Regularizing RNNs for Caption Generation by Reconstructing The Past with The Present –Supplementary Material

1. Image Captioning Examples

As shown in Fig.1, we provide more examples generated by the attentive encoder-decoder models with and without our ARNet. Note that the model fine-tuned with our proposed ARNet gives more detailed and meaningful descriptions, such as the words or phases "`soccer`", "`fireplace`", "`black and white photo`", and so on.

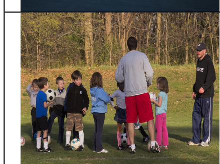
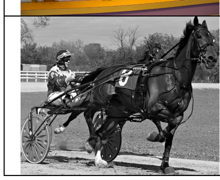
Images	Generated Captions	Ground Truth Captions
	**Attentive Encode-Decoder**: a bus that is sitting in the street. **Attentive Encode-Decoder-AR**: a white bus driving down a street next to a building.	1. a black and white bus some bushes and building. 2. a white decorated bus is next to a building. 3. a large white bus that is by a building. 4. a large bus parked in a parking lot. 5. a white bus driving past a tall building.
	**Attentive Encode-Decoder**: a group of people standing in a field. **Attentive Encode-Decoder-AR**: a group of young children playing with soccer.	1. a man instructing a group of kids on a soccer field. 2. dad coaches talking to the little soccer team players on the field. 3. a soccer coach is instructing the children on the field. 4. a group of young children standing around a field. 5. pair of adult males with group of small children with soccer balls.
	**Attentive Encode-Decoder**: a living room with a flat screen tv. **Attentive Encode-Decoder-AR**: a living room with a television and a fireplace.	1. interior of a living room with furniture, plant, fireplace and a tv. 2. white furniture and fireplace with a tv over it decorate this living room. 3. a chair and a couch in a room. 4. a tv mounted above a fireplace in a nicely furnished living room. 5. a tv sitting above a fire place in a living room.
	**Attentive Encode-Decoder**: a display case filled with lots of donuts. **Attentive Encode-Decoder-AR**: a display case filled with lots of different types of donuts.	1. a bakery with boxes of donuts and bread. 2. a selection of donuts and pastries at an oriental bakery. 3. this is a display of donuts on a couple shelves. 4. assorted bakery goods are on display in a cabinet. 5. fifteen different varieties of doughnuts in a display case.
	**Attentive Encode-Decoder**: two men riding on a horse drawn carriage. **Attentive Encode-Decoder-AR**: a black and white photo of a man riding a horse drawn carriage.	1. a man on a cart racing on a back of a horse. 2. a man is riding in a horse drawn carriage. 3. a man rides behind a horse during a race. 4. the man is driving the horse fast. 5. a black and white photo of a horse running on a track with a man being pulled.

Figure 1. Image captions generated by the attentive encoder-decoder with and without our proposed ARNet, along with their corresponding ground truth captions.

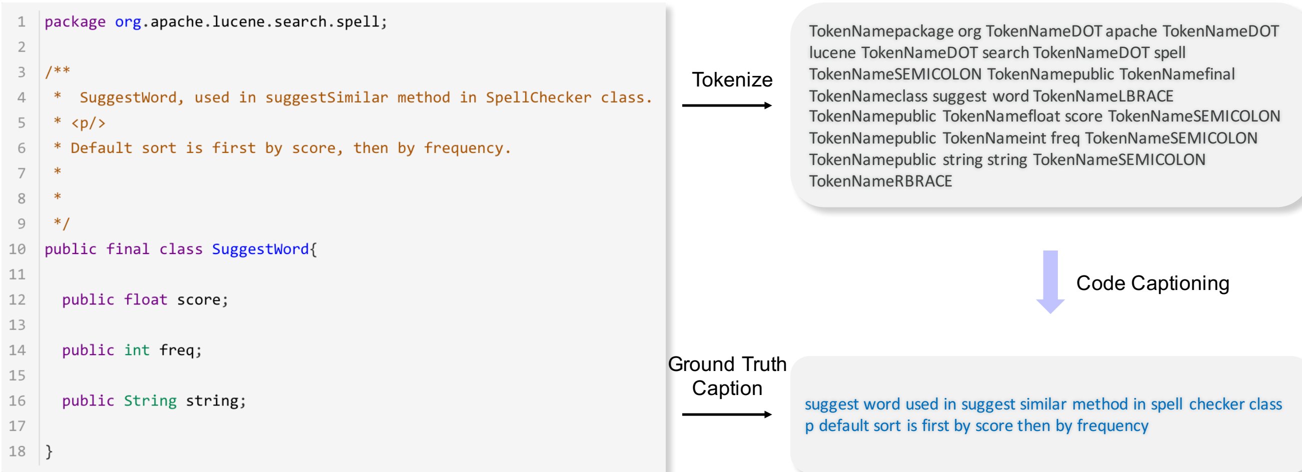


Figure 2. An example of the code captioning task. A source code file is tokenized using the Eclipse JDT compiler tools. Meanwhile, the comment text of this source code file was extracted as ground truth caption. The aim is of this task to generate the meaning of the tokenized source code file.

2. Code Captioning

Code captioning was proposed in ReviewNet which used HabeasCorpus dataset. As illustrated in Fig.2, the aim is to produce a condensed representation of the source code file[1] that captures its core meaning.

2.1. Evaluation Scores

As SPICE and CIDEr metrics are proposed specially for evaluating image captioning results, they are not suitable for the task of code captioning. The performances on the HabeasCorpus dataset are illustrated in Table 1. It can be observed that our proposed ARNet can significantly boost the performance.

Table 1. Performance comparison on the testing split of the HabeasCorpus dataset. The best results among all models are highlighted in boldface.

Model Name	BLEU-1	BLEU-2	BLEU-3
Review Net	0.192	0.105	0.074
Encoder-Decoder	0.183	0.093	0.063
Encoder-Decoder + Zoneout	0.182	0.080	0.063
Encoder-Decoder + Scheduled Sampling	0.186	0.098	0.067
Encoder-Decoder + ARNet	0.196	0.107	0.075
Attentive Encoder-Decoder	0.228	0.140	0.106
Attentive Encoder-Decoder + Zoneout	0.227	0.140	0.105
Attentive Encoder-Decoder + Scheduled Sampling	0.229	0.142	0.108
Attentive Encoder-Decoder + ARNet	**0.255**	**0.173**	**0.139**

2.2. Examples

As shown in Fig. 3, Fig. 4, and Fig. 5, we provide examples of code captioning, with side-by-side comparisons of the ground truth captions and the captions produced by our ARNet.

[1] https://lucene.apache.org/core/3_6_2/api/all/org/apache/lucene/search/spell/SuggestWord.html

Original Source Code (tokenized and truncated): TokenNamepackage org TokenNameDOT apache TokenNameDOT lucene TokenNameDOT analysis TokenNameDOT en TokenNameSEMICOLON TokenNameclass k stem data0 TokenNameLBRACE TokenNameprivate k stem data0 TokenNameLPAREN TokenNameRPAREN TokenNameLBRACE TokenNameRBRACE TokenNamestatic string TokenNameLBRACKET TokenNameRBRACKET data TokenNameEQUAL TokenNameLBRACE TokenNameStringLiteral TokenNameCOMMA TokenNameStringLiteral

---

Reference Caption:
a list of words used by kstem

Attentive Encoder-decoder + ARNet:
a list of words used by kstem

Attentive Encoder-decoder:
this class implements link svgfe func a light stemmer with a link token filter

Figure 3. Code captions generated by attentive encoder-decoder and attentive encoder-decoder-ARNet. Compared to the caption generated by the vanilla attentive encoder-decoder model, the model with our ARNet completely predicts the meaning of this code.

Original Source Code (tokenized and truncated): TokenNamepackage org TokenNameDOT apache TokenNameDOT batik TokenNameDOT dom TokenNameDOT svg00 TokenNameSEMICOLON TokenNameimport org TokenNameDOT apache TokenNameDOT batik TokenNameDOT dom TokenNameDOT abstract document TokenNameSEMICOLON TokenNameimport org TokenNameDOT apache TokenNameDOT batik TokenNameDOT dom TokenNameDOT svg TokenNameDOT svg graphics element TokenNameSEMICOLON TokenNameimport org TokenNameDOT apache TokenNameDOT batik TokenNameDOT util TokenNameDOT sv g00 constants TokenNameSEMICOLON TokenNameimport org TokenNameDOT w0c TokenNameDOT dom TokenNameDOT node TokenNameSEMICOLON TokenNamepublic TokenNameclass svgom flow root element TokenNameextends svg graphics element TokenNameLBRACE TokenNameprotected svgom flow root element TokenNameLPAREN TokenNameRPAREN TokenNameLBRACE TokenNameRBRACE TokenNamepublic svgom flow root element TokenNameLPAREN string prefix TokenNameCOMMA abstract document owner TokenNameRPAREN TokenNameLBRACE TokenNamesuper TokenNameLPAREN prefix TokenNameCOMMA owner TokenNameRPAREN TokenNameSEMICOLON TokenNameRBRACE TokenNamepublic string get local name TokenNameLPAREN TokenNameRPAREN TokenNameLBRACE TokenNamereturn sv g00 constants TokenNameDOT svg flow root tag TokenNameSEMICOLON TokenNameRBRACE TokenNameprotected node new node TokenNameLPAREN TokenNameRPAREN TokenNameLBRACE TokenNamereturn TokenNamenew svgom flow root element TokenNameLPAREN TokenNameRPAREN TokenNameSEMICOLON TokenNameRBRACE TokenNameRBRACE

Reference Caption:
this class implements a regular polygon extension to svg author a hrefmailtothomasdeweesekodakcom thomas de weesea version id svgom flow root elementjava

Attentive Encoder-decoder + ARNet:
this class implements a regular polygon extension to svg author a hrefmailtothomasdeweesekodakcom thomas de weesea version id svgom flow region elementjava

Attentive Encoder-decoder:
this class implements link svg font face element author a hrefmailtostephanehillionorg stephane hilliona version id svgom font face elementjava

Figure 4. Code captions generated by attentive encoder-decoder and attentive encoder-decoder-ARNet. Red denote incorrect tokens in the generated captions. As we can see, the attentive encoder-decoder model with our ARNet achieve more accuracy than the vanilla one.

Original Source Code (tokenized and truncated): TokenNamepackage org TokenNameDOT apache TokenNameDOT batik TokenNameDOT anim TokenNameDOT values TokenNameSEMICOLON TokenNameimport org TokenNameDOT apache TokenNameDOT batik TokenNameDOT dom TokenNameDOT anim TokenNameDOT animation target TokenNameSEMICOLON TokenNameimport org TokenNameDOT w0c TokenNameDOT dom TokenNameDOT svg TokenNameDOT svg angle TokenNameSEMICOLON TokenNamepublic TokenNameclass animatable angle or ident value TokenNameextends animatable angle value TokenNameLBRACE TokenNameprotected TokenNameboolean is ident TokenNameSEMICOLON TokenNameprotected string ident TokenNameSEMICOLON TokenNameprotected animatable angle or ident value TokenNameLPAREN animation target target TokenNameRPAREN TokenNameLBRACE TokenNamesuper TokenNameLPAREN target TokenNameRPAREN TokenNameSEMICOLON TokenNameRBRACE TokenNamepublic animatable angle or ident value TokenNameLPAREN animation target target TokenNameCOMMA TokenNamefloat v TokenNameCOMMA TokenNameshort unit TokenNameRPAREN TokenNameLBRACE TokenNamesuper TokenNameLPAREN target TokenNameCOMMA v TokenNameCOMMA unit TokenNameRPAREN TokenNameSEMICOLON TokenNameRBRACE TokenNamepublic animatable angle or ident value TokenNameLPAREN animation target target TokenNameCOMMA string ident TokenNameRPAREN TokenNameLBRACE TokenNamesuper TokenNameLPAREN target TokenNameRPAREN TokenNameSEMICOLON TokenNamethis TokenNameDOT ident TokenNameEQUAL ident TokenNameSEMICOLON TokenNamethis TokenNameDOT is ident TokenNameEQUAL TokenNametrue TokenNameSEMICOLON TokenNameRBRACE TokenNamepublic TokenNameboolean is ident TokenNameLPAREN TokenNameRPAREN TokenNameLBRACE TokenNamereturn is ident TokenNameSEMICOLON TokenNameRBRACE TokenNamepublic string get ident TokenNameLPAREN TokenNameRPAREN TokenNameLBRACE TokenNamereturn ident TokenNameSEMICOLON TokenNameRBRACE TokenNamepublic TokenNameboolean can pace TokenNameLPAREN TokenNameRPAREN TokenNameLBRACE TokenNamereturn TokenNamefalse TokenNameSEMICOLON TokenNameRBRACE TokenNamepublic TokenNamefloat distance to TokenNameLPAREN animatable value other TokenNameRPAREN TokenNameLBRACE TokenNamereturn TokenNameFloatingPointLiteral TokenNameSEMICOLON TokenNameRBRACE TokenNamepublic animatable value get zero value TokenNameLPAREN TokenNameRPAREN TokenNameLBRACE TokenNamereturn TokenNamenew animatable angle or ident value TokenNameLPAREN target TokenNameCOMMA TokenNameIntegerLiteral TokenNameCOMMA svg angle TokenNameDOT svg angletype unspecified TokenNameRPAREN TokenNameSEMICOLON TokenNameRBRACE TokenNamepublic string get css text TokenNameLPAREN TokenNameRPAREN TokenNameLBRACE TokenNameif TokenNameLPAREN is ident TokenNameRPAREN TokenNameLBRACE TokenNamereturn ident TokenNameSEMICOLON TokenNameRBRACE TokenNamereturn TokenNamesuper TokenNameDOT get css text TokenNameLPAREN TokenNameRPAREN TokenNameSEMICOLON TokenNameRBRACE TokenNamepublic animatable value interpolate TokenNameLPAREN animatable value result TokenNameCOMMA animatable value to TokenNameCOMMA TokenNamefloat interpolation TokenNameCOMMA animatable value accumulation TokenNameCOMMA TokenNameint multiplier TokenNameRPAREN TokenNameLBRACE animatable angle or ident value res TokenNameSEMICOLON TokenNameif TokenNameLPAREN result TokenNameEQUAL EQUAL TokenNamenull TokenNameRPAREN TokenNameLBRACE res TokenNameEQUAL TokenNamenew animatable angle or ident value TokenNameLPAREN target TokenNameRPAREN TokenNameSEMICOLON TokenNameRBRACE TokenNameelse TokenNameLBRACE res TokenNameEQUAL TokenNameLPAREN

Reference Caption:
an svg angleoridentifier value in the animation system author a hrefmailtocam00mcc0eid0eau cameron mc cormacka version id animatable angle or ident valuejava

---

Attentive Encoder-decoder + ARNet:
an svg font value in the animation system author a hrefmailtocam00mcc0eid0eau cameron mc cormacka version id animatable color valuejava

---

Attentive Encoder-decoder:
this class provides a manager for the colorinterpolation property values author a hrefmailtostephanehillionorg stephane hilliona version id messagesjava

Figure 5. Code captions generated by attentive encoder-decoder and attentive encoder-decoder-ARNet. The attentive encoder-decoder model with our ARNet generates most of the true tokens in the ground truth caption successfully and achieves better performance than the vanilla model obviously.