To solve our problem of classifying crashes, first we have to think how represent them in our machine learning algorithm. Server gets the report with several fields containing:
1) The name and version of crashed application, along with exit code
2) System version information (kernel and system version, installed modules)
3) stderr output, consisting of several lines of text

First two are easy to feed to the classifier, as they are primarily numbers or proper names (libraries and applications), but the third, as important as them, is just a variable length blob of text. This is where in my opinion paragraph2vec (extension of word2vec) algorithm comes to use.

My idea is to use paragraph2vec algorithm on text data, then extend the vector using information from 1) and 2) and then use constrained spectral clustering to obtain labels.

A comparison [3] shows that the spectral analysis is currently one of the best clustering algorithms, and with constrained SC [4] we can incorporate prior knowledge. Our data is high-dimensional, but due to the use of spectral clustering it shouldn't be much of a problem (PCA step). We will probably be forced to modify the original approach described in the paper, since our task will require providing “cannot-link constraints” (as opposed to “must-link constrains” described in [4], indicating that two elements are in the same cluster).

Paragraph2vec (doc2vec) is already implemented in gensim package (python), I couldn't find any python constrained spectral clustering algorithm, so it is possible we'll have to implement it ourselves.

Links:

Short explanation of spectral analysis: https://www.youtube.com/watch?v=P-LEH-AFovE
Word2vec introduction paper http://arxiv.org/pdf/1411.2738v1.pdf
Word2vec introduction ipython notebook
https://github.com/fbkarsdorp/doc2vec/blob/master/doc2vec.ipynb

[1] A p2v algorithm with w2v algorithm description http://arxiv.org/pdf/1405.4053v2.pdf
[2] A comparison between p2v and other text analysis algorithms + an improvement idea for p2v http://arxiv.org/pdf/1507.07998v1.pdf
[3] A comparison between different clustering algorithms http://arxiv.org/pdf/1511.09123v1.pdf
[4] Constrained spectral clustering overview https://dl.acm.org/citation.cfm?id=1148241&dl=ACM&coll=DL&CFID=759388251&CFTOKEN=72271786