صورت جداگانه برای مجموعه سؤال‌ها و مجموعه جواب‌ها نیز به کار گرفته شد که در هر ویژگی توضیح داده شده است.

- ویژگی نخست:

N-گرم‌ها یکی از مشهورترین مدل‌های آماری زبان هستند. در این مدل‌ها ارتباطات زنجیره واژگان در نظر گرفته می‌شود. به عبارت دیگر، مدل‌های ۱۱-گرم براساس همپیوندی و کنار هم قرار گرفتن نویسه‌های لغات در پردازش متن عمل می‌کنند. ابتدا ۱۱-گرم‌های مشترک را شمرده با یکدیگر جمع و بر تعداد کل ۱۱-گرم‌ها تقسیم می‌کنیم (رابطه ۱):

$$\text{Count_N} = \sum_{S_i \in \text{conv}} \frac{\sum_{ngram \in S_i} \text{Count}_{match}(gram_n)}{\sum_{ngram \in S_i} \text{Count}(gram_n)} \quad (1)$$

که در این رابطه S_i ، n -امین جمله از هر مجموعه مکالمه و n طول هر ۱۱-گرم است. فرض کنید یک مکالمه شامل N سؤال و پاسخ باشد. ۱۱-گرم‌های مشترک بین هر سؤال جواب را شمرده با یکدیگر جمع و بر مجموع تعداد ۱۱-گرم‌ها تقسیم می‌کنیم. این کار به ازای $n=1, 2, 3$ برای مجموعه‌های پرسش و پاسخ (Q-A)، مجموعه سؤال‌ها (Q-Q) و مجموعه جواب‌ها (A-A) از هر مکالمه صورت پذیرفت.

⁴ Longest Common Substring(LCS)

```

<exact-answer> سه شانزدهم <exact-answer>
</Query9>
.
.
.
<Query81>
<Question> مبلغ وام تحصیلی ممتاز چیست <General-answer> وام موارد خاص بیماری‌های خاص و برهزینه مبلغ وام ۲۰۰۰۰۰۰ ریال. وام تحصیلی ممتاز و نمونه حداکثر ۲ <General-answer> برابر وام تحصیلی. <acceptable-answer> وام تحصیلی ممتاز و نمونه حداکثر ۲ <acceptable-answer> برابر وام تحصیلی. <exact-answer> حداکثر ۲ برابر وام تحصیلی. </exact-answer>
<answer>

```

نتایج حاصل از این بهینه‌سازی در [۳] ارائه شد.

۳-۲- پیش‌پردازش

یکی از مراحلی که بر روی متون استخراج شده از سامانه‌های IQA صورت گرفت، نرمال‌سازی اطلاعات بود. در این مرحله متن ورودی به ساختاری قابل پردازش برای مراحل بعد تبدیل شد. مراحل نرمال‌سازی به ترتیب صورت گرفته شامل موارد زیر است:

۱- مشخص کردن مرز جمله‌ها: در بیشتر مواقع، تعیین مرز جمله‌ها از طریق بررسی علائم جداکننده از قبیل فضای خالی، "،"!، "؟، "؛" و غیره انجام می‌شود که با این علائم به تنهایی کافی نیست؛ لذا ما برای متون انگلیسی، علاوه‌بر این علائم از تجزیه‌کننده استنتفورد^۱ نیز استفاده کردیم.

۲- ریشه‌یابی: در این حالت یک واژه به شکل عمومی خود کاهش می‌یابد که این شکل عمومی باید برای همه واژگان هم ریشه یکسان باشد. برای دادگان انگلیسی از ریشه‌یاب استنتفورد و مجموعه دادگان فارسی از ابزار پردازش زبان دانشگاه فردوسی^۲ استفاده شد.

۳- حذف واژگان و واژه‌های غیر مهم: فهرستی مشتمل بر دویست واژه ایستواژه^۳ آماده (واژگانی که در محتوای اصلی متن تأثیری ندارند) و از مکالمه‌ها حذف شد.

۴- شناسایی مقادیر عددی: بعد از شناسایی اعدادی که به صورت حروف در مکالمه‌ها یادشده بودند، این واژگان برچسب مقدار عددی دریافت کردند.

۵- یکسان‌سازی متن‌ها: در متون انگلیسی تمامی واژگانی که با حروف بزرگ بودند به حروف کوچک تبدیل شدند و در

¹ <https://nlp.stanford.edu/software/lex-parser.html>

² wtlab.um.ac.ir

³ Stop-words

- ویژگی دوم:

در یک مکالمه برای n های بزرگ‌تر، هر چه تعداد n -گرم‌های مشترک بیشتر باشد، امتیاز آن مکالمه بیشتر خواهد بود و احتمال پیوستگی متن مکالمه بیشتر خواهد شد [6]. بر این اساس، در این ویژگی پیشنهادی، هر کدام از n -گرم‌ها، بر اساس ارزش یک ضریب وزنی برای هر n -گرم به ارزش W_k با یکدیگر جمع می‌شوند تا مقدار این ویژگی به دست آید (رابطه ۲):

$$\text{Count_Weight_N} = \frac{1}{M} \times \sum_{i=1}^M \frac{\sum_{n\text{-grams}_i} W_k \times \text{Count}_{\text{match}}(\text{gram}_n)}{\sum_{n\text{-grams}_i} \text{Count}(\text{gram}_n)} \quad (2)$$

که در آن M تعداد عضوهای مجموعه برای میانگین‌گرفتن و W_k ضریب تأثیر هر n -گرم و مقدار آن متناسب با عدد n است. این ویژگی نیز برای $n=1, 2, 3$ محاسبه شد. نحوه محاسبه این رابطه مانند ویژگی نخست است با این تفاوت که به‌ازای n -گرم‌های مشترک ضریبی برابر با n به هر n -گرم مشترک نسبت داده می‌شود.

- ویژگی سوم:

متناسب با رابطه (۲)، معادله (۳) پیشنهاد شد، در این ویژگی، ابتدا به‌ازای هر جفت سؤال-جواب، ابتدا n -گرم‌های مشترک را به‌دست آورده و در ارزش هر n -گرم ضرب، سپس بر مجموع تعداد n -گرم‌ها تقسیم کرده، بیشینه بین آنها را در نظر گرفته و درنهایت پاسخ به‌دست‌آمده از M تا جفت درون یک مکالمه آن را با استفاده از میانگین‌گیری نرمالیزه می‌کنیم:

$$\text{Count_Weightmax_N} = \frac{1}{M} \times \sum_{i=1}^M \text{argmax}_{n\text{-grams}_i} \left(\frac{W_k \times \text{Count}_{\text{match}}(\text{gram}_n)}{\sum_{n\text{-grams}_i} \text{Count}(\text{gram}_n)} \right) \quad (3)$$

- ویژگی چهارم:

در n -گرم‌ها طول تعريف نمی‌شود؛ زیرا بزرگ‌ترین رشته مشترک در نظر گرفته می‌شود. بنابراین انطباق پشت سر هم در سطح جمله به‌طور معمول در n -گرم‌ها دیده می‌شود. در حالی که در بزرگ‌ترین زیررشته مشترک، نیاز نیست انطباق پشت سر هم باشد؛ همچنین برای اینکه مسئله هم‌رخدادی در جملات در نظر گرفته شود، از معادله (۶) استفاده کردیم. در رابطه تعريف شده، برای یک مکالمه ابتدا یک جفت سؤال-پاسخ را در نظر گرفته، سپس برای هر جفت بازیابی و دقت را محاسبه و برای تمامی جفت سؤال-پاسخ این کار را انجام می‌دهیم. درنهایت پاسخ به‌دست‌آمده را در رابطه (۶) قرار داده و امتیاز هر مکالمه را محاسبه می‌کنیم:

$$R_{LCS} = \frac{1}{M} \times \sum_{i=1}^M \frac{\text{LCS}(Q_i, A_i)}{L_{Q_i}} \quad (4)$$

$$P_{LCS} = \frac{1}{M} \times \sum_{i=1}^M \frac{\text{LCS}(Q_i, A_i)}{L_{A_i}} \quad (5)$$

$$F_{LCS} = \frac{(1+\beta^2)R_{LCS}P_{LCS}}{R_{LCS} + \beta^2 P_{LCS}} \quad (6)$$

که در آن M تعداد جفت سؤال-پاسخ هر مکالمه، $\beta = \frac{P_{LCS}}{R_{LCS}}$ و L_{Q_i} طول سؤال i -مین مکالمه و L_{A_i} طول جواب i -مین مکالمه است.

- ویژگی پنجم:

در این ویژگی اجتماع بزرگ‌ترین زیررشته مشترک بین Q_i و A_i مجموعه جواب‌ها را محاسبه کردیم که هر چه این عدد بزرگ‌تر باشد، ارتباط بین جملات در مکالمه بیشتر است. به‌طور مثال فرض کنید جمله Q_1 شامل واژگان $w_1w_2w_3w_4w_5$ باشد و A_1 شامل $w_1w_2w_6w_7w_8$ و پاسخ A_2 شامل واژگان $w_1w_3w_8w_9w_5$ باشد، بنابراین $\text{LCS}(Q_1, A_1) = 4/5$ در رابطه بین Q_1 و A_2 برابر است؛ لذا اجتماع بین Q_1 و A_1 و A_2 برابر با $w_1w_2w_3w_5$ است که یک مکالمه این کار را انجام دادیم. بنابراین روابط به صورت زیر پیشنهاد شد:

$$R_{LCS} = \frac{1}{M} \times \sum_{i=1}^M \frac{\sum_{j=1}^U \text{LCS}_U(Q_i, A_j)}{P} \quad (7)$$

$$P_{LCS} = \frac{1}{M} \times \sum_{i=1}^M \frac{\sum_{j=1}^V \text{LCS}_U(Q_i, A_j)}{n} \quad (8)$$

$$F_{LCS} = \frac{(1+\beta^2)R_{LCS}P_{LCS}}{R_{LCS} + \beta^2 P_{LCS}} \quad (9)$$

که در آن M تعداد سؤالات در یک مکالمه و P مجموعه سؤالات خواهد بود. در این ویژگی Q_i شامل U جمله با P واژه و A_i مجموعه جواب‌ها شامل V جمله با n واژه است.

- ویژگی ششم:

در این ویژگی، بزرگ‌ترین زیررشته مشترک بین هر سؤال و مجموعه جواب‌ها را یافته و درون یک مجموعه قرار داده، سپس در بین همه اعضای این مجموعه بزرگ‌ترین زیررشته حاصل شده را انتخاب می‌کنیم. روابط به صورت زیر تعريف شد:

$$R_{LCS} = \frac{1}{N} \times \sum_{i=1}^N \max \left(\sum_{j=1}^P \frac{\text{LCS}(Q_i, A_j)}{L_{Q_i}} \right) \quad (10)$$

$$Score_{w_i} = \frac{\deg(w)}{freq(w)} \quad (16)$$

$$Score_{conv} = \frac{1}{N} \times \sum_{j=1}^N Score_{w_i} \quad (17)$$

- ویژگی نهم (فاصله همینگ):

استفاده از فاصله همینگ برای اندازه گیری تشابه بین جملات یک مکالمه یکی دیگر از ویژگی هایی است که مورد استفاده قرار گرفت. هر جمله در مکالمه، از تعدادی واژه تشکیل شده است؛ با بدست آوردن میزان شباهت بین دو واژه، میزان شباهت بین جملات را محاسبه می کنیم. فاصله همینگ دو واژه تعداد حروف متناظر نامشابه است. بنابراین این معیار میزان تفاوت را نشان می دهد. برای محاسبه شباهت عدد حاصل را بر طول واژگان تقسیم و از عدد یک کسر می کیم. این کار را برای کل واژگان جمله انجام داده و درنهایت بر اساس اعداد بدست آمده میزان شباهت بین دو جمله محاسبه می شود. به طور مثال فرض کنید جمله Q_i شامل واژگان $w_1 w_2 w_3 w_4 w_5$ باشد، تعداد حالت های ممکن ترکیب m و n است. معیار شباهت بین دو واژه و یک جمله به صورت زیر محاسبه می شود:

$$Similarity_{words} = 1 - \frac{\text{Hamming Distance}(A, B)}{\text{Max}(|A|, |B|)} \quad (18)$$

$$Similarity_{sen} = \frac{1}{C(m, n)} \times \sum_{j=1}^{C(m, n)} Similarity_{words_j} \quad (19)$$

که در آن A و B واژگان مربوط به جملات، m تعداد واژگان جمله نخست و n تعداد واژگان جمله دوم است. بر اساس روابط تعریف شده برای هر ویژگی، ۲۲ ویژگی حاصل شد. جدول (۲) شماره رابطه استفاده شده برای استخراج هر ویژگی را نمایش می دهد.

(جدول-۲): فهرست ویژگی های استخراج شده از روابط (Table-2): List of extracted properties from Equations

توضیحات	شماره رابطه	شماره ویژگی	توضیحات	شماره رابطه	شماره ویژگی
پیشنهادی	۳	۱۲	به روزرسانی شده	۱	۱
به روزرسانی شده	۶	۱۳	به روزرسانی شده	۱	۲
پیشنهادی	۹	۱۴	به روزرسانی شده	۱	۳
پیشنهادی	۱۲	۱۵	به روزرسانی شده	۱	۴
پیشنهادی	۱۵	۱۶	به روزرسانی شده	۱	۵
پیشنهادی	۱۵	۱۷	به روزرسانی شده	۱	۶
پیشنهادی	۱۵	۱۸	پیشنهادی	۲	۷
پیشنهادی	۱۵	۱۹	پیشنهادی	۲	۸
پیشنهادی	۱۵	۲۰	پیشنهادی	۲	۹
پیشنهادی	۱۸	۲۱	پیشنهادی	۳	۱۰
پیشنهادی	۲۲	۲۲	پیشنهادی	۳	۱۱

$$P_{LCS} = \frac{1}{N} \times \sum_{i=1}^N \max\left(\sum_{j=1}^p \frac{LCS(Q_i, A_j)}{L_{A_j}}\right) \quad (11)$$

$$F_{LCS} = \frac{(1+\beta^2)R_{LCS}P_{LCS}}{R_{LCS} + \beta^2 P_{LCS}} \quad (12)$$

که در این رابطه $\beta = 1$. P تعداد جوابها و N تعداد سؤالات است.

- ویژگی هفتم:

برای محاسبه امتیاز هر مکالمه معادله (۱۵) پیشنهاد شد. در این ویژگی فرض شد که دو مجموعه S_i و S_j داریم که S_i از N جمله با K واژه و S_j با P جمله با T واژه هستند. بنابراین با به روزرسانی روابط قبلی، معادلات جدید به صورت زیر معرفی شدند:

$$R_{LCS} = \frac{1}{N} \times \sum_{S_i \in S_1} \max_{S_j \in S_2} (LSC(S_i, S_j)) \quad (13)$$

$$P_{LCS} = \frac{1}{P} \times \sum_{S_i \in S_1} \max_{S_j \in S_2} (LSC(S_i, S_j)) \quad (14)$$

$$F_{LCS} = \frac{(1+\beta^2)R_{LCS}P_{LCS}}{R_{LCS} + \beta^2 P_{LCS}} \quad (15)$$

که در رابطه (۱۵) مقدار $\beta = 1$ در نظر گرفته شد. همچنین مجموعه های S_1 و S_2 را به صورت $(Q_i, Q_{i+1}, A_i, A_{i+1})$ و $(Q_i, A_{i+1}, Q_{i+1}, A_i)$ در نظر گرفتیم. بنابراین به ازای هر مجموعه مقدار این ویژگی نیز محاسبه شد.

- ویژگی هشتم (محاسبه امتیاز واژگان):

به طور معمول در یک مکالمه درباره موضوع خاصی بین کاربر و سامانه صورت می پذیرد، این باور وجود دارد که طرفین مکالمه از واژگان معینی برای ادامه بحث یا تشریح دقیق جنبه های مختلف موضوع استفاده و یا از تکرار آنها استفاده می کنند. بنابراین در این ویژگی برای هر مکالمه گراف هم رخداد^۱ واژگان ترسیم می شود. ویژگی این گراف در این است که در این حالت واژگان هم رخداد در متن بدون استفاده از اندازه پنجه مشخص تعیین می شوند. با توجه به اینکه تعداد رخداد هر واژه می تواند به عنوان عامل تعیین درجه اهمیت واژگان مورد استفاده قرار گیرد. بنابراین در این گراف، تعداد تکرار هر واژه در مکالمه و این که هر واژه با چه واژه دیگری آمده است نمایش داده می شود. با توجه به گراف حاصل، از روی آن، فرکانس کلمه، درجه واژه در گراف و نسبت درجه به فرکانس واژه محاسبه می شود. نسبت درجه به فرکانس را به عنوان امتیاز نهایی هر واژه مشخص کرده مجموع امتیازات هر واژه به عنوان امتیاز هر مکالمه در نظر گرفته می شود.

^۱ Co-occurrence

۴-۳- مدل‌سازی

تحلیل رگرسیون، روشی آماری، برای بررسی و مدل‌سازی ارتباط بین متغیر وابسته و متغیر مستقل بوده با هدف پیش‌بینی متغیر وابسته از روی متغیر و یا متغیرهای مستقل است. هدف روش پیشنهادی، ارائه یک مدل آماری جهت ارزیابی سامانه‌های IQA است. بنابراین، با استفاده از رگرسیون به دنبال پیش‌بینی امتیاز در نظر گرفته شده توسط ارزیاب‌ها هستیم. در روش پیشنهادی، ویژگی‌های استخراج شده به عنوان متغیرهای وابسته و نظرات انسانی به عنوان متغیر مستقل در نظر گرفته شد. مدل‌های مختلف از رگرسیون اعم از خطی، چندگانه، غیرخطی و غیره مورد بررسی و آزمایش قرار گرفت تا بهترین مدلی که بتواند داده‌ها را توصیف کند، استخراج و بر اساس معیار ارزیابی انتخاب شود.

۴-۴- رگرسیون برای پیش‌بینی نظرات

رگرسیون ساده خطی، برای بررسی رابطه یک متغیر مستقل (پیش‌بین) و یک متغیر وابسته استفاده می‌شود؛ در حالی که اگر تعداد متغیرهای مستقل در این رابطه خطی بیش از یک عدد شود، مدل رگرسیون، خطی چندگانه نامیده می‌شود. معادله رگرسیون خطی ساده به شکل $Y = AX + B$ و $Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n$ است. که در آن X ها، ویژگی‌های استخراج شده و a ضرایب تأثیر هر یک از ویژگی‌ها باشد. مقدار حاصله برای X بیان گر مقادیر پیش‌بینی شده Y با ثابت‌ماندن مقادیر X است. بنابراین این ضرایب باید طوری محاسبه شدند که معیار کمینه مربعات خطرا را تأمین کند. از طرفی با مقایسه اندازه مقادیر ضرایب a با یکدیگر، میزان تأثیر و اولویت هر یک از عوامل مشخص می‌شود. علامت ضرایب هم بر تغییرات متغیر وابسته تأثیرگذارند. رگرسیون غیرخطی مدل‌های مختلفی دارد که از جمله آن می‌توان به مدل‌های درجه دو به بالا، چندجمله‌ای نمایی، توانی و غیره اشاره کرد که متناسب با مدل انتخاب شده معادلات آن‌ها متفاوت خواهد بود. تعدادی از معادلات رگرسیون غیرخطی در نظر گرفتیم و بر روی اطلاعات استخراج شده، مورد آزمایش قرار گرفته تا بهترین پاسخ انتخاب شود.

۴-۵- معیار ارزیابی

برای ارزیابی نتایج حاصل از مدل به دست آمده با داده‌های واقعی سه سنجه آماری ضریب تعیین^۱، مجدور میانگین مربعات خطرا (RMSE^۲) و درصد میانگین مطلق خطرا (MAPE^۳) مورد استفاده قرار گرفت. روابط مربوط به این

معیارها در زیر نشان داده شده است.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n}} \quad (20)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{e_i}{Y_i} \right| \times 100 \quad (21)$$

که در این روابط n تعداد پیش‌بینی‌ها و e_i خطای پیش‌بینی است که از تفاوت مقادیر پیش‌بینی شده و مقادیر واقعی به دست می‌آید و Y_i مقادیر واقعی است.

۵- نتایج آزمایشات

با توجه به وجود انواع مختلف رگرسیون، نیاز به استفاده از نرم‌افزارهای آماری برای استفاده از رگرسیون است؛ لذا ما برای شبیه‌سازی نتایج از نرم‌افزار مطلب استفاده کردیم. همچنین برای رگرسیون غیرخطی از جعبه‌ابزار apm^۳ استفاده شد. نتایج حاصل از مدل‌سازی نظرات با استفاده رگرسیون در ادامه آورده شده است.

۱-۳-۵- پایگاه داده

به دلیل نبود پایگاه داده استاندارد در زمینه ارزیابی سامانه‌های IQA، نیاز به ایجاد یک پایگاه داده مناسب از سؤالات ردوبدل شده بین سامانه و کاربر با برچسب‌گذاری مناسب بود. با توجه به این موضوع، علاوه بر سامانه تعاملی پایه طراحی شده، سه سامانه دیگر تعاملی موجود در نظر گرفته شد. در این راستا تعداد ۱۲۰ کاربر برای پنج موضوع مختلف با سامانه کار کردن و با توجه به موضوع مکالمات هر یک به صورت جداگانه ذخیره شد. از این مجموعه ششصدتایی، ۵۴۰ نمونه توسط فرد خبره به عنوان نمونه مناسب‌تر انتخاب شد. شرکت‌کنندگان در این دوره ارزیابی شامل دانشجویان دانشگاه و زبان‌آموزان یک مؤسسه بودند. با توجه به اینکه یک سامانه با زبان فارسی و سه سامانه دیگر با زبان انگلیسی عمل می‌کرد، برای یکپارچه‌سازی شرایط کار با این سامانه‌ها و راحتی کاربران، سامانه‌ای تحت وب طراحی شد.

این سامانه متن تبادل شده و امتیاز داده شده توسط کاربران به سامانه‌ها را به صورت خودکار در پایگاه داده‌ای ذخیره می‌کرد. شکل (۲) نمایی یکی از صفحات سامانه طراحی شده و جدول (۳)، یک نمونه از خروجی متن ذخیره شده (بدون فرمت) را از تعامل با سامانه پایه نشان می‌دهد.

^۱ Root Mean Square Error

^۲ Mean Absolute Percent Error

^۳ <http://apmonitor.com/che263/index.php/Main/MatlabDataRegression>

قابل قبول بودن ضرایب به دست آمده برای ویژگی‌ها است. ضرایب رگرسیون خطی به دست آمده با توجه به معادله رگرسیون خطی چندگانه در شکل (۳) نمایش داده شده است، که ضریب تعیین R^2 با توجه به ضرایب به دست آمده برای معادله برابر 0.76 ، MSE برابر 0.24 و $MAPE$ برابر 7% حاصل شد. همان‌طور که در شکل (۳) نشان داده شده است؛ مقادیر به دست آمده دارای علامت و مقدار متفاوتی برای هر متغیر هستند؛ بنابراین می‌توان در مورد نقش و تأثیر هر کدام از ویژگی‌ها در مدل سازی نظرات بحث کرد. ویژگی‌هایی که دارای مقدار کمی باشند، دارای تأثیر کمتر در نظرات انسانی و آن‌هایی که دارای مقادیر بزرگ‌تر هستند، حاکی از تأثیر بیشتر این ویژگی‌ها در خروجی هستند؛ همچنین با توجه به ضرایب به دست آمده، می‌توان تغییرات نظرات انسانی را نسبت به متغیرهای مستقل نیز به دست آورد.

۳-۵-۳- نتایج حاصل از پیاده‌سازی رگرسیون غیرخطی

در رگرسیون غیر خطی از معادلات متعددی برای مدل سازی و ضرایبی با مقادیر اولیه متفاوتی آزمایش شد تا بهترین مدل انتخاب شود. با توجه به آزمایش‌های متعدد انجام شده، بهترین مدلی که برای داده‌ها استخراج شد، مدل سری توانی رگرسیون غیرخطی بود. معادله آن برای ارزیابی نظرات به صورت معادله (۲۲) حاصل شد.

$$Y = 1.6X_1^{1.80} + 0.9X_2^{-0.25} + 0.85X_3^{-1.32} + \dots + 1.01X_{22}^{0.11} \quad (22)$$

ضرایب مربوط به این معادله در شکل (۴) و (۵) آورده شده است. قابل توجه است که ارزیابی صورت گرفته بر اساس سنجه‌های آماری در نظر گرفته، مقدار R^2 برابر 0.83 ، MSE برابر 0.13 و $MAPE$ برابر 6% حاصل شد. شکل (۶) مقایسه بین نتایج حاصل از ارزیابی‌ها را برای رگرسیون خطی و غیرخطی نمایش می‌دهد. نتایج حاکی از ارائه دقت بیشتر رگرسیون غیرخطی سری توانی نسبت به بقیه معادلات غیرخطی و برتری آن نسبت به رگرسیون خطی است. در روش پیشنهادی سعی شد ویژگی‌هایی پیشنهاد شود تا بتوانند توصیف بهتری از ارزیابی انسانی بر اساس متن خروجی تولید شده از سامانه‌های IQA داشته باشد؛ به طوری که خروجی مدل پیشنهادی بر اساس ویژگی‌های تعریف شده بیشترین شباهت را به نظرات انسان داشته باشد. روش پیشنهادی در هیچ یک از کارهای قبلی مشاهده نشد با این حال می‌توان به



(شکل-۲): نمایی از یک صفحه سامانه تحت وب طراحی شده
(Figure-2): A screenshot of a web page design system

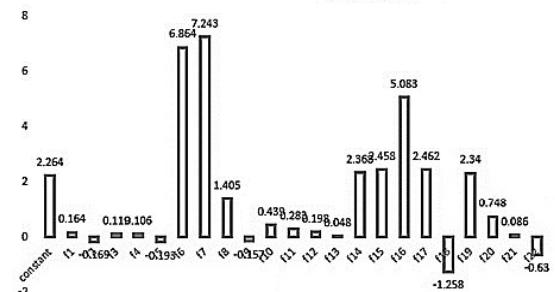
(جدول-۳): نمونه تولید شده از یک مکالمه در سامانه IQA
(Table-3): Sample of generated conversation in IQA system

سامانه	کاربر
دانشجو مربوط به چه مقطعی است؟	اگر معدل دانشجو کمتر از ۱۴ شود چه باید کرد؟
منظور شما معدل کل هست؟	دانشجو کارشناسی ارشد
در صورت کمتر شدن معدل از ۱۴ در هر نیمسال دانشجو مشروط تلقی می‌شود.	خیر، معدل یک ترم
دانشجوی کارشناسی چطور؟	دانشجوی کارشناسی چطور؟
دانشجو کارشناسی یا ارشد؟	چند ترم متواالی مشروط شدن دانشجو اخراج می‌شود؟
در صورتی که دانشجو کارشناسی سه ترم متواالی یا چهار ترم غیرمتواالی مشروط شود در آستانه اخراج قرار می‌گیرد.	دانشجو کارشناسی

۳-۵-۲- نتایج حاصل از پیاده‌سازی رگرسیون خطی

در روش رگرسیون خطی، ویژگی‌های استخراج شده به عنوان متغیر مستقل و نظرات انسانی به عنوان متغیر وابسته در نظر گرفته شدند.

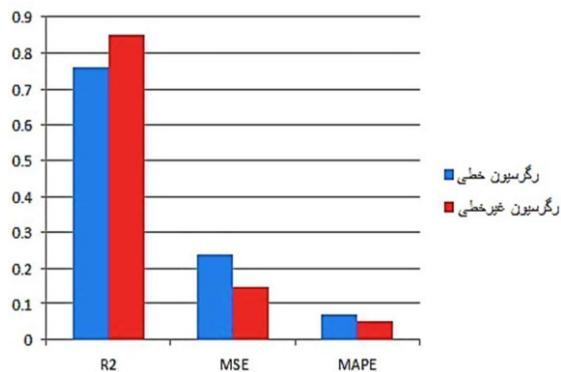
مقادیر ویژگیها



(شکل-۳): ضرایب معادله رگرسیون خطی
(Figure-3): Linear regression equation coefficients

مقدار پارامتر sig برای تمامی ضرایب به دست آمده در رگرسیون خطی چندگانه کمتر از 0.05 بود. که این به معنای

جدید پیشنهاد شد که بتوانند به خوبی در مدل سازی نظرات کارایی داشته باشند. همان طور که در قسمت استخراج ویژگی بیان شد، ویژگی های پیشنهادی، بر روی زوج های پرسش-پاسخ (Q_i-A_i), مجموعه جوابها (A_i-A_{i+1}) و مجموعه پاسخ ها (Q_i-Q_{i+1}) اعمال شد تا تعاملی که بین جملات را بدل شده در یک مکالمه وجود داشته و نقش به سازی در امتیازدهی توسط کاربر دارد، در نظر گرفته شود؛ سپس از روی متن های موجود در پایگاه داده، ویژگی ها استخراج و براساس آن ماتریس ویژگی تشکیل شد. در ادامه برای پیش بینی امتیاز داده شده توسط کاربر (برای مدل سازی نظرات) از رگرسیون خطی و غیر خطی بهره گرفته شد.



(شکل-۶): مقایسه نتایج حاصل از مدل سازی با رگرسیون خطی و غیر خطی

(Figure-6): Comparison of the results of modeling with linear and nonlinear regression

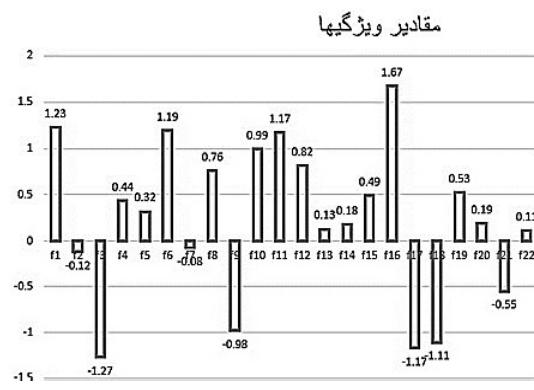
نتایج ارائه شده بر اساس معیارهای ارزیابی R^2 و MSE رگرسیون غیرخطی سری توانی دارای دقت بالاتری بود که نشان دهنده پایداری مدل پیشنهادی است. برای کارهای آینده، پیشنهاد می شود با توجه به ضرایب بدست آمده و مقادیر آنها، می توان تأثیر هر یک از ویژگی ها را بر روی خروجی مشخص کرد و با توجه به همبستگی بین ویژگی ها، به کاهش ویژگی ها پرداخت تا پیچیدگی معادلات بدست آمده به مراتب کمتر و یا برای رسیدن به معادله رگرسیونی از الگوریتم های هوشمند مانند برنامه نویسی ژن و یا ژنتیک استفاده شود.

5- References

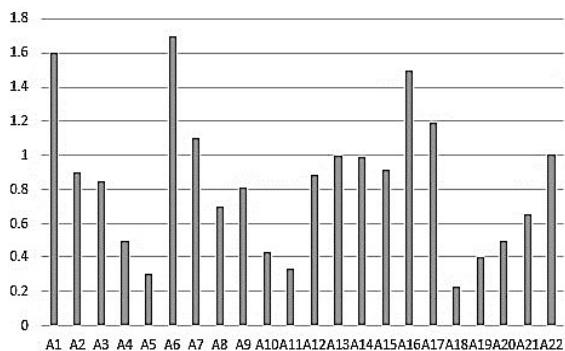
5- مراجع

- [۱] شهرآیینی، سلیمه، زاهدی، مرتضی، "سیستم پاسخگوی تعاملی با استفاده از تکنیک های هوش مصنوعی"، دانشگاه صنعتی شاهرود، دانشکده کامپیوتر و فناوری اطلاعات، پایان نامه ارشد، ۱۳۹۴.

مقاله [۱۶] که در زمینه ارزیابی سامانه های IQA ارائه شده است، اشاره کرد. در این مقاله به بررسی سؤالات تعاملی کنفرانس TREC و یک سامانه QA مربوط به پژوهشی پرداخته و درنهایت بیشتر تمرکز خود را در معرفی عواملی که می توانند در ارزیابی یک سامانه پرسش و پاسخ پژوهشی تأثیرگذار باشند، پرداخته است. ارزیابی توسط دو گروه از دانشجویان صورت پذیرفته است. نتایج نشان داد که عواملی مانند سن، جنسیت، تجربه در استفاده از رایانه، نگرش نسبت به رایانه و چندین ویژگی دیگر جزو عوامل تعیین کننده در موقوفیت یک سامانه QA است. درصورتی که روش پیشنهادی به معرفی ویژگی های آماری که بتواند جایگزین ارزیابی کیفی سامانه های پرسش و پاسخ تعاملی شود، پرداخته ایم.



(شکل-۴): ضرایب توان معادله رگرسیون غیر خطی
(Figure-4): Non Linear regression equation coefficients



(شکل-۵): ضرایب تأثیر هر ویژگی مربوط به رگرسیون غیر خطی

(Figure-5): The effect coefficients of each characteristic on nonlinear regression

۴- نتیجه گیری

در این مقاله، روشی خودکار برای مدل سازی نظرات ارزیابی انسانی بر اساس متن خروجی یک سامانه IQA ارائه شده. که مشابه با روش ارائه شده، در هیچ یک از کارهای قبلی مشاهده نشد. در روش پیشنهادی ابتدا مجموعه ای از ویژگی های آماری





- [12] M. Mansoori, and H. Hassanpour, "Boosting passage retrieval through reuse in question answering," *International Journal of Engineering* 25, no. 3, pp.187-196, 2012.
- [13] برشبان، یاسمون، یوسفی نسب، حامد، میرروشنل، سید ابوالقاسم "ارایه یک پیکره پرسش و پاسخ مذهبی در زبان فارسی"، فصلنامه پردازش علائم و داده‌ها، دوره ۱۵، شماره ۱، صفحات ۸۷-۱۰۲، ۱۳۹۷.
- [14] Y. Boreshban, H. Yousefinasab, S. A. Mirroshandel, "Providing a Religious Corpus of Question Answering System in Persian", *Journal of Signal and Data Processing (JSDP)*, Vol 15, no 1, pp.87-102, 2018.
- [15] N. Wacholder, S. G. Small, B. Bai, D. Kelly, R. trittman, S. Ryan, R. Salkin, "Designing a Realistic Evaluation of an End-to-end Interactive Question Answering System." In *LREC*. 2004.
- [16] Quarterni, Silvia and S. Manandhar, "Designing an interactive open-domain question answering system," *Natural Language Engineering* 15, no. 1, pp. 73-95,2009.
- [17] S. Ying, P. B. Kantor and E. L. Morse, "Using cross-evaluation to evaluate interactive QA systems." *Journal of the Association for Information Science and Technology* 62, no. 9, pp. 1653-1665, 2011.



محمد‌مهدی حسینی در حال حاضر هیأت علمی دانشگاه آزاد اسلامی واحد شهرroud هستند. ایشان دوره دکترای خود را در دانشگاه صنعتی شهرroud گذرانده و بر روی سامانه‌های پرسش و پاسخ تعاملی پژوهش کرده است. زمینه‌ها و علایق پژوهشی ایشان پردازش متن، سامانه‌های پرسش و پاسخ تعاملی و پردازش تصویر است. نشانی رایانامه ایشان عبارت است از:

Hosseini_mm@shahroodut.ac.ir



مرتضی زاهدی در حال حاضر عضو هیأت علمی دانشکده مهندسی کامپیوتر دانشگاه صنعتی شهرroud است. وی پژوهش‌هایی را در زمینه تعامل انسان و رایانه، شناسایی الگو،

پردازش تصویر و ویدئو، و بینایی ماشین در دست اجرا دارد که در آن‌ها به طور معمول از اطلاعات و دانش آماری استفاده می‌شود. وی دارای دکترای تخصصی رایانه از دانشگاه RWTH-Aachen آلمان است. تألیف کتب و مقالات علمی و همچنین سرپرستی پژوهش‌های دانشگاهی و صنعتی در ایران و کشورهای اروپایی در کارنامک کاری

- [1] S. Shahriini, S. Zahedi, "Interactive Question answering System Using Artificial Intelligence Techniques", Senior Thesis, Shahrood University of Technology, Faculty of Computer and Information Technology, 2015.

- [۲] حسینی، محمد Mehdi، زاهدی، مرتضی، "بهبود پاسخ ارائه شده در سیستم‌های پرسش و پاسخ تعاملی با استفاده از شبکه عصبی"، هشتمین کنفرانس بین‌المللی فناوری اطلاعات و دانش، صفحات ۸۴-۹۱، ۱۳۹۵.

- [2] M.M. Hosseini, M. Zahedi, "Improvement of the response provided in interactive question answering systems using neural network", *Eighth International Conference on Information and Knowledge Technology*, pp. 84-91, 2016.
- [3] Bouziane, Abdelghani, Bouchiha, Doumi, and Malki, "Question Answering Systems: Survey and Trends", *Procedia Computer Science*, pp. 366-375, 2015.
- [4] Bao, Junwei, Nan Duan, Ming Zhou, and Tiejun Zhao, "Knowledge-based question answering as machine translation," *Cell* 2, no. 6, 2014.
- [5] C. Guinaudeau, M. Strube, "Graph-based Local Coherence Modeling", *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pp. 93–103, 2013.
- [6] Hartawan, Andrei, and Derwin Suhartono, "Using Vector Space Model in Question Answering System", *Procedia Computer Science*, pp. 305-311, 2015.
- [7] Höffner, Konrad, S. Walter, E. Marx, R. Usbeck, J. Lehmann, and A. Ngomo, "Survey on challenges of question answering in the semantic web", *Semantic Web* 8, no. 6, pp.895-920, 2017.
- [8] Kelly, Diane, P. B. Kantor, E. L. Morse, J. Scholtz, and Y. Sun, "Questionnaires for eliciting evaluation data from users of interactive question answering systems," *Natural Language Engineering* 15, no. 1, , pp. 119-141, 2009.
- [9] L. C. Yew, "Rouge: A package for automatic evaluation of summaries," In *Text summarization branches out: Proceedings of the ACL-04 workshop*, vol. 8, 2004.
- [10] L. Vanessa, V. Uren, M. Sabou and E. Motta. "Is question answering fit for the semantic web? A survey," *Semantic Web* 2, no. 2, pp.125-155, 2011.
- [11] M. Amit, and S. K. Jain, "A survey on question answering systems with classification," *Journal of King Saud University-Computer and Information Sciences* 28, no. 3, pp. 345-361,2016.

ایشان دیده می‌شود. زمینه‌ها و علایق پژوهشی ایشان تعامل انسان و رایانه، پردازش متن، شناسایی الگو، پردازش تصویر و ویدئو، بینایی ماشین است.

نشانی رایانامه ایشان عبارت است از:
Zahedi@ganjineh.co.ir

حمید حسن‌پور مدرک دکترای خود را از دانشگاه صنعتی کوئینزلند استرالیا در گرایش پردازش سیگنال در سال ۱۳۸۳ دریافت کرده‌اند. ایشان مدرک کارشناسی

ارشد خود را در گرایش هوش مادین در سال ۱۳۷۵ از دانشگاه صنعتی امیرکبیر، و مدرک کارشناسی خود را در سال ۱۳۷۲ در گرایش سختافزار از دانشگاه علم و صنعت ایران اخذ کرده‌اند. دکتر حسن‌پور در طی سال‌های ۱۳۸۴ تا ۱۳۸۶ به عنوان عضو هیئت علمی در دانشکده مهندسی برق و رایانه دانشگاه صنعتی بابل فعالیت داشتند؛ سپس به دانشکده کامپیوتر و فناوری اطلاعات دانشگاه صنعتی شاهroud منتقل شدند. زمینه‌ها و علایق پژوهشی ایشان پردازش سیگنال، پردازش تصویر، داده‌کاوی، معماری کامپیوتر و پردازش متن است.

نشانی رایانامه ایشان عبارت است از:
h.hassanpour@shahroodut.ac.ir



فصل پنجم

سال ۱۳۹۸ شماره ۳ پیاپی ۴۱

۸۴

