SVM Classification of Neonatal Facial Images of Pain

Sheryl Brahnam[1], Chao-Fa Chuang[2], Frank Y. Shih[2], and Melinda R. Slack[3]

[1] Missouri State University, Computer Information Systems, 901 South National, Springfield MO 65804, USA
Shb757f@smsu.edu
[2] New Jersey Institue of Technology, University Heights, Newark, NJ 07102
{cxc1235, shih}@njit.edu
[3] Medical Director of Neonatology, St. John's Hospital, 1235 E. Cherokee, Springfield, MO 65894, USA
Melinda_slack@pediatrix.com

Abstract. This paper reports experiments that explore performance differences in two previous studies that investigated SVM classification of neonatal pain expressions using the Infant COPE database. This database contains 204 photographs of 26 neonates (age 18-36 hours) experiencing the pain of heel lancing and three nonpain stressors. In our first study, we reported experiments where representative expressions of all subjects were included in the training and testing sets, an experimental protocol suitable for intensive care situations. A second study used an experimental protocol more suitable for short-term stays: the SVMs were trained on one sample and then evaluated on an unknown sample. Whereas SVM with polynomial kernel of degree 3 obtained the best classification score (88.00%) using the first evaluation protocol, SVM with a linear kernel obtained the best classification score (82.35%) using the second protocol. However, experiments reported here indicate no significant difference in performance between linear and nonlinear kernels.

1 Introduction

Accurate assessment of pain in neonates is a difficult yet crucial task. The clinical definition of pain assumes the person experiencing pain has the ability to articulate the location, duration, quality, and intensity of their pain experience. Although nonverbal self reporting methods have been devised that allow preverbal children to indicate their pain levels by pointing to abstract renditions of facial expressions expressive of increasing levels of discomfort, neonates must rely exclusively on the proxy judgments of others [3].

Several pain assessment measures have been developed to assist clinicians in diagnosing neonatal pain. Most of these instruments rely on the neonate's facial displays. Facial displays are considered the gold standard of pain assessment [4] because they are the most specific and consistent indicators of pain. The facial characteristics of neonatal pain displays include prominent forehead, eye squeeze, naso-labial furrow, taut tongue, and an angular opening of the mouth [5]. Despite the fact that neonatal facial displays of pain are the most reliable source of pain assessment, instruments

I. Bloch, A. Petrosino, and A.G.B. Tettamanzi (Eds.): WILF 2005, LNAI 3849, pp. 121 – 128, 2006.

based on facial displays are unsatisfactory because clinicians tend to underrate pain intensity [6] and often fail to utilize all the information available to them in the infants facial signals [7].

In an attempt to bypass the unreliable observer, our research group is investigating the potential benefits face recognition technology would offer pediatric clinicians in diagnosing neonatal pain. Applying face recognition techniques to medical problems is a novel application area. Gunaratne and Sato [17] have used a mesh-based approach to estimate asymmetries in facial actions to determine the presence of facial motion dysfunction for patients with Bell's palsy, and Dai et al. [12] have proposed a method for observing the facial expressions of patients in hospital beds. The facial images used in the Dai et al. study, however, were not of actual patients but rather of subjects responding to verbal cues suggestive of medical procedures and conditions. Our work with neonatal pain expressions is the only other research we are aware of that uses face recognition techniques to diagnose medical problems.

We began work on this problem by developing the Infant COPE database. The facial displays of 26 neonates between the ages of 18 hours and 3 days old were photographed experiencing the pain of a heel lance and a variety of stressors, including transport from one crib to another, an air stimulus on the nose, and friction on the external lateral surface of the heel.

In our initial study [1], three face classification techniques, Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Support Vector Machines (SVMs), were used to classify the faces into two categories: pain and nonpain. The training and testing sets contained multiple samples of each subject in each expression category. No two samples were identical as each varied slightly in angle and facial configuration. While, ideally, as is the case with speech recognition software, samples of individual subjects would be available to personalize the classifier, in a clinical setting this is not practical as the typical newborn's stay is short-term. The evaluation protocol used in our first study would probably only be applicable in intensive care situations where neonates have longer stays that present opportunities for collecting facial samples. It is more realistic to assume that the classifier will need to be trained on one set of subjects and then applied out of the box to future newborns. In [2], an evaluation protocol was developed that evaluated trained classifiers using unknown subjects.

Results of the two studies were contradictory in terms of the best kernel to use with SVM. An SVM with polynomial kernel of degree 3 obtained the best classification score (88.00%) in the first study, and an SVM with a linear kernel obtained the best classification score (82.35%) in the second study. Sampling error caused by the small number of images in the sample pool is one possible explanation for this discrepancy. A set of new experiments using the first protocol was designed to explore sample error. The results of these experiments, reported in section 4, suggest that there is no significant difference in the performance of an SVM with a linear kernel and an SVM with a polynomial kernel of degree 3.

In section 2, we describe of the facial displays in the infant COPE database more completely. In section 3, we outline the two experimental protocols, designated A and B, used in the earlier studies. In section 4, we compare SVM classification rates