

**14th International Conference on Frontiers in Handwriting Recognition
(ICFHR2014)**
1-4 September, Crete island, Greece

Text/Non-text Classification in Online Handwritten Documents with Recurrent Neural Networks

Truyen Van PHAN and Masaki NAKAGAWA
Tokyo University of **Agriculture & Technology** (TUAT), Japan

Background – Trends

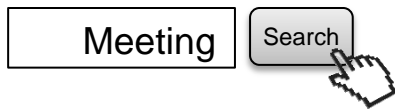
- **Pen-based, touch-based devices/applications** have become popular.



- **People can create notes, make diagrams, draw sketches** on their mobile devices.

Background – Needs

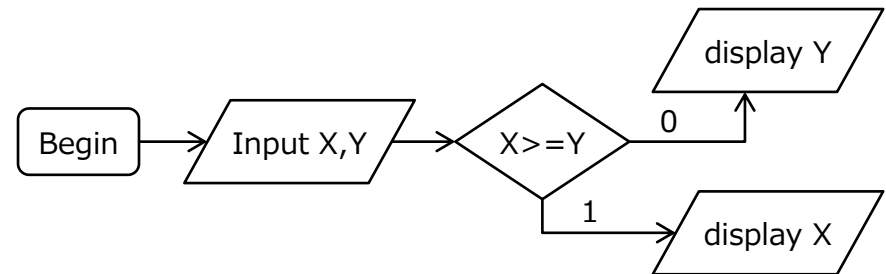
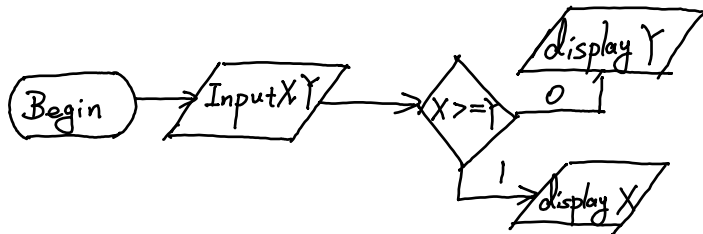
■ Text search on hand-written notes, schedules,...



Weekly Schedule Table

	Schedule	Notes
Mon	10:00 - 12:00	Meeting
Tue	13:00 - 14:00	Lunch
Wed	18:00 - 20:00	Yoga
Thu	9:00 - 11:00	Lecture
Fri	14:00 - 16:00	Class
Sat	15:00 - 20:00	Sleep

■ Hand-drawn sketch (diagrams, flowcharts, ...) recognition/interpretation.



Ref. to presentation of “**Recognition System for On-line Sketched Diagrams**” by Martin Bresler, on Thursday, Session 9, 10:00 ~ 10:20

Text/Non-text Classification

- To classify handwritten ink strokes into two categories: **text** and **non-text**.

$= -\log [10^{-14}] = 14$

Skindarts	skin	southpaw	return
evidence	4751	7605	9
discharges	85	30	1783
sup	517	2	3510
recipient	5	3	32
milk	447	496	7576

The rules and policies to be applied in this process of course must be based on objectives which represent what is to be desired if radio service is to be of maximum use to the Nation. To provide service of local origin to as many listeners as possible.



$= -\log [10^{-14}] = 14$

Skindarts	skin	southpaw	return
evidence	4751	7605	9
discharges	85	30	1783
sup	517	2	3510
recipient	5	3	32
milk	447	496	7576

The rules and policies to be applied in this process of course must be based on objectives which represent what is to be desired if radio service is to be of maximum use to the Nation. To provide service of local origin to as many listeners as possible.

Text/Non-text Classification

- It can be used as a preprocessing step for **text search**, text recognition, or diagram interpretation.

objectives Search

Anticline Monocline
Acclinal Sineclise

$= -\log [10^{-4}] = 14$

skudants	sein	sordyraw	retinn
evidence	4751	7603	9
discharges	85	30	7883
sup	517	2	7510
recipients	5	3	32
mils	447	496	7576

The rules and policies to be applied in this process of course must be based on **objective** which represent what is to be desired if radio service is to be of maximum use to the Nation. To provide service of local origin to as many listeners as possible.

Text/Non-text Classification

- It can be used as a preprocessing step for text search, text recognition, or diagram interpretation.

$= -\log[10^{-14}] = 14$

Students	skin	southpaw	return
evidence	4751	7603	9
discharges	85	90	1783
Sop	517	2	3510
recipients	5	3	32
miles	847	496	7576

The rules and policies to be applied in this process of course must be based on objectives which represent what is to be desired if radio service is to be of maximum use to the Nation. To provide service of local origin to as many listeners as possible.



$= -\log[10^{-14}] = 14$

Students	skin	southpaw	return
evidence	4751	9603	9
discharges	85	90	1783
Sop	517	2	3510
recipients	5	3	32
miles	847	496	7576

The rules and policies to be applied in this process of course must be based on objectives which represent what is to be desired if radio service is to be of maximum use to the Nation. To provide service of local origin to as many listeners as possible .

Related work

■ Two approaches

- ◆ Context-integrated classification: single stroke classification (SVM, MLP) → context integration (HMM, MRF, CRF)
- ◆ Sequence classification by using BLSTM.

■ Remarkable results

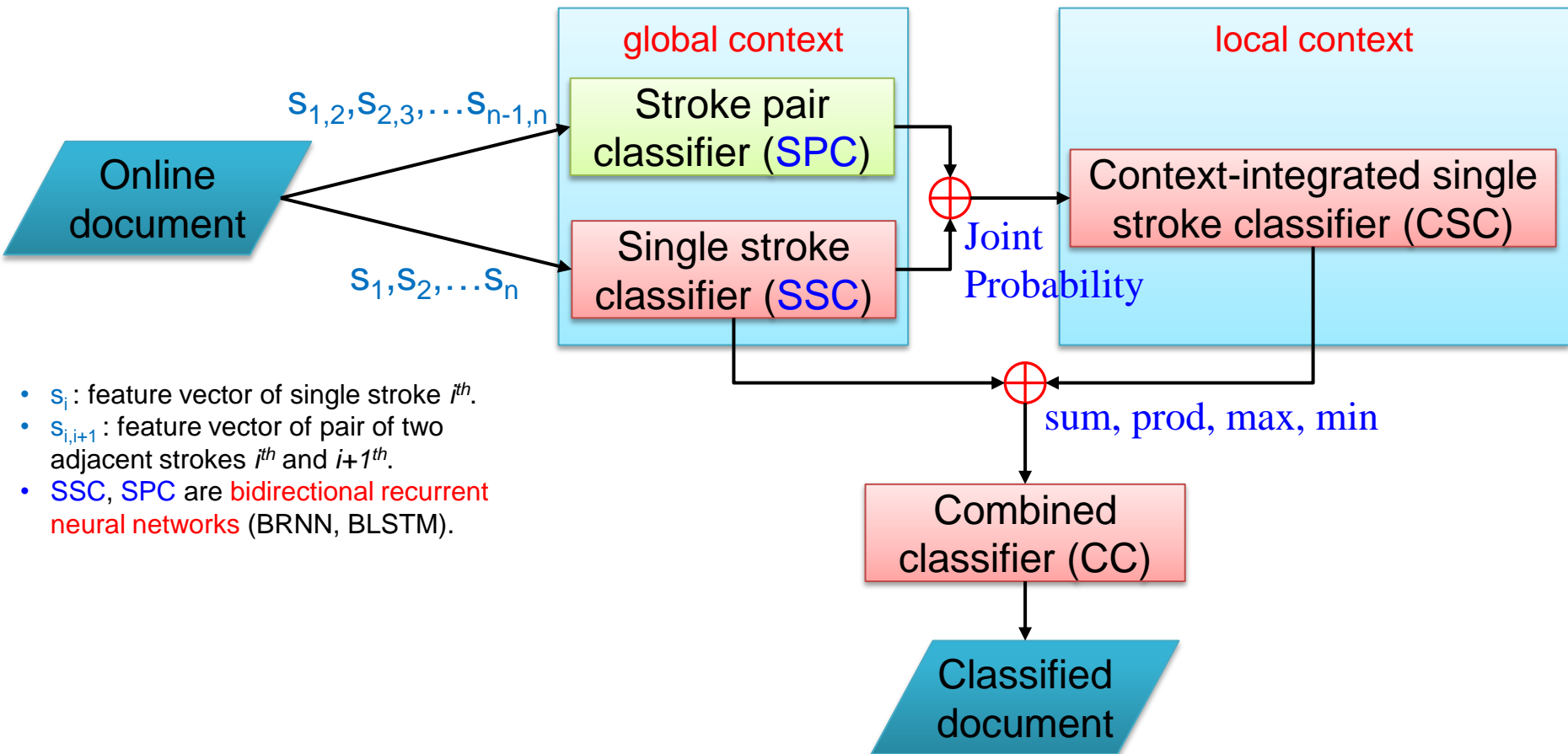
Publication	Method	Data Set	Accuracy
Zhou and Liu [3]	SVMs + HMM SVMs + MRF	Kondate	94.48 96.61
Indermuhle et al. [8]	BLSTM	IAMonDo	97.01
Delaye et al. [4]	SVMs + CRF	IAMonDo	97.23

Consideration

- In order for text/non-text classification to be practical, it should be **highly accurate** and **quick**.
- **State-of-the-art performances are not satisfied.**
 - ◆ Current accuracies are only about 97%.
 - ◆ SVMs + MRF/CRF methods are **slow** because of SVMs.
 - ◆ BLSTM method is **quick but not accurate** since it considers interactions between points only.

Outline of our method

- It's a **context-integrated sequence classification method** using both **global and local contexts**.



- s_i : feature vector of single stroke i^{th} .
- $s_{i,i+1}$: feature vector of pair of two adjacent strokes i^{th} and $i+1^{th}$.
- SSC, SPC are **bidirectional recurrent neural networks** (BRNN, BLSTM).

Global context model

- **Make use of bidirectional recurrent neural networks to gain access to the global context of the whole document.**
- **To transcribe a sequence of feature vectors to a sequence of labels.**

	Single stroke classifier (SSC)	Stroke pair classifier (SPC)
Input (feature)	Unary (extracted from each stroke)	Binary (extracted from pair of two temporally adjacent strokes)
Output (label)	2 categories <ul style="list-style-type: none">• text (T)• non-text (N)	3 categories <ul style="list-style-type: none">• text:text (TT)• text:non-text (TN) or non-text:text (NT)• non-text:non-text (NN)
Sequence length	N S_1, S_2, \dots, S_n	N-1 $S_{1,2}, S_{2,3}, \dots, S_{n-1,n}$

N: number of strokes in document

Local context model

- Use **marginal distribution** to integrate context of neighboring strokes.

$$P(X = x) = \sum_y P(X = x|Y = y)P(Y = y)$$

Probability of a stroke i^{th} being text is calculated by



$$P_{PM}(T_i) = P(TT_{i,i-1})P(T_{i-1}) + P(TN_{i,i-1})P(N_{i-1})$$



$$P_{SM}(T_i) = P(TT_{i,i+1})P(T_{i+1}) + P(TN_{i,i+1})P(N_{i+1})$$



$$P_{BM}(T_i) = P_{PM}(T_i) + P_{SM}(T_i)$$

Combined classifier

■ Use four basic combination rules

◆ Sum rule (SUM) $l_i^* = \operatorname{argmax} \left\{ \sum_{k=1}^K f_k(l_i | s_i), l_i \in \{T, N\} \right\}$

◆ Product rule (PROD) $l_i^* = \operatorname{argmax} \left\{ \prod_{k=1}^K f_k(l_i | s_i), l_i \in \{T, N\} \right\}$

◆ Max rule (MAX) $l_i^* = \operatorname{argmax} \left\{ \max_{k=1}^K f_k(l_i | s_i), l_i \in \{T, N\} \right\}$

◆ Min rule (MIN) $l_i^* = \operatorname{argmax} \left\{ \min_{k=1}^K f_k(l_i | s_i), l_i \in \{T, N\} \right\}$

where $K = 2$,

- $f_1(l_i | s_i)$ is the probability distribution of l_i calculated by the single stroke classifier (SSC),
- $f_2(l_i | s_i)$ is one of the three probability distributions of l_i calculated by the context-integrated single stroke classifier (CSC).

Experiment

■ Data

Database	Subset	# pages	# strokes	% T	% N
Kondate (Japanese)	<i>Training</i>	210	41,190	83.53	16.47
	<i>Validation</i>	100	18,525	84.89	15.11
	<i>Testing</i>	359	71,846	85.44	14.56
IAMonDo (English)	<i>Training</i>	403	143,350	80.96	19.04
	<i>Validation</i>	200	68,726	83.60	16.40
	<i>Testing</i>	203	70,927	81.26	18.74

■ Evaluation

- ◆ Single stroke classifiers (SSCs)
- ◆ Stroke pair classifiers (SPCs)
- ◆ Context-integrated single stroke classifiers (CSCs)
- ◆ Combined classifiers (CCs)
- ◆ Computational time

Data samples – Kondate

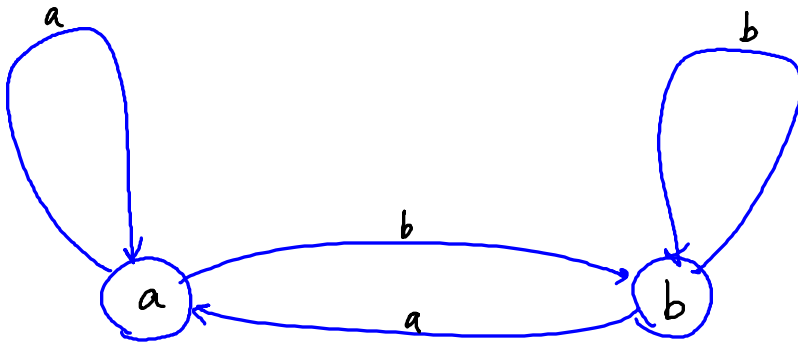
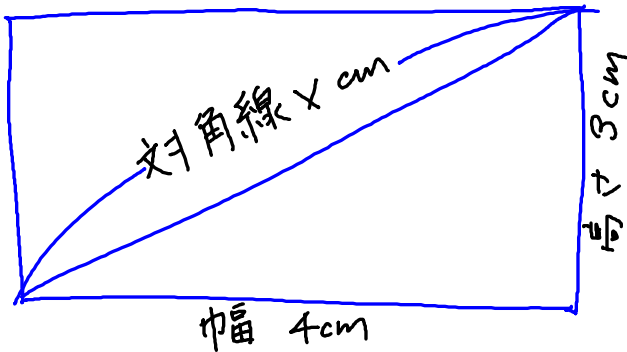


図1 バイグラムの確率有限オートマトンによる表現

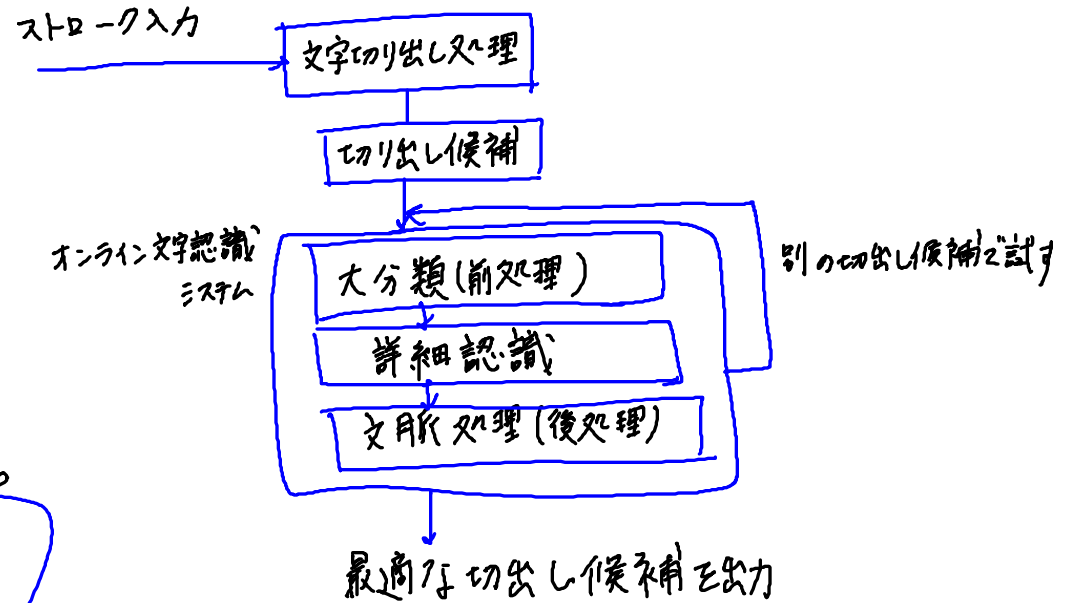
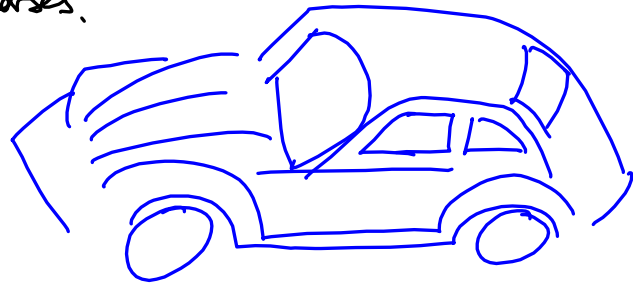
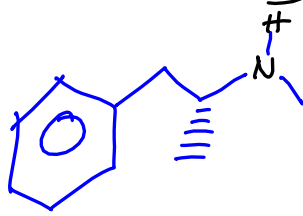


図5 オンライン 枠外文字認識システムの大まかな流れ

Data samples – IAMonDo

And he had a feeling – thanks to the girl –
 that things would get worse before they get
 better. wehe
amen! they had the house
 [cleaned up] by room,
 and Wilson sent
 the boy out to the meadow to bring in the
 horses.



He stood on the porch
 and watched him struggling
 with the heavy harness,
 and finally went over to help him.

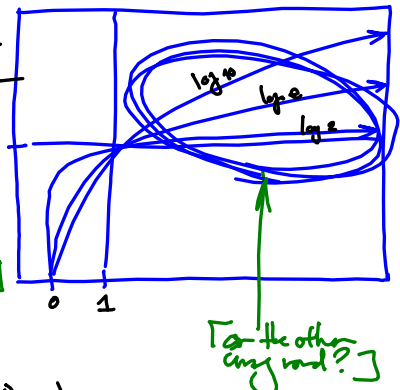
- tender
- fanciers
- daughter
- daughter

| finally

12/9/2008

Template 282

For few of the case (the
 Amrocants of the amrocs)
 the python cannot all
 included for lack of
 data) – there is also a
 connection between
 size of amrocs &
 maximum length.



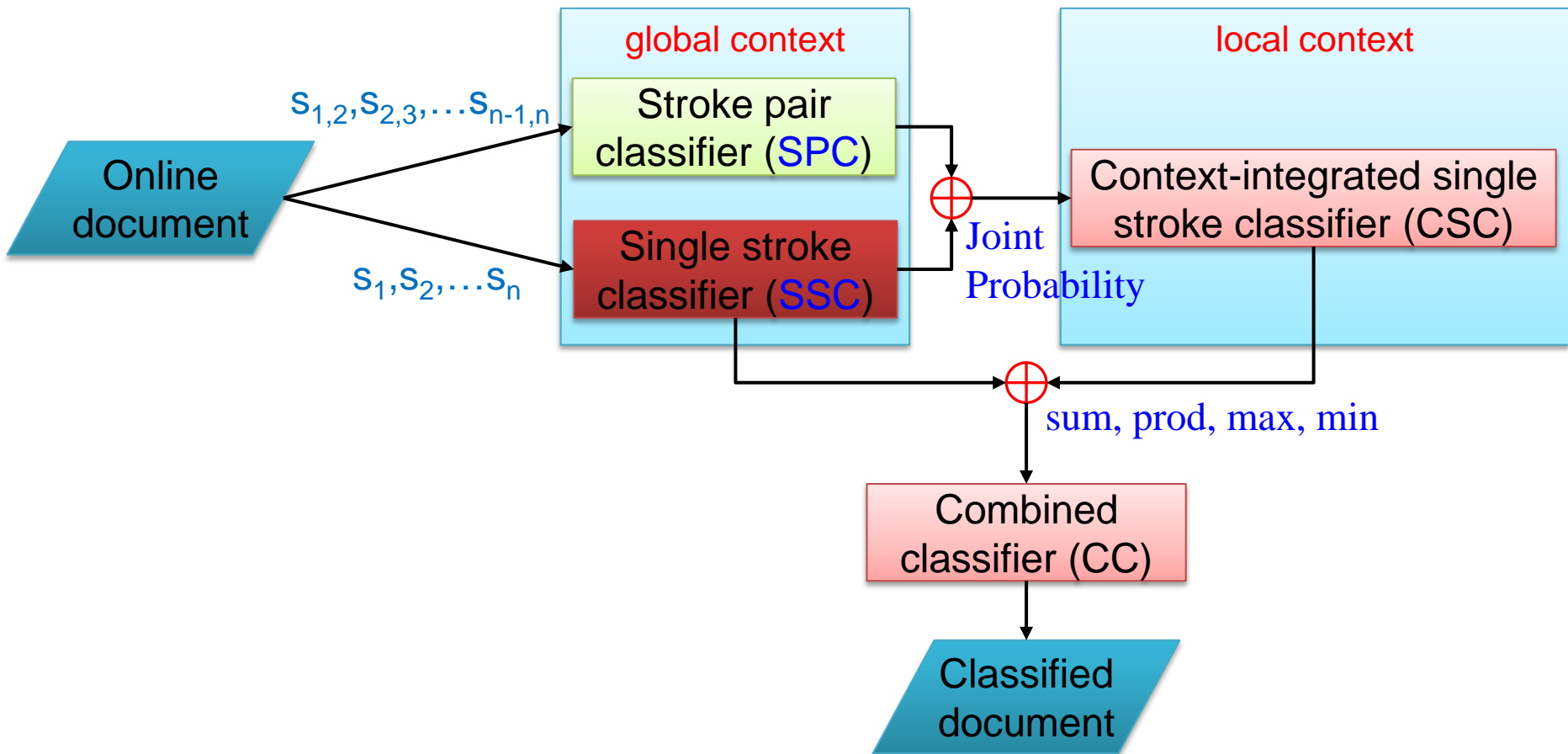
$$P_n(x) = \frac{f^{(n)}(\xi)}{n!} (x-\xi)^{n-1} - P(x-a)^m$$

The pro constraint very flexible and the
 inclusion python ~~python~~ ^{comp} ~~herb~~ ^{size} at the
 former stage.

WEEK	rooms	sentin	order
Sarth	1061	28	611
water	2	612	437
law	1	53 53	79
Bros	64	41	6081
hbrn	7	635	358

Experimental result

■ Single stroke classifiers (SSCs)



Experimental result

■ Single stroke classifiers (SSCs)

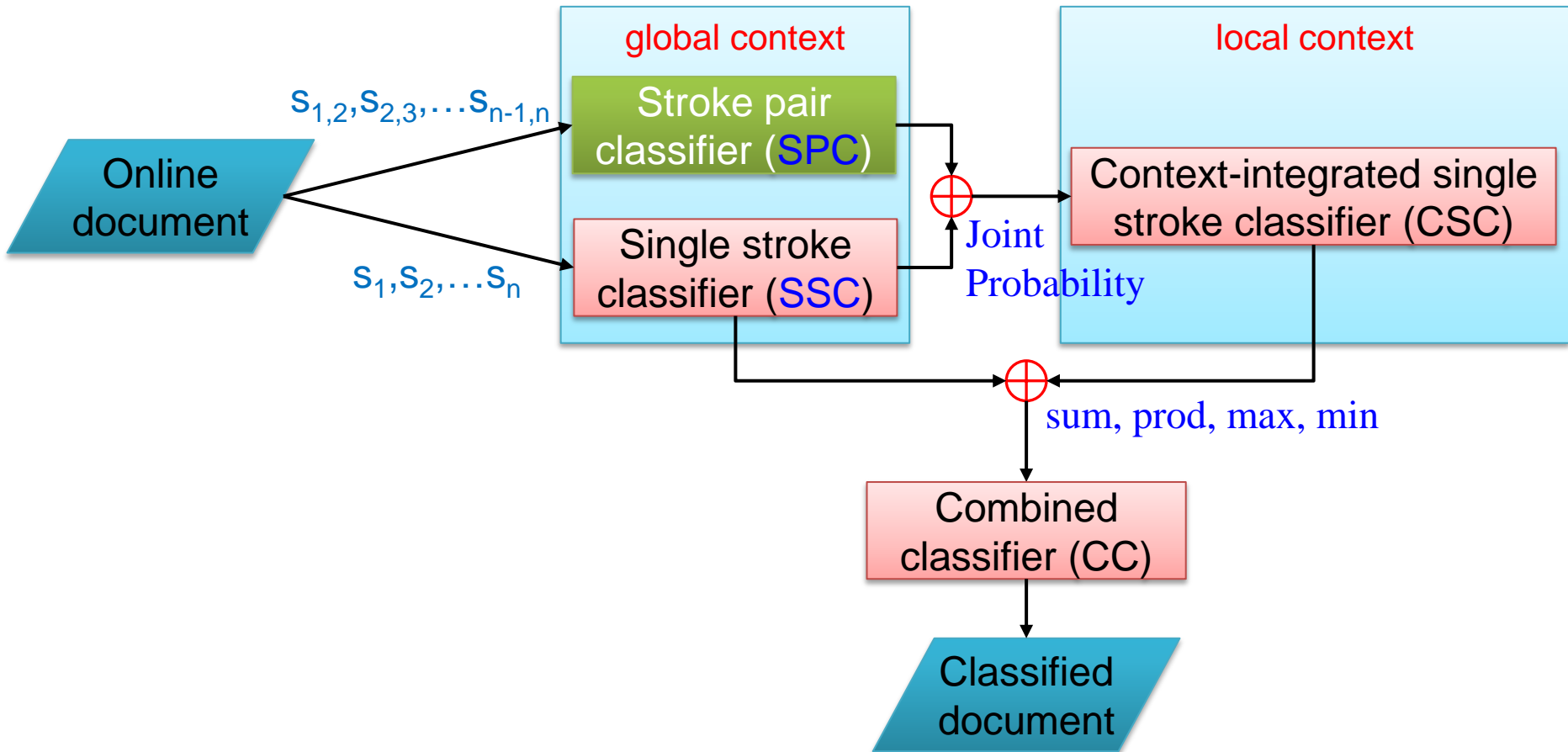
Database	Classifier	Accuracy (%)			Total time (s)
		Overall (average (min ~ max))	Text (average)	Non-text (average)	
Kondate (Japanese)	SSC11_RNN	95.39 (94.33~96.36)	98.37	77.90	0.38
	SSC11_LSTM	96.13 (95.48~96.57)	98.53	82.05	1.53
	SSC19_RNN	96.43 (94.65~97.45)	98.58	83.79	0.39
	SSC19_LSTM	97.01 (95.68~ 97.62)*	98.71	87.08	1.56
IAMonDo (English)	SSC11_RNN	95.62 (94.76~96.28)	97.87	85.87	0.36
	SSC11_LSTM	96.41 (96.15~96.73)	98.15	88.87	1.49
	SSC19_RNN	96.62 (96.15~97.09)	98.13	90.09	0.37
	SSC19_LSTM	96.93 (96.65~ 97.34)**	98.38	90.63	1.50

* Zhou and Liu [3]: 96.61%

** Delaye et al. [4]: 97.23%

Experimental result

■ Stroke pair classifiers (SPCs)



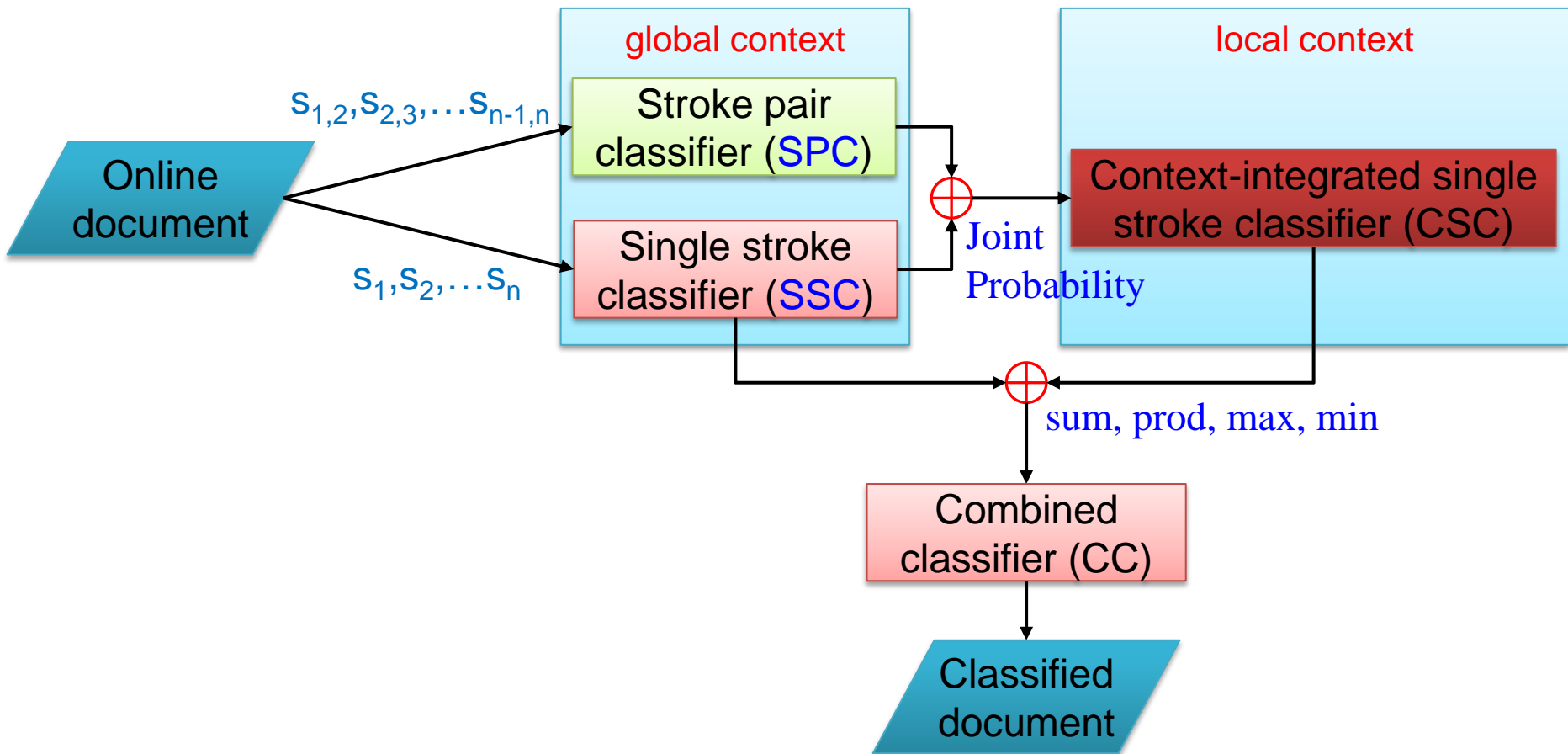
Experimental result

■ Stroke pair classifiers (SPCs)

Database	Classifier	Accuracy (%)				Total time (s)
		Overall <i>(average (min ~ max))</i>	TT <i>(average)</i>	TN <i>(average)</i>	NN <i>(average)</i>	
Kondate (Japanese)	SPC_RNN	95.54 (93.00~96.79)	98.16	68.16	87.46	0.38
	SPC_LSTM	96.71 (95.71~ 97.38)	98.63	80.65	89.35	1.51
IAMonDo (English)	SPC_RNN	92.90 (90.55~94.52)	97.86	28.04	82.65	0.53
	SPC_LSTM	94.46 (93.67~ 95.03)	97.72	47.47	88.55	1.48

Experimental result

■ Context-integrated single stroke classifiers (CSCs)



Experimental result

■ Context-integrated single stroke classifiers (CSCs)

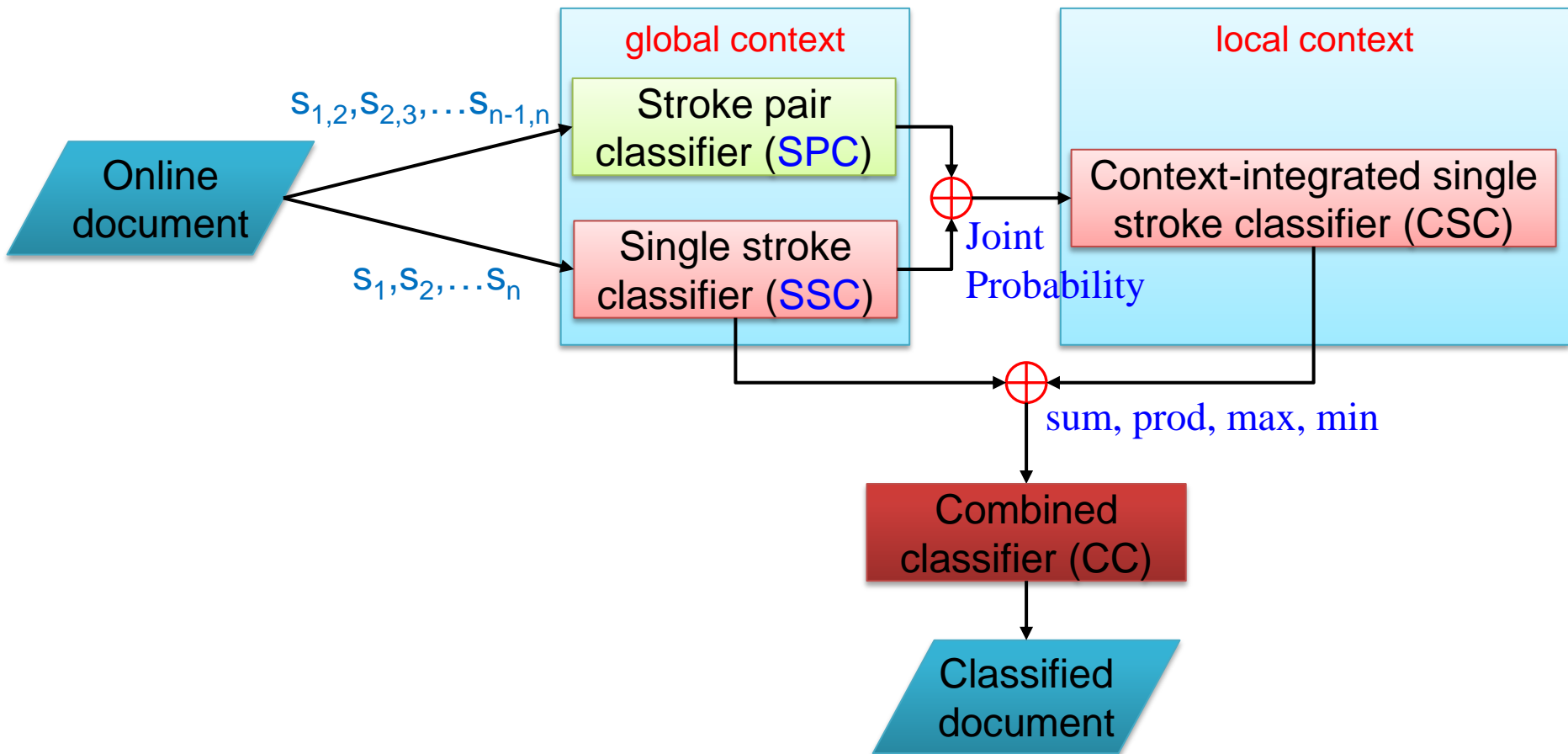
Database	SPC Classifier	SPC Accuracy (%)	SSC Classifier	SSC Accuracy (%)	CSC Accuracy (%)		
					PM	SM	BM
Kondate (Japanese)	SPC_RNN	96.79	SSC11_RNN	96.36	97.13	97.13	97.72
			SSC19_RNN	97.45	97.43	97.27	97.91
	SPC_LSTM	97.38	SSC11_LSTM	96.57	97.55	97.59	98.18
			SSC19_LSTM	97.62	97.96	97.89	98.43*
IAMonDo (English)	SPC_RNN	94.52	SSC11_RNN	96.28	96.26	96.29	96.69
			SSC19_RNN	97.09	96.51	96.50	96.89
	SPC_LSTM	95.03	SSC11_LSTM	96.73	96.63	96.59	97.03
			SSC19_LSTM	97.34**	96.85	96.75	97.19

* Zhou and Liu [3]: 96.61%

** Delaye et al. [4]: 97.23%

Experimental result

■ Combined classifiers (CCs)



Experimental result

■ Combined classifiers (CCs)

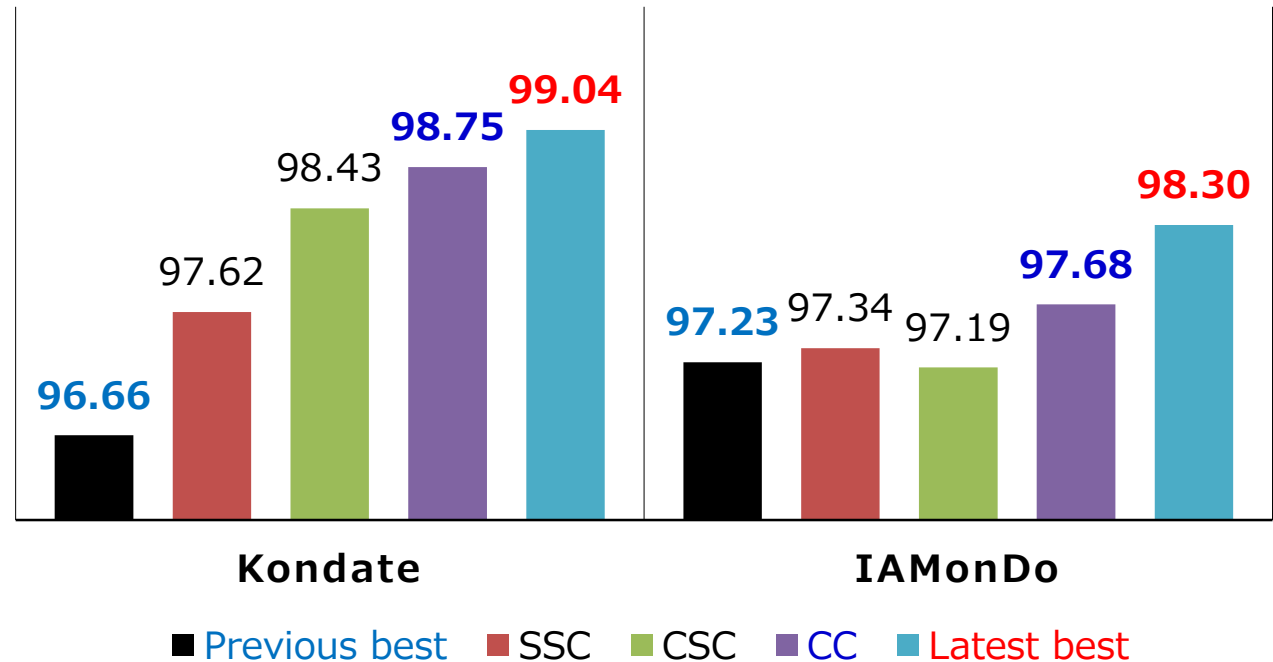
Database	Type	SSC Accuracy (%)	CSC Accuracy (%)	CC Accuracy (%)			
				SUM	PROD	MAX	MIN
Kondate (Japanese)	<i>11_RNN</i>	96.57	98.18	97.72	98.23	97.58	98.29
	<i>19_LSTM</i>	97.62	98.43	98.38	98.73	98.28	98.75*
IAMonDo (English)	<i>11_RNN</i>	96.73	97.03	97.19	97.32	97.10	97.36
	<i>19_LSTM</i>	97.34	97.19	97.55	97.65	97.50	97.68**

* Zhou and Liu [3]: 96.61%

** Delaye et al. [4]: 97.23%

Experimental result

■ Summary



■ Computational time

Time (s)	Kondate 359 pages	IAMonDo 203 pages
Feature Extraction (Unary + Binary)	1.1	4.6
Classification (Single stroke + Stroke pair)	3.2	3.2
Context Integration	0.01	0.01
Total	4.3	7.8
Average (per page)	0.012	0.038*

* Delaye et al. [4]: 1.53 (s)

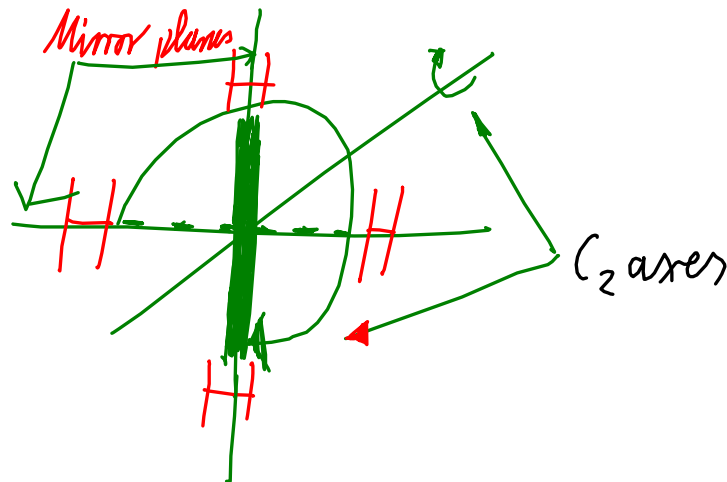
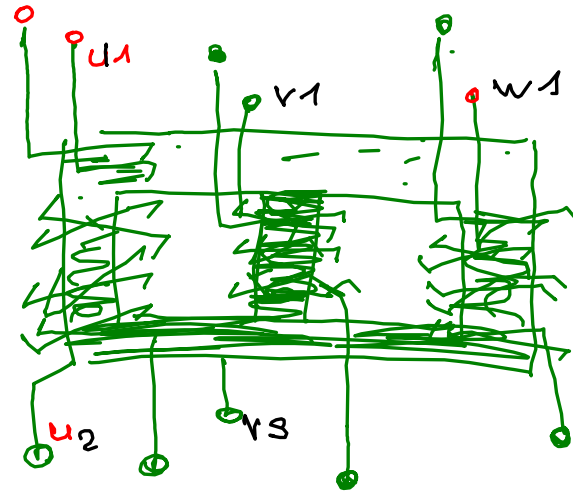
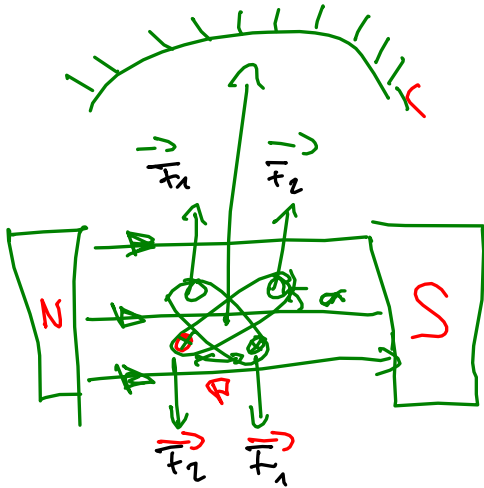
Conclusion

- We propose a **context-integrated sequence classification** method for text/non-text classification in online handwritten documents.
 - ◆ Recurrent neural networks for sequence classification.
 - ◆ Marginal distribution and simple combination rules (sum, prod, max, min) for context integration.

- We achieve classification rates of **98.75% (99.04%)** on Kondate and **97.68% (98.30%)** on IAMonDo.

Future work

■ Solve these kinds of problems



Demo
if you requested