

Score (harsh) vs Confidence Threshold: train, p1qa

