

دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

دانشکده مهندسی کامپیوتر و فناوری اطلاعات

طرح سخنرانی پایانی

پرسش و پاسخ بصری با استفاده از شبکه های عصبی کانولوشنی و بازگشتی عمیق

ارائه دهنده:

علی غلامی

استاد راهنما:

دکتر محمد رحمتی

هدف خاص سخنرانی

درک چگونگی عملکرد ماشین در فهم تصاویر دنیای واقعی و نیز جملات زبان طبیعی و پاسخ متقابل به سوالات مرتبط با آن تصاویر.

موضوع اصلی سخنرانی

آشنایی با کاربردهای بینایی ماشین در زمینه پرسش و پاسخ بصری و بررسی معماری و زیر ساخت شبکه عصبی مورد نیاز جهت پاسخگویی به سوالات مطرح در مورد تصاویر.

عنوان سخنرانی

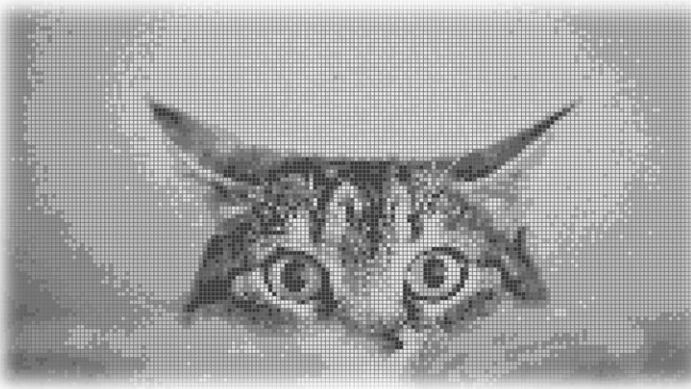
صحبت های یک ماشین درباره جهان اطراف!

1. مقدمه

1.1. آنچه انسان می بیند:



1.2. آنچه ماشین می بیند:

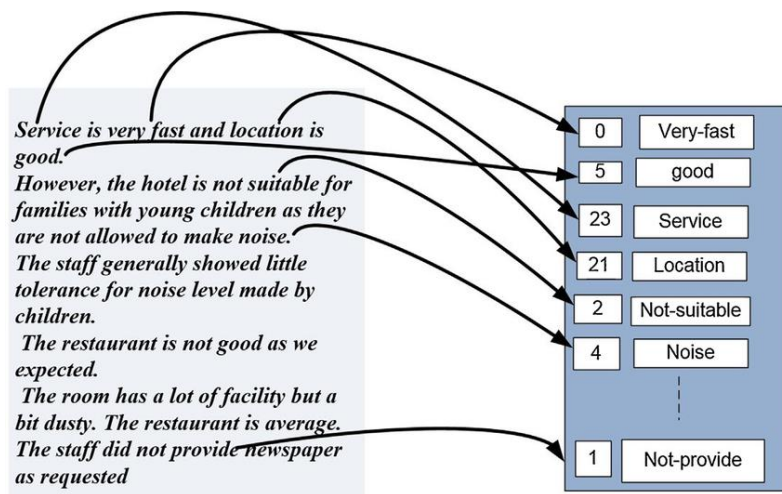


1.3. آنچه انسان می شنود:



(پخش صدای انسان)

1.4. آنچه ماشین می شنود:



1.5. ولتاژ به منزله ی 0 و 1.

1.6. قصد داریم با این زبان با ماشین ها صحبت کنیم و تصاویر و جملات را به آنها بفهمانیم.

2. متن

- 2.1. برای ماشین ها، تصاویر به شکل ماتریس تداعی می شوند.
 - 2.1.1. این ماتریس ها شامل خانه هایی هستند به نام پیکسل.
 - 2.1.2. آنچه برای انسان قابل درک است، با اعداد متفاوت در ماتریس ها نمایش داده می شود. این اعداد لزوماً بزرگ نیستند و ممکن است مقادیر کم گویای مطالب مهم باشند.
 - 2.1.3. می توان از حاصل ضرب ماتریس های از پیش تعیین شده (فیلتر ها)، مفاهیم مورد انتظار را برای کاربرد های خاص استخراج کرد.
 - 2.1.3.1. این مفاهیم باید از قبل تعیین شده باشند بنابراین کاربرد های محدودی دارند.
 - 2.1.3.2. راه حل جدید تر استفاده از شبکه های عصبی عمیق است.
- 2.2. شبکه های عصبی عمیق قابلیت یادگیری مفاهیم را دارند.
 - 2.2.1. تصاویر به عنوان ورودی به این شبکه ها خورنده می شوند و آنها با مکانیزمی ریاضی مفاهیم را یاد می گیرند.
 - 2.2.2. به عنوان مثال اگر هدف طراحی یک سیستم تشخیص چهره باشد، کافی است تصاویر متعدد از حالت های مختلف چهره انسان به این شبکه ها داده شود تا آنها یاد بگیرند که چهره انسان چه ویژگی هایی دارد.
- 2.3. می توان برای متن نیز همین مفهوم را پیاده سازی کرد.
 - 2.3.1. به شبکه های عصبی مورد استفاده در کاربرد متن، RNN یا شبکه عصبی بازگشتی عمیق گفته می شود.
- 2.4. ایده ی شبکه های عصبی از مغز می آید.
 - 2.4.1. در مغز، نرون ها واحد های پردازشی هستند که در صورت ورود اطلاعات آشنا برای انسان، برانگیخته می شوند.
 - 2.4.2. این برانگیختگی به صورت سریال در طول تمام این نرون ها جابجا می شود و نهایتاً به مفهوم خاصی منجر می شود.
 - 2.4.3. به عنوان مثال زمانی که انسان چهره ی فردی را که می شناسد می بیند، پس از فعال شدن حدود 2 میلیون نرون، فرد مورد نظر تشخیص داده می شود.
 - 2.4.4. این مفهوم در سیستم شنیدار انسان نیز پیدا است.
 - 2.4.5. می توان از نحوه ی عملکرد مغز برای بهبود فهم و درک ماشین ها کمک گرفت. اما پیش نیاز مهمی وجود دارد و آن مدلسازی ریاضی است!
- 2.5. یکی از کاربرد های بروز و جدید شبکه های عصبی کانولوشنی و بازگشتی عمیق پرسش و پاسخ بصری است.
 - 2.5.1. در پرسش و پاسخ بصری، تصاویر متعددی (موضوع محدود) به این شبکه های خورنده می شود.

2.5.2. در ادامه با زبان طبیعی نیز از ماشین سوال پرسیده می شود. این سوال مرتبط با آنچه که در تصویر وجود دارد است.

2.5.3. ماشین با استفاده از مکانیزم تعریف شده، تصویر و جملات را هضم کرده و پاسخ برای آن تولید می کند.

2.5.4. اندازه گیری دقت پاسخ های تولید شده توسط ماشین یکی از چالش های بزرگ در این امر می باشد.

2.6. آشنایی با اسامی معماری های مختلف شبکه های عصبی کاربردی در تصاویر و زبان طبیعی.

2.7. کاربرد های پرسش و پاسخ بصری در صنعت و بهبود زندگی انسان

3. نتیجه گیری

4. تشکر

5. مراجع مورد نظر