

۱- طرح موضوع

امروزه در حوزه پردازش داده و یادگیری ماشین با حجم وسیعی از توانمندی‌ها و پژوهش‌ها در کشور روبرو هستیم که با شناخت روزافزون این شاخه از علم، تعداد آنها رو به افزایش خواهد بود. از این رو، این نیروی عظیم خبره را می‌توان از طریق طرح مناسب موضوع به سمت حل مسائلی پیش برد که زمینه کاربرد و بازار وسیعی دارند. هدف خاص ستاد در این چالش حل مسائلی است که از ترکیب داده‌های بزرگ، یادگیری ماشین و علوم شناختی حاصل می‌گردد.

از طرف دیگر از این نکته نمی‌توان غافل شد که داده‌های اجتماعی، فرهنگی و اقتصادی روزانه در حال افزایش هستند و بازار استفاده از آنها در سراسر دنیا حجم قابل توجهی را شامل می‌شود. مثال‌های متعددی از این دست وجود دارند که برخی از آنها همانند ترجمه‌های ماشینی بر اساس یادگیری عمیق، بدون در نظر گرفتن شاخصه‌های شناختی و تنها متکی بر روش‌ها و الگوهای یادگیری ماشین انجام می‌گیرد و کیفیت خروجی نیز قابل عرضه است. یکی از اهداف ستاد ترکیب این مسائل با علوم شناختی در جهت ارتقای نتیجه است.

به عنوان نمونه می‌توان به حوزه هوش مصنوعی اشاره کرد که تا چندی پیش هیچ کس آن را جدی نمی‌گرفت، در صورتی که امروزه همه جا به گونه‌ای اغراق‌آمیز صحبت از هوش مصنوعی است. حال باید دید آیا در خصوص علوم شناختی نیز چنین اتفاقی رخ خواهد داد یا خیر. آنچه مسلم است افزایش قدرت پردازش و حجم داده با سرعتی بسیار چشم‌گیر است که می‌تواند در روند پیشرفت این علوم تأثیر بسزایی داشته باشد. به طور مثال یکی از موارد موفق هوش مصنوعی در سال‌های اخیر ترجمه ماشینی است. بدین صورت که یک جمله انگلیسی نه چندان ساده به مترجم گوگل داده می‌شود و بلافاصله معادل صحیح آن در زبان مورد نظر دریافت می‌گردد. این امر که مترجم گوگل از چه ترفندی استفاده نموده است و شیوه کار آن به شیوه کار مغز شباهت دارد یا خیر، اهمیت چندانی ندارد و تنها شیوه صحیح و هوشمندانه انجام این کار برای ما کافی است. با این وجود، این مسأله برای بسیاری افراد زمانی جالب توجه است که این رفتارهای به ظاهر هوشمندانه به فرایندهای شناختی و سازوکارهای مغزی شباهت بیشتری داشته باشند.

نمونه بارز این مسأله رویکرد یادگیری عمیق^۱ است که با وجود فاصله بسیار با مشابه زیستی آن، این روزها طرفداران بسیاری پیدا کرده است. البته یادگیری عمیق تنها یکی از شاخه‌های یادگیری ماشین است و تاکنون برتری خاصی نسبت به روش‌های هم‌نوع خود نداشته است؛ اما این امر که مقوله یادگیری که از مبانی علوم

^۱ Deep Learning

شناختی است، تا این اندازه در فناوری‌های روز مطرح شده است، پیشرفت قابل توجهی است. مقوله یادگیری در حوزه هوش مصنوعی قدیم و نمادین^۲ دغدغه اصلی نبود. همچنین یادگیری به شیوه شبکه‌های عصبی نیز تا دو دهه پیش تا این اندازه پرتعداد نبود، اما امروزه از نظر بسیاری افراد به شیوه غالب تبدیل شده است. در واقع این فرضیه قوت گرفته که روش‌های معمول هوش مصنوعی در بسیاری حوزه‌ها ممکن است در کوتاه‌مدت جوابگو باشد اما در درازمدت پیچیدگی فرایندهای شناختی و طبیعی چالش‌های دشوارتری می‌طلبد که نیازمند شناخت بیشتر سازوکارهای مغزی است.

به عنوان نمونه می‌توان به مسأله تبدیل یک تصویر به متنی اشاره کرد که حاوی اطلاعات آن تصویر است. کاری که در بسیاری موارد برای انسان به طور طبیعی ساده است اما برنامه‌هایی که در قالب یادگیری عمیق طراحی شده‌اند، هنوز مرتکب اشتباهات فاحشی می‌گردند. به نظر می‌رسد قوه توجه در انسان در شناسایی و استخراج اطلاعات مفید تصویر سهم مهمی دارد و به همین علت است که نحوه پیاده‌سازی این توانایی به مسأله‌ای داغ در یادگیری عمیق تبدیل شده است. البته لازم به ذکر است که پیشرفت فناوری بر حسب شاخص‌های اقتصادی و به اصطلاح معیارهای بازار تعیین می‌گردد. به عنوان مثال اگر قرار بر این بود که مترجم ماشینی گوگل ابتدا معنی جمله را متوجه شود و سپس آن را ترجمه کند، فناوری امروز به ثمر نمی‌رسید. اما مباحث شناختی و معناکاوی به همان اندازه مفید است و پیشرفت فناوری ترجمه ماشین در آینده نزدیک بر چنین چالش‌هایی متکی است. علاوه بر اینها ویژگی بارز شناخت در انسان دسترسی مستقیم به احوالات مغزی است و همین اساس همدلی است که انسان را نسبت به هیجانات و عواطف هم‌نوع خود حساس می‌کند. این روزها ابزارهایی که باعث برانگیختگی احساسات می‌شود، کاربرد فراوانی دارند. به ویژه شایعاتی که سازوکارهای مغزی عواطف را فعال می‌کند، بیشترین تأثیر را در انتخاب و تصمیم‌گیری ما دارند. چه انتخاب سهام باشد، چه ارزیابی کالای جدید یا نظرخواهی مشتریان و ...، سمت و سوی تجزیه و تحلیل داده‌های موجود بدون در نظر گرفتن یافته‌ها و چالش‌های شناختی در این حوزه میسر نیست.

۲- مواردی از حضور آشکار علوم شناختی در داده کاوی هوشمند

از نظر خیلی‌ها ترکیب علوم شناختی و علوم داده در آینده نزدیک به قطب جدیدی از فعالیت در حوزه داده کاوی هوشمند تبدیل می‌شود. سوال اصلی این است که داده کاوی هوشمند با توجه به ساختار فعلی تا چه اندازه می‌تواند الگوهای رفتاری و فکری انسان را شبیه‌سازی کند. به لحاظ رفتاری شاید الگوهای تصمیم‌گیری بیشترین اهمیت را داشته باشد و خیلی مهم است که این الگوهای تصمیم‌گیری بر چه اساسی تغییر می‌کند. آن چه امروز به فناوری

^۲ Symbolic

اقتناع^۳ موسوم شده ارتباط نزدیکی با فرآیندهای شناختی تاثیر گزار در اقتصاد رفتاری دارد. به طور مثال تاثیر عوامل هیجانی-احساسی از موضوعات داغ تحقیق در اقتصاد رفتاری است و فناوری اقتناع به طور محسوسی تاثیر این عوامل را در تصمیم گیری هدف قرار داده است. تاثیر این عوامل از داده های تصویری گرفته تا صوتی و متنی بسیار گسترده است. یک نمونه از چنین کاربردی روش هایی است که احساسات را از روی حرکات صورت اندازه گیری می کند و از استخراج چنین داده هایی میزان ذوق زدگی نسبت به تبلیغات یک محصول و یا میزان احساسی که مردم نسبت به گفتار یک مبلغ سیاسی نشان میدهند را ارزیابی میکند. همین طور نقش شایعات و تاثیر آن در برانگیختن احساسات و نهایتا اتخاذ تصمیمات از قبل طراحی شده حتی به صورت نا آگاه می تواند چالشی فرا گیر در داده کاوی هوشمند تلقی شود. این نگرش روزنه کوچکی ایست به حوزه مهمی از علوم شناختی تحت عنوان هوشمندی احساسی و شامل توسعه ابزار هایی است که بتواند چنین اطلاعاتی را از داده های موجود استخراج کند. از سوی دیگر، این ابزارها مکمل فناوری شناخته شده ایست که به سیستم تحلیل احساسات^۴ موسوم است و با داده های مربوط به حال و هوای کاربران در مواردی چون رضایت مشتریان یا مرورگران کتاب ها سرو کار دارد. این تحلیل ها که عمدتاً با داده های متنی سرو کار دارد به طور روز افزونی مبتنی ایست بر پردازش و فهم زبان طبیعی. اما با وجود همه کوشش ها هنوز تا تحلیل معنایی و معنایابی فاصله زیادی باقی است. در حال حاضر شرکت هایی چون گوگل، فیس بوک یا تویتر در زمینه سلیقه یابی کاربران موفقیت زیادی به دست آورده اند. داده های مقیاس بزرگ این امکان را فراهم کرده که بتوان کالاهای مشابه و افراد مشابه را با معیار های از قبل تعیین شده دسته بندی کرد و از این طریق رابطه معقولی بین افراد و کالای مورد علاقتان به دست آورد. به عبارتی محتوای علایق و تمایلات کاربران به طور خود کار قابل تنظیم و انتخاب است. این فناوری به اصطلاح محتوا یابی خودکار^۵ به حق باعث بروز نگرانی های بسیاری شده است تا جای که ممکن است اصل اختیار و اقتدار کاربران در اتخاذ تصمیم مورد سوال واقع شود. موفقیت بزرگ داده کاوی هوشمند در یافتن بهینه این الگو های مشابهت بوده اما بیشتر در خدمت سود آوری شرکت ها و کمتر به نفع مصرف کنندگان. چالش بزرگ تری که در همین جهت با معنا یابی در پردازش متن سرو و کار دارد پردازش مبتنی بر بافتار^۶ متن است. بافتار یابی خودکار^۷ هنوز قدم های اولیه را بر میدارد و از مسایل بنیادی هوش مصنوعی و علوم شناختی است. کاری که برای انسان به سادگی قابل انجام است ولی برای ماشین نه. الگوهای تصمیم گیری بر حسب بافتار فرهنگی، اجتماعی، هیجانی همواره تغییر می کند. هر چه تصمیم گیری به بافتار هیجانی بیشتر وابسته باشد کمتر قابل پیش بینی است و به همین

^۳ persuasive technology

^۴ sentiment analysis

^۵ automated content selection

^۶ context

^۷ self-automated context selection

جهت استخراج بافتار هیجانی از داده های متنی در مسایلی چون بازار یابی و رضایت مشتریان اهمیت زیادی دارد. در رویکرد یاد گیری عمیق پیشنهاد شده که آنچه باید به عنوان بافتار در نظر گرفته شود خود موضوع یادگیری باشد. اما مساله به این سادگی نیست و چالش های شناختی زیادی را میطلبد. مطلب دیگر اهمیت دادن به موضوع سازو کار توجه است. هنوز سطح دانش عموم از سازو کار توجه از آنچه که طی چند ده اخیر در حوزه علوم شناختی به دست آمده بسیار اندک است. فقط این نیست که استفاده از تلفن همراه هنگام رانندگی چه خطراتی به دنبال دارد. میتوان میزان تمرکز کاربران را حین دنبال کردن متون اینترنتی مورد ارزیابی قرار داد. بزودی تلفن های همراه اطلاعات خوبی از حرکات چشم و داده های مربوط به آن به دست خواهند داد. تحلیل چنین داده هایی با استفاده از دانش علوم شناختی میتواند اطلاعات مفیدی از سازو کار توجه و به کار گرفتن آن در فناوریهای دیجیتالی در اختیار کاربران قرار دهد. به طور مثال استفاده از این سازوکار در تبدیل اطلاعات تصویری به متن به موضوع داغی در یادگیری عمیق تبدیل شده است. در آخر، یاد آور شویم که این روزها با جستجوی مختصری درباره فناوریهای شناختی، معنا یابی و داده کاوی هوشمند میتوان به موضوعات جالبی از این دست برخورد کرد. نوشته حاضر بخش کوچکی از توان شناختی است و امید است چالشی که تحت عنوان "داده کاوی هوشمند به توان شناختی" آغاز شده فصل تازه ای در فعالیت بین رشته ای علوم داده، علوم شناختی، هوش مصنوعی و یادگیری ماشین ایجاد کند.

۳- تعریف

چالشی که در آن به تفسیر شناختی نتایج تحلیل هوشمند دادگان پرداخته می شود و به وسیله آن می توان کیفیت سرویس ها و محصولات موجود را ارتقاء بخشید، محصولات و خدمات جدیدی ایجاد نمود که ارزش اقتصادی آن مورد توجه بخش های مختلف صنعت و جامعه است، و همچنین به درک بهتری از مسائل اجتماعی، اقتصادی و فرهنگی دست یافت.

منظور از شناختی بودن در چارچوب این چالش ایده یا روشی است که به مدل تصمیم گیری، مدل باور و شبکه مفاهیم و باورها، شناخت لایه های زیرین و پنهان تفکر^۸ می پردازد. ایده یا روش مورد نظر می تواند قابلیت ها یا ویژگی های شناختی را از داده مورد نظر استخراج نماید، نظر کاوی^۹ یا احساس کاوی^{۱۰} کند و به تحلیل نظر، احساس، ارزیابی و نگرش افراد درباره یک سرویس یا موضوع مشخص در داده مورد نظر بپردازد (مخاطب شناسی

^۸ Deep Structure

^۹ Opinion Mining

^{۱۰} Sentiment Analysis – Attitude Analysis

کند؛ یا بتواند موضوعات یا عناصری که توجه افراد را جلب می کند، از متن نظرات یا گفتگوها استخراج نماید. همچنین شامل مسائلی می شود که میزان اطمینان^{۱۱} فرد را از نظری که ارائه داده است، مشخص می نماید.

۳-۱- بانک دادگان در دسترس

- دیجی کالا (اطلاعات خرید، قیمت ها، کامنت ها و ...)
- تپسی (اطلاعات سفرها، رد یا قبول سفر توسط راننده یا مسافر، جابجایی ها و ...)
- سایت دیوار
- داده های فضای مجازی (تلگرام، توئیتر، اینستاگرام)
- سایر (به انتخاب شرکت کنندگان)

۳-۲- تعریف مسأله

تعریف مسأله در این چالش به انتخاب شرکت کنندگان است. تیم می تواند با استفاده از یکی از دادگان فوق یا ترکیب آنها طرح مسأله شناختی نماید و یا از بانک دادگان خود استفاده کند.

۳-۲-۱. معیارهای ورود تیم ها به چالش

شرکت های دانش بنیان، اساتید و پژوهشگران دانشگاه ها و موسسات آموزش عالی، مخترعان و گروه های دانشجویی می توانند در این چالش شرکت کنند. تیم های شرکت کننده بایستی:

۱- دارای اعضای مسلط به حوزه علوم داده و داده کاوی هوشمند باشد. بدیهی است تیم های بین رشته ای و آشنا با علوم شناختی مرجح است.

۲- دادگان را به انتخاب خود و یا از داده هایی که از طریق سایت در دسترس است انتخاب کنند.

¹¹ Confidence

۳-۲-۲. فرایند ارزیابی تیم‌ها

طرح ارزیابی بر اساس تست پلن خواهد بود که توسط تیم داوری طراحی و ارائه می‌شود. در این تست پلن موارد زیر مورد ارزیابی قرار خواهند گرفت:

- گزارش ارائه شده بایستی از متدولوژی مناسب علمی برخوردار باشد.
- اثربخشی نتایج از یک یا چند منظر زیر باعث کسب امتیاز بیشتر خواهد شد:
 - ✓ اقتصادی و کارآفرینی؛
 - ✓ راهبردی؛
 - ✓ دانش و فناوری؛
 - ✓ برتری نسبت به راه‌حل‌های موجود.
- نوآوری از منظر متدولوژی یا تعریف مسأله موجب کسب امتیاز بیشتر خواهد شد.

شرکت‌کنندگان می‌توانند برای ثبت نام در پیش رویداد، حداکثر تا تاریخ 25 مهر 98 از طریق سایت Hamava.ir/cdm/ اقدام نمایند. در این پیش رویداد نیم روزه که 1 آبان 98 در هم‌آوا در کارخانه نوآوری آزادی برگزار خواهد شد، شرکت‌کنندگان با اهداف و چگونگی چالش آشنا می‌گردند. پس از آن تا 15 آبان 98 فرصت دارند پروپوزال مورد نظر را طبق فرمتی که اعلام می‌شود ارسال نمایند.

به تیم‌های برگزیده علاوه بر جوایز نقدی، فرصت استفاده از امکانات شتابدهی توسط بخش شتابدهی هم‌آوا و جذب سرمایه‌گذار داده خواهد شد.

اطلاعات تکمیلی در سامانه Hamava.ir/cdm/ قرار گرفته است. علاقه‌مندان می‌توانند در صورت داشتن هرگونه پرسش از طریق ایمیل cdm@hamava.ir با ما در ارتباط باشند.

چالش داده کاوی هوشمند به توان شناختی:

معنایابی در داده های ایرانی

۳-۲-۳. برگزارکنندگان و حامیان

