



دانشکده مهندسی کامپیوتر

پرسش و پاسخ تصویری

گزارش سمینار برای دریافت درجه کارشناسی ارشد در رشته مهندسی کامپیوتر
گرایش هوش مصنوعی

مریم سادات هاشمی

استاد راهنما

سید صالح اعتمادی

آبان ۱۳۹۹



فهرست مطالب

۱	فصل ۱:
۱-۱	بررسی کلی و تعریف مبحث پرسش و پاسخ تصویری
۱-۲	کاربرد و اهمیت این مسئله
۱-۳	بررسی چالشهای موجود در این مسئله
۱-۴	بررسی مجموعه دادگان مطرح و مسابقات مطرح این حوزه
۱-۴-۱	مجموعه داده DAQUAR
۱-۴-۲	مجموعه داده VQA
۱-۴-۳	مجموعه داده Madlibs Visual
۱-۴-۴	مجموعه داده wV Visual
۱-۴-۵	مجموعه داده CLEVR
۱-۴-۶	مجموعه داده Tally-QA
۱-۴-۷	مجموعه داده KVQA
۱-۵	بررسی فازهای مختلف مسئله پرسش و پاسخ تصویری
۱-۵-۱	فاز ۱: استخراج ویژگی از تصویر و سوال
۱-۵-۲	فاز ۲: درک مشترک تصویر و سوال
۱-۵-۳	فاز ۳: تولید جواب
۱-۶	معیارهای ارزیابی مسئله پرسش و پاسخ تصویری
۱-۷	چگونگی ساخت مجموعه داده حاوی پرسش و پاسخ به زبان فارسی

فصل ۱

۱-۱ بررسی کلی و تعریف مبحث پرسش و پاسخ تصویری

باید تکمیل شود.

۲-۱ کاربرد و اهمیت این مسئله

باید تکمیل شود.

۳-۱ بررسی چالشهای موجود در این مسئله

باید تکمیل شود.

۴-۱ بررسی مجموعه دادگان مطرح و مسابقات مطرح این حوزه

باید تکمیل شود.

۱-۴-۱ مجموعه داده DAQUAR

DAQUAR مخفف Dataset for Question Answering on Real World Images است که توسط مالدینوفسکی منتشر شده است. این اولین مجموعه داده ای است که برای مسئله VQA منتشر شده است. تصاویر از مجموعه داده

NYU-Depth V2 گرفته شده است. اندازه این مجموعه داده کوچک است و در مجموع ۱۴۴۹ تصویر دارد. DAQUAR شامل ۱۲۴۶۸ زوج پرسش و پاسخ با ۲۴۸۳ سوال منحصر به فرد است. برای تولید پرسش و پاسخ ها از دو روش مصنوعی و انسانی استفاده شده است. در روش مصنوعی پرسش و پاسخ ها به صورت خودکار از الگوهای موجود در جدول فلان تولید شده است. در روش دیگر از ۵ نفر انسان خواسته شده است تا پرسش و پاسخ تولید کنند. تعداد پرسش و پاسخ های آموزشی در این مجموعه داده ۶۷۹۴ و تعداد پرسش و پاسخ های تست ۵۶۴ است و به طور میانگین برای هر عکس تقریباً ۹ پرسش و پاسخ وجود دارد. این مجموعه داده با مشکل بایاس روبه رو است زیرا تصاویر این مجموعه تنها مربوط به داخل خانه است و بیش از ۴۰۰ مورد وجود دارد که اشیایی مثل میز و صندلی در پاسخ ها تکرار شده است.

۱-۴-۲ مجموعه داده VQA

باید تکمیل شود.

۱-۴-۳ مجموعه داده Madlibs Visual

باید تکمیل شود.

۱-۴-۴ مجموعه داده wV Visual

باید تکمیل شود.

۱-۴-۵ مجموعه داده CLEVR

باید تکمیل شود.

۱-۴-۶ مجموعه داده Tally-QA

باید تکمیل شود.

۱-۴-۷ مجموعه داده KVQA

باید تکمیل شود.

۱-۵ بررسی فازهای مختلف مسئله پرسش و پاسخ تصویری

باید تکمیل شود.

۱-۵-۱ فاز ۱: استخراج ویژگی از تصویر و سوال

باید تکمیل شود.

۱-۵-۲ فاز ۲: درک مشترک تصویر و سوال

باید تکمیل شود.

۱-۵-۳ فاز ۳: تولید جواب

باید تکمیل شود.

۱-۶ معیارهای ارزیابی مسئله پرسش و پاسخ تصویری

باید تکمیل شود.

۱-۷ چگونگی ساخت مجموعه داده حاوی پرسش و پاسخ به زبان فارسی

باید تکمیل شود.