# Classifier Combination

Computational Social Intelligence - Lecture 20

Prof. Alessandro Vinciarelli
School of Computing Science &
Institute of Neuroscience and Psychology

http://www.dcs.gla.ac.uk/vincia
Alessandro.Vinciarelli@glasgow.ac.uk

University of Glasgow

EPSRC
Engineering and Physical Sciences Research Council

FNSNF

This lecture is based on the following text (available on Moodle):

- Vinciarelli & Esposito, "Multimodal Analysis of Social Signals", in "The Handbook of Multimodal-Multisensor Interfaces", Oviatt et al. (eds.), 203-226, ACM, 2018

# Outline

- Quick Recap
- Late Fusion (Sum Rule)
- Variants of Late Fusion
- Conclusion

# Outline

- Quick Recap
- Late Fusion (Sum Rule)
- Variants of Late Fusion
- Conclusion

There are no changes for the priors (they do not depend on the input vectors)

$$\mathcal{C}^* = \arg\max_{\mathcal{C}_k \in \mathcal{C}} p(\vec{x}_1, \ldots, \vec{x}_R | \mathcal{C}_k) p(\mathcal{C}_k)$$

The likelihood must be changed to reflect the presence of multiple feature vectors

The assumption is that the input vectors are statistically independent given the class

Product over all sensors

$$p(\vec{x}_1, \ldots, \vec{x}_R | C_k) = \prod_{j=1}^{R} p(\vec{x}_j | C_k)$$

If one term is close to zero, the entire product is close to zero

# Outline

- Quick Recap
- Late Fusion (Sum Rule)
- Variants of Late Fusion
- Conclusion

The Bayes Theorem

$$p(\vec{x}_j|C_k) = \frac{p(C_k|\vec{x}_j)p(\vec{x}_j)}{p(C_k)}$$

The posterior is assumed to approximate the prior

The absolute value is significantly smaller than one

$$p(\mathcal{C}_k|\vec{x}_j) \simeq p(\mathcal{C}_k)(1+\delta_{jk})$$

The expression of the posterior can be changed

The Bayes Theorem

$$p(\vec{x}_j|C_k) = \frac{p(C_k|\vec{x}_j)p(\vec{x}_j)}{p(C_k)}$$

$$p(\vec{x}_j|C_k) = \frac{p(\cancel{C_k})(1+\delta_{jk})p(\vec{x}_j)}{p(\cancel{C_k})}$$

The assumption is that the input vectors are statistically independent given the class

Product over all sensors

$$p(\vec{x}_1, \ldots, \vec{x}_R | C_k) = \prod_{j=1}^{R} p(\vec{x}_j | C_k)$$

If one term is close to zero, the entire product is close to zero

$$p(\vec{x}_1, \ldots, \vec{x}_R | \mathcal{C}_k) = \prod_{j=1}^{R} (1 + \delta_{jk}) p(\vec{x}_j)$$

The factors of the product are rearranged

This term does not depend on the class

$$\prod_{j=1}^{R}(1+\delta_{jk})p(\vec{x}_j) = \prod_{j=1}^{R}(1+\delta_{jk})\prod_{j=1}^{R}p(\vec{x}_j)$$

This product can be neglected because it has always the same value for all classes

This term can be neglected because the delta's are small

$$\prod_{j=1}^{R}(1+\delta_{jk}) = (1+\delta_{1k}+\delta_{2k}+\delta_{1k}\delta_{2k})\prod_{j=3}^{R}(1+\delta_{jk}) \simeq (1+\delta_{1k}+\delta_{2k})\prod_{j=3}^{R}(1+\delta_{jk})$$

The terms of the product that include several delta's can be neglected

$$\prod_{j=1}^{R}(1+\delta_{jk}) \simeq 1+\sum_{j=1}^{R}\delta_{jk}$$

The posterior is assumed to approximate the prior

The absolute value is significantly smaller than one

$$p(\mathcal{C}_k|\vec{x}_j) \simeq p(\mathcal{C}_k)(1 + \delta_{jk})$$

$$\delta_{jk} = \frac{p(\mathcal{C}_k | \vec{x}_j)}{p(\mathcal{C}_k)} - 1$$

The expression of the delta's is replaced in the product of the likelihoods

$$\prod_{j=1}^{R} p(\vec{x}_j|\mathcal{C}_k) = 1 + \sum_{j=1}^{R}\left[\frac{p(\vec{C}_k|\vec{x}_j)}{p\mathcal{C}_k} - 1\right]$$

There are no changes for the priors (they do not depend on the input vectors)

$$\mathcal{C}^* = \arg\max_{\mathcal{C}_k \in \mathcal{C}} p(\vec{x}_1, \ldots, \vec{x}_R | \mathcal{C}_k) p(\mathcal{C}_k)$$

The likelihood must be changed to reflect the presence of multiple feature vectors

# The Sum Rule

$$\mathcal{C}^* = \arg\max_k (1-R)p(\mathcal{C}_k) + \sum_{j=1}^{R} p(C_k|\vec{x}_j)$$

The sum of the
posteriors when using
the individual modalities

# Outline

- Quick Recap
- Late Fusion (Sum Rule)
- Variants of Late Fusion
- Conclusion

The product of the posteriors is bound by the minimum of the posteriors

$$\prod_{j=1}^{R} p(\mathcal{C}_k|\vec{x}_j) \leq \min_j p(\mathcal{C}_k|\vec{x}_j) \leq$$

$$\leq \frac{1}{R}\sum_{j=1}^{R} p(\mathcal{C}_k|\vec{x}_j) \leq \max_k p(\mathcal{C}_k|\vec{x}_j)$$

The sum of the posteriors is bound by the maximum of the posteriors

$$
\begin{aligned}
\mathcal{C}^* &= \arg\max_k (1-R)p(\mathcal{C}_k) + \sum_{j=1}^{R} p(\mathcal{C}_k|\vec{x}_j) \\
&= \arg\max_k (1-R)p(\mathcal{C}_k) + R\max_k p(\mathcal{C}_k|\vec{x}_j)
\end{aligned}
$$

The Max Rule

When the priors are
uninformative

$$\mathcal{C}^* = \arg\max_{jk} p(\mathcal{C}_k | \vec{x}_j)$$

Max Rule when the
priors are uninformative

There are no changes for the priors (they do not depend on the input vectors)

$$\mathcal{C}^* = \arg\max_{\mathcal{C}_k \in \mathcal{C}} p(\vec{x}_1, \ldots, \vec{x}_R | \mathcal{C}_k) p(\mathcal{C}_k)$$

The likelihood must be changed to reflect the presence of multiple feature vectors

$$\mathcal{C}^* = \arg\max_k p(\mathcal{C}_k) \prod_{j=1}^{R} p(\vec{x}_j | \mathcal{C}_k)$$

The likelihood is rewritten using the Bayes Theorem

The Min Rule

$$C^* = \arg\max_k \prod_{j=1}^{R} p(C_k|\vec{x}_j) =$$

$$= \arg\max_k \min_j p(C_k|\vec{x}_j)$$

The class where the minimum of the posteriors is the highest

# The Sum Rule

$$\mathcal{C}^* = \arg\max_k (1 - R)p(\mathcal{C}_k) + \sum_{j=1}^{R} p(C_k|\vec{x}_j)$$

The sum of the
posteriors when using
the individual modalities

$$\mathcal{C}^* = \arg\max_k (1-R)p(\mathcal{C}_k) + \frac{1}{R}\sum_{j=1}^{R} p(\mathcal{C}_k|\vec{x}_j)$$

The average can be
noisy when R is small

$$\mathcal{C}^* = \arg\max_k (1 - R) p(\mathcal{C}_k) + M_j p(\mathcal{C}_k | \vec{x}_j)$$

The median of the posteriors for a given class

# Outline

- Quick Recap
- Late Fusion (Sum Rule)
- Variants of Late Fusion
- Conclusion

# Conclusions

- The combination of multiple classifiers is the methodology underlying multimodal approaches;
- The early fusion works when the number of feature vectors is the same across multiple modalities;
- The late fusion works when the number of feature vectors is different for different modalities.