ESRM algorithm. Clearly, the results show the effectiveness of the proposed ESRM algorithm for the summarization process. The validation of this is accomplished on public data set Microsoft Research Paraphrase (MSRP)[2] corpus. The MSRP consists of 1,725 test pairs and 4,076 training pairs. The pairs were automatically collected from thousands of news sources. Then subsequently labeled by two human annotators who determined whether the two sentences in a pair were semantically equivalent or not.

On the other hand, Document Understanding Conference (DUC)[3] has organized yearly evaluation of document summarization. The standard summarization benchmark DUC2005, DUC2006 and DUC2007 data sets are used for validating the proposed AM-DES framework. DUC 2005 used in these experiments is partitioned into 50 topic sets, each containing 25–50 documents. DUC2006 contains 50 document sets while DUC2007 contains 45 document sets. Every document set in DUC2006, and DUC2007 has 25 news articles. Each document set consists of several articles written by various authors, which is also the ground truth of the evaluation. Every sentence is either used in its entirety or not at all for constructing a summary. The length of a result summary is limited by 250 tokens [61].

### 4.3 Experimental Results

In this section, the results of the experiments are analyzed in details. Two experiments are done using the evaluation data sets outlined in the previous section. Experiment 1are used to validate the similarity, and they prove that the proposed ESRM algorithm performs the best results among the compared methods. As well, Experiment 2 is used to validate the effectiveness of the proposed AM-DES framework. The efficiency is achieved through selecting the most relevant sentences with the least redundancy.

#### 4.3.1 Experiment 1

In MSRP, the proposed ESRM determines the number of correctly identified paraphrase pairs in the corpus and compares the result with STS approach [59] and LG approach [64]. The effectiveness of the proposed ESRM algorithm is measured by two quantities and one combined measure, named “recall” and “precision” rate. Fig 8 depicts the precision, recall, and F-measure versus similarity threshold practical values.

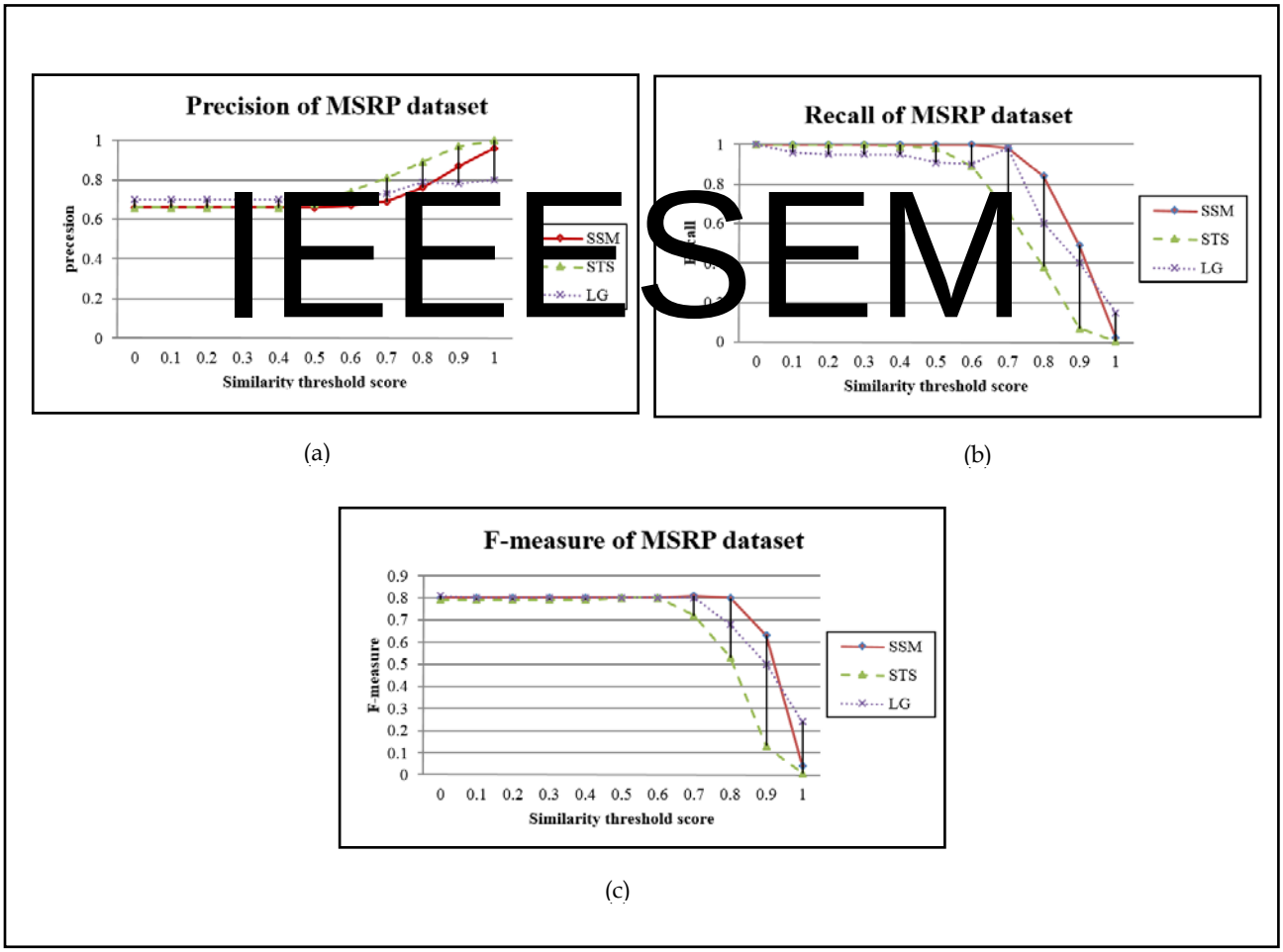


Figure 8 Comparison results on MSRP data set

[2] http://research.microsoft.com/en-us/downloads/607d14d9-20cd-47e3-85bc-a2f65cd28042/
[3] http://duc.nist.gov/data.html