April 15, 2019 - PAKDD 2019

Keyword Extraction with Character-level Convolutional Neural Tensor Networks

Zhe-Li Lin and Chuan-Ju Wang

Research Center for Information Technology Innovation,
Academia Sinica, Taipei, Taiwan
{joli79122, cjwang}@citi.sinica.edu.tw

Outline

- Introduction
- Methodology: charCNTN
- Experiments
- Conclusion
- Future Work

Introduction

- Harry Potter is a series of fantasy novels written by British author J. K. Rowling. The novels chronicle the lives of a young wizard, Harry Potter, and his friends Hermoione Granger and Ron Weasley, all of whom are students at Hogwarts School of Witchcraft and Wizardry.

Introduction

- Harry Potter is a series of fantasy novels written by British author J. K. Rowling. The novels chronicle the lives of a young wizard, Harry Potter, and his friends Hermoione Granger and Ron Weasley, all of whom are students at Hogwarts School of Witchcraft and Wizardry.

Keyword Extraction

- Keyword extraction is the automatic identification of important, representative terms which accurately and concisely capture the main topics of a given document.
- Two kinds of strategies:
  - Supervised approach
  - Unsupervised approach.

Supervised Approach

- Supervised approach
  - Treats keyword extraction as a binary classification
  - E.g. Naive Bayes, decision trees and SVM
  - Contains *unknown words* problem
  - Needs handcraft features such as term frequency, lexical features and syntactics patterns

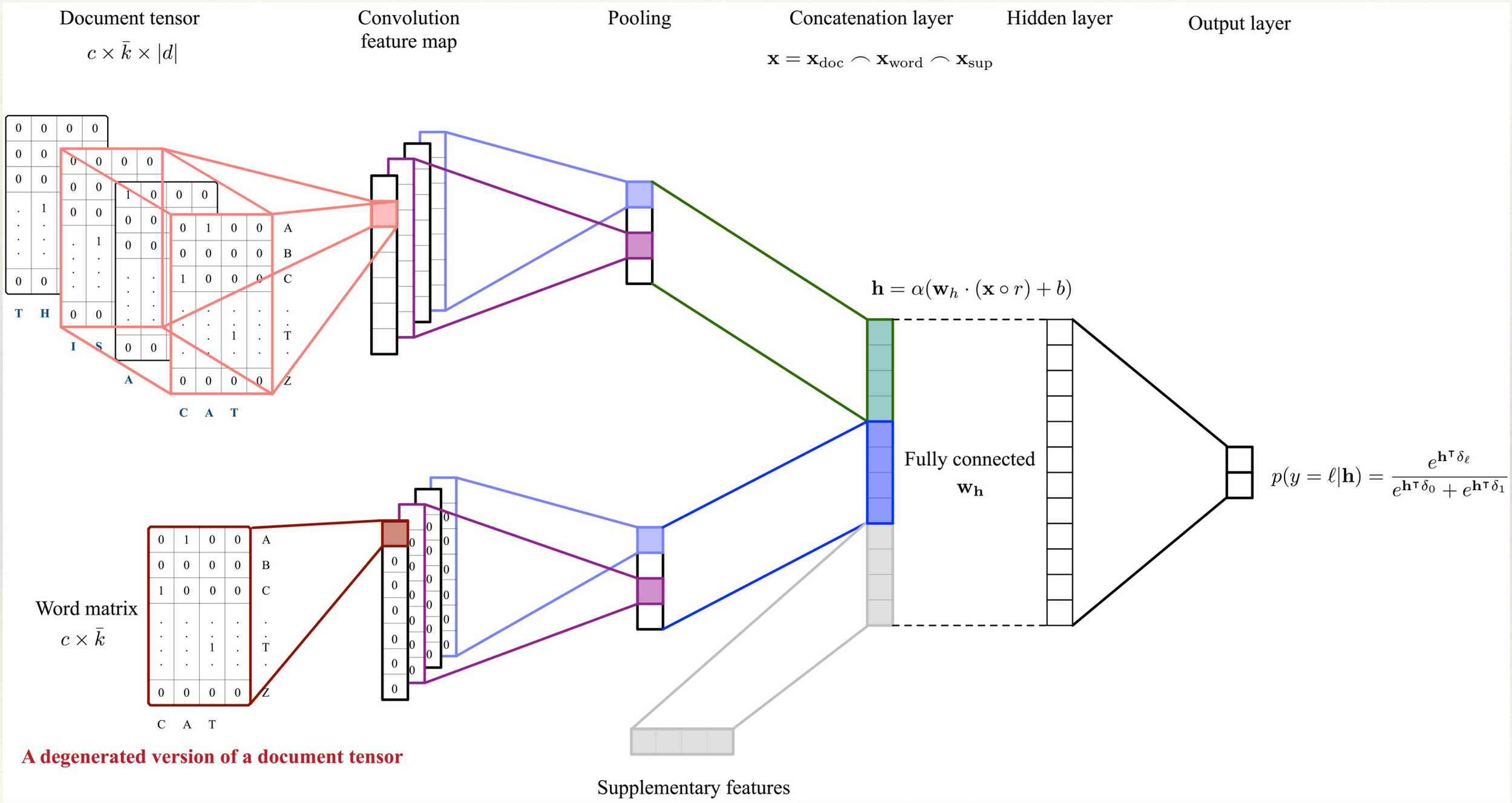
Unsupervised Approach

- Unsupervised approach
  - Uses a graph to describe a corpus
  - Generates keywords by learning the structure of the given keywords
    - Is difficult to incorporate deeper background knowledge extracted from external databases
- The graph-based ranking methods such as TextRank are trained using the graph which set vertices as words and edges as the co-occurrence relation of words.

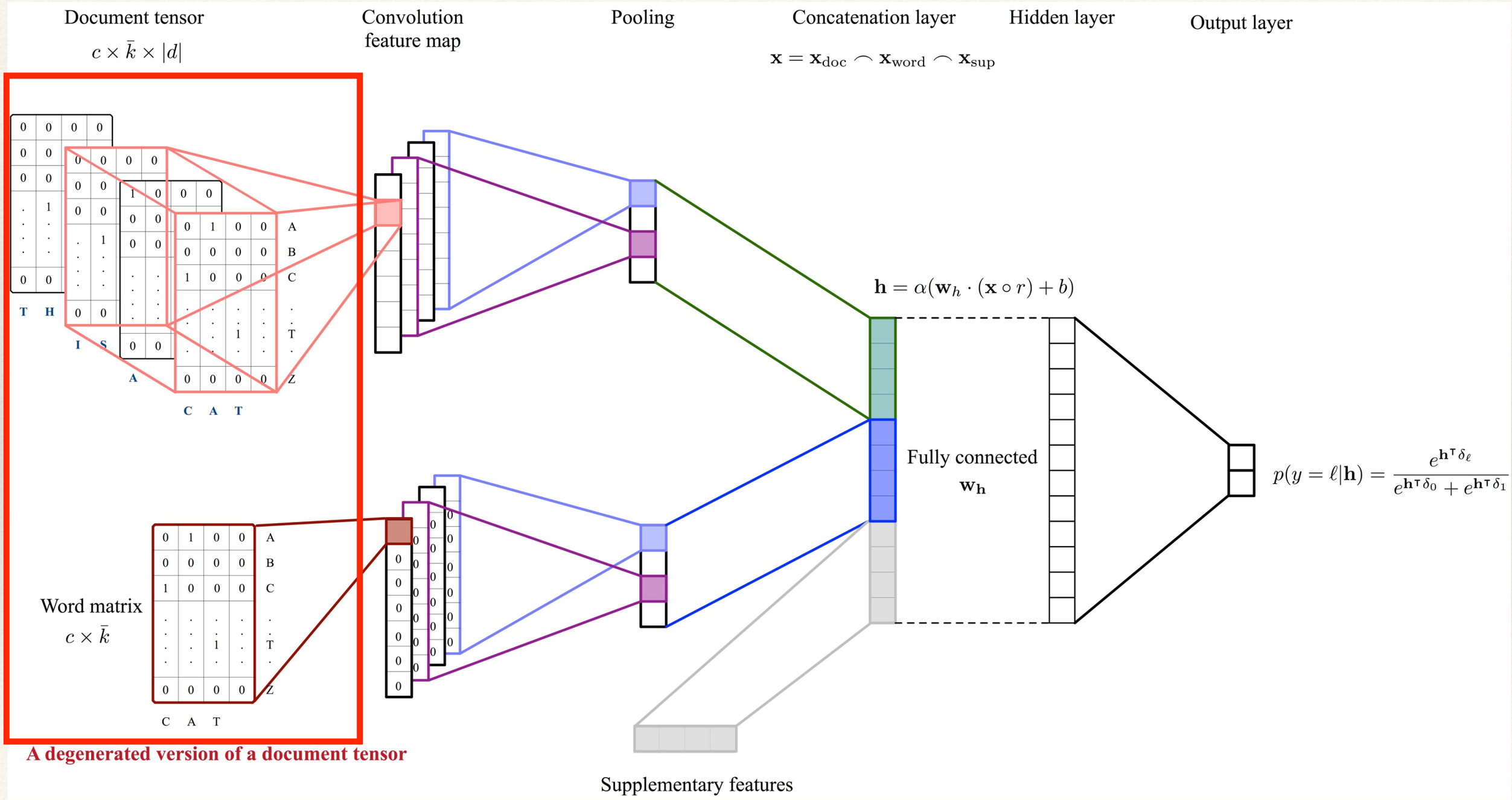
charCNTN

- The proposed **character-level Convolutional Tensor Network (charCNTN)** is a supervised approach.
- Solves the *unknown words* problem
  - Character level modeling
- Automatically extracts features
  - Distributed representations

charCNTN


Document tensor
$c \times \bar{k} \times |d|$
Convolution feature map
Pooling
Concatenation layer
$\mathbf{x} = \mathbf{x}_{\text{doc}} \frown \mathbf{x}_{\text{word}} \frown \mathbf{x}_{\text{sup}}$
Hidden layer
Output layer
T H I S A C A T
$\mathbf{h} = \alpha(\mathbf{w}_h \cdot (\mathbf{x} \circ r) + b)$
Fully connected
$\mathbf{w_h}$
$p(y = \ell|\mathbf{h}) = \frac{e^{\mathbf{h}^\intercal \delta_\ell}}{e^{\mathbf{h}^\intercal \delta_0} + e^{\mathbf{h}^\intercal \delta_1}}$
Word matrix
$c \times \bar{k}$
C A T
A degenerated version of a document tensor
Supplementary features

charCNTN


Document tensor
$c \times \bar{k} \times |d|$
Convolution feature map
Pooling
Concatenation layer
$\mathbf{x} = \mathbf{x}_{\text{doc}} \frown \mathbf{x}_{\text{word}} \frown \mathbf{x}_{\text{sup}}$
Hidden layer
Output layer
$\mathbf{h} = \alpha(\mathbf{w}_h \cdot (\mathbf{x} \circ r) + b)$
Fully connected
$\mathbf{w_h}$
$p(y = \ell|\mathbf{h}) = \frac{e^{\mathbf{h}^\intercal \delta_\ell}}{e^{\mathbf{h}^\intercal \delta_0} + e^{\mathbf{h}^\intercal \delta_1}}$
Word matrix
$c \times \bar{k}$
A degenerated version of a document tensor
Supplementary features

Document Tensor

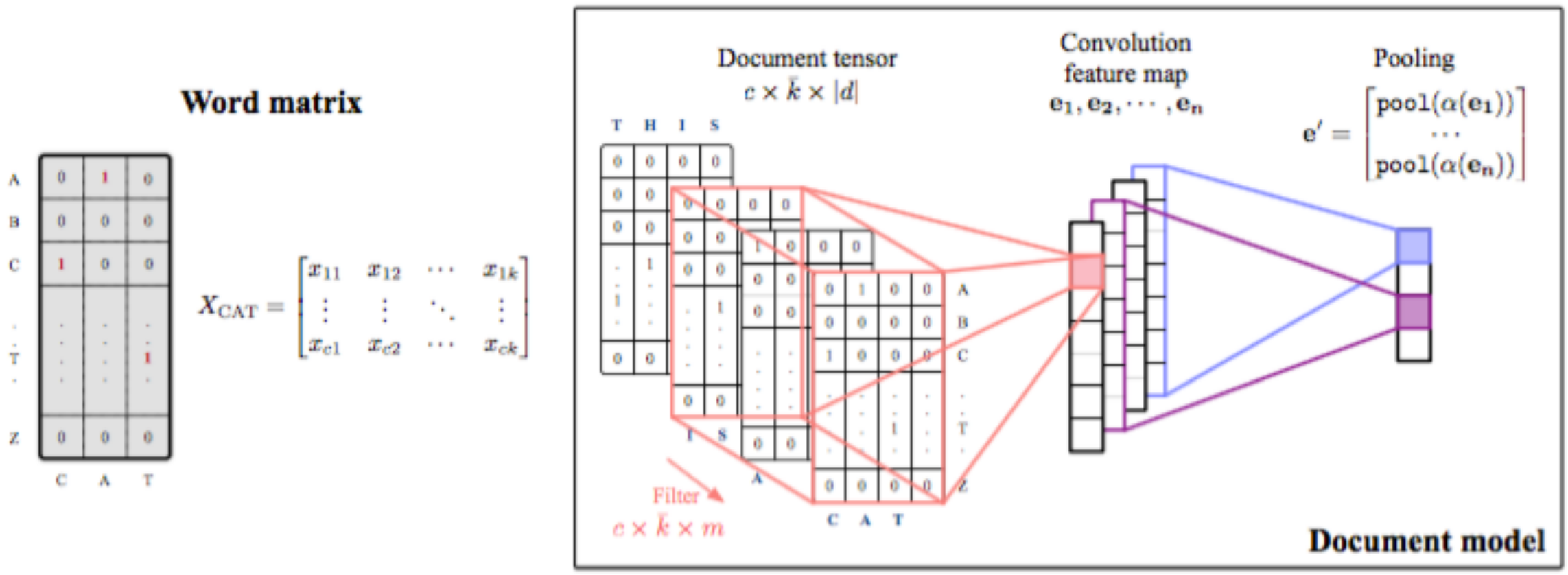
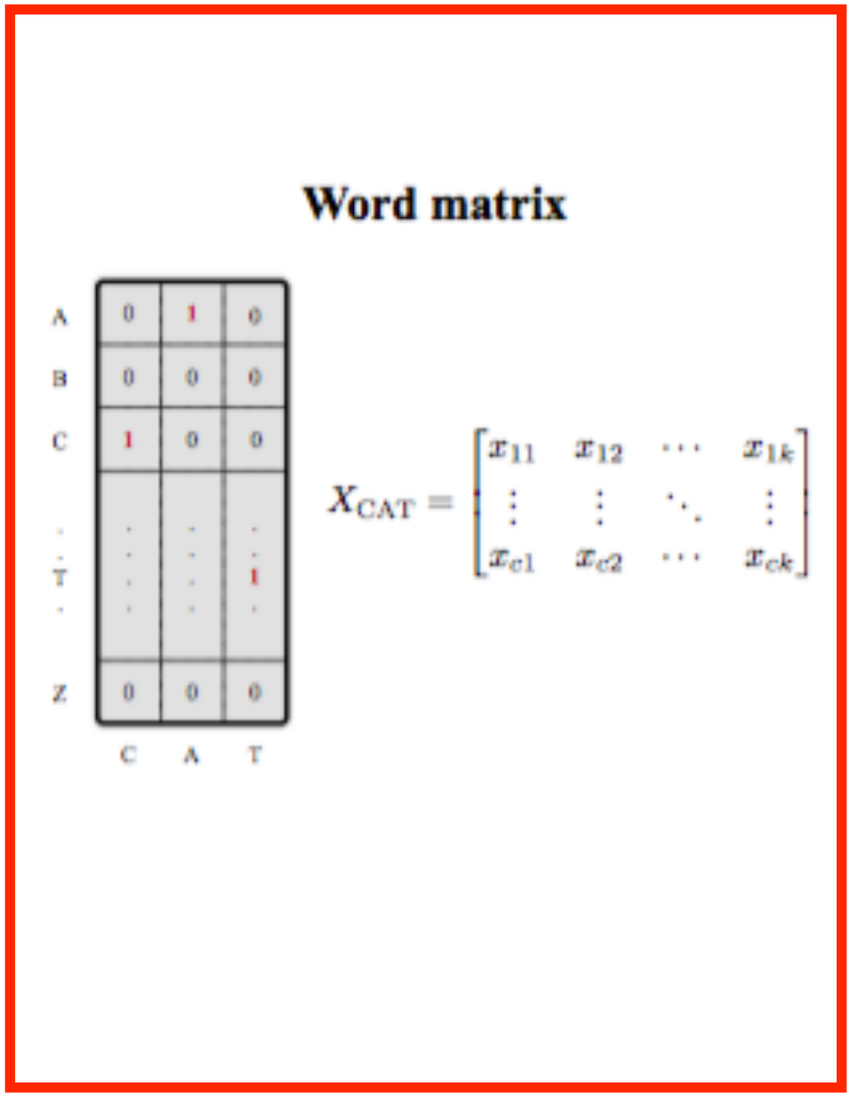


Fig. 1. Document model

Document Tensor



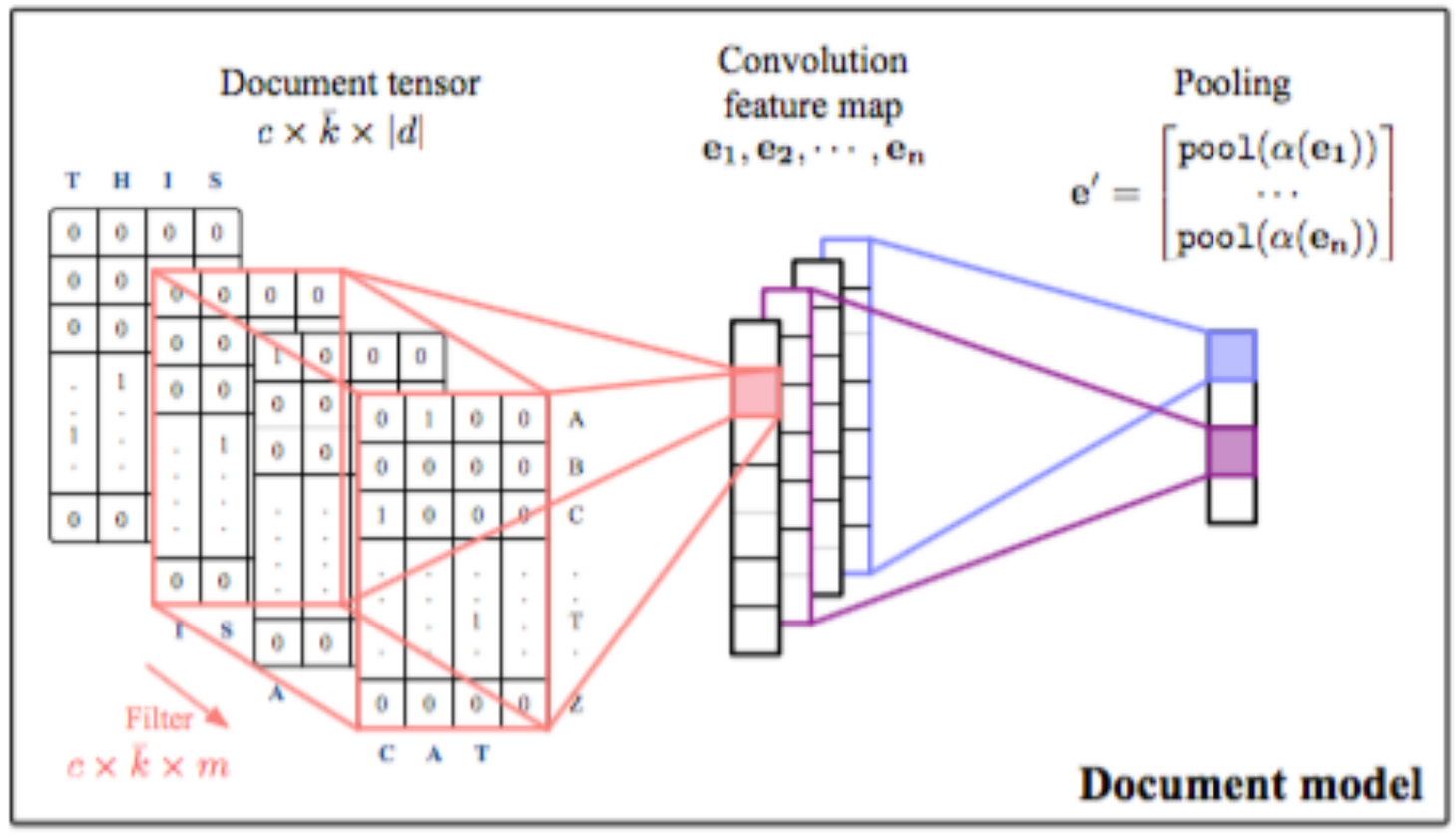
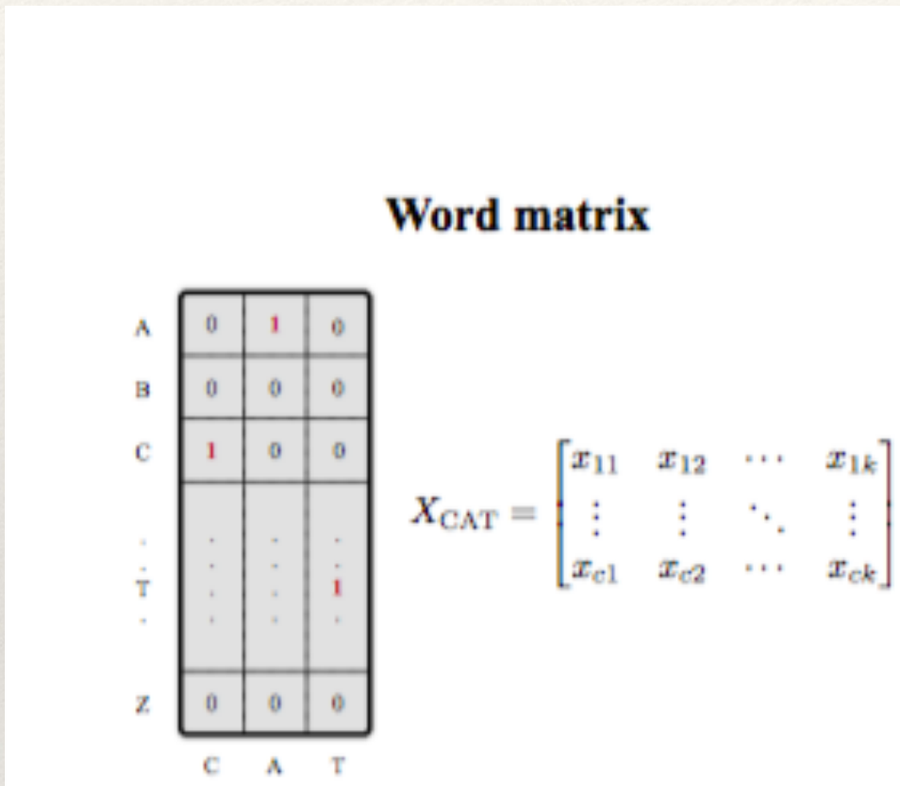


Fig. 1. Document model

Document Tensor



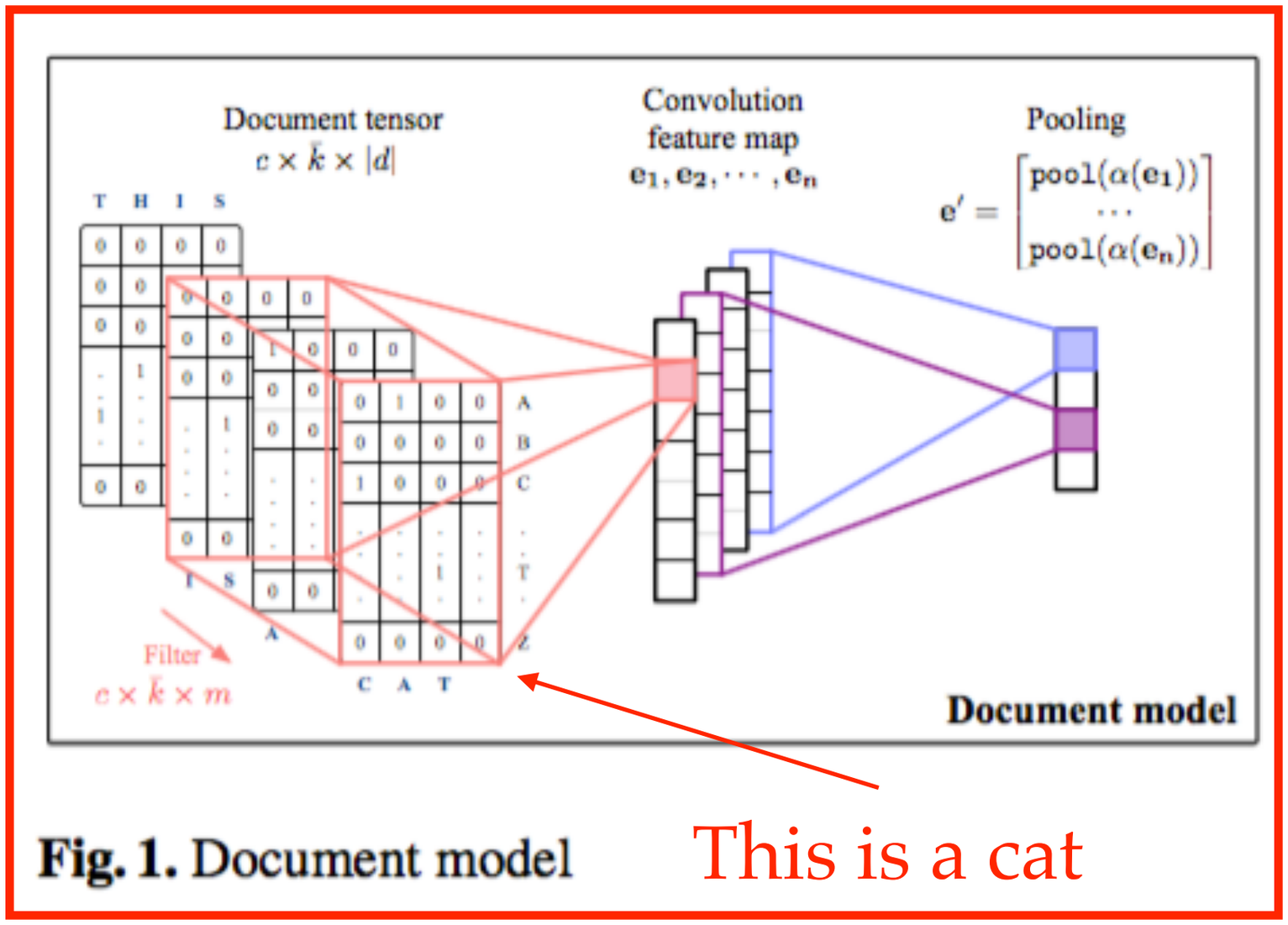
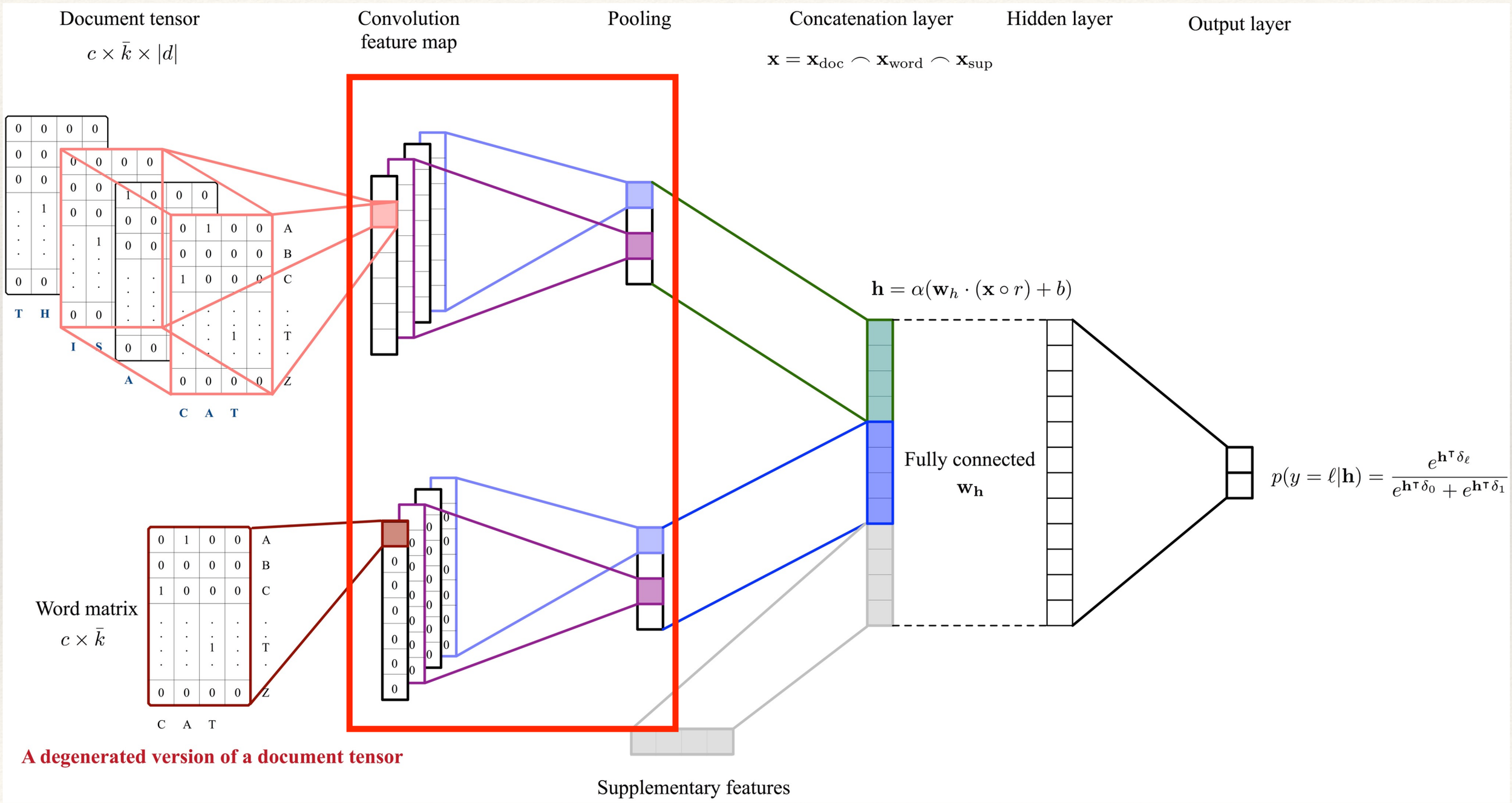


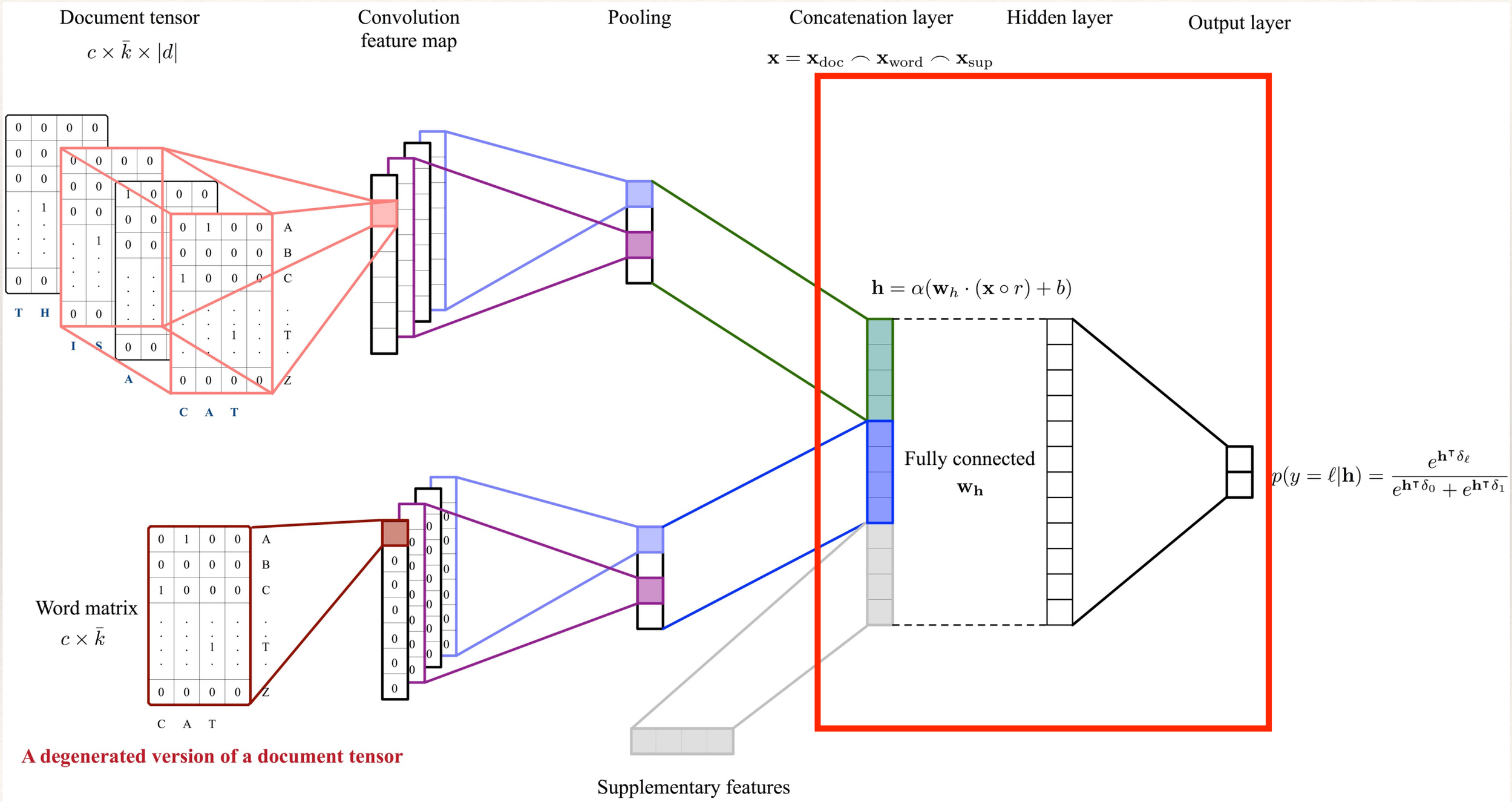
Fig. 1. Document model

This is a cat

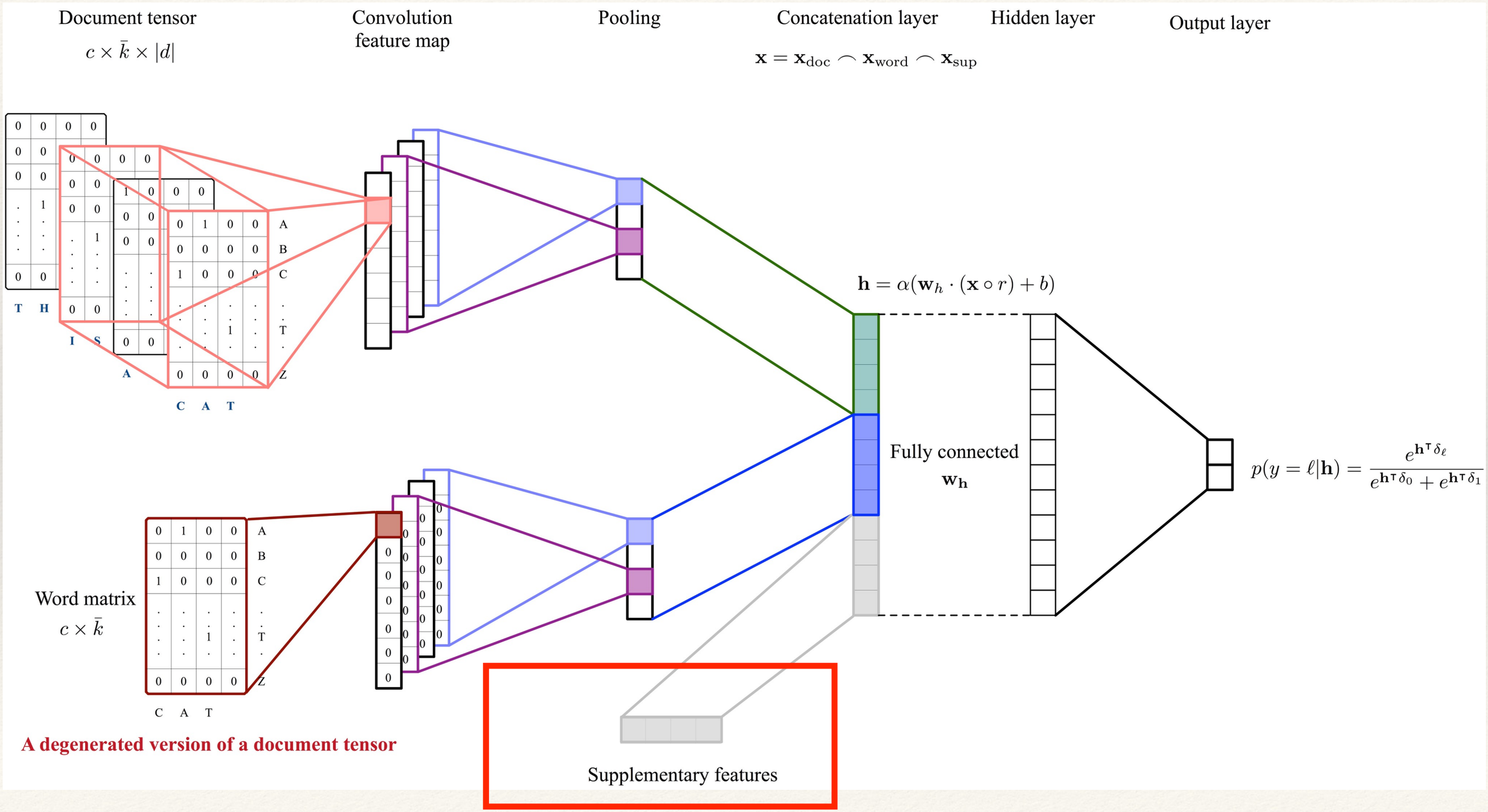
charCNTN


Document tensor
$c \times \bar{k} \times |d|$
Convolution feature map
Pooling
Concatenation layer
$\mathbf{x} = \mathbf{x}_{\text{doc}} \frown \mathbf{x}_{\text{word}} \frown \mathbf{x}_{\text{sup}}$
Hidden layer
Output layer
$\mathbf{h} = \alpha(\mathbf{w}_h \cdot (\mathbf{x} \circ r) + b)$
Fully connected
$\mathbf{w_h}$
$p(y = \ell|\mathbf{h}) = \frac{e^{\mathbf{h}^\intercal \delta_\ell}}{e^{\mathbf{h}^\intercal \delta_0} + e^{\mathbf{h}^\intercal \delta_1}}$
Word matrix
$c \times \bar{k}$
C A T
A degenerated version of a document tensor
Supplementary features

charCNTN


Document tensor
$c \times \bar{k} \times |d|$
Convolution feature map
Pooling
Concatenation layer
Hidden layer
Output layer
$\mathbf{x} = \mathbf{x}_{\text{doc}} \frown \mathbf{x}_{\text{word}} \frown \mathbf{x}_{\text{sup}}$
$\mathbf{h} = \alpha(\mathbf{w}_h \cdot (\mathbf{x} \circ r) + b)$
Fully connected
$\mathbf{w_h}$
$p(y = \ell|\mathbf{h}) = \frac{e^{\mathbf{h}^\intercal \delta_\ell}}{e^{\mathbf{h}^\intercal \delta_0} + e^{\mathbf{h}^\intercal \delta_1}}$
Word matrix
$c \times \bar{k}$
C A T
A degenerated version of a document tensor
Supplementary features

charCNTN


Document tensor
$c \times \bar{k} \times |d|$
Convolution feature map
Pooling
Concatenation layer
$\mathbf{x} = \mathbf{x}_{\text{doc}} \frown \mathbf{x}_{\text{word}} \frown \mathbf{x}_{\text{sup}}$
Hidden layer
Output layer
$\mathbf{h} = \alpha(\mathbf{w}_h \cdot (\mathbf{x} \circ r) + b)$
Fully connected
$\mathbf{w_h}$
$p(y = \ell|\mathbf{h}) = \frac{e^{\mathbf{h}^\intercal \delta_\ell}}{e^{\mathbf{h}^\intercal \delta_0} + e^{\mathbf{h}^\intercal \delta_1}}$
Word matrix
$c \times \bar{k}$
C A T
Supplementary features
A degenerated version of a document tensor

Experiments

- Datasets
- Experimental setting
- Results

Datasets

Table 1. Datasets

	Inspec		SemEval	
	Train	Test	Train	Test
Documents	1,000	500	144	40
Avg. length of documents	150	156	67	69
Unique words	9,258	5,660	2,047	1,047
Unique words (stemmed)	6,804	4,219	1,564	846
Unique keywords	6,377	3,792	1,310	680
Unique keywords (stemmed)	4,954	3,010	1,150	617
Unique keyphrases	-	4,913	-	621
Training instances	30,231	-	2,112	-

Experimental Setting

- Baselines
  - Supervised approach
    - TF-IDF
    - TextRank
    - CoreRank, dens, inf
  - Unsupervised approach
    - SVM
    - CNN with word2vec

Result

Table 2. Keyword extraction

		Unsupervised methods					Supervised methods			
		TD-IDF†	TextRank†	CoreRank†	dens†	inf†	SVM†	CNN†	charCNTN	charCNTN+
Inspec	Precision	43.18	63.45	63.59	48.06	48.18	60.10	54.35	57.32	69.31
	Recall	84.90	39.96	39.98	72.29	72.51	56.96	55.21	69.25	68.08
	F1-measure	55.45	47.20	47.25	55.35	55.62	56.18	50.12	60.80*	**61.45***
SemEval	Precision	33.12	50.89	48.86	36.66	37.21	49.61	51.27	41.09	42.63
	Recall	50.81	25.21	24.28	43.20	42.41	30.49	28.79	42.27	39.72
	F1-measure	38.68	32.75	31.49	37.96	37.72	36.71	33.84	**40.02***	39.61*

† denotes baseline methods; * denotes the statistical significance at $p < 0.05$ with respect to all baselines.

Result

Table 2. Keyword extraction

		Unsupervised methods					Supervised methods			
		TD-IDF†	TextRank†	CoreRank†	dens†	inf†	SVM†	CNN†	charCNTN	charCNTN+
Inspec	Precision	43.18	63.45	63.59	48.06	48.18	60.10	54.35	57.32	69.31
	Recall	84.90	39.96	39.98	72.29	72.51	56.96	55.21	69.25	68.08
	F1-measure	55.45	47.20	47.25	55.35	55.62	56.18	50.12	60.80*	**61.45***
SemEval	Precision	33.12	50.89	48.86	36.66	37.21	49.61	51.27	41.09	42.63
	Recall	50.81	25.21	24.28	43.20	42.41	30.49	28.79	42.27	39.72
	F1-measure	38.68	32.75	31.49	37.96	37.72	36.71	33.84	**40.02***	39.61*

† denotes baseline methods; * denotes the statistical significance at $p < 0.05$ with respect to all baselines.

Result

Table 2. Keyword extraction

		Unsupervised methods					Supervised methods			
		TD-IDF†	TextRank†	CoreRank†	dens†	inf†	SVM†	CNN†	charCNTN	charCNTN+
Inspec	Precision	43.18	63.45	63.59	48.06	48.18	60.10	54.35	57.32	69.31
	Recall	84.90	39.96	39.98	72.29	72.51	56.96	55.21	69.25	68.08
	F1-measure	55.45	47.20	47.25	55.35	55.62	56.18	50.12	60.80*	**61.45***
SemEval	Precision	33.12	50.89	48.86	36.66	37.21	49.61	51.27	41.09	42.63
	Recall	50.81	25.21	24.28	43.20	42.41	30.49	28.79	42.27	39.72
	F1-measure	38.68	32.75	31.49	37.96	37.72	36.71	33.84	**40.02***	39.61*

† denotes baseline methods; * denotes the statistical significance at $p < 0.05$ with respect to all baselines.

Result

Table 3. Performance on unknown words and keyphrase extraction

		Unknown Words				Keyphrase		
		SVM	CNN	charCNTN	charCNTN+	TextRank	charCNTN	charCNTN+
Inspec	Precision	38.52	35.86	42.78	42.41	30.40	24.43	23.43
	Recall	71.24	52.37	78.20	74.67	16.02	35.81	36.01
	F1-measure	50.01	42.57	**55.29**	54.05	19.91	**28.03**	27.37
SemEval	Precision	38.13	31.95	37.05	35.95	19.38	14.25	14.40
	Recall	41.22	27.84	71.16	61.31	6.05	17.15	18.77
	F1-measure	39.62	29.76	**48.62**	45.24	8.71	14.94	**15.62**

Result

Table 3. Performance on unknown words and keyphrase extraction

		Unknown Words				Keyphrase		
		SVM	CNN	charCNTN	charCNTN+	TextRank	charCNTN	charCNTN+
Inspec	Precision	38.52	35.86	42.78	42.41	30.40	24.43	23.43
	Recall	71.24	52.37	78.20	74.67	16.02	35.81	36.01
	F1-measure	50.01	42.57	**55.29**	54.05	19.91	**28.03**	27.37
SemEval	Precision	38.13	31.95	37.05	35.95	19.38	14.25	14.40
	Recall	41.22	27.84	71.16	61.31	6.05	17.15	18.77
	F1-measure	39.62	29.76	**48.62**	45.24	8.71	14.94	**15.62**

Result

Table 3. Performance on unknown words and keyphrase extraction

		Unknown Words				Keyphrase		
		SVM	CNN	charCNTN	charCNTN+	TextRank	charCNTN	charCNTN+
Inspec	Precision	38.52	35.86	42.78	42.41	30.40	24.43	23.43
	Recall	71.24	52.37	78.20	74.67	16.02	35.81	36.01
	F1-measure	50.01	42.57	**55.29**	54.05	19.91	**28.03**	27.37
SemEval	Precision	38.13	31.95	37.05	35.95	19.38	14.25	14.40
	Recall	41.22	27.84	71.16	61.31	6.05	17.15	18.77
	F1-measure	39.62	29.76	**48.62**	45.24	8.71	14.94	**15.62**

Result

Table 3. Performance on unknown words and keyphrase extraction

		Unknown Words					Keyphrase	
		SVM	CNN	charCNTN	charCNTN+	TextRank	charCNTN	charCNTN+
Inspec	Precision	38.52	35.86	42.78	42.41	30.40	24.43	23.43
	Recall	71.24	52.37	78.20	74.67	16.02	35.81	36.01
	F1-measure	50.01	42.57	**55.29**	54.05	19.91	**28.03**	27.37
SemEval	Precision	38.13	31.95	37.05	35.95	19.38	14.25	14.40
	Recall	41.22	27.84	71.16	61.31	6.05	17.15	18.77
	F1-measure	39.62	29.76	**48.62**	45.24	8.71	14.94	**15.62**

Conclusion

- We present an efficient supervised deep neural network for keyword extraction, in which models semantics down to the character level to capture the morphological information about words.
- The proposed approach effectively mitigates the unknown word problem.
- The experiment results show the proposed charCNTN outperforms both supervised and unsupervised baselines.

Future Work

- We plan to investigate how to integrate the proposed charCNTN architecture with sequential model like attention mechanism to better capture word sequence behavior.
- Extend the proposed architecture to directly handle keyphrase extraction, which is also a challenging problem.

Thanks for your listening.

Q&A