

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده مهندسی برق و کامپیوتر

استخراج استراتژی‌های موثر در یک بازی شناختی در حوزه‌ی توجه و تمرکز و بررسی تاثیر آموزش آن بر افراد مبتدی

نگارش

الهه ابوالحسنی شهرضا

استاد راهنما

دکتر مجید نیلی

پایان‌نامه برای دریافت درجه کارشناسی ارشد در رشته
مهندسی کامپیوتر - گرایش هوش مصنوعی

شهریور ۱۳۹۶



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده مهندسی برق و کامپیوتر

پایان نامه برای دریافت درجه کارشناسی ارشد در رشته مهندسی کامپیوتر

عنوان:

**استخراج استراتژی‌های موثر در یک بازی شناختی در حوزه‌ی توجه و تمرکز و
بررسی تاثیر آموزش آن بر افراد مبتدی**

نگارش: الهه ابوالحسنی شهرضا

این پایان‌نامه در تاریخ ۱۳۹۶/۰۶/۱۲ در مقابل هیأت داوران دفاع گردید و مورد تصویب قرار گرفت.

معاون آموزشی و تحصیلات تکمیلی پردیس دانشکده‌های فنی: دکتر جلیل آقا راشد محصل

رئیس دانشکده مهندسی برق و کامپیوتر: دکتر مجید نیلی

معاون پژوهشی و تحصیلات تکمیلی: دکتر بابک نجار اعرابی

استاد راهنما: آقای دکتر مجید نیلی

استاد راهنمای دوم: آقای دکتر هادی مرادی

عضو هیأت داوران: خانم دکتر فتانه تقی‌یاره

عضو هیأت داوران: آقای دکتر بهروز مینایی

تعهدنامه اصالت اثر

اینجانب الهه ابوالحسنی شهرضا تایید می‌کنم که مطالب مندرج در این پایان نامه حاصل کار و پژوهش اینجانب بوده و به دستارودهای پژوهشی دیگران که در این نوشته از آنها استفاده شده است مطابق مقررات ارجاع گردیده است. به‌علاوه این پایان نامه قبلاً برای احراز هیچ مدرک هم سطح یا بالاتر ارائه نشده است. کلیه حقوق مادی و معنوی این اثر متعلق به دانشکده فنی دانشگاه تهران است.

نام و نام خانوادگی دانشجو: الهه ابوالحسنی شهرضا

امضای دانشجو:

تقدیر و تشکر

مراتب تشکر و قدردانی خود را نسبت به تمام کسانی که مراد انجام این پایان نامه یاری کرده اند، خصوصاً اساتید گرامی و ارجمندم، جناب آقای دکتر نیلی و جناب

آقای دکتر مرادی که رهنمودهای ایشان بهواره راحلشای پجیدگی های این پژوهش بوده است، ابراز می دارم.

همچنین، از تمامی دوستانم در آزمایشگاه رباتیک شناختی که با حضور و محبتشان مرایاری نمودند تشکر و قدردانی می نمایم.

چکیده

با توجه به رشد روزافزون اطلاعات موجود در وب، موتورهای جست‌وجو در بازیابی اطلاعات مورد نیاز کاربران از میان حجم زیادی از اطلاعات نقشی اساسی ایفا می‌کنند. با بررسی رفتار کاربر در اینترنت مشاهده شده است که بیشترین بازدید از یک صفحه وب، به واسطه نتایج اولیه بازیابی شده توسط موتورهای جست‌وجو می‌باشد. با توجه به این امر، ایده هرزنویسی در وب با هدف افزایش رتبه صفحات هرز در میان نتایج موتورهای جست‌وجو مطرح شد. برای شناسایی و مقابله با این صفحات روش‌هایی ارائه شده است که می‌توان آن‌ها را به سه دسته کلی روش‌های مبتنی بر محتوا، روش‌های مبتنی بر پیوند و روش‌های مبتنی بر داده‌های جانبی تقسیم نمود. در این پژوهش تمرکز بر روی دو روش اصلی مبتنی بر محتوا و مبتنی بر پیوند و همچنین ترکیب این دو روش به منظور شناسایی وب‌گاه‌های هرز می‌باشد.

از آنجایی که عملکرد موتورهای جست‌وجو در شناسایی وب‌گاه‌های هرز فارسی پایین می‌باشد، در این پژوهش پس از ساخت یک مجموعه داده‌ای مناسب شامل وب‌گاه‌های هرز و معتبر فارسی، به بررسی و تحلیل تعدادی از ویژگی‌های محتوایی برای شناسایی وب‌گاه‌های هرز فارسی می‌پردازیم. سپس با ارائه چندین ویژگی محتوایی جدید و استفاده از روش‌های انتخاب ویژگی، کارایی رده‌بندی وب‌گاه‌ها را افزایش می‌دهیم. در ادامه، یک سامانه جدید شناساگر هرز وب فارسی را ارائه می‌دهیم که از مدل بهبود یافته کیف کلمات برای استخراج ویژگی‌ها استفاده می‌نماید و نسبت به روش‌های محتوایی پیشین کارایی بالاتری دارد. با توجه به گسترش استفاده از الگوریتم‌های مبتنی بر پیوند در روش‌های هرزنویسی، تعدادی از الگوریتم‌های مهم در این زمینه را مورد بررسی قرار داده و دو الگوریتم جدید ارائه می‌دهیم که بسیاری از نقاط ضعف الگوریتم‌های پیشین را ندارند. در الگوریتم اول برای بهبود انتشار امتیاز اعتماد در گراف وب، از سه سیاست انتخاب بهینه گره‌های بذر، وزن‌دهی به یال‌های گراف برای مشخص کردن میزان اعتبار یال‌ها، و بسط دوره‌ای گره‌های بذر استفاده می‌شود. در الگوریتم دوم با استفاده از انتشار امتیاز هرز، هم‌زمان به صورت پیش‌رو و پس‌رو در سراسر گراف وب، کیفیت رتبه‌بندی وب‌گاه‌های هرز را بهبود می‌دهیم. در آخر نیز به منظور بهبود کیفیت رتبه‌بندی وب‌گاه‌ها روشی پیشنهاد داده می‌شود که برای انتشار امتیاز وب‌گاه‌ها، از احتمال اعتبار و هرز بودن محتوایی وب‌گاه‌ها در تمام بخش‌های گراف استفاده می‌نماید.

در پایان این پژوهش، به منظور ارزیابی روش‌ها و بررسی میزان کارایی آن‌ها، آزمایش‌های مربوطه انجام شده است. نتایج آزمایش‌ها نشان می‌دهد که روش‌های ارائه شده در مقایسه با روش‌های قبلی، از کارایی و دقت بالاتری برخوردار هستند.

واژه‌های کلیدی: هرزنویسی، هرز وب، شناسایی هرز، انتشار برجسب، ویژگی‌های محتوایی.

فهرست مطالب

۱	مقدمه	۱
۵	پژوهش‌های پیشین	۲
۵	۱.۲ بازی‌های شناختی	۲
۶	۲.۲ استفاده از آموزش استراتژی در بازی‌ها	۲
۶	۳.۲ پژوهش‌های انجام شده روی فعالیت ردیابی همزمان چند شیء	۲
۷	روش تحقیق	۳
۷	۱.۳ مقدمه	۳
۸	۲.۳ انتخاب بازی	۳
۸	۱.۲.۳ بازی‌های انتخاب شده	۳
۹	۲.۲.۳ طراحی آزمایش برای انتخاب بازی	۳
۱۰	۳.۳ طراحی آزمایش اصلی	۳
۱۰	۱.۳.۳ بخش اول - استخراج استراتژی	۳
۱۳	ارزیابی	۴
۱۵	جمع‌بندی و نکته‌های پایانی	۵
۱۷	مراجع	

۱۸

واژه‌نامه انگلیسی به فارسی

۲۰

واژه‌نامه فارسی به انگلیسی

فهرست جدول‌ها

۱۱	۱.۳	تعداد مراحل هر بخش
----	-------	-----	--------------------

فهرست شکل‌ها

فصل ۱

مقدمه

طبق تعریف بازی‌های شناختی^۲ بازی‌هایی هستند که هدف آنها بهبود توانمندی‌های شناختی^۳ بازیکنان است. در این بازی‌ها سعی می‌شود توانمندی‌های شناختی مانند توجه و تمرکز^۴، حافظه^۵ و حل مساله^۶ بهبود پیدا کنند. این بازی‌ها به منظور استفاده‌ی عموم مردم طراحی شده‌اند. با وجود توسعه و استفاده‌ی روزافزون از این بازی‌ها نتایج برخی از تحقیقات انجام شده ([۱]، [۲]) نشان می‌دهند در بسیاری از موارد تاثیرگذاری مورد انتظار را نداشته‌اند.

هدف نهایی بازی‌های شناختی بهبود توانمندی‌های شناختی افراد است. بهبود این توانمندی‌ها در هر فرد باعث بهبود کیفیت زندگی او می‌شود و آسیب‌های شناختی احتمالی ناشی از کهولت سن یا حوادث را به تعویق می‌اندازد. در اصل این بازی‌ها نوعی ورزش مغزی محسوب می‌شوند.

پرسشی که ایجاد می‌شود این است که آیا می‌توان با تغییر دادن بازی‌های شناختی و شخصی سازی آنها به بهبود تاثیرگذاری این بازی‌ها کمک کرد؟ آیا می‌توان با توجه به نقاط ضعف و قوت بازیکن به نحوی بازی را تغییر داد که بیشترین تاثیرگذاری ممکن اتفاق بیافتد؟

²Cognitive Games

³Cognitive Abilities

⁴Attention

⁵Memory

⁶Problem Solving

یکی از روش‌های رایج به منظور بهبود عملکرد افراد در حوزه‌های مختلف، مانند توانبخشی شناختی [۱]، آموزش زبان دوم [۲] و یا عملکرد دانشگاهی [۳] آموزش استراتژی^۱ است. منظور از استراتژی در این پژوهش، استراتژی یادگیری^۲ است. استراتژی یادگیری به فرآیندهایی گفته می‌شود که وقتی با نیازمندی‌های یک تمرین مطابقت پیدا می‌کنند باعث بهبود عملکرد می‌شوند [۳].

استراتژی‌های مربوط به بهبود حافظه شناخته‌شده‌تر هستند. به عنوان مثال می‌توان به استراتژی تکرار کردن^۳، برقرار کردن ارتباط معنایی^۴ یا گروه کردن^۵ اشاره کرد [۴]. اما روی استراتژی‌های مربوط به بهبود توجه کمتر کار شده است. از استراتژی‌های شناخته شده در حوزه‌ی توجه می‌توان به تماس چشمی^۶، توضیح دادن^۷ یا گفتگو با خود^۸ اشاره کرد [۵].

در این پژوهش هدف نهایی طراحی یک بازی شناختی تطبیق‌پذیر با بازیکن است که می‌تواند با توجه به نحوه‌ی عملکرد او روش بازی‌اش را استخراج کند و سپس استراتژی‌هایی را به او آموزش دهد که باعث بهبود عملکرد وی در بازی و نهایتاً در زندگی واقعی می‌شود.

استراتژی‌های مربوط به توجه تا به امروز مورد توجه قرار نگرفته بودند و استراتژی‌های بسیار محدودی برای آن مطرح شده بود. در این پژوهش سعی شده است مجموعه‌ای از استراتژی‌های مربوط به «توجه تقسیم شده»^۹ معرفی شوند.

علاوه بر این، این پژوهش یک چارچوب جهت استفاده از آموزش استراتژی در بازی‌های مختلف ارائه می‌دهد که می‌توان از آن برای بازی‌های دیگر نیز استفاده کرد.

به منظور دستیابی به اهداف این پژوهش یک بازی شناختی با نام «ابر باران‌زا» انتخاب شد که هدف اصلی آن بهبود شاخه‌ی «توجه تقسیم‌شده» از زیرشاخه‌های «توجه» است.

این پژوهش شامل دو فاز عمده است. فاز اول را فاز استخراج استراتژی و فاز دوم را فاز انتقال استراتژی می‌نامیم.

¹Strategy

²Learning Strategy

³Rehearsal

⁴Semantic

⁵Grouping

⁶Eye contact

⁷Paraphrasing

⁸self-talk

⁹Divided Attention

در فاز اول دو هدف پیگیری می‌شوند. هدف اول گردآوری استراتژی‌هایی است که افراد در این بازی استفاده می‌کنند و هدف دوم بررسی میزان اثرگذاری این استراتژی‌ها است. به این معنا که هر کدام از این استراتژی‌ها به صورت میانگین چقدر توانسته‌اند برای این بازیکن امتیاز به دست بیاورند.

در فاز دوم هدف بررسی تاثیر انتقال این استراتژی‌ها به افرادی است که عملکرد ضعیف‌تری داشته‌اند.

(تکمیل شود)

فصل ۲

پژوهش‌های پیشین

در این فصل، پژوهش‌های پیشین را سه بخش ارائه می‌دهیم: بازی‌های شناختی، استفاده از آموزش استراتژی در بازی‌ها، پژوهش‌های انجام شده روی فعالیت ردیابی همزمان چندین شیء^۲

۱.۲ بازی‌های شناختی

بازی‌های شناختی بازی‌هایی هستند که تلاش می‌کنند توانمندی‌های شناختی افراد را تقویت کنند. توانمندی‌های شناختی مهارت‌های ذهنی هستند که برای انجام دادن ساده‌ترین تا پیچیده‌ترین کارها مورد نیاز هستند. این مهارت‌ها شامل ادراک^۳، توجه^۴، حافظه^۵، مهارت‌های حرکتی^۵ و کارکردهای اجرایی^۶ هستند. یکی از مهم‌ترین مسائلی که در رابطه با بازی‌های شناختی مطرح می‌شود مساله‌ی میزان تاثیرگذاری این بازی‌ها است. سوالاتی که در این زمینه مطرح می‌شوند از این دست هستند: آیا بازی کردن با یک بازی مخصوص حافظه باعث می‌شود حافظه‌ی فرد بازیکن بهبود پیدا کند؟ چه مدت باید این بازی صورت بگیرد؟ بازیکن

^۲Multiple Object Tracking

^۳Perception

^۴Attention

^۵Motor Skills

^۶Executive Functions

باید چه شرایطی داشته باشد؟ مدت زمان تاثیرگذاری بازی چه مدت است؟ در سال ۲۰۱۱ جائگی [۶] تاثیرگذاری طولانی مدت و کوتاه مدت تمرین‌های شناختی را بررسی کرد. همچنین در تاثیرگذاری تمرین در یک حوزه شناختی بر بهبود عملکرد در یک حوزه شناختی دیگر بررسی شده است و نتیجه گرفته شده است که این دو حوزه بر یکدیگر اثرگذار هستند. به این نوع اثرگذاری در ادبیات این پژوهش انتقال گفته می‌شود. در [۶] انجام دادن تمرین‌هایی مرتبط با حوزه حافظه‌ی کاری انجام گرفته است و در نهایت عملکرد افراد در تمرین‌های مرتبط با هوش سیال ارزیابی شده است. نتیجه‌ی نهایی این است که افرادی که تمرین‌های مربوط به حافظه‌ی کاری را انجام داده‌اند در تمرین‌های مربوط به هوش سیال بهتر عمل کرده‌اند و در نتیجه انتقال اتفاق افتاده است. [۶] همچنین به بررسی تاثیر طولانی (۲ ماه) مدت این انتقال پرداخته است و نتیجه گرفته است که این تاثیرات در طولانی مدت نیز وجود داشته‌اند.

۲.۲ استفاده از آموزش استراتژی در بازی‌ها

۳.۲ پژوهش‌های انجام شده روی فعالیت ردیابی همزمان چند شیء

فصل ۳

روش تحقیق

۱.۳ مقدمه

هدف از انجام این پژوهش بررسی تاثیرگذاری آموزش استراتژی بر کارایی کاربر در یک بازی شناختی در حوزه‌ی توجه و تمرکز است. در حوزه‌های دیگر مانند حافظه کارهای مشابه صورت گرفته است (رفرنس) بنابراین یکی از دلایل انتخاب حوزه‌ی توجه و تمرکز شناخته شده نبودن استراتژی‌های مطرح در این حوزه بود. استراتژی‌های حوزه‌ی حافظه به قدری شناخته شده هستند که بسیاری از ما هنگام امتحانات مدرسه از آنها استفاده کرده‌ایم. یعنی به صورت عمومی افراد جامعه از آنها استفاده می‌کنند. اما در مورد توجه و تمرکز با اینکه تقاضا زیاد است اما استراتژی‌های مرتبط با آن به اندازه‌ی حافظه شناخته شده نیست. علاوه بر این این حوزه به خودی خود از اهمیت بالایی برخوردار است. بسیاری از افراد از نداشتن تمرکز به هنگام انجام کارهای روزمره‌ی خود شکایت دارند. همچنین اختلالات زیادی در حوزه‌ی توجه و تمرکز وجود دارند (مانند اختلال کمبود توجه^۲ یا بیش‌فعالی^۳).

شاید در این نقطه بگوییم بسیاری از حوزه‌های شناختی مانند توجه هستند که هم استراتژی‌های آنها ناشناخته است و هم از درجه اهمیت بالایی برخوردار هستند. ویژگی دیگری که حوزه‌ی توجه را از سایر حوزه‌های

^۲ADD

^۳ADHD

شناختی متمایز می‌کند این است که فعالیت‌ها^۱ و بازی‌های متعددی برای توجه و تمرکز طراحی شده و از آنها استفاده می‌شود. به همین دلیل نیازی به طراحی یک بازی جدید و صحت‌سنجی مجدد آن نیست. با توجه به مجموع این عوامل حوزه‌ی توجه و تمرکز انتخاب شد.

۲.۳ انتخاب بازی

بازی‌های متعددی در حوزه‌ی توجه و تمرکز توسعه پیدا کرده‌اند. این بازی‌ها روی فاکتورهای مختلف مانند توجه انتخابی، توجه تقسیم‌شده، توجه پایدار و فراخنای توجه کار می‌کنند. بازی انتخاب شده روی فاکتور توجه تقسیم شده کار می‌کند و بر تقویت توانایی انسان برای تمرکز همزمان روی چند عامل تاکید می‌نماید. هرچقدر این نوع از توجه بهتر باشد یک فرد بهتر می‌تواند چندین کار را به صورت همزمان با هم انجام دهد. مهم‌ترین معیار انتخاب این بود که بازی استراتژی‌های متنوع داشته باشد. به این معنا که افراد از استراتژی‌ها مختلف برای انجام بازی استفاده کنند. با توجه به این معیار ۴ بازی به عنوان کاندید انتخاب شدند.

۱.۲.۳ بازی‌های انتخاب شده

بازی ابر باران‌زا: این بازی روی توجه تقسیم‌شده کار می‌کند. بازی به این صورت است که تعدادی ابر باران‌زا و تعدادی ابر عادی در صفحه وجود دارند. این ابرها شروع به حرکت می‌کنند و کم‌کم ابرهای باران‌زا تبدیل به ابرهای عادی می‌شوند. در نهایت وقتی ابرها از حرکت ایستادند کاربر باید ابرهایی را که در ابتدا باران‌زا بودند مشخص کند. بازی سنگ-کاغذ-قیچی: این بازی روی ... کار می‌کند. این بازی مشابه بازی سنگ-کاغذ-قیچی مرسوم است با این تفاوت که کاربر با کامپیوتر بازی می‌کند و کامپیوتر برای بازی کردن الگوی مشخصی دارد. کاربر باید الگوی کامپیوتر را بفهمد و سپس با توجه به الگو به نحوی بازی کند که بیشترین امتیاز را به دست بیاورد. بازی مسافرخانه زنبوری: این بازی روی کار می‌کند. بازی به این صورت است که در ابتدا تعدادی کندوی زنبور عسل نمایش داده می‌شود که همه خالی هستند.

¹task

سپس زنبورها در کندوها رفت و آمد می‌کنند. این رفت و آمد شامل سه حرکت است: از بیرون به داخل کندو می‌روند، از کندو خارج می‌شوند یا بین کندوها جابجا می‌شوند. پس از اینکه حرکت زنبورها به پایان رسید کاربر باید مشخص کند در هر کندو چند زنبور وجود دارد. کتابخانه: این بازی روی کار می‌کند. بازی به این صورت است که تعدادی کتاب نمایش داده می‌شود. کاربر باید مشخص کند جلد هر کتاب با کتاب قبلی یکسان بوده است یا خیر. در مراحل بالاتر به جای کتاب قبلی باید دو کتاب یا سه کتاب قبلی را در نظر بگیرد.

۲.۲.۳ طراحی آزمایش برای انتخاب بازی

برای انتخاب بازی یک آزمایش ساده طراحی شد. آزمایش به این صورت بود که آزمون‌دهنده ابتدا دستورالعمل بازی انتخاب شده را مطالعه می‌کرد و سپس یکی از آنها را انتخاب می‌کرد. هر بازی به سه بخش مجزا تقسیم شده بود که بخش اول شامل مراحل ساده، بخش دوم شامل مراحل کمی سخت‌تر و بخش سوم شامل مراحل بسیار سخت بود. از آزمون‌دهنده خواسته می‌شد که بخش‌های مختلف را به ترتیب بازی کند و پس از اتمام هر بخش استراتژی‌های مورد استفاده‌ی خود را یادداشت کند. در انتهای بازی نیز پرسشنامه‌ای راجع به ویژگی‌های فردی خود را تکمیل می‌کرد. (پرسشنامه در پیوست آمده است) در این آزمون ۵ نفر در بازی ابرباران‌زا، ۷ نفر در بازی سنگ- کاغذ- قیچی، ۸ نفر در بازی مسافرخانه زنبوری و ۴ نفر در بازی کتابخانه شرکت کردند. (عددها رو درست کنم) در بازی ابرباران‌زا ۱۰ استراتژی، در بازی سنگ- کاغذ- قیچی ۲ استراتژی، در بازی مسافرخانه زنبوری ۳ استراتژی و در بازی کتابخانه ۴ استراتژی در مجموع گزارش شد. با توجه به نتایج به دست آمده دیده می‌شود که بازی ابرباران‌زا تنوع استراتژی بالاتری دارد و برای اهداف ما در این پژوهش مناسب‌تر است. بنابراین در نهایت بازی ابرباران‌زا انتخاب شد.

۳.۳ طراحی آزمایش اصلی

آزمایش از دو بخش اصلی تشکیل شده است که بخش اول پیش‌نیاز بخش دوم است. بخش اول مربوط به استخراج استراتژی‌ها و بخش دوم مربوط به انتقال استراتژی است. در بخش اول دو هدف وجود دارد. اول اینکه مجموعه‌ای از استراتژی‌های مورد استفاده‌ی افراد در بازی مورد نظر استخراج شود و دوم اینکه یک رده‌بندی برای استراتژی‌های موجود استخراج شود. به این معنا که مشخص شود کدام استراتژی(ها) به صورت میانگین کارایی بهتری داشتند. در بخش دوم هدف اصلی این است که تاثیر انتقال استراتژی سنجیده شود. به این معنا که استراتژی به افرادی که عملکرد ضعیفی داشتند منتقل می‌شود و میزان اثرگذاری آن سنجیده می‌شود.

۱.۳.۳ بخش اول - استخراج استراتژی

ساختار آزمون

آزمونی که در این بخش طراحی شد عمدتاً مشابه آزمونی بود که برای انتخاب بازی انجام دادیم ولی چند تفاوت عمده داشت. اولین تفاوت مهم این بود که بازی را به ۷ بخش تقسیم کردیم. بخش اول مرحله‌ی بودند که یک ابر بارانی داشتند، بخش دوم دو ابر بارانی داشتند و الی آخر. آزمون به این صورت است که ابتدا آزمون‌دهنده یک بخش را کامل بازی می‌کند و سپس استراتژی‌های خود در آن بخش را یادداشت می‌کند. بازی از نظر زمانی محدود است. هر فرد ۱۰ دقیقه برای انجام بازی فرصت دارد. نحوه‌ی انتقال بین مراحل به این صورت است که اگر آزمون‌دهنده تمام ابرهای باران‌زا را به درستی تشخیص دهد به مرحله‌ی بعدی می‌رود. اما اگر حتی یکی از آنها را اشتباه انتخاب کند در همان مرحله باقی می‌ماند. محدودیت تکرار هر مرحله ۲۰ بار است. یعنی اگر فردی بعد از ۲۰ بار تکرار یک مرحله نتواند آن را با موفقیت پشت سر بگذارد امتیازی از آن مرحله به او تعلق نخواهد گرفت. بازی در مجموع ۴۲ مرحله است. تعداد مراحل در هر بخش در جدول ۱.۳ نمایش داده شده است.

بین هر دو بخش توقفی وجود دارد تا آزمون‌دهنده فرصت داشته باشد استراتژی‌های خود را بنویسد.

جدول ۱.۳: تعداد مراحل هر بخش

تعداد مراحل	شماره بخش
۴	۱
۵	۲
۶	۳
۶	۴
۷	۵
۷	۶
۷	۷

در نهایت پس از اتمام زمان از آزمون دهنده تقاضا می شود پرسشنامه ی اطلاعات فردی را تکمیل کند. آزمون با استفاده از نرم افزار adobe flash cs6 و با استفاده از زبان برنامه نویسی 3 action script طراحی شد. برای اجرای آزمون از یک لپ تاپ (Lenovo ThinkPad E460) استفاده شد و شرکت کنندگان با استفاده از نشانگر جواب خود را انتخاب می کردند.

ثبت داده

در این مرحله اطلاعات را با استفاده از دو ابزار مختلف ثبت می کنیم. ابزار اول استفاده از اطلاعات ثبت شده از نحوه ی بازی کردن آزمون دهنده است. به ازای هر مرحله این اطلاعات شامل مکان ابرهای بارانی و عادی، تعداد ابرهایی که به درستی انتخاب شدند، تعداد ابرهایی که به اشتباه انتخاب شدند، مکان نشانگر در هر لحظه و ویژگی های آن مرحله از بازی است. ابزار دیگری که برای ثبت اطلاعات از آن استفاده کردیم یک دستگاه ردیاب چشم بود. (توضیح ویژگی های دستگاه) هدف استفاده از این دستگاه ثبت نقطه ی نگاه آزمون دهنده و تطبیق آن با استراتژی های گزارش شده توسط وی بود.

شرکت کنندگان

در این مرحله، آزمون به صورت یک مسابقه برگزار شد. مجموعاً ۵۷ نفر در آزمون شرکت کردند که از بین آنها اطلاعات ۴۵ نفر با توجه به پرسشنامه ها قابل استفاده بود. به عنوان جایزه به دو نفری که بیشترین امتیاز را کسب کردند یک فلش مموری با ظرفیت ۳۲ گیگابایت داده شد و به دو نفر نیز به قید قرعه یک فلش

مموری با ظرفیت ۱۶ گیگابایت داده شد.

نحوه محاسبه امتیاز

دو معیار برای محاسبه‌ی امتیاز اهمیت دارند. اولین معیار آخرین مرحله‌ای است که شرکت‌کننده موفق شده به آن برسد و معیار دوم میزان توقف وی در مراحل دیگر است. به عنوان مثال فردی که توانسته همه‌ی مراحل را با یک بار بازی کردن پشت سر بگذارد و تا مرحله‌ی ۳۰ جلو برود باید امتیاز بیشتری از فردی بگیرد که تا مرحله‌ی ۳۰ جلو رفته اما هر مرحله را ۲ بار انجام داده است. علاوه بر این هزینه‌ی خطاها در مراحل بالاتر بیشتر است. به این معنی که فردی که مرحله‌ی ۱ را ۵ بار تکرار کرد امتیاز بیشتری می‌گیرد نسبت به فردی که مرحله‌ی ۲۰ را ۵ بار تکرار کرده است (با فرض اینکه بقیه‌ی مراحل را مشابه هم بازی کرده باشند). با توجه به این موضوع معیار امتیاز دهی را به این صورت تعیین کردیم که شماره‌ی مرحله ضرب امتیاز آن مرحله باشد و تعداد تکرارهای هر مرحله از ۲۱ کم می‌شود و در ضرب آن مرحله ضرب می‌شود. در نهایت امتیاز همه‌ی مراحل با هم جمع می‌شوند.

$$score = \sum_{level=1}^{lastLevelReached} level(21 - numOfLevelRepeat) \quad (1.3)$$

فصل ۴

ارزیابی

در این فصل به ارزیابی روش‌های ارائه شده در شناسایی هرزوب می‌پردازیم. بدین منظور، ابتدا به معرفی مجموعه‌های داده‌ای مورد استفاده در این پژوهش پرداخته و سپس، معیارهایی را که برای ارزیابی هر یک از روش‌های معرفی شده در این پژوهش استفاده شده‌اند معرفی می‌نماییم. در انتها نیز نتایج آزمایش‌هایی را که در راستای انجام این پژوهش صورت گرفته است ارائه داده و با سایر روش‌های پیشین مربوطه مقایسه می‌نماییم.

فصل ۵

جمع‌بندی و نکته‌های پایانی

در این پژوهش، سعی کردیم تاثیر انواع ویژگی‌های محتوایی را در شناسایی وب‌گاه‌های هرز فارسی بررسی کرده و یک روش با کارایی بالا برای شناسایی این نوع از وب‌گاه‌ها ارائه دهیم. همچنین، با توجه به خصوصیات خاص وب‌گاه‌های هرز، دو روش مبتنی بر پیوند برای رتبه‌بندی وب‌گاه‌ها ارائه دادیم. در نهایت برای بهبود کیفیت رتبه‌بندی، روشی را معرفی کردیم که از ویژگی‌های محتوایی نیز در کنار ویژگی‌های پیوندی استفاده می‌کند. در این فصل، ابتدا مرور مختصری بر دستاوردهای این پژوهش داشته و در ادامه، به منظور بهبود و گسترش این پژوهش، پیشنهادهایی را برای کارهای آینده ارائه می‌دهیم.

مراجع

- [1] M. Melby-Lervåg and C. Hulme, "Is working memory training effective? a meta-analytic review.," *Developmental Psychology*, vol.49, no.2, p.270–291, 2013.
- [2] T. S. Redick, Z. Shipstead, T. L. Harrison, K. L. Hicks, D. E. Fried, D. Z. Hambrick, M. J. Kane, and R. W. Engle, "No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study.," *Journal of Experimental Psychology: General*, vol.142, no.2, p.359–379, 2013.
- [3] A. Donker, H. D. Boer, D. Kostons, C. D. V. Ewijk, and M. V. D. Werf, "Effectiveness of learning strategy instruction on academic performance: A meta-analysis," *Educational Research Review*, vol.11, p.1–26, 2014.
- [4] A. B. Morrison, G. M. Rosenbaum, D. Fair, and J. M. Chein, "Variation in strategy use across measures of verbal working memory," *Memory & Cognition*, vol.44, p.922–936, Jan 2016.
- [5] E. W. Twamley, G. N. Savla, C. H. Zurhellen, R. K. Heaton, and D. V. Jeste, "Development and pilot testing of a novel compensatory cognitive training intervention for people with psychosis," *American Journal of Psychiatric Rehabilitation*, vol.11, p.144–163, Oct 2008.
- [6] S. M. Jaeggi, M. Buschkuhl, J. Jonides, and P. Shah, "Short- and long-term benefits of cognitive training," *Proceedings of the National Academy of Sciences*, vol.108, no.25, pp.10081–10086, 2011.

Abstract

In recent years, due to the increasing amount of data available on the internet, the use of search engines to retrieve relevant information from the World Wide Web has become pervasive. Among the huge number of websites, the ones which succeed to appear more frequently and in higher ranks of search engine results would receive more visitors. So, spammers struggle to achieve a higher than deserved rank for their websites using some illegal techniques called web spamming. Although various methods have been used for combatting web spamming, we could basically categorize them into three groups: content-based methods, link-based methods, and the methods based on miscellaneous data. In this thesis, we focus on content-based and link-based methods, and also their combination.

Despite the existence of many spam detection methods, the search engines do not perform well in detecting Persian spam websites. Thus, in this thesis, after preparing a corpus of spam and non-spam Persian websites, we analyze the effectiveness of many previously proposed content-based features on detecting Persian spam websites. To improve the performance of classification, we present a number of new content-based features and examine a number of feature selection method. As another approach, we propose a new Persian spam detection system which uses an improved version of bag-of-words model and has better performance in detecting Persian web spam. Due to the prevalence of link-based spamming methods, we analyze some of these methods and propose two new algorithms which do not have the weaknesses of previous methods. In the first algorithm, to improve the process of label propagation, we use three mechanisms: optimized seed selection, edge weighting, and seed expansion. In the second algorithm, we improve the quality of websites ranking, using label propagation in both forward and backward directions. Finally, we propose a combined method, which uses the content-based probability of being spam (non-spam) to propagate the spam (non-spam) score of websites. Using this method, we increase the performance of ranking websites.

Finally, to evaluate the proposed methods and compare their performance with the existing methods for this task, we have conducted several experiments on different datasets. Experiment results indicate that the proposed methods have a good performance in detecting web spam.

Keywords: *Spamming, Web Spam, Spam Detection, Label Propagation, Content-Based Features*



University of Tehran
School of Electrical and Computer Engineering

Detecting Persian Spam Web Pages

By
Elahe Rabbani

Supervisors:
Dr. Azadeh Shakery

A thesis submitted to the Graduate Studies Office
in partial fulfillment of the requirements
for the degree of Master of Science

in

Computer Engineering

September 2014