Utilizing Consistency Context for Handwritten Mathematical Expression Recognition

Kang Kim, Taik Heon Rhee, Jae Seung Lee and Jin Hyung Kim
Department of Computer Science, Korea Advanced Institute of Science and Technology, Korea
{kkim,three,jslee,jkim}@ai.kaist.ac.kr

Abstract

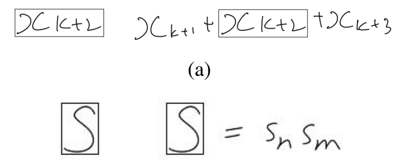
This paper presents a rule-based approach that utilizes some types of contextual information to improve the accuracy of handwritten mathematical expression(ME) recognition. Mining context from corpus is not practical for ME recognition due to the complexity originated from 2-D nature of MEs. For practicality, we identify typical types of consistencies that are often found in customary usage and general patterns in MEs. We aim to increase these consistencies in recognition results by correcting symbol labels and/or spatial relationships among symbols. Such consistencies are easily encoded as condition-action pairs. Preliminary interpretations generated by the base recognizer are reordered by increasing or decreasing scores by the rules. Although our approach is not complete, it easily implements even global context among distant symbols. Experimental results show that our approach is useful to increase the accuracy of handwritten ME recognition.

1. Introduction

Due to two-dimensional structures of mathematical expression(ME), inputting mathematical expressions by keyboard and mouse interface is inconvenient. Handwriting would be convenient to input MEs if it is supported by a good recognizer. Although several handwritten ME recognition systems have been developed [1, 6, 9, 11, 12], it is still hard to obtain robust recognition results for complex inputs.

It is well-known that context in MEs can help to improve recognition accuracy by resolving ambiguities in recognition results. Although there are some attempts [2, 5, 10], its implementation is still a challenge due to the complexity originated from 2-D nature of MEs.

Context may be grouped into three categories. First one is what we call syntactic context which is often encoded as grammars. It can be used to resolve ambiguities associated with special symbols like fractions, parentheses and operators. Second one is what we call semantic context which is spread over an entire ME. Some ambiguities in this category would be resolved by checking other parts of the ME for the coherence of similar symbols or structures. Last one is the context from physical form of handwritten input. We often observe consistencies in a users̋ writing style such as sizes and shapes. As shown in Fig.1 (a), the expression in the box is ambiguous whether it is subscripted or plain. However, by looking at the other parts of the expression, it can be easily identified as a subscripted expression. Similarly upper case and lower case ambiguity is also easily resolved by examining the size of other symbols. In natural language processing(NLP) fields, several machine learning techniques have been suggested for context analysis based on a large corpus [3, 4, 7, 8]. All these work targeted local context, such as examining fixed number of adjacent words of a target word. It is because local context is more practical in NLP and examining global context requires large computation and has to overcome data sparsity.



(a)

(b)

Figure 1. Examples of context. (a) Semantic context. The local ambiguity of the handwriting in the solid box whether x_{k+2} and $xk + 2$ is resolved in a global view. (b) Physical context. Case ambiguity of the single symbol in the solid box is resolved in a global view

In the ME recognition fields, using a corpus of printed MEs collected from web, Smirnova et al. counted frequent symbol sequences up to length 5, and used the N-gram approach to correct character recognition results [10]. Garain et al. [2] encoded several knowledge of MEs as hard decision rules such as integrals($\int$) with differentials(d). The

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.140

limitation of these approaches is that they targeted only correction of the symbol level recognition result. Structure modification has not been attempted.

Miller and Viola tried to utilize stochastic context-free grammars with A-star search [5]. But they can handle only a few syntactic context.

A general way of obtaining context is to learn from large corpus. However, learning of ME context in training seems impractical because of several difficulties. Huge amount of tagged data and computation are required due to the two dimensional nature of MEs. Furthermore, global context associated with two or more symbols located far apart may not be effectively picked up from the current data mining techniques.

Consistencies in MEs are general patterns and customary usage that commonly appear in MEs. In this paper, we propose an approach to utilize consistency-based context for improving accuracy of handwritten ME recognition. To avert from data related impracticality, our approach focuses on some consistencies that can be easily identified in customary usage and general patterns when reading or writing MEs.

Although the completeness of covering all useful consistencies cannot be guaranteed, several benefits come from this approach. The first is that there is no need of a huge corpus for training because well-known usage and patterns in MEs can be captured by human knowledge. The second is that these consistencies can cover even global context such as relationships among distant symbols. The last is that these consistencies can be easily encoded by condition-action pair rules. These favorable features provide us a practical way to utilize both of local and global context as well as structure-related consistencies.

By investigating well-known usage and patterns in MEs, we identified some consistencies suitable to encode. Then, an appropriate score revision criterion for each of the consistencies is developed and encoded as a condition-action pair. For the recognition of given handwriting input, the rules are applied for modifying recognition scores derived from the base recognition system. Consequently, rankings of alternative interpretations are changed. This demonstrates that our condition-action pair implementation is simple and effective.

The rest of this paper is organized as follows. Section 2 describes several types of consistencies in MEs that we found. Section 3 explains our representation scheme of a ME interpretation. Section 4 describes how to utilize these consistencies to a real ME recognition system. The experimental results are shown in section 5, and section 6 discusses conclusions.

Table 1. Consistencies and grouping

Form Consistency	*Size*
	Style
Frequency Consistency	*Avoiding Similar Patterns*
	Repetition
	Order
	Sequence
Structure Consistency	*Subscript*

2. Consistencies in MEs

There are frequent ME patterns found from usual practice and customary usage for convenience. We found seven consistencies from these patterns and categorized them into three groups.[1]

2.1. Form Consistency

From a handwritten input, we can obtain contextual information related physical form such as relative sizes and shapes of symbols.

Size: When writing alphabet symbols, the writer writes intentionally in a different size for each case to avoid ambiguity of uppercase and lowercase for the symbols that have similar shapes for both cases, as in c,o,s,u,v,w,x,z. We can determine unknown case by comparing the size of the unknown with unambiguous symbols.

Style: If very similar symbols appear in a single ME (such as 2,z), the writer writes each symbol with different styles to distinguish one from another, as shown in Fig.2. In other words, if two symbols in this category have different style, then we may safely assume that each symbol has different labels. From this observation we can keep the style consistency by checking the styles and the labels of all the symbols in a ME.

In order to check the style consistency, we should determine whether two styles are similar or different. Since the input data is handwritten, no two traces are exactly same. So we need a criterion to determine the style similarity of two symbols. We use five features: the number of intersections, the number of cusps, the number of strokes, and the location index of initial and end positions as [11]. We consider two styles are different if more than two features are different among five.

2.2. Frequency Consistency

Generally people choose or avoid some patterns in writing variable names and indexes.

[1] We omit one category named syntax consistency since most of previous studies have covered.

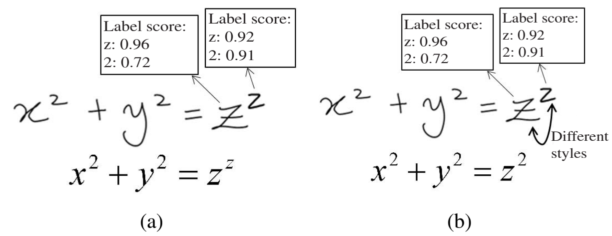


Figure 2. An example of style consistency (a) Without style consistency. Ambiguity occurs from scores of z(0.92) and 2(0.91) for the superscript of z . (b) With style consistency. Since styles for the two symbols are disparate, the symbols may have different labels.

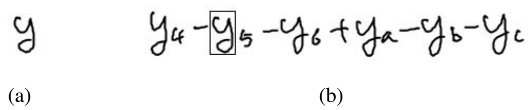


Figure 3. An example of repetition consistency (a) Ambiguity between y and g in a local view (b) Ambiguity is resolved by repetition consistency in a global view

Avoiding Similar Patterns: There are symbol pairs that the shapes and sizes are similar, such as p,P,k,K,y,Y,C,(,etc. So, people tend to avoid selecting similar symbols together.

Repetition: As MEs become longer, same labels tend to be reused. People tend to read an ambiguous symbol as one of the similar repeated labels as shown in Fig.3. We can simulate this tendency by giving additional score for frequent labels.

Order: Writers generally tend to choose alphabetically or numerically ordered labels in MEs as shown in Fig.4. If some labels make an order we can expect the rest also be ordered. From this idea we measure the degree of order in MEs and modify the symbol labels to increase the order consistency. For example, suppose a ME has five symbols. The labels are x,+,y,= respectively, but the label of the last symbol is ambiguous between z and 2. In this case, we can guess safely the label as z, since it makes an order with other alphabet labels x,y.

Sequence: Certain groups of labels unlikely follow other groups of labels with certain spatial relationship. For example, when we write y times 2, it is likely to write $2y$ instead

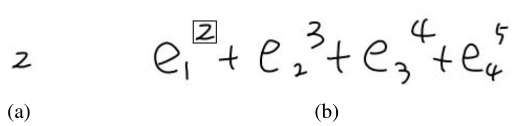


Figure 4. An example of order consistency (a) Ambiguity between 2 and z in a local view (b) Ambiguity is resolved by order consistency in a global view.

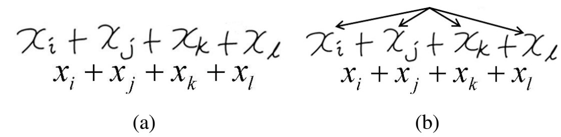


Figure 5. An example of subscript consistency (a) Without subscript consistency. The spatial relationship score for the subscript of x and i (0.7) is lower than the score for the plain(0.75). (b) With subscript consistency. All same labels must (not) have subscripts.

of $y2$. We can check the sequence consistency by examining the sequence of alphabet and numeric.

2.3. Structure Consistency

Subscript: Subscripts are often used with a common label, as shown in Fig.5. With this consistency we revise the recognition result when a ME has symbols of a common label and some of them have subscripted but others not.

3. Representation of Handwritten ME Interpretations

We adopt the representation concept of handwritten ME interpretations in the work of Rhee et al. [6]. As in the base system, An interpretation of a handwritten input is represented as Fig.6. Each node of a tree represents a single symbol. A link between two symbols represents their spatial relationships, such as *over*, *under*, *inside*, *superscript*, *subscript*, *right*(plain). A symbol is a unit of segmentation and character recognition. Each symbol is associated with a group of strokes in an input handwritten sequence. A score for a ME o can be written as

$$S(o) = \alpha_Y \sum_n S_Y(y_n) + \alpha_R \sum_m S_R(r_m)$$

where $S_Y(y_n)$ is a cost for the symbol y_n and $S_R(r_m)$ is one for the spatial relationship r_m, and α_Y and α_R are coefficients for adjusting relative portions. We can view the ME recognition process as finding the best-scored ME for given input.

4. Utilizing Consistencies to ME Recognition System

As the base recognizer represents each ME interpretation with a score, we added a new term for the context score.

$$S(o) = \alpha_Y \sum_n S_Y(y_n) + \alpha_R \sum_m S_R(r_m) + \alpha_C \sum_k S_{C_k}(o)$$

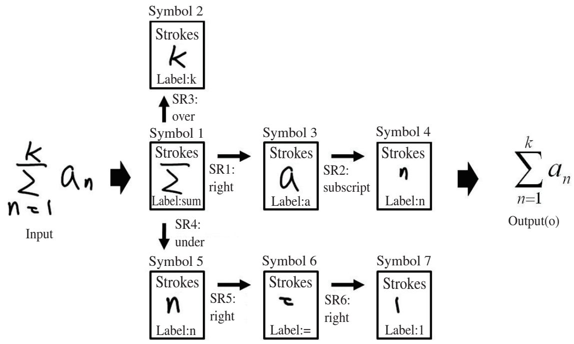


Figure 6. An example of a ME tree structure

where each $S_{C_k}(o)$ is a score of each consistency for a ME interpretation o and α_C is a revision factor that determines weight of the context score for the total interpretation score. The additional term gives a relative score to a ME according to how much it preserves consistencies.

The strategy of the score revision depends on the property of each consistency. The *repetition* and *order* consistencies are represented as a degree of complying with each consistency. The repetition degree of a ME is defined as the sum of the degrees of each symbol repetition. The degree of a symbol repetition is defined, scaled from 0 to 1, as the proportion to the number of same labels in a ME.

When a group of symbols are positioning at the same level of the structure such as a series of subscripts, we compute the degree of being ordered for the level in a ME. The order degree of a ME is sum of the order degrees of each level. The order degree of a level, from 0 to 1, is in proportion to the number of ordered symbols in the level.

The other five consistencies share the property that they are not strict but tend to be preserved in general MEs. So we define penalty for a ME interpretation in proportion to how many symbols or structures violate the consistencies.

Since the degree of impact of a context to the recognition score depends on the original score from the base recognizer, proper setting of the revision factor α_C is the key for the successful utilization of context. The sensitivity of the revision factor should be carefully examined to determine the proper value. It becomes insensitive if the factor is too small, and forces toward a wrong correction if it is too large. We did select the factor by several trials and evaluations.

Our implementation approach is revising scores and reordering the possible interpretations. After collecting multiple interpretations with associated scores, the context score is calculated for each interpretation and ranked again according to the revised scores.

The advantage of this approach is that it does not need to consider the order of the computation. The shortcoming of this approach is the time consumed for seeking multiple interpretations for checking consistency.

5. Experiments

5.1. Proposed System

The base system used for evaluation is a handwritten ME recognizer developed by Rhee et al. [6]. The base system based on a layered structure search reported its recognition accuracy as 87.7% in symbol labeling including segmentation and structuring, and 38.7% in ME level for KME-I database [6].

The consistency checking module, which sits on top of the base system, consists of seven rules corresponding to the seven consistencies mentioned earlier. To check consistencies, we modified the search module of the base system for collecting multiple interpretations. After collecting multiple interpretations, the consistency check module revises scores for the interpretations according to the rules.

The size rule checks the average sizes of unambiguous symbols and reports the number of symbols whose labels are not appropriate for their symbol sizes. The style rule computes the five features for each matched template and returns the number of symbol pairs that have same labels but different styles.

Avoiding-Similar-Pattern rule returns the number of symbol pairs that have similar sizes and shapes but different labels. The repetition rule returns the sum of degrees of each symbol repetition, where the degree of each symbol repetition has the maximum degree of 1 when the number of occurrence of the label exceeds five. The order rule returns the sum of degrees of each level, where the degree of each level has the maximum degree of 1 when more than five symbols in the level are ordered. The sequence rule returns the number of neighboring symbol pairs that the sequence of labels are variable and numeric and linked as the plain spatial relationship.

The subscript rule returns the number of labels not having subscripts where others have subscript.

The consistency check module finally returns the context score as the sum of the return values. It adds values from repetition and order rules and subtracts values from other five rules.

After the revision factor is multiplied, the context score is added to the score of the base recognizer. The system reorders candidate interpretations with new scores and output the interpretation associated with the highest score as the final recognition result.

5.2. Data Set

Since no public database is available for handwritten ME evaluation, we constructed our own database. We collected 1500 ME data consists of 15 sets of 100 high school level

Table 2. Comparison of system performance

	Incorrect symbols	Incorrect spatial relationships	Incorrect MEs
Base system	1375	878	815
Proposed system	941(31.6% reduction)	674(23.2% reduction)	635(22.0% reduction)

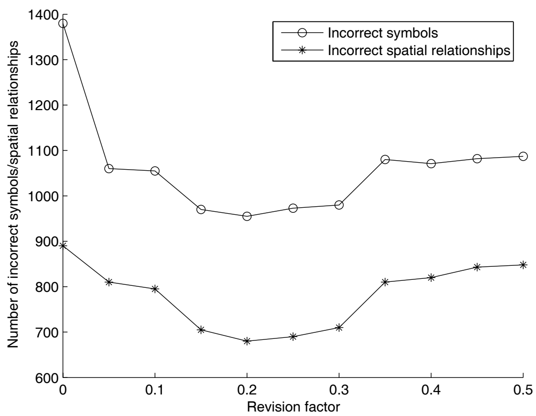


Figure 7. Effects of revision factor

ME examples from 15 writers. Each ME contains 5 to 25 symbols.

5.3. Experiment 1: Effects of Context

Our first experiment is to verify effectiveness of our context processing module. We compared the accuracies of the base and the proposed systems with the data set, allowing 1 minute for recognition of each individual ME. The proposed system collected maximum 200 output interpretations from the base recognizer for reordering. As results, 1478 MEs were recognized and 20403 of symbols and spatial relationships were retrieved. We set the revision factor α_C as 0.15. Table 2 shows recognition results of both systems. The proposed system reduced 22.0% of errors in the ME level evaluation.

5.4. Experiment 2: Effects of Revision Factor

The second experiment is to find the best revision factor for score correction with respect to consistencies. We varied α_C and verified the effects of the factor. Fig.7 shows the result of the experiment. The best results was obtained when α_C was set to around 0.2.

6. Conclusions

We have presented a rule-based approach to utilize the contextual information for handwritten ME recognition. With several rules of recognition score revision, typical types of consistencies are implemented that are often found in customary usage and general patterns in MEs. Experimental results show that our approach is useful to increase the accuracy of handwritten ME recognition.

Acknowledgment

This work is partially supported by the grant of Microsoft Research Asia under the regional theme of Mobile Computing in Education 2008.

References

[1] K.-F. Chan and D.-Y. Yeung. Mathematical expression recognition: a survey. *International Journal on Document Analysis and Recognition*, 2000.

[2] U. Garain and B. B. Chaudhuri. On machine understanding of online handwritten mathematical expressions. In *International Conference on Document Analysis and Recognition*, 2003.

[3] A. R. Golding and D. Roth. A winnow-based approach to context-sensitive spelling correction. *Machine Learning*, 1999.

[4] L. Mangu and E. Brill. Automatic rule acquisition for spelling correction. In *International Conference on Machine Learning*, 1999.

[5] E. Miller and E. Miller. Ambiguity and constraint in mathematical expression recognition. In *Association for the Advancement of Artificial Intelligence*, 1998.

[6] T. H. Rhee and J. H. Kim. Efficient search strategy in structural analysis for handwritten mathematical expression recognition. *Pattern Recognition*, 2009.

[7] J. Schaback and F. Li. Multi-level feature extraction for spelling correction. In *Workshop on Analytics for Noisy Unstructured Text Data*, 2007.

[8] M. Schierle, S. Schulz, and M. Ackermann. From spelling correction to text cleaning - using context information. In *Data Analysis, Machine Learning and Applications*, 2006.

[9] Y. Shi, H. Y. Li, and F. K. Soong. A unified framework for symbol segmentation and recognition of handwritten mathematical expressions. In *International Conference on Document Analysis and Recognition*, 2007.

[10] E. Smirnova and S. M. Watt. Context sensitive mathematical character recognition. In *International Conference on Frontiers in Handwriting Recognition*, 2008.

[11] S. M. Watt and X. Xie. Recognition for large sets of handwritten mathematical symbols. In *International Conference on Document Analysis and Recognition*, 2005.

[12] R. Yamamoyo, S. Sako, T. Nishimoto, and S. Sagayama. On-line recognition of handwritten mathematical expressions based on stroke-based stochastic context-free grammar. In *International Workshop on Frontiers in Handwriting Recognition*, 2006.