**Problem and Idea**

- Training neural networks when we have a large set of data with weak labels and a small amount of data with true labels.
- We train two neural networks: a target network and a confidence network. The target network is optimized to perform a given task, trained on a large set of unlabeled data that are weakly annotated. We propose to control the magnitude of the gradient updates to the target network using the scores provided by the second confidence network trained on a small amount of supervised data.

**Model Architecture and Training**

- Both networks are trained in a multi-task fashion alternating between the full supervision and the weak supervision mode. In the full supervision mode, the parameters of the confidence network get updated using batches of instances from training set with true labels, while in the weak supervision mode the parameters of the target network are updated using training set with weak labels and confidence score from the confidence network. The representation learning layer is shared between two networks.

**Experimental Setups**

- We apply our method to sentiment classification task.
  - The weak annotator is a simple unsupervised lexicon-based method which averages over predefined sentiment score of words in the sentence.
  - Dataset: For train/test our model, we use SemEval-13, SemEval-14, SemEval-15, twitter sentiment classification task. We use a large corpus containing 50M tweets collected during two months an unlabeled set.
  - The target network is a convolutional model in which the inputs are first passed through an embedding layer mapping the sentence to a matrix, followed by a series of 1d convolutional layers with max-pooling, followed by feed-forward layers and a softmax output layer which returns the probability distribution over all the classes.
  - Baselines:
    - (WA) Weak Association: i.e. the unsupervised method that we used for evaluating the unlabeled data.
    - (WS) Weak Supervision Only: i.e. the target network trained only on weakly labeled data.
    - (WSO) Full Supervision Only: i.e. the target network trained only on true labeled data.
    - (WSF) Weak Supervision + Fine Tuning: i.e. the target network trained on the weakly labeled data and then tuned on true labeled data.
    - (CWS) Confidence Weighted Supervision: this is similar to our proposed method but without the confidence network giving the confidence scores as the training instances, there is a robust gradient descent which is trained on set V to train the weak labels of the instances in V to the true labels. The new labels are then used as the target labels of the target network.
    - (LWA)Our model with different training setup: Weak Association. i.e. we consider the confidence network as a separate network, without sharing the representation learning layer, and train it as i.e. We then train the target network on the combined weak supervision signals.
    - (LWAF)Our model with different training setup: Weak Association + Fine Tuning. i.e. we consider the confidence network as a separate network, without sharing the representation learning layer, and train it as i.e. We then train the target network on the combined weak supervision signals.

**Results**

- Controlling the effect of supervision to train neural networks not only improves the performance, but also provides the network with more solid signals which speeds up the training process.

---

**Table 1:** Performance of the baseline models as well as our proposed model on different datasets in terms of F1-score.

<table>
<thead>
<tr>
<th>Method</th>
<th>SemEval-14</th>
<th>SemEval-15</th>
</tr>
</thead>
<tbody>
<tr>
<td>WA Performance</td>
<td>0.5141</td>
<td>0.4471</td>
</tr>
<tr>
<td>WSO</td>
<td>0.6319</td>
<td>0.5060</td>
</tr>
<tr>
<td>WSO</td>
<td>0.6307</td>
<td>0.5811</td>
</tr>
<tr>
<td>WSO+FT</td>
<td>0.7008*</td>
<td>0.6441*</td>
</tr>
<tr>
<td>LJ</td>
<td>0.7131*</td>
<td>0.6433*</td>
</tr>
<tr>
<td>LJWFA</td>
<td>0.7158*</td>
<td>0.6501*</td>
</tr>
<tr>
<td>LJWAFT</td>
<td>0.7162*</td>
<td>0.6436*</td>
</tr>
<tr>
<td>SemEval+</td>
<td>0.7362*</td>
<td>0.6618*</td>
</tr>
</tbody>
</table>

**Additional Notes:**

- The loss of the target network and the confidence network compared to the loss of the weakly supervised model is consistent of 50% of the loss on training data in sentiment classification.

---

**Figure:**

- Performance of the model on different datasets in terms of F1-score.