Downloaded from jipm.irandoc.ac.ir at 13:46 IRDT on Sunday June 13th 2021

Determining Reliability in Interactive Question Answering Systems by Regression

Mohammad Mehdi Hosseini*
PhD in Computer Engineering; Assistant Professor; Department of Computer Engineering; Shahrood Branch; Islamic Azad University; Shahrood, Iran Email: hosseini_mm@yahoo.com

Alireza Jalali
PhD in Computer Engineering; Assistant Professor; Department of Computer Engineering; Shahrood Branch; Islamic Azad University; Shahrood, Iran Email: dr.jalali16@gmail.com

Iranian Journal of Information Processing and Management

Received: 15, Jul. 2019 **Accepted: 02, Nov. 2020**

Iranian Research Institute for Information Science and Technology (IranDoc)
ISSN 2251-8223
eISSN 2251-8231
Indexed by SCOPUS, ISC, & LISTA
Vol. 36 | No. 3 | pp. 817-834
Spring 2021

Abstract: Nowadays, the increasing amount of information available online has led scientists to develop technologies to use and deal with it, of which one is Interactive Question Answering (IQA) systems. One of the most important criteria in the assessment of these systems is reliability. Research results in this field have shown that evaluation of reliability in IQA plays an important role to determine the effectiveness of these systems. During the process of reliability evaluation of an IQA system, several features are considered. Few attempts have been made to measure the reliability and accurate determination of reliability criteria in interactive question answering systems.

In this paper, a novel method is proposed to determine the percentage of reliability by using a set of criteria affecting including qualitative and quantitative criteria. In the proposed method, first, a list of qualitative and quantitative characteristics is collected and according to the criteria set, eight quantitative characteristics are selected. Then, based on the prepared questionnaire, the coefficients of the impact of each question are calculated. According to the proposed relationship and by combining these characteristics the percentage of reliability is measured. The evaluation of the performance of the selected criteria, on four interactive question answering indicates that by applying this set of criteria the reliability of an interactive question answering system can be used to automate the calculation of this feature in the evaluation process.

Keywords: Reliability, Evaluation Criteria, Qualitative Software Features, Interactive Question Answering Systems

* Corresponding Author

Downloaded from jipm.irandoc.ac.ir at 13:46 IRDT on Sunday June 13th 2021

تعیین میزان قابلیت اطمینان در سیستم‌های پرسش و پاسخ تعاملی به کمک رگرسیون خطی

محمدمهدی حسینی
دکتری کامپیوتر؛ استادیار؛ واحد شاهرود؛ دانشگاه آزاد اسلامی؛ شاهرود، ایران؛
پدیدآور رابط hosseini_mm@yahoo.com

علیرضا جلالی
دکتری کامپیوتر؛ استادیار؛ واحد شاهرود؛ دانشگاه آزاد اسلامی؛ شاهرود، ایران؛
dr.jalali16@gmail.com

پژوهش‌نامهٔ پردازش و مدیریت اطلاعات

دریافت: ۱۳۹۸/۰٤/۲٤ **پذیرش:** ۱۳۹۹/۰۸/۱۲ مقاله برای اصلاح به مدت ۲۳ روز نزد پدیدآوران بوده است.

نشریهٔ علمی | رتبهٔ بین‌المللی
پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)
شاپا (چاپی) ۸۲۲۳-۲۲۵۱
شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در LISTA، ISC، و SCOPUS
jipm.irandoc.ac.ir
دورهٔ ۳٦ | شمارهٔ ۳ | ص‌ص ۸۱۷-۸۳٤
بهار ۱٤۰۰

چکیده: ارزیابی سیستم‌های پرسش و پاسخ تعاملی به‌منظور تعیین و ارتقای کارایی آن‌ها از اهمیت زیادی برخوردار است. با وجود این، هنوز روش استاندارد و مخصوصی برای ارزیابی این سیستم‌ها ارائه نشده است. در ارزیابی یک سیستم پرسش و پاسخ تعاملی معیارهای زیادی تأثیرگذار است. یکی از این معیارها قابلیت اطمینان است. تلاش‌های اندکی در راستای اندازه‌گیری اتوماتیک و تعیین دقیق معیارهای قابلیت اطمینان در سیستم پرسش و پاسخ تعاملی با به‌کارگیری مدل‌های گوناگون صورت گرفته است. در این مقاله، یک روش جدید جهت اندازه‌گیری قابلیت اطمینان با استفاده از ایجاد مجموعه معیارهای تأثیرگذار بر قابلیت اطمینان اعم از معیارهای کیفی و کمی به کمک رگرسیون ارائه شده است. در این روش، ابتدا لیستی از ویژگی‌های کیفی و کمی تأثیرگذار جمع‌آوری و سپس، با توجه به ضوابط تعیین‌شده هشت ویژگی کمی انتخاب شدند. سپس، بر اساس پرسشنامهٔ تهیه‌شده، ضرایب تأثیر سؤالات تعیین و با ترکیب این ویژگی‌ها طبق رابطهٔ پیشنهادی قابلیت اطمینان اندازه‌گیری شد. ارزیابی عملکرد مجموعه معیارهای انتخابی بر روی چهار سیستم پرسش و پاسخ تعاملی نشان داد که با به‌کارگیری این مجموعه معیارها، امکان اندازه‌گیری قابلیت اطمینان برای یک سیستم پرسش و پاسخ فراهم شده است.

کلیدواژه‌ها: قابلیت اطمینان، معیار ارزیابی، خصوصیات کیفی نرم‌افزار، سیستم پرسش و پاسخ تعاملی

Downloaded from jipm.irandoc.ac.ir at 13:46 IRDT on Sunday June 13th 2021

۱. مقدمه

بسیاری از کاربران دنیای مجازی تلاش می‌کنند که با باز کردن صفحات مختلف و بررسی محتوای آن‌ها پاسخ مطلوب خود را پیدا کنند. این کار معمولاً تصادفی یا با استفاده از دانشی که کاربر قبلاً از مشاهدهٔ صفحات مشابه کسب کرده صورت می‌پذیرد (Iyyer, Yih and Chang 2017). با توجه به حجم زیاد صفحات بازگردانده‌شده توسط موتورهای جست‌وجو در اینترنت، بررسی تمامی آن‌ها توسط کاربران امکان‌پذیر نیست. بنابراین، این احتمال وجود دارد که با صرف زمان زیاد، کاربر به پاسخ مطلوب و قابل قبول خود دست پیدا نکند. از این رو، برای تسهیل جست‌وجو، سیستم‌های بازیابی اطلاعات کلاسیک در قالب موتورهای جست‌وجو طراحی شدند (Baudiš 2015). با وجود این، مشکلاتی در بازیابی اطلاعات همچنان باقی است. یکی از مشکلات این نوع بازیابی اطلاعات عدم نمایش پاسخی مستقیم، قابل فهم و اجمالی به کاربر است (Webber and Webb 2010). برای رفع این مشکل، سیستم‌های پرسش و پاسخ[1] طراحی شدند. این سیستم‌ها سؤال کاربر را به شکل زبان طبیعی دریافت کرده و پس از تجزیه و تحلیل، پاسخی مناسب و دقیق در قالب یک متن کوتاه یا یک جمله یا یک عبارت و یا یک کلمه ارائه می‌کنند (Dwivedi and Singh 2013). پرسش در این سیستم‌ها به‌صورت مستقل از پرسش‌های پیشین مطرح می‌شود تا سیستم با استفاده از تکنیک‌های موجود پاسخ احتمالی را بدهد. محدودیت‌های بسیاری در این سیستم‌ها وجود دارد که سبب می‌شود در بسیاری از موارد کاربران در یافتن پاسخ مطلوب با مشکل مواجه باشند. از جملهٔ این مشکلات می‌توان به وابستگی بعضی از این سیستم‌ها به پرسش مطرح‌شده، نبودِ حافظه، عدم امکان سؤالات پیچیده و دیگر موارد اشاره کرد (همان). بزرگ‌ترین مشکل موجود در سیستم‌های پرسش و پاسخ، عدم امکان تعامل بین کاربر و سیستم بوده و کاربران قادر به ارائهٔ توضیح اضافی دربارهٔ نیاز اطلاعاتی خود به سیستم نیستند. از طرفی، سیستم نیز نمی‌تواند با طرح پرسش از کاربر جهت رفع ابهامات احتمالی در سؤالات مطرح‌شده اقدام نماید. بنابراین، برای بهتر شدن عملکرد سیستم‌های پرسش و پاسخ می‌بایست سطح تعامل به این سیستم‌ها اضافه گردد (Iyyer, Yih and Chang 2017). سیستم‌های پرسش و پاسخ تعاملی[2] این امکان را به کاربر می‌دهند که در ابتدا سؤالات پرسیده شود و سپس، سیستم با بهره‌گرفتن از بازخورد مربوط به سؤالات طوری عمل می‌نماید که کاربر پاسخ

1. question answering (QA) 2. nteractive question answering (IQA)

Downloaded from jipm.irandoc.ac.ir at 13:46 IRDT on Sunday June 13th 2021

مطلوب را بیابد. بنابراین، می‌توان این طور تصور نمود که در سیستم IQA یک فرایند تکراری (مطرح‌شدن سؤال و پاسخ متوالی) اتفاق می‌افتد (Webber and Webb 2010).

ارزیابی، نقش مهمی در پیشبرد اهداف مد نظر محققان این حوزه ایفا می‌نماید. به‌رغم این موضوع، در زمینهٔ ارزیابی این سیستم‌ها روش خاص و استانداردی تقریباً وجود ندارد که به ارزیابی کلی سیستم‌های IQA پرداخته باشد (حسینی و زاهدی ۱۳۹۷). بیشتر روش‌های به‌کار رفته در ارزیابی سیستم‌های IQA، روش‌هایی است که در سیستم‌های QA و دیالوگ، مورد استفاده قرار گرفته است و از کارایی مناسبی برخوردار نیستند (Dwivedi and Singh 2013). این است که برای ارزیابی مناسب‌تر این سیستم‌ها از انسان بهره گرفته می‌شود. خود این مسئله فرایند ارزیابی را دچار چالش‌های متعددی نموده است. بنابراین، ارائهٔ یک روش خودکار که بتواند جایگزین انسان شود، می‌تواند کمک زیادی به ارتقای روش‌های ارزیابی سیستم‌های IQA نماید. در فرایند ایجاد یک مدل خودکار تعیین پارامترهای تأثیرگذار در ارزیابی دارای اهمیت است. یکی از پارامترهایی که در ارزیابی نقش مهمی ایفا می‌کند، قابلیت اطمینان است. از این رو، در این مقاله یک روش خودکار جهت مدل کردن نحوهٔ اندازه‌گیری قابلیت اطمینان برای سیستم‌های IQA ارائه شده است.

ساختار مقالهٔ پیشنهادی بدین‌صورت است که در بخش اول و دوم به مرور سیستم‌های QA و IQA پرداخته شده است. مرور بر روش‌های ارزیابی این سیستم‌ها و روش‌های موجود برای اندازه‌گیری قابلیت اطمینان در بخش سوم صورت گرفته است. در بخش چهارم، سیستم پایهٔ تولیدشده در آزمایشگاه که برای صحت روش پیشنهادی در نظر گرفته شده، تشریح خواهد شد. در نهایت، در بخش آخر، روش پیشنهادی و نتایج بیان شده است.

۲. سیستم پرسش و پاسخ

سیستم‌های پرسش و پاسخ شکل پیچیده‌تری از سیستم‌های بازیابی اطلاعات هستند که به‌جای ارائهٔ کل سند، تنها بخش‌های خاصی از اطلاعات سند را به‌عنوان پاسخ بازمی‌گردانند (Webber and Webb 2010). چارچوب کلی یک سیستم QA شامل سه مرحله است:

پردازش سؤال: تجزیهٔ نحوی، کلاس‌بندی سؤال و فرمول‌بندی مجدد پرس‌وجو؛

تحلیل متن: استخراج متون مناسب و تشخیص پاسخ بر اساس بازیابی اطلاعات از پایگاه دادهٔ موجود؛

Downloaded from jipm.irandoc.ac.ir at 13:46 IRDT on Sunday June 13th 2021

تحلیل پاسخ: استخراج جواب‌های مناسب و رتبه‌بندی بهترین پاسخ و ارائهٔ آن به کاربر.

روش‌های متعددی برای هر یک از این مراحل در سیستم‌های QA مطرح شده است که هر یک از آن‌ها می‌تواند مورد توجه و بررسی قرار گیرد (Dwivedi and Singh 2013).

در سیستم پرسش و پاسخ تعاملی (IQA) حداقل یک تبادل بین کاربر و سیستم پشتیبانی کرده و به کاربر اجازه می‌دهد تا کنترل نسبی روی محتوای نمایش داده‌شده و اقدامات صورت‌گرفته داشته باشد. با این حال، محققان در گزارشات مختلف خود استدلال کرده‌اند که اگر سیستم QA می‌خواهد تعاملی باشد، باید بیش از یک تبادل داشته باشد (حسینی و زاهدی ۱۳۹۷). به همین دلیل، می‌توانیم سیستم‌های پرسش و پاسخ تعاملی را یک قدم جلوتر از سیستم‌های پرسش و پاسخ قلمداد کنیم که به کاربران این امکان را می‌دهد که برای یافتن اطلاعات مورد نیازشان به شکلی دقیق‌تر عمل نمایند. سیستم‌های IQA دقیق‌تر از سیستم‌های QA معمولی هستند (Dwivedi and Singh 2013). علت این امر آن است که سیستم‌های IQA به سازه‌های مبهم زبان، مانند عبارت حرف اضافه یا ساختارهای نحوی پیچیده می‌پردازند. افزون بر این، هنگامی که سیستم IQA با چنین سازه‌های مبهمی روبه‌رو می‌شود، می‌تواند مکالمه‌ای را آغاز نماید تا درخواست کاربر را شفاف سازد. با این حال، چالش‌های متعددی در این راه وجود دارد که محققان در حال توسعه و مطالعهٔ این سیستم‌ها هستند. یکی از مسائل چالش‌برانگیز در رابطه با این سیستم‌ها ارزیابی آن‌هاست (همان).

۳. پیشینهٔ پژوهش

از آنجا که پاسخ صحیح در یک سیستم IQA و ارزیابی جواب‌های ارائه‌شده با روش‌های متفاوتی بیان می‌شود، روش‌های گوناگونی برای ارزیابی هر سیستم ارائه شده و یک چارچوب کلی و واحدی برای انجام فرایند ارزیابی ارائه نشده است. بنابراین، اکثر سیستم‌های موجود برای ارزیابی، از ارزیابی انسان بهره می‌گیرند. بیشتر روش‌های مورد استفاده برای ارزیابی سیستم‌های IQA روش‌های ارزیابی است که برای سیستم‌های دیالوگی و QA به‌کار رفته است. بنابراین، ارزیابی سیستم‌های QA بسته به ارزیابی سؤالات پیچیده یا ساده متفاوت است (Webber and Webb 2010). یکی از این روش‌ها استفاده از مجموعهٔ سؤال‌ها و پاسخ‌ها به‌نام «مجموعه استاندارد طلایی» است که توانایی سیستم با استفاده از میزان منطبق‌بودن آن با این مجموعه استاندارد طلایی سنجیده می‌شود (Breck et al. 2000).

Downloaded from jipm.irandoc.ac.ir at 13:46 IRDT on Sunday June 13th 2021

البته، این روش برای سؤالات پیچیده و مبهم قابل استفاده نیست. بیشتر ارزیابی سیستم‌های QA توسط کنفرانس سالانهٔ بازیابی متون (TREC)[1] انجام شده است. این ارزیابی‌ها بیشتر به‌جای این‌که مبتنی بر یک سیستم باشد، بر اساس کاربر صورت پذیرفته است (ibid). در حقیقت، بیشتر کارهای صورت‌گرفتهٔ ارزیابی در زمینهٔ استخراج پاسخ و نحوهٔ تعامل و استفاده از آن انجام شده است. بنابراین، روشی که هم به ارزیابی سیستم و هم به ارزیابی کاربر پرداخته باشد، وجود نداشته و در صورت وجود، همهٔ ویژگی‌های ارزیابی را در نظر نگرفته است (حسینی و زاهدی ۱۳۹۷). یکی از معیارهای مهمی که معمولاً انسان‌ها در ارزیابی یک سیستم پرسش و پاسخ تعاملی در نظر می‌گیرند، قابلیت اطمینان است (Salata et al. 2015). روش‌های محدودی در این زمینه وجود دارد که تنها به‌صورت کیفی به ارزیابی و محاسبهٔ این ویژگی پرداخته‌اند Breck et al. (2000). سیستمی به نام Eavar را برای ارزیابی خودکار توسعه دادند. سیستم پیشنهادی آن‌ها جواب‌های به‌دست‌آمده را با جواب‌های ارزیابی ارائه‌شده توسط TREC مقایسه می‌نمود. این سیستم به‌صورت خودکار جواب‌های غلط را که از سایرین متفاوت بودند، شناسایی می‌کرد. این روش ارزیابی، مفهوم‌محور نبود و نمی‌توانست جواب‌های مبدأ را در لیست TREC شناسایی کند (ibid). «بورستاین و چودوروف» یک سیستم درجه‌بندی اتوماتیک برای ارزیابی سؤالات متن باز ارائه کردند. آن‌ها موفقیت سیستم را در مقایسه با قضاوت انسانی بررسی کردند و نتایج حاصل، حاکی از موفقیت ۹۲ درصدی روش آن‌ها بود. روش ارائه‌شده به نحوهٔ آماده‌سازی هر سؤال در مجموعهٔ نمونهٔ آزمایشی با درجات مختلف کیفی وابسته بود. آن‌ها در کارشان از الگوریتم‌های یادگیری ماشین برای درجه‌بندی خودکار مجموعهٔ ویژگی‌ها استفاده کردند. روش‌های ارزیابی خودکار زمانی که جواب‌ها شناخته‌شده یا ثابت باشند، مفیدند و قادر نیستند جواب‌ها در ارزیابی را شناسایی کنند (Burstein & Chodorow 2009). «اسپیترز» و همکارانش، ویژگی‌های مربوط به پرسش و کامل‌کردن پرسشنامه را مورد بررسی قرار داده و کیفیت سیستم را بر اساس تکمیل‌شدن پرسشنامه درجه‌بندی نمودند. سؤالاتی که ایشان در نظر گرفته بودند، شامل درخواست ارزیابی سیستم بود که چنین روشی خود قابل بحث و بررسی است (Spitters et al. 2009). «هاراباگوئی» و همکارانش در حیطهٔ عبارات ردوبدل شده در سیستم IQA دیالوگ را مورد ارزیابی قرار دادند. آن‌ها کارایی مؤثر سیستم را وابسته به تعداد سؤالات که کاربر برای یافتن اطلاعاتش مطرح

1. Text Retrieval Evaluation Conference

Downloaded from jipm.irandoc.ac.ir at 13:46 IRDT on Sunday June 13th 2021

می‌نماید، تعریف کرد (Harabagiu et al. 2005). «دریکی» و همکارانش چارچوبی را برای سیستم پرسش و پاسخ معرفی نمودند که می‌توانست به سؤالات دانش‌آموزان پاسخ دهد. در چارچوب پیشنهادی تجزیه و تحلیل سؤالات و بخش‌های مهم بر اساس روش‌های مبتنی بر قانون و آماری صورت می‌گرفت؛ در حالی که یک سؤال با استفاده از یک وابستگی مورد تجزیه و تحلیل قرار می‌گرفت، ولی معیار قابلیت اطمینان و پوشانندگی مناسبی را ارائه می‌نمود (Derici et al. 2018). «موته» و همکارانش سیستمی را ارائه دادند که به‌طور پویا قابلیت اطمینان اسناد چندرسانه‌ای را با اعمال مجموعه‌ای از اسناد و بر اساس ارتباط با حداقل یک مجموعه اسناد از منابع چندرسانه‌ای منتخب مرتبط محاسبه می‌نمود. سپس، با پیدا کردن ارتباط بین اسناد چندرسانه‌ای از منابع دیگر و با محاسبهٔ امتیازی مشروط، میزان قابلیت اطمینان سیستم را گزارش می‌نمود (Motte et al. 2019). «بهاتیا و مان» یک رویکرد جدید برای اندازه‌گیری قابلیت اطمینان به‌عنوان یک مشخصهٔ کیفیت نرم‌افزاری ارائه نمودند. رویکرد مطرح‌شده بر اساس معیارهای (CBO[1]، NOC[2]، DIT[3]) ارائه گردیده بود (Bhatia and Mann 2008). «پالیوال، شریواستاوا و تیواری» روشی برای محاسبهٔ قابلیت اطمینان از برنامه‌های شئ‌گرا ارائه کردند. آن‌ها در روش خود استفادهٔ مجدد را به‌عنوان یکی از خصوصیات کیفی معرفی کرده و برخی از ویژگی‌های مؤثر بر استفادهٔ مجدد، مانند ویژگی‌های پیچیدگی، پیچیدگی رابط، اندازهٔ کلاس و وابستگی را بررسی نمودند (Paliwal, Shrivastava & Tiwari 2014). «کارل، بانسیا» قابلیت استفادهٔ مجدد را با چهار ویژگی اندازهٔ طراحی، اتصال، انسجام و پیام‌دهی بررسی کرده‌اند (Carl and Bansiya 2013). «سورش، پاتی و راث» در مقالهٔ خود به بررسی قابلیت اطمینان خودکار در ارزشیابی و یا مؤثر بودن آن در توسعهٔ نرم‌افزار پرداختند. آن‌ها نشان دادند که شناسایی عوامل مؤثر در قابلیت اطمینان می‌تواند در هزینهٔ توسعهٔ نرم‌افزار صرفه‌جویی کند (Suresh, Pati & Rath 2012). بیشتر مطالعات صورت‌گرفته در زمینهٔ ارزیابی سیستم‌های تعاملی به بررسی و ارزیابی کلی این سیستم‌ها پرداخته و در بیشتر موارد از انسان برای ارزیابی استفاده نموده‌اند. یکی از کارهای مهمی که می‌تواند در این راستا صورت پذیرد، خودکار کردن ویژگی‌هایی است که انسان در فرایند ارزیابی مورد توجه قرار می‌دهد. قابلیت اطمینان، یکی از معیارهایی است که همواره در ارزیابی‌های سیستم‌های تعاملی مورد توجه قرار گرفته است. این

1. Coupling Between Object Classes
2. Number Of Children
3. Depth Of Inheritance Tree

Downloaded from jipm.irandoc.ac.ir at 13:46 IRDT on Sunday June 13th 2021

است که در این مقاله روشی جهت خودکار کردن این معیار ارائه شده تا بتوان گامی در جهت ارائهٔ یک روش کاملاً خودکار برای ارزیابی سیستم‌های تعاملی و جایگزین نمودن آن به‌جای انسان برداشت. با توجه به مطالعات انجام‌شده در حوزهٔ مهندسی نرم‌افزار و سیستم‌های پرسش و پاسخ تعاملی، برای اندازه‌گیری قابلیت اطمینان در سیستم‌های IQA تعریف آن بدین‌صورت انجام شد. منظور از قابلیت اطمینان در سیستم‌های IQA، توانایی سیستم در انجام کارکردهای خواسته‌شده از سیستم تحت شرایط محیطی یکسان برای مدت زمان خاص و با در نظر گرفتن دقت سیستم است. نحوهٔ اندازه‌گیری این معیار بر اساس تعریف ارائه‌شده صورت گرفت.

٤. روش پژوهش

با توجه به این‌که یک سیستم پرسش و پاسخ تعاملی از دو موجودیت سیستم و کاربر تشکیل شده و گستردگی ویژگی‌ها از جنبه‌های مختلف امکان ایجاد یک روش استاندارد و جامع را بسیار سخت می‌سازد، جایگزین کردن مرحله به مرحله و محاسبهٔ خودکار ویژگی‌ها تا حدود زیادی می‌تواند به فرایند خودکار کردن ارزیابی سیستم‌ها کمک نماید. یکی از این ویژگی‌ها قابلیت اطمینان یک سیستم پرسش و پاسخ است که در توسعهٔ سامانه‌های پرسش و پاسخ تعاملی، محاسبهٔ خودکار آن جهت خودکارسازی فرایند ارزیابی ضروری به نظر می‌رسد. با توجه به این‌که در بیشتر ارزیابی‌های صورت‌گرفته از معیار کیفی یا کمی استفاده شده است، در روش پیشنهادی، برای محاسبهٔ معیار قابلیت اطمینان از هر دو معیار کیفی و کمی بهره گرفته شد تا مقدار حاصل برای قابلیت اطمینان دقیق‌تر باشد. برای ترکیب این دو معیار رابطهٔ پیشنهادی به‌صورت معادلهٔ ۱ آورده شده است.

$$\mathrm{Re}liability = \alpha R_{quntity} + \sum_{i=1}^{n} \beta_i \times Score_i \qquad (۱)$$

که در آن $R_{quntity}$ میزان قابلیت اطمینان کمی به‌دست‌آمده از سیستم و $Score_i$ امتیاز اامین پرسش از پرسشنامهٔ ارزیابی کیفی، α و β ضرایب تأثیر و n تعداد سؤالات پرسشنامه است. لازم به ذکر است که برای محاسبهٔ قسمت کیفی رابطهٔ پیشنهادشده از رگرسیون بهره گرفته شد که در ادامه، نحوهٔ محاسبه بیان شده است.

Downloaded from jipm.irandoc.ac.ir at 13:46 IRDT on Sunday June 13th 2021

٤-١. سیستم تعاملی پایه

برای محاسبهٔ قابلیت اطمینان سیستم‌های IQA، نیاز به دسترسی کامل به کلیهٔ قسمت‌های سخت‌افزاری و نرم‌افزاری این سیستم‌هاست. از آنجا که این امکان برای سیستم‌های موجود فراهم نبود، یک سیستم تعاملی پایه در آزمایشگاه تکنولوژی زبان‌های طبیعی[1] دانشگاه «شاهرود» طراحی شد و از آن، جهت ارزیابی روش پیشنهادی بهره گرفته شد. این سیستم بر خلاف سیستم‌های پیشین که عمدتاً از تجزیه و تحلیلی معنایی و گرامری استفاده می‌کنند، مستقل از زبان طراحی شده است؛ یعنی این سیستم با در اختیار داشتن پایگاه دادگان مناسب هر زبان می‌تواند به سؤالات مطرح‌شده به آن زبان پاسخ دهد. در سیستم طراحی‌شده از سه پایگاه داده[2] با نام WMPR-QA2-2015 ،WMPR-QA1-2015و WMPR-QA3-2015 جهت آموزش و ارزیابی استفاده شد (شهرآیینی و زاهدی ۱۳۹۵). برای بهبود عملکرد سیستم پایه، تغییراتی در نحوهٔ بازیابی نتایج آن به کمک شبکهٔ عصبی ایجاد گردید تا پاسخ ارائه‌شده از دقت بالاتری برخوردار باشد. نتایج حاصل از این بهینه‌سازی در مقالهٔ «حسینی و زاهدی (۱۳۹۵) آورده شده است.

٤-٢. انتخاب ویژگی در ارزیابی کمی

از آنجا که سامانهٔ طراحی‌شده برای سیستم پرسش و پاسخ تعاملی و دیگر سیستم‌ها یک سامانهٔ تحت است، بنابراین، برای ارزیابی کمّی آن می‌بایست یک مجموعهٔ بهینه از معیارهای ارزیابی قابلیت اطمینان ایجاد شود که اولاً همهٔ جوانب لازم برای این ارزیابی را پوشش دهد و ثانیاً از نظر پیچیدگی و سربارهای ارزیابی کمینه باشد (Carl and Bansiya 2013). برای تحقق این هدف رویکردهای گوناگونی قابل به‌کارگیری هستند که رویکرد مورد استفاده در این مقاله گزینشی است. در این رویکرد، ابتدا مجموعهٔ معیارهای موجود به‌دقت تجزیه و تحلیل و بررسی شد. سپس، بر اساس یک‌سری ضوابط، معیارهای موجود گزینش می‌شوند تا مجموعهٔ بهینه از بین معیارها برای ارزیابی قابلیت اطمینان انتخاب گردند. مجموعهٔ معیارهایی که به‌عنوان مبنا در نظر گرفته شدند تا انتخاب نهایی از بین آن‌ها صورت پذیرد، معیارهای C&K[3] و MOOD هستند. این دو مجموعه، برای ارزیابی از معیارهای شئ‌گرا بهره می‌گیرند (Carl and Bansiya 2013). معیارهای C&K برای

1. www.hltech.ir
2. www.hltech.ir/fa/51
3. Chidamber and Kemerer

Downloaded from jipm.irandoc.ac.ir at 13:46 IRDT on Sunday June 13th 2021

اندازه‌گیری پیچیدگی طراحی بر اساس تأثیر آن‌ها روی ویژگی‌های کیفیت، مانند قابلیت فهم، قابلیت نگهداری و قابلیت استفادهٔ مجدد تعریف شده‌اند و مجموعهٔ معیارهای MOOD به یک مکانیسم ساختاری شئ‌گرا، مانند نهان‌سازی، ارث‌بری، چندریختی و اتصال به‌کار گرفته می‌شوند (Laheeb & Ibrahim 2013). برای انتخاب و تشکیل مجموعهٔ بهینهٔ ویژگی‌ها بر اساس مطالعات قبلی، از پنج ضابطهٔ قابلیت سنجش ابعاد کیفی، قابلیت سنجش خصوصیت کیفی استفادهٔ مجدد (منظور ضابطه‌ای که یکی از ویژگی‌های اتصال، انسجام، وراثت یا پیچیدگی را بررسی کند)، اعتبارسنجی تجربی، قابلیت استفادهٔ زودهنگام و تعریف دقیق معیار بهره گرفته شد که معیارهای C&K و معیارهای MOOD این چند ضابطه را پوشش می‌دادند (Suresh, Pati & Rath 2012). با توجه به این ضوابط، هشت معیار WMC[1]، DIT[2]، NOC[3]، RFC[4]، CBO[5]، LOCM[6]، AIF[7] و MIF[8] تحت عنوان یک مجموعهٔ کمینه برای سنجش قابلیت اطمینان مورد استفاده قرار گرفتند. توضیحات مربوط به هر یک از این ویژگی‌ها در جدول ۱، آورده شده است. با توجه به آزمایشات و بررسی‌های صورت‌گرفته، تأثیر هر یک از این معیارها به‌صورت مجزا برای محاسبهٔ قابلیت اطمینان مورد بررسی قرار گرفت که در نهایت، برای WMC، DIT، RFC، LOCM، CBO ارتباط معکوس و برای NOC رابطهٔ مستقیم در نظر گرفته شد.

جدول ۱. ویژگی‌های کمی در نظر گرفته‌شده برای ارزیابی سیستم پرسش و پاسخ تعاملی

ویژگی	توصیف عملکرد
WMC	تعداد روش‌های هر کلاس را اندازه‌گیری می‌کند.
DIT	طول طولانی‌ترین مسیر ارث‌بری از یک کلاس به کلاس ریشه
RFC	تعداد روش‌های پاسخ‌دهنده به پیام دریافتی یکی از اشیای کلاس
LOCM	نبود چسبندگی داخل کلاس
CBO	تعداد کلاس‌های پیوسته به یک کلاس خاص
NOC	تعداد کلاس‌هایی که از یک کلاس ارث می‌برند
AIF و MIF	میزان استفاده از تکنیک وراثت در ساختار اشیا

1. Weighted Methods Per Class
2. Depth of Inheritance Tree
3. Number Of Children
4. Response For a Class
5. Coupling Between Object Classes
6. Lack Of Cohesion in Method
7. Attribute Inheritance Factor
8. Method Inheritance Factor

Downloaded from jipm.irandoc.ac.ir at 13:46 IRDT on Sunday June 13th 2021

بر این اساس، برای اندازه‌گیری میزان قابلیت اطمینان در ارزیابی کمی بعد از بررسی حالت‌های مختلف، رابطهٔ زیر پیشنهاد گردید تا بیانگر بهتر ارتباط بین هر کدام از ویژگی‌ها با میزان قابلیت اطمینان باشد.

$$R_{quantity} = K \times \frac{1}{WMC \times DIT \times RFC \times LCOM \times CBO} + \log(10 + NOC) \quad (۲)$$

در این رابطه، K ضریب تأثیر است. مقدار این ضریب می‌تواند بین بازهٔ ۱۰ تا ۲۰ باشد که در روش پیشنهادی مقدار K، در عدد ۱۵ نتایج مطلوب‌تری را از خود نشان داد.

٤-۳. نتایج ارزیابی کمی

همان‌طور که در بخش قبل اشاره شد، هر یک از معیارها در محاسبهٔ قابلیت اطمینان می‌تواند تأثیری مستقیم یا معکوس و یا حتی خنثی به نرم‌افزار داشته باشد. بنابراین، با توجه به آزمایش‌های صورت‌گرفته، ارتباط بین هر یک از ویژگی‌ها با خروجی تعیین و با به‌کارگیری رابطهٔ ۲ میزان قابلیت اطمینان محاسبه گردید. به‌عنوان نمونه، ارزیابی ساختار مدیریت بانک اطلاعاتی سیستم طراحی‌شده در جدول ۲، نشان داده شده است.

جدول ۲. نمونه‌ای از ارزیابی انجام‌شده بر روی بانک اطلاعاتی سیستم پرسش و پاسخ طراحی‌شده

متریک	تأثیر	کلاس DBMS	کلاس ارائه‌دهنده	کلاس امنیتی
WMC	معکوس	۲	۳	۲/۳
DIT	معکوس	۳	۲/۹	۴
RFC	معکوس	۴	۳/۸	۳/۵
LOCM	معکوس	۴	۴	۳/۸
CBO	معکوس	۴	۴	۳/۷
NOC	مستقیم	۴	۳	۱
قابلیت اطمینان		۳/۶۷	۴/۲۵	۳/۵

همان‌طور که در جدول ۲، نمایش داده شده، به ازای هر متریک تأثیر آن بیان شده است و بنابراین، با توجه به مقدار هر یک و تأثیر آن و با قرار دادن در معادلهٔ ۲ می‌توان مقدار قابلیت اطمینان را محاسبه نمود. به‌طور مثال، برای پارامتر WMC چون رابطهٔ معکوس

Downloaded from jipm.irandoc.ac.ir at 13:46 IRDT on Sunday June 13th 2021

دارد و مقدار آن ۴ است، برای مقدار آن در رابطهٔ ۲ عدد یک (حداکثر مقدار ۵ امتیاز در نظر گرفته شده است) در نظر گرفته می‌شود. قابل ذکر است که قابلیت اطمینان برای ساختار بانک اطلاعاتی، ساختار مدیریت خطا و استفاده از نرم‌افزار به‌صورت جداگانه برای سیستم پایه تعاملی در آزمایشگاه محاسبه گردید. عدد به‌دست آمده برای هر قسمت (که امتیازی بین ۱ تا ۵ است)، نرمالیزه شده و به‌ترتیب، برای هر قسمت گزارش گردید. برای محاسبهٔ نهایی قابلیت اطمینان میانگین این سه عدد حاصل در نظر گرفته شد. نتایج به‌دست آمده به تفکیک در جدول ۳، آورده شده است.

جدول ۳. مقادیر حاصل از قابلیت اطمینان برای بخش‌های مختلف

	ساختار بانک اطلاعاتی به‌طور متوسط	ساختار مدیریت خطا	استفاده از نرم‌افزار
مقدار قابلیت اطمینان	۳/۸	۴/۲	۴/۸

٤-٤. ارزیابی کیفی و نتایج آن

ارزیابی کمی متأسفانه اطلاعات کافی دربارهٔ کیفیت تعامل سیستم با کاربر و این‌که آیا تعامل موفقیت‌آمیز خاتمه یافته است یا خیر، ارائه نمی‌کند. به همین دلیل، تنها از طریق ارزیابی کمی نمی‌توان به قابلیت اطمینان یک سیستم به‌صورت صحیح دست یافت (حسینی و زاهدی ۱۳۹۷). بنابراین، در روش پیشنهادی از ارزیابی کیفی نیز برای محاسبهٔ مناسب‌تر قابلیت اطمینان سیستم بهره گرفته شد. برای ارزیابی کیفی پرسشنامه‌ای تهیه شده و در اختیار کاربران قرار گرفت تا بر اساس امتیاز به‌دست آمده از آن، میزان قابلیت اطمینان سیستم محاسبه شود. «هاراباگویی» و همکارانش، بر اندازه‌گیری سه ویژگی از سیستم‌های پرسش و پاسخ تعاملی به‌وسیلهٔ پرسشنامه تأکید می‌کنند. این سه ویژگی عبارت‌اند از (Harabagiu et al. 2005):

الف) کارآمدی[1]: تعداد پرسش‌هایی که کاربر باید برای دستیابی به اطلاعات مورد نظر مطرح کند؛

ب) تأثیرگذاری[2]: صحت پاسخ‌های ارائه‌شده توسط سیستم؛

ج) رضایت عمومی کاربر از عملکرد سیستم.

1. efficiency 2. effectiveness

Downloaded from jipm.irandoc.ac.ir at 13:46 IRDT on Sunday June 13th 2021

بر اساس این سه ویژگی، پرسشنامه‌ای مشتمل بر ۱۷ سؤال توسط افراد خبره در ارزیابی سیستم پرسش و پاسخ تعاملی تهیه گردید. قبل از این‌که پرسشنامه برای ارزیابی سیستم‌ها در اختیار کاربران قرار گیرد، پایایی آن با استفاده از روش بازآزمایی مورد بررسی قرار گرفت. بدین‌صورت که آزمون ارزیابی بیش از یک‌بار در یک گروه ۲۰ نفره تحت شرایط یکسان مورد بررسی قرار گرفت و با توجه به نمرات داده‌شده به سؤالات، ضریب همبستگی محاسبه و سؤالات با وابستگی بالا حذف گردید. در نهایت، تعداد ۱۳ پرسش مورد تأیید بود و در آزمون ارزیابی مورد استفاده قرار گرفت. از کاربرانی که با سیستم کار می‌کردند، خواسته شد تا در انتهای کار در پاسخ به این پرسش‌ها امتیازی بین ۱ تا ۵ به‌عنوان قابلیت اطمینان سیستم اختصاص دهند. به‌طور نمونه، تعداد ۵ پرسش از ۱۳ پرسشی که در پرسشنامه مورد استفاده قرار گرفته بود، در جدول ۴، آورده شده است.

جدول ٤. نمونه‌ای از سؤالات پرسشنامه

شماره سؤال	سؤال
۱	تا چه حدی سناریوی در نظر گرفته‌شده واقع‌گرایانه بود؟ به‌عبارت دیگر، آیا این سناریو به وظایفی که شما می‌توانستید در اجرای کار تصور کنید، شباهت داشت؟
۳	آیا دسترسی به اطلاعات مورد نظرتان برای شما ساده بود؟
۴	آیا کار با این سیستم به‌گونه‌ای بود که در آینده مجدداً از این سیستم استفاده نمایید؟
۱۲	طول مدت زمانی را که برای انجام این کار به‌خود اختصاص دادید را چگونه ارزیابی می‌کنید؟
۱۳	اگر شما مجبور به انجام وظیفه‌ای شبیه به آن چیزی که در سناریو توضیح داده شده، بودید، فکر می‌کنید دسترسی داشتن به سیستم آزمایشگاه به شما کمک می‌کرد؟ الف) گزارش نهایی شما را بهبود می‌بخشید؟ ب) به سؤالات خاصی که شما به‌طور جاری در پاسخ دادن به آن‌ها مشکل داشتید، پاسخ می‌داد؟ ج) سرعتی را که شما با آن اطلاعات را می‌یابید، افزایش می‌دهد؟ د) اطلاعاتی را که شما نسبت به آن‌ها آگاهی ندارید، می‌یابد؟

جدول ۵، نتایج حاصل از فراوانی نمرات کلی شرکت‌کنندگان در ارزیابی را نشان داده است. تعداد شرکت‌کنندگان ۲۵۰ نفر از دانشجویان دانشگاه صنعتی «شاهرود» بودند که نسبت به نحوهٔ استفاده از سیستم آگاهی کامل داشتند.

Downloaded from jipm.irandoc.ac.ir at 13:46 IRDT on Sunday June 13th 2021

جدول ۵. فراوانی نمرات داده‌شده

پرسش	امتیاز ۱	امتیاز ۲	امتیاز ۳	امتیاز ٤	امتیاز ٥	امتیاز متوسط
۱	–	–	۲۱	۵۶	۱۷۳	۴/۶۰۸
۲	۱۲	۱۷	۵۱	۷۵	۹۵	۳/۸۹۶
۳	–	۲۶	۶۷	۸۴	۷۳	۳/۸۱۶
۴	–	۱۲	۳۸	۱۲۷	۷۳	۴/۰۴۴
۵	–	۲۷	۳۳	۹۱	۹۹	۴/۰۴۸
۶	–	۲۰	۸۴	۱۰۹	۳۷	۳/۶۵۲
۷	–	۳۵	۳۴	۸۱	۱۰۰	۳/۹۸۴
۸	۱۰	۴۳	۵۷	۶۴	۷۶	۳/۶۱۲
۹	–	۲۴	۳۱	۷۹	۱۱۶	۴/۱۴۸
۱۰	–	۳۳	۷۲	۷۷	۶۸	۳/۷۲۰
۱۱	۱۰	۱۹	۵۲	۷۸	۹۱	۳/۸۸۴
۱۲	–	–	۲۸	۱۰۷	۱۱۵	۴/۳۴۸
۱۳الف	–	۴۶	۷۳	۵۴	۷۷	۳/۶۴۸
۱۳ب	–	۱۱	۵۴	۹۹	۸۶	۴/۰۴۰
۱۳ج	–	۲۱	۶۲	۸۹	۸۸	۴/۰۹۶
۱۳د	–	۲۶	۷۴	۷۹	۷۱	۳/۷۸۰

٤-۵. رگرسیون خطی

تحلیل رگرسیون، تکنیکی آماری برای بررسی و مدل‌سازی ارتباط بین متغیر وابسته و متغیر مستقل بوده و هدف آن پیش‌بینی متغیر وابسته از روی متغیر و یا متغیرهای مستقل است. در رگرسیون سادهٔ خطی به بررسی رابطهٔ یک متغیر مستقل (پیش‌بین) و یک متغیر وابسته پرداخته می‌شود. حال، اگر تعداد متغیرهای مستقل در این رابطهٔ خطی بیش از یک عدد شود، مدل رگرسیون، خطی چندگانه نامیده می‌شود. معادلهٔ رگرسیون خطی چندگانه به‌صورت $Y = a + b1X1 + b2X2 + \ldots + bnXn$ است. برای پیش‌بینی ضرایب مربوط به هر امتیاز با توجه به امتیاز کلی داده‌شده توسط هر کاربر از رگرسیون چندگانه بهره گرفتیم. ضرایب به‌دست آمده برای هر سؤال در جدول ۶، آورده شده است.

Downloaded from jipm.irandoc.ac.ir at 13:46 IRDT on Sunday June 13th 2021

جدول ٦. ضرایب به‌دست‌آمده برای هر سؤال

شماره سؤال	ضریب	شماره سؤال	ضریب
۱	۰/۸۲۰۴	۱۰	- ۰/۰۸۳۵
۲	۰/۰۲۲۹	۱۱	- ۰/۰۵۴۳
۳	- ۰/۹۵۵۹	۱۲	۱/۶۰۶۲
۴	۱/۱۰۴۸	۱۳a	۱/۰۹۳۰
۵	۱/۰۰۷۳	۱۳b	- ۰/۶۷۳۱
۶	- ۱/۰۱۹۰	۱۳c	- ۱/۱۹۲۹
۷	۰/۳۶۳۹	۱۳d	۴/۶۲۷۲
۸	- ۰/۴۸۲۲	مقدار ثابت معادله	۱/۴۴۴۵
۹	۰/۹۰۷۲	مقدار α	۱/۵۲۷

٤-٦. معیار ارزیابی

برای ارزیابی نتایج حاصل از مدل به‌دست آمده با داده‌های واقعی از سه سنجهٔ آماری ضریب تعیین R2، جذر میانگین مربعات خطا[1] (RMSE) و درصد میانگین مطلق خطا[2] (MAPE) استفاده گردید. روابط مربوط به این معیارها در روابط ۳ تا ۵ نشان داده شده است.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} e_i}{n}} \quad (۳)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{e_i}{Y_i}\right| \times 100 \quad (۴)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - y)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \quad (۵)$$

در این روابط n تعداد پیش‌بینی‌ها و ei خطای پیش‌بینی است که از تفاوت مقادیر

1. Root Mean Square Error
2. Mean Absolute Percentage Error

Downloaded from jipm.irandoc.ac.ir at 13:46 IRDT on Sunday June 13th 2021

پیش‌بینی شده و مقادیر واقعی به‌دست می‌آید و Yi مقادیر واقعی است. همان‌طور که در جدول ۶، نشان داده شده، مقدار تعدادی از ضرایب مثبت و مقدار تعدادی از آن‌ها منفی است. علامت به‌دست آمده برای این مقادیر به معنای تأثیر هر یک از این ویژگی‌ها در معادله است. یعنی ویژگی‌هایی که دارای ضریب مثبت هستند، تأثیر مستقیم و آن‌هایی که دارای علامت منفی هستند، تأثیر معکوس دارند. همچنین، معادلهٔ خط رگرسیونی به‌دست آمده می‌بایست مورد ارزیابی قرار می‌گرفت تا صحت معادلهٔ به‌دست آمده مورد بررسی قرار گیرد. با توجه به معیارهای ارزیابی در نظر گرفته شده، نتیجهٔ حاصل از این معیارها در جدول ۷، نشان داده شده است.

جدول ۷. نتایج حاصل از معیارهای ارزیابی

نوع معیار	ضریب R^2	MSE	MAPE
اندازه	۰/۸۵	۰/۱۵	۵ درصد

همان‌طور که در بالا توضیح داده شد، ضرایب به‌دست‌آمده برای هر پرسش، حاکی از میزان تأثیر آن در ارزیابی است. بنابراین، هرچه مقدار قدر مطلق این ضریب بیشتر، تأثیرگذاری آن بیشتر و هرچه مقدار آن کمتر باشد، تأثیر آن در خروجی کمتر خواهد بود. در ادامه، برای بررسی بیشتر این ضرایب، ویژگی‌ها تک‌تک حذف و یا اضافه شدند تا تأثیر آن‌ها در خروجی مورد بررسی قرار گیرد. در نهایت، با در نظر گرفتن یک مقدار آستانهٔ برابر با ۰/۵، سؤالاتی را که ضریب آن‌ها (بدون در نظر گرفتن علامت) کمتر از این مقدار بود، حذف گردید. بر این اساس، سؤالات ۲، ۷، ۸، ۱۰ و ۱۱ از مجموعهٔ پرسشنامه حذف و دوباره مدل‌سازی انجام پذیرفت. نتایج به‌دست آمده بیانگر بهبود نتایج نسبت به حالت قبلی بود (جدول ۸).

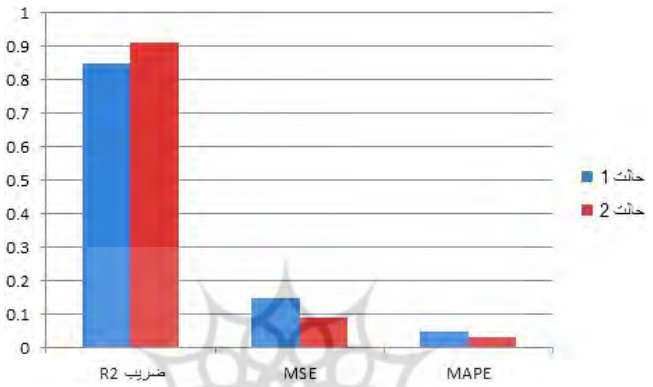
جدول ۸. نتایج حاصل از معیارهای ارزیابی بعد از حذف پرسش‌ها

نوع معیار	ضریب R^2	MSE	MAPE
اندازه	۰/۹۱	۰/۰۹	۳ درصد

نتایج حاصل از مقایسهٔ این دو حالت در شکل ۱، نمایش داده شده است. در این شکل با حذف‌شدن پنج سؤال از سؤالات پرسشنامه (تقریباً معادل حذف ۳۸ درصد سؤالات) توانستیم به نتیجهٔ مطلوب دست پیدا کنیم. در نهایت، با به‌دست آمدن ضرایب هر سؤال

Downloaded from jipm.irandoc.ac.ir at 13:46 IRDT on Sunday June 13th 2021

و ترکیب آن‌ها با یکدیگر و همچنین، قراردادن این امتیاز کمّی سیستم و نرمالیزه‌کردن این اعداد (عدد بین صفر و یک) می‌توان مقدار قابلیت اطمینان را مطابق با معادلهٔ ۱ برای یک سیستم محاسبه نمود. بنابراین، با توجه به نتایج گزارش‌شده می‌توان ادعا نمود که روش پیشنهادی بیانگر ارائهٔ یک روش مناسب جهت اندازه‌گیری قابلیت اطمینان در سیستم‌های پرسش و پاسخ تعاملی است.



شکل ۱. نمودار مقایسه بین معیارهای ارزیابی برای حالت حذف و بدون حذف پرسش‌های ۲، ۷، ۸، ۱۰، ۱۱

۵. نتیجه‌گیری

با توجه به تحقیقات بسیار اندک در حوزهٔ ارزیابی سیستم‌های IQA و ارائهٔ یک مدل خودکار جهت ارزیابی این‌گونه سیستم‌ها، در این مقاله روشی برای محاسبهٔ قابلیت اطمینان (که یکی از ویژگی‌های مورد ارزیابی سیستم‌های تعاملی است)، ارائه شده است تا بتوان گامی در جهت خودکارنمودن فرایند ارزیابی سیستم‌های تعاملی برداشت. استفاده از کاربران واقعی، وظایف واقعی و سیستم واقعی از ویژگی‌های این پژوهش است. روش ارائه‌شده تا حد قابل قبولی می‌تواند به‌عنوان یک ابزار ارزیابی مقطعی به اندازه‌گیری‌های قابلیت اطمینان با تعداد محدود کاربر به‌کار گرفته شود. از طرف دیگر، با توجه به رعایت‌کردن استانداردهای لازم در ایجاد پرسشنامه‌ها، این چارچوب می‌تواند به درک فرایندی که کاربران به شکل واقعی وظیفه‌ای را در یک محیط کار حقیقی انجام می‌دهند تا به یک سیستم پرسش و پاسخ تعاملی اطمینان کنند، کمک نماید. در نهایت، این‌که با مدل‌سازی دیگر ویژگی‌های موجود در فرایند ارزیابی و کمک به حذف انسان می‌توان گام بزرگی در فرایند ارزیابی سیستم‌های تعاملی برداشت.

Downloaded from jipm.irandoc.ac.ir at 13:46 IRDT on Sunday June 13th 2021

فهرست منابع

حسینی، محمدمهدی، و مرتضی زاهدی. ۱۳۹۵. بهبود پاسخ ارائه‌شده در سیستم‌های پرسش و پاسخ تعاملی با استفاده از شبکه عصبی. هشتمین کنفرانس بین‌المللی فناوری اطلاعات و دانش. همدان. صفحات ۸۴-۹۱.

_____. ۱۳۹۷. ارائه یک مدل احتمالاتی جهت تعیین انسجام متن در سیستم‌های پرسش و پاسخ تعاملی. *مجله پردازش و مدیریت اطلاعات* ۳۳ (۴): ۱۷۳۷-۱۷۶۰.

_____، حمید حسن‌پور. ۱۳۹۷. مجموعه‌ای از ویژگی‌های آماری جدید برای ارزیابی سیستم‌های پرسش و پاسخ تعاملی. *مجله مهندسی برق تبریز*.

شهرآیینی، سلیمه‌سادات، و مرتضی زاهدی. ۱۳۹۴. سیستم پاسخگوی تعاملی با استفاده از تکنیک‌های هوش مصنوعی. پایان‌نامهٔ کارشناسی ارشد. دانشکدهٔ کامپیوتر، دانشگاه صنعتی شاهرود.

References

Baudiš, Petr. 2015. YodaQA: a modular question answering system pipeline", In POSTER 2015-19th International Student Conference on Electrical Engineering, pp. 1156-1165. Prague.

Bhatia, Pradeep K. and Mann, Rajbeer. 2008. An Approach to Measure Software Reusability of OO Design. Proceedings of 2nd National Conference on Challenges & Opportunities in Information Technology, RIMT-IET, Mandi Gobindgarh, pp. 26-30. India.

Breck, E., J. D. Burger, L. Ferro, L. Hirschman, D. House, M. Light, and I. Mani. 2000. How to evaluate your question answering system every day and still get real work done. In Proceedings of the Second International Conference on Language Resources and Evaluation (LREC), Greece. pp. 51-57.

Burstein, J., & M. Chodorow. 2009. Automated essay scoring for non-native English speakers. In Proceedings of the ACL Workshop on Computer Mediated Language Assessment and Evaluation of Natural Language Processing. Mexico.

Carl G. Davis, and Jagdish Bansiya. 2013. A Hierarchical Model for Object-Oriented Design Quality Assessment. *IEEE Transactions on Software Engineering* 28 (1): 1-31.

Derici, Caner, Yiğit Aydin, Çiğdem Yenialaca, Nihal Yağmur Aydin, Günizi Kartal, Arzucan Özgür, and Tunga Güngör. 2018. A closed-domain question answering framework using reliable resources to assist students. *Natural Language Engineering* 24 (5): 725-762.

Dwivedi, S. K., and V. Singh. 2013. Research and Reviews in Question Answering System. *Procedia Technology*, pp. 424-417.

Harabagiu, S., A. Hickl, J. Lehmann, and D. Moldovan. 2005. Experiments with interactive question-answering. In Proceedings of the 43rd annual meeting on Association for Computational Linguistics, Ann Arbor, USA, pp. 205-214.

Iyyer, Mohit, Wen-tau Yih, and Ming-Wei Chang. 2017. Search-based neural structured learning for sequential question answering. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1821-1831. Vancouver, Canada.

Laheeb M. Al-Zobaidy, & Khalil A. Ibrahim. 2013. Existing Object Oriented Design Metrics a Study and Comparison. *Journal of Computer* 10: 137-147.

Motte, Stanislas, Ramon Ruti, Arnaud Jacolin, and Pierre-Albert Ruquier. 2019. Apparatus, systems and methods for scoring and distributing the reliability of online information. U.S. Patent 10, 169, 424, issued January 1, 2019.

Paliwal, Nagesh, Vivek Shrivastava, & Ketki Tiwari. 2014. An Approach to Find Reusability of Software Using Objet Oriented Metrics. *International Journal of Innovative Research in Science, Engineering and Technology* 3 (3): 10371-10378.

Salata, Ferdinando, Andrea de Lieto Vollaro, Roberto de LietoVollaro, and Lorenzo Mancieri. 2015. Method for energy optimization with reliability analysis of a trigeneration and teleheating system on urban scale: A case study. *Energy and Buildings* 86: 118-136.

Spitters, M., M. De Boni, J. Zavrel, and R. Bonnema. 2009. Learning effective and engaging strategies for advice-giving human-machine dialogue. *Natural Language Engineering* 15 (3): 378-355.

Suresh Y, J. Pati, Ku Rath. 2012. Effectiveness of Software Metrics For Object-Oriented System. *Procedia Technology* 6: 420-427.

Webber, B., and N. Webb. 2010. Question answering, the handbook of computational linguistics and natural language processing. Blackwell Publishing. PP. 630-654, 2010.

محمدمهدی حسینی

متولد ۱۳۶۲ دارای مدرک تحصیلی دکتری در رشتهٔ کامپیوتر گرایش هوش مصنوعی از دانشگاه صنعتی شاهرود است. ایشان هم‌اکنون استادیار گروه مهندسی کامپیوتر دانشگاه آزاد اسلامی واحد شاهرود است.

پردازش متن، پردازش تصویر و حل مسائل بهینه‌سازی از جمله علایق پژوهشی ایشان است.

علیرضا جلالی

متولد ۱۳۵۷ دارای مدرک مهندسی کامپیوتر گرایش نرم‌افزار از دانشگاه مالایا مالزی است. ایشان هم‌اکنون استادیار گروه مهندسی کامپیوتر دانشگاه آزاد اسلامی واحد شاهرود است.

طراحی و مهندسی نرم‌افزار و تحلیل و طراحی سیستم‌های اطلاعات مدیریت از جمله علایق پژوهشی ایشان است.

Downloaded from jipm.irandoc.ac.ir at 13:46 IRDT on Sunday June 13th 2021