We address weakly supervised action alignment and segmentation in videos, where only the order of occurring actions is available during training. We propose