Dear Jinpeng Li:

We are pleased to inform you that your paper listed below has been accepted for ORAL presentation at the 11th International Conference on Document Analysis and Recognition (ICDAR 2011), to be held in Beijing, China, on September 18-21, 2011. Congratulations!

67: Symbol Knowledge Extraction from a Simple Graphical Language

The author kit to be used in preparing the camera-ready version of your paper will be sent to you within a few days. You must follow these formatting guidelines. You should also take into account the feedback you have received from the reviewers below as you work to prepare the final version of your paper.

Please be aware that there is a 5-page limit for your paper in the proceedings. Additional pages are permitted, but each extra page will incur an additional $100 (650CNY) charge, payable when you register for the conference. Your final camera-ready paper and copyright transfer form are due by July 1, otherwise your paper will not be included in the proceedings. At least one author from each paper must register for the conference by July 1, otherwise the paper will not be included in the proceedings.

Congratulations once again! We look forward to seeing you in Beijing in September.

The ICDAR 2011 Program Chairs:
Chew Lim Tan,
Daniel Lopresti,
Thomas Breuel

---------- REVIEW 1 ----------
PAPER: 67
TITLE: Symbol Knowledge Extraction from a Simple Graphical Language
AUTHORS: Jinpeng Li, Harold Mouchere and Christian Viard-Gaudin

OVERALL RATING: 1 (weak accept)
Relevance (potential interest to the ICDAR community): 3 (fair)
Originality: 4 (good)
Technical soundness: 4 (good)
Experimental evaluation: 3 (fair)
Presentation: 4 (good)

The authors present an unsupervised learning method for extracting automatically the lexicon of simple mathematical
expressions.
The approach seems very interesting, however, this method is applied to a very simple online mathematical handwriting
dataset (composed of digits and operators). As the authors state in the conclusion, the current method might not be able
to cope with more complex mathematical expressions. I encourage the authors to add variables (eg. x,y,z...), fractions,
exponents, square roots, etc. in the mathematical dataset, and improve the proposed method to cope with this kind of
expressions. In this sense, graph grammars could be helpful for the recognition and description of complex relations
among the elements in the mathematical expressions. Please consider the use of offline datasets too.

//Jinpeng: For this remark, I will do that in future work.

Please check the grammar of the following sentence:
- Page 1, column 1, second paragraph, second row: "which will defined..."

---------- REVIEW 2 ----------
PAPER: 67

TITLE: Symbol Knowledge Extraction from a Simple Graphical Language
AUTHORS: Jinpeng Li, Harold Mouchere and Christian Viard-Gaudin

OVERALL RATING: 2 (accept)
Relevance (potential interest to the ICDAR community): 5 (excellent)
Originality: 5 (excellent)
Technical soundness: 4 (good)
Experimental evaluation: 3 (fair)
Presentation: 4 (good)

Please address the following points in your review.

(1) Please describe how the work is novel. Be specific.

I'm not 100% sure this approach is novel in an absolute sense, but it definitely is in the domain of document analysis. It combines "classical" clustering/classification approaches with information theoretical metrics to achieve a potentially "optimal" graphical language vocabulary.

(2) When appropriate, provide concrete ways in which the paper could be improved.

The main criticism one could express is that the experimental basis is currently very simple (simplistic ?). It is clear that the authors want to make a case and validate a proof of concept. It will be interesting to follow this work and see how it can handle the hurdles of more complex situations (full 2D spatial relations, extensive vocabulary, more complex shapes, stroke combinations/variations/orderings for identical tokens ...)

//Jinpeng: For this remark, I will do that in future work.

It needs to be established with good experimentation setup. And of course, state-of-the-art methods must be compared.

(3) If the paper is accepted, would you attend this talk at ICDAR 2011?

Yes

(4) Other comments regarding this paper.

OVERVIEW:
The paper focuses on symbol knowledge extraction using graphical language. Graphical language uses graphemes and spatial relations. Grapheme refers to graphical units. The method is validated by using simple on-line mathematical expressions.

COMMENTS:

Question about DTW matching of strokes. Would the method correctly distinguish between '-' and '/' ? Since both are straight strokes (just slanted differently)

//Jinpeng: Yes. Each point is represented by seven features (the coordinates, the local direction, the local curvature, and the binary pen-up or pen-down information).

Section IV (B)
pp3. paragraph 1: 'A new expression e is transformed ... from the relational graph, e ? {sqk }.'
Is it sqj or sqk?

//Jinpeng: It's a new expression. Thus we use sqk instead of sqj.

Typos/Spelling

p.1 col.1 last but one paragraph "In this respect, we propose to extract automatically the relevant patterns which will 'defined' the lexical units of the
language.' replace to 'define'.

The caption of Fig. 1: 'Lexicon extraction overview' is probably better.

//Jinpeng: OK

---------- REVIEW 3 ----------
PAPER: 67
TITLE: Symbol Knowledge Extraction from a Simple Graphical Language
AUTHORS: Jinpeng Li, Harold Mouchere and Christian Viard-Gaudin

OVERALL RATING: 2 (accept)
Relevance (potential interest to the ICDAR community): 4 (good)
Originality: 4 (good)
Technical soundness: 3 (fair)
Experimental evaluation: 3 (fair)
Presentation: 3 (fair)

Please address the following points in your review.

(1) Please describe how the work is novel. Be specific.

INcreasing efficiency in symbol labelling

(2) When appropriate, provide concrete ways in which the paper could be improved.

See below

(3) If the paper is accepted, would you attend this talk at ICDAR 2011?

Possibly

(4) Other comments regarding this paper.

I cannot quite make up my mind on one fundamental assumption which seems to underpin this paper - trying to correlate with human performance is always potentially dangerous, because different individuals can often return rather different outcomes, and I am not sure how much this is a factor here.

The experimentation carried is focused on a rather narrow application area, but one which I think will be of interest to the ICDAR community. The reults presented do not seem to me to be especially impressive (I cannot be too categoric on this point because I do not know this area well enough) but no detailed comparisons are provided to help us to judge this. That said, I thought this was an interesting and potentially useful paper.