Graphical Object Recognition using Statistical Language Models

Laura Keyes, Andrew O'Sullivan and Adam Winstanley

School of Informatics and Engineering, Institute of Technology Blanchardstown, Dublin 15,
Department of Computer Science, NUI Maynooth, Maynooth, Co. Kildare
Laura.keyes/Andrew.O'Sullivan@itb.ie; *Adam.Winstanley@nuim.ie*

Abstract

This paper describes a proposed system for the recognition and labeling of graphical objects within architectural and engineering documents that integrates Statistical Language Models (SLMs) with traditional classifiers. SLMs are techniques used with success in Natural Language Processing (NLP) for use in such tasks as Speech Recognition and Information Retrieval. This research proposes the adaptation of SLMs for use with graphical notation i.e. Statistical Graphical Language Model (SGLMs). Reasoning of the similarities between natural language and technical graphics is presented and the proposed use of SGLM for graphical object recognition is described.

1. Introduction

Graphics recognition involves the recognition and structuring of geometry such as points, lines, text, symbols on graphical documents into meaningful objects for use in graphical information systems. Graphics recognition is a sub-field of pattern recognition and includes classification and recognition of graphical data based on shape description of primitive components, structure matching of composite objects and semantic analysis of whole documents. There are many types of graphical information systems including Computer-Aided Design (CAD) systems, Geographical Information Systems (GIS), Multimedia Systems and systems for document processing and analysis. These systems can be used for digital mapping, engineering drawings, graphical documents for libraries, information transmission. Common to all of these systems is the need to automatically capture, store, access and manipulate large volumes of graphical data.

The field of graphics recognition has seen many solutions proposed, applied and evaluated. Examples of such work include: Liu and Luo [1], they proposed an interactive approach using geometric constraints to represent generic graphical knowledge; Dosch et al. [2] presented a complete system for analyzing architectural drawings that makes use of several recognition techniques such as polygonal approximation, structural texture analysis and a network of constraints to recognise symbols and Cordella and Vento [3] provide a review of some of the major techniques used for symbol recognition. This paper describes a proposed framework that will apply statistical models to graphical languages based on the associations between different classes of shape [4] in a drawing to automate the structuring of graphical data.

'Shape' is an important feature that conforms to the way humans interpret and interact with real world objects. Shape description methods provide a measurement of shape that characterises an object type. Traditionally shape description was used for the recognition and classification of very specific shapes such as the classification of a particular make or model of aircraft. They can also be easily applied to vector graphical shapes. Shape information can form a significant part of the semantic content of an object within a graphical information system.

Statistical Language Models are used in Natural Language Processing (NLP), for recognising textual data. SLMs estimate the probability distributions of letters, words, sentences and whole documents within text data. These models can be used for Speech Recognition [5] and Information Retrieval [6]. Similarities can be drawn between natural language and graphical data. For instance, both consist of discrete objects that have physical form; both have a semantic component; are classified according to function and can be formed into larger components, (discussed in section 3). This work will investigate further this analogy for graphical notation used in architectural and engineering domains (for example (CAD system) to determine if SLMs may have applicability to

improve the classification of graphic objects as they do for natural language processing applications.

In this paper SGLMs for graphics recognition are introduced and discussed. Section 2 describes SLMs and *N*-grams, the most popular and powerful SLM, while Section 3 discusses the similarities between natural language and technical graphical data that support the application of SLM to graphical notations. It also describes SGLM and shows how *N*-grams can be used to build SGLM for recognition of unknown objects within architectural drawings. Section 4 describes the overall system. Section 5 concludes and discusses future work.

2. Statistical Language Models

Statistical Language Models (SLMs) are estimates of probability distributions, usually over natural language phenomena such as sequences of letters, words, sentences or whole documents. First used by Andrei A. Markov at the beginning of the 20th century to model letter sequences in Russian literature [7], they were then developed as a general statistical tool, primarily for Natural Language Processing (NLP). Automatic Speech Recognition is arguably the area that has benefited the most from SLMs [5] but they have also been used in many other fields including machine translation, optical character recognition, handwriting recognition, information retrieval, augmentative communication systems [8].

For NLP, SLMs make use of large corpora of examples of natural language. These corpora can consist of thousands or millions of words from a language. In order to be as representative as possible of a language, a corpus usually has text from a wide variety of sources. For example the derived *Brown Corpus* [7] contains one million words taken from fifteen different sources such as legal text, scientific text and press reportage. However, a corpus can be built to just include a particular sub-set of language, if so required for a particular task. Generally the larger the corpus the better it will be for statistical language models [7].

There are different types of SLMs that can be used. These include *N-gram* models that use an estimate of the probability distribution of the occurrence of sequences of words. In general they are represented as a conditional probability of each specific word given the previous sequence of words. *Decision Tree* models [9], which assign probabilities to each of a number of choices based on the context of decisions. Some SLM techniques are derived from grammars commonly used by linguists. For example Sjilman et al. [10] use a declarative grammar to generate a language model in order to recognise hand-sketched digital ink. Other methods include *Exponential* models and *Adaptive* models. Rosenfeld [11] suggests that some other SLM techniques such as *Dependency* models,

Dimensionality reduction and *Whole Sentence* models show significant promise. However this research will focus on the most powerful of these models, *N*-grams and their variants.

2.1. *N*-gram models for SLMs

N-gram models are the most widely used SLM technique. In NLP, *N*-grams are used to predict words based on the previous *N*-1 words. They estimate the probability distribution *P(s)* of words within a corpus by using the frequency of the words and the frequency of their co-occurrences with other words. Generally a *bigram* model (*N*=2) or a *trigram* model (*N*=3) is used. A bigram models estimates the probability

$$P(w_i \mid w_{i-1}) \qquad (1)$$

where w_i is the word to be predicted and w_{i-1} is the previous word. The probability is estimated using relative frequencies so

$$P(w_i \mid w_{i-1}) = C(w_{i-1}w_i)/C(w_{i-1}) \qquad (2)$$

where C is the frequency of words within the corpus. Likewise a tri-gram model estimates the probability

$$P(w_i \mid w_{i-1}, w_{i-2}) \qquad (3)$$

by computing

$$C(w_{i-2}, w_{i-1}w_i)/C(w_{i-2}, w_{i-1}) \qquad (4)$$

For example, if the next word in the sentence

"*I was walking the*"

is missing, a trigram model can predict what the next word will be. The trigram estimates the probabilities of phrases of words within a corpus being the missing word by computing

$$P(w_i \mid walking, the) =$$
$$C(walking, the, w_i)/C(walking, the) \qquad (5)$$

This computes the frequency of three word phrases beginning with "*walking the*" divided by the frequency of the phrase "*walking the*". The word which results in the highest frequency and hence the highest probability is judged to be the next word in the sentence.

3. SLM for Graphical Notations

SLMs have previously been used almost exclusively for NLP. There are sufficient similarities between natural language and graphical notations that suggest that adapting SLMs to become SGLMs is a worthwhile approach [12]. Recent work applied SLMs to the automatic structuring of topographic data [13] for GIS. In their work Winstanley and Salaik characterise the similarities that can be drawn between topographic data and natural language. A similar analogy can be used for natural language and graphical data found on architectural or engineering plans. Both consist of discrete objects (words, graphical objects) and these objects:

- have a physical form (for example spelling, object shape);
- have a semantic component (meaning, graphical object label);
- are classified according to function (part of speech, object class) and
- are also formed into larger components (sentences/paragraphs, diagrams/documents).

There are of course differences between natural language and graphical notation. One major difference which has to be addressed when developing SGLMs, is that language is naturally a one-dimensional sequence of symbols, whereas graphics are inherently multi-dimensional. This difference is significant in the context of using *N*-grams because with natural language the choice of words to use for prediction is an easy one, that is, the previous words. With graphics however, the choice is not so easy. One approach of dealing with this is to use adjacency relationships between objects on a document. For example, if *N*-grams are being used to predict an unknown object on a CAD diagram the decision must be made as to which of the adjacent objects to use for the prediction. Adjacency can be determined by using neighbouring objects (objects on documents that physically touch each other) or by the connection line between objects (for example the connection line between a capacitor and fuse). The topology or relationship between objects on graphical document is important for developing SGLM. Using the concept of adjacency *N*-gram phrases can be extracted and tables of frequencies constructed.

3.1 Statistical Graphical Language models

In order to use SLM techniques such as *N*-gram models a corpus of graphical data must first be constructed. For NLP applications the corpora consist of a large amount of text data. For this research the corpus used will consist of engineering and architectural documents. The model is used to increase the effectiveness of other shape (and other possible) recognisers and as such the data will also contain some unclassified and misclassified data.
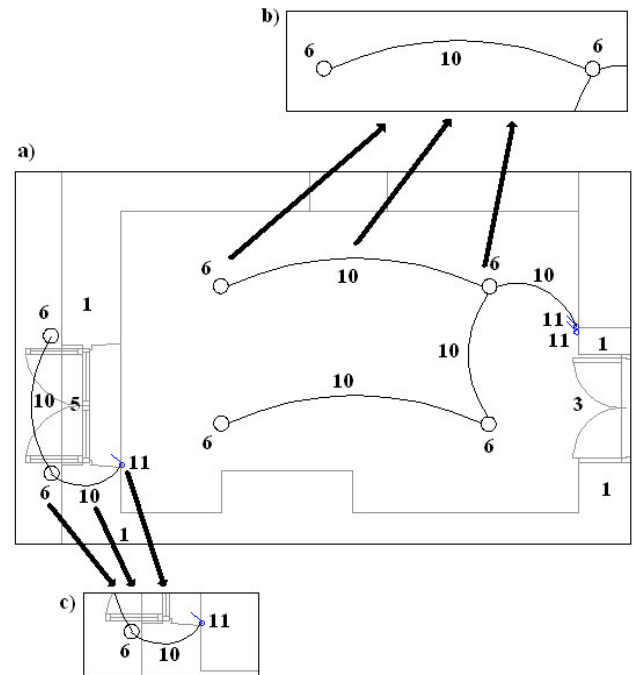

Figure 1. a) Sample of document with objects labeled, b) *Light bulb – Wire – Light bulb* phrase, c) *Light bulb – Wire – Switch* phrase

A recognition system produces probabilities for candidate classes of each object based on their shape. The SLM, built from analysis of another data set, uses the probabilities to construct "phrases" of objects. Figure 1 depicts a sample architectural document with objects labelled 1:Lightbulb, 2:Wire, 3:Wall, 4:Double Door, 5: Double Window, 6: Switch. Figure 1 b) and c) show example tri-gram phrases that can be constructed from the objects in a). The phrases are

Light bulb – Wire – Light bulb
Wall – Double Door – wall.

All possible phrases such as these are searched for in the corpus of data and their frequencies are recorded in the N-gram tables. When the N-gram tables have been constructed, they can be used for the recognition of unknown graphical objects.

In this example, as there are fourteen objects being used it consists of fourteen two-dimensional tables. Each table contains the frequencies of phrases beginning or ending with one of the objects. For example Table 1

below stores the frequencies of phrases that either end or begin with a *Wall* object. The value 64 in cell (10,11) means that there are 64 occurrences of the phrase:

"Wire – Switch – Wall" (or *"Wall – Switch –Wire")* within the corpus.

Table 1. Frequencies of phrases ending or beginning with a *Wall* object

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	36	4	29	6	0	0	0	0	0	0	0	0	2
2	0	0	0	0	0	0	0	0	0	1	0	0	0	0
3	0	0	4	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	4	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	5	0	0	10	0	0	6	0	64	16	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0

One problem associated with *N*-grams is the data sparseness problem. This occurs because no matter how big the corpus, there will always be events that do not occur within it but that may occur in the future. For example, all the zeroes within Table 1 above represent phrases of objects that did not occur within the corpus. It should be noted that the high amount of zeroes is because of the limited size of the corpus used for this example but whatever the size there will always be such zero frequencies. These phrases may however occur in the future but as their probability is zero they will not be considered in any prediction process. To solve this problem *Smoothing* is used. There are several *Smoothing* techniques available but here *Add-One Smoothing* is used [7]. This is a simple technique where the value '1' is added to all the frequencies stored in the trigram table.

Table 2. Total frequencies of pairs of objects that occur in triples

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	14	50	27	43	20	24	14	14	20	19	78	30	14	16
2	14	14	14	14	14	14	14	14	14	27	14	14	14	14
3	15	14	18	14	14	14	14	14	14	18	14	14	14	14
4	14	14	14	18	14	20	14	14	14	20	14	14	14	14
5	14	14	14	14	14	17	14	14	14	15	14	14	14	14
6	14	14	14	14	14	14	14	14	14	185	14	14	14	14
7	14	14	14	14	14	14	14	14	14	18	14	14	14	14
8	14	14	14	14	14	14	14	14	14	16	14	14	14	14
9	14	14	14	14	14	14	14	14	14	18	14	14	14	14
10	14	14	19	14	14	33	14	14	20	14	78	30	15	14
11	14	14	14	14	14	14	14	14	14	83	14	14	14	14
12	14	14	14	14	14	14	14	14	14	25	14	14	14	14
13	14	14	14	14	14	14	14	14	14	20	14	14	14	14
14	18	14	14	14	14	14	14	14	14	14	14	14	14	14

The frequencies of phrases within the corpus are now known but in order to compute the probabilities of phrases a table of total frequencies must be constructed. These total frequencies are the frequencies of pairs of objects occurring as part of triples of objects within the corpus. Table 2 shows the total frequency table that has been computed by summing up the fourteen frequency tables. The value 185 in cell (6,10) means that there are 185 occurrences of triples of objects that begin with:

Light bulb –Wire (or that end in *Wire – Light bulb*).

The frequency tables computed can now be used to estimate probability values for unclassified objects.

4. Graphical Object Recognition System

Within the graphical object recognition system, the underlying classifier is based on shape recognition techniques. Shape analysis methods play an important role in systems for object recognition and representation. Previous work has evaluated the recognition and labeling of objects and components on drawings and plans based on their shape [4]. Shape description methods used include Fourier Descriptors (FD) and Moment Invariants (MI) and Scalar Descriptors (for example, area, elongation, number of corners etc.). These techniques are applied to object boundaries extracted from drawings represented as vector descriptions. The output obtained by the description methods provides a measurement of shape that characterises the object type. This provides a list of candidate classes of each object.
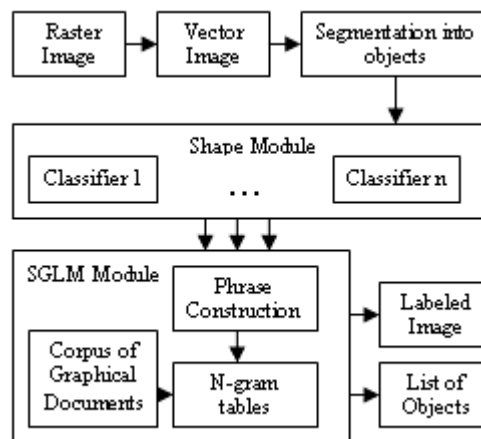

Figure 2. Graphical object recognition system

The SGLM model is combined with this score to improve the likelihood that the classification is correct or re-classify incorrect or misclassified features. Figure 2

shows the configuration of the recognition system and the role of SCLM within this system. Combination of shape scores with SGLM will be evaluated on different methodologies. Classifier combination techniques from pattern recognition such as *Borda Count*, and *Borda Count Ratio* [4] will be tested. Also an extension of *N*-grams used in NLP is to count the *part-of-speech* of the word (noun, verb and so on) rather than the word itself. This *N*-gram part-of-speech tagging model can be used with shape for graphical data, where the tag is some descriptive classification of the graphical object. For example:

$$P(object\ shape\ |\ tag)*P(tag\ |\ neighbouring\ k\ tags)$$

That is, the shape probability of an object belonging to a particular class combined with the likelihood that the class would have the observed neighbouring class (of neighbours up to k deep). So in this way SGLMs may be used to improve the performance of other recognisers applied to the data.

5. Discussions and Future Work

This paper has proposed the adaptation of Statistical Language Models for recognition of graphical objects within architectural and engineering documents. Previously used for Natural Language Processing there exists similarities between natural language and technical graphical data that suggest that Statistical Graphical Language Models is a worthwhile approach. Digitised CAD drawings are processed to extract their component objects from which shape descriptions are built. These feed into several description and matching algorithms, each of which produces one or more candidate categories to which each object may belong. A fusion algorithm produces an overall consensus decision giving a ranked list of candidate types. The SGLM module can then be used to improve the performance of the recognisers.

An example to demonstrate how *N*-grams can be used to build a *trigram* table for the purpose of recognising unclassified graphical objects was shown. This example used a simple corpus of two architectural documents but the future work will build and utilise a substantial corpus. It is proposed to develop SGLMs and evaluate rigorously their applicability for graphical recognition in combination with existing recognisers. Other SLM techniques such as Hidden Markov Models and Part-of-Speech tagging will be also investigated for use with graphical notations.

This work has relevance to sectors that collect, supply or use graphical data in digital form. There are enormous amounts of data in paper form, examples come from surveying, mapping, architecture, engineering and multimedia systems. It is envisaged that this research will result in a system that can be used in various configurations for different application domains.

5. References

[1] Y. Liu, W. Luo, "Interactive Recognition of Graphical Objects in Engineering Drawings", GREC 2003, LNCS 3088, pp.128-141, 2004

[2] P. Dosch, K. Tombre, C Ah-Soon, G. Masini, "A complete system for the analysis of architectural drawings", IJDAR (2000) 3: 102-116

[3] L.P Cordella, M.Vento, "Symbol recognition in documents: a collection of techniques?", *IDJAR* (2000) 3: 73-88

[4] Keyes, L., Winstanley, A., "Shape Description for Automatically Structuring Graphical Data", in Josep Llados, Y.B. Kwon (eds), Graphics Recognition – Recent Advances and Perspectives, LNCS 3088, 353-262, Springer-Verlag, 2004

[5] F. Jelinek, *Statistical Methods for Speech Recognition*. MIT Press 1997

[6] J.M. Ponte, W.B. Croft, "A language modeling approach to information retrieval", *Proceedings of the 6th international conference on research and development in information retrieval (SIGIR'98)* 1998, 276-281

[7] Manning, C.,D., and Schutz, H., *Foundations of Statistical Natural Language Processing*, MIT Press, Cambridge, 2001.

[8] Jurafsky, D. and .Martin, J.H., *Speech and Language Processing*, Prentice-Hall, 2000.

[9] Bahl, L R., Brown, P. F., Peter V. de Souza and R. L. Mercer., "A tree-based statistical language model for natural language speech recognition." *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37:1001-1008, July 1989.

[10] Shilman, M., Pasula, H., Russell, S. and Newton, R., "Statistical Visual Language Models for Ink Parsing." *AAAI Spring 2002 Symposium on Sketch Understanding*, 2002.

[11] Rosenfeld, R., "Two Decades of Statistical Language Modeling: Where Do We Go From Here?", *Proceedings of the IEEE*, 88 (8), pp 1270-1278, 2000.

[12] Andrews, J.H., Maps and language, A metaphor extended, *Cartographic Journal*, 27, 1–19, 1990.

[13] Winstanley, A., B. Salaik, L. Keyes: "Statistical Language Models For Topographic Data Recognition", *IEEE international Geoscience and Remote Sensing Symposium* (IGARSS'03), July 2003.

IEEE
COMPUTER
SOCIETY