

Internal distribution code:

- (A) [-] Publication in OJ
(B) [-] To Chairmen and Members
(C) [-] To Chairmen
(D) [X] No distribution

**Datasheet for the decision
of 16 November 2015**

Case Number: T 0022/12 - 3.5.01

Application Number: 99930560.0

Publication Number: 1090368

IPC: G06F17/60

Language of the proceedings: EN

Title of invention:

A TECHNIQUE WHICH UTILIZES A PROBABILISTIC CLASSIFIER TO
DETECT "JUNK" E-MAIL

Applicant:

Microsoft Technology Licensing, LLC

Headword:

Spam classification/MICROSOFT

Relevant legal provisions:

EPC 1973 Art. 56
EPC Art. 52(2)

Keyword:

Inventive step - mixture of technical and non-technical
features
Inventive step - mathematical method
Inventive step - technical effect (no)

Decisions cited:

T 1358/09, T 0258/03



Beschwerdekammern
Boards of Appeal
Chambres de recours

European Patent
Office
D-80298 MUNICH
GERMANY
Tel. +49 (0) 89 2399-0
Fax +49 (0) 89
2399-4465

Case Number: T 0022/12 - 3.5.01

D E C I S I O N
of Technical Board of Appeal 3.5.01
of 16 November 2015

Appellant: Microsoft Technology Licensing, LLC
(Applicant) One Microsoft Way
Redmond, WA 98052 (US)

Representative: Grünecker Patent- und Rechtsanwälte
PartG mbB
Leopoldstraße 4
80802 München (DE)

Decision under appeal: Decision of the Examining Division of the
European Patent Office posted on 16 June 2011
refusing European patent application No.
99930560.0 pursuant to Article 97(2) EPC.

Composition of the Board:

Chairman W. Chandler
Members: P. Scriven
I. Beckedorf

Summary of Facts and Submissions

- I. This is an appeal against the decision of the Examining Division to refuse the application EP99930560.0 for lack of inventive step.
- II. The Examining Division considered that the method steps in claim 1 of the main request defined a linguistic analysis using mathematical algorithms, which, as subject-matter excluded from patentability, could not support the presence of an inventive step. The implementation of the method on a computer was considered straightforward, using notoriously known techniques. The first and second auxiliary requests were rejected for similar reasons.
- III. In the statement setting out the grounds of appeal, the appellant requested that the decision under appeal be set aside and that a patent be granted on the basis of a new main or first or second auxiliary request. As a further auxiliary request, the appellant requested oral proceedings.
- IV. In a communication pursuant to Article 15(1) RPBA, the Board set out its preliminary view that the classification of messages as a function of their content was not technical *per se*, and that, if there were a technical effect, it could only reside in the implementation of the classification using a computer.

The Board also cited the document:

Joachims T: "Text Categorization with Support Vector Machines: Learning with Many Relevant Features", Proceedings of the European Conference on Machine Learning, 21 April 1998, pages 137-142 (A1).

V. In response to the Board's communication, the appellant submitted arguments in favour of the main request and auxiliary requests I and II. The appellant also filed new auxiliary requests III and IV.

VI. Oral proceedings before the Board took place on 16 November 2015, for the course of which reference is made to the minutes of the oral proceedings. The appellant requested that the decision under appeal be set aside and that a patent be granted on the basis of the main request, or one of auxiliary requests I and II filed with letter dated 26 October 2011 and auxiliary requests III and IV filed with letter dated 13 October 2015.

VII. Claim 1 of the main request reads as follows:

"A method of classifying an incoming electronic message (205) performed by a computer program (130) executing on a computer, as a function of content of the message, into one of a plurality of predefined classes,

wherein the classes comprise first and second classes for first and second predefined categories of messages, respectively;

the method comprising the steps of:

detecting whether each of a first group of predefined handcrafted features exists in the incoming message (205) so as to yield first output data;

analyzing text in the incoming message (205) so as to break the text into a plurality of constituent tokens;

ascertaining, using a word-oriented indexer and in response to said tokens, whether each of a second group of predefined word-oriented features exists in the incoming message so as to yield second output data, said first and second groups collectively defining an n-element feature space;

forming, in response to the first and second output data, an N-element feature vector which specifies whether each of N pre-defined features exists in the incoming message; and

applying the N-element feature vector as input to the probabilistic classifier (210) so as to yield the output confidence level for the incoming message, which specifies a probability that the incoming message belongs to said one class;

wherein the classifier (210) has been trained, on past classifications of message content for a plurality of messages (234) that form a training set and belong to said one class, to recognize said N features in the training set; and

classifying, in response to a magnitude of the output confidence level, the incoming message as a member of said one class of messages; said classifying comprising:

comparing the output confidence level for the incoming message to a predefined probabilistic threshold value so as to yield a comparison result; and

distinguishing said incoming message, in a predefined manner associated with the first class, from

messages (232) associated with the second class if the comparison result indicates that the output confidence level equals or exceeds the threshold level

wherein:

the applying step comprises the step of yielding the output confidence level for said incoming message (205) through a support vector machine; and

the comparing step comprises the step of thresholding the output confidence level through a predefined sigmoid function to produce the comparison result for the incoming message (205)."

VIII. Auxiliary request I differs from the main request in that it comprises the following, additional feature at the end of claim 1:

"wherein the characteristics of the sigmoid function may be adjusted using constants A and B as given by the equation

$$f(u) = \frac{1}{1 + e^{Au+B}}$$

IX. Claim 1 of auxiliary request II reads:

"A method of classifying an incoming electronic message (205) performed by a computer program (130) executing on a computer, as a function of content of the message, into one of a plurality of predefined classes, the method comprising the steps of:

determining whether each one of a pre-defined set of N features is present in the incoming message (205) so as to yield feature data associated with the

message;

applying the feature data to a probabilistic classifier (210) so as to yield an output confidence level for the incoming message which specifies a probability that the incoming message belongs to said one class;

wherein the classifier (210) has been trained, on past classifications of message content for a plurality of messages (234) that form a training set and belong to said one class, to recognize said N features in the training set;

classifying, in response to a magnitude of the output confidence level, the incoming message as a member of said one class of messages; and

updating, from a remote server, the probabilistic classifier (210)."

- X. Claim 1 according to auxiliary request III differs from the main request in that:

the detecting step ends with: "the handcrafted features determined through human judgment alone, wherein the handcrafted features comprise features correspondingly related to formatting, authoring, delivery or communication attributes that characterize a message as belonging to the first class";

the analysing step has the word "tokens" replaced by: "textual components that are separated from another such component by a blank or white space or leading or following punctuation mark";

the ascertaining step has the word "token" replaced by "textual components" and after the words "second output data," is inserted "wherein each of the word-oriented features defines a word";

and, to the wherein clause, the following text is appended after "training set": ", wherein the classifier is a modified support vector machine".

XI. Auxiliary request IV differs from auxiliary request III in that the form of sigmoid function is specified, as in auxiliary request I.

XII. The appellant's arguments can be summarized as follows:

The method according to claim 1 of the main request had the following novel features over A1:

- the text classification was applied to emails;
- the classification was based on handcrafted features in addition to word-oriented features; and
- the use of a sigmoid function for thresholding the output confidence level yielded by the support vector machine (SVM).

The invention had a technical effect and purpose and was particularly suitable for implementation on a computer.

The use of handcrafted features reduced the processing load compared to using only word-oriented features for extracting the same information. Thus, the handcrafted features allowed a more practical computer implementation.

Performing the method in two stages, using first an SVM

and then applying a sigmoid function to its output, had the technical effect of reducing processing load. This was because it was possible to adjust the result of the classification by adjusting the only parameters of the sigmoid, without having to retrain the SVM. The low complexity of the method made it particularly suitable for implementation on a computer. Thus, the invention was based on technical considerations.

The skilled person had no incentive to modify the classification method in A1 to use handcrafted features determined by a user, or to apply a sigmoid function to the output of the SVM.

Updating the classifier, for example by changing the weights, from a remote server, avoided the need to retrain the classifier on each local machine. This was a technical effect that counted towards inventive step. Furthermore, at the priority date, it was neither known nor obvious to use remote updates in connection with text classification.

Reasons for the Decision

1. *The invention*
- 1.1 The invention concerns the classification of emails, e.g. as either spam or legitimate mail (page 10, lines 18 to 21 of the published application).
- 1.2 An incoming email is first analysed to determine whether it contains one or more features in a set of

predetermined features that are particularly characteristic of spam (page 10, line 28 to page 11, line 3). Two types of feature are used: word-oriented and "handcrafted". The former refers to the presence of particular words, or stems of words, the latter to features determined through human judgement alone (page 22, line 25 to page 23, line 26). Examples of handcrafted features are multi-word phrases and non-word distinctions such as formatting attributes, sender address, and delivery attributes. For instance, most spam messages are sent at night from ".com" or ".net" domains (page 23, lines 11 and 18 - 19).

- 1.3 An N-dimensional feature vector, with one element for each feature in the set, is produced for each email (page 23, lines 28 to 30). The feature vector is input to a probabilistic classifier which generates an "output confidence level". The output confidence level is then compared with a threshold to produce an indication of whether or not the email is spam (page 24, lines 3 to 16).

- 1.4 In the claims of all the requests, the probabilistic classifier is a Support Vector Machine (SVM). An SVM is a learning algorithm for assigning objects one of two distinct classes. It is trained using objects with known classifications (for example, a set of emails that have been classified manually) to define a hyperplane, or hypersurface, that provides maximum separation between the two classes in feature vector space (page 39, lines 23 to 28). Objects to be classified are mapped to that same space and assigned to a class based on which side of the hyperplane they fall on.

SVMs can be implemented efficiently, because the

hyperplane can normally be defined in terms of a small subset of feature vectors.

2. *Main request - inventive step (Article 56 EPC)*

2.1 The Examining Division considered that the invention, in claim 1 according to the main request before it, did not involve an inventive step over a notoriously known computer, since the only contribution was to a field excluded from patentability under Article 52(2) EPC, namely mathematical algorithms used for the purpose of linguistic analysis.

2.2 The Board agrees with the Examining Division that the classification of messages as a function of their content is not technical *per se*. In this regard, it is immaterial whether the messages are electronic messages, because, even though an email has technical properties, it is the content of the email that is classified. Furthermore, mathematical methods as such are not technical and the application of a mathematical method as such in a non-technical analysis of message content does not change that. Thus, if there is a technical effect, it can only reside in the automation of the email classification using a computer. The technicality of the computer is not enough to establish a technical effect of any method that it executes.

2.3 The appellant argued that classification based on a combination of "handcrafted features" and "word-oriented features", in claim 1 of the main request, had the technical effect of reducing processing load. In the prior art, for example in A1, where the classification was based on automatically-detectable, word-oriented features, more processing would be

required to extract the same information as that provided by the handcrafted features. For example, detecting the phrase "weight loss" (a phrase that seems to occur frequently in spam email), using only word-oriented features, would require the evaluation of all combinations of two words. Since the handcrafted features allowed a more compact representation of features, they contributed to a less complex computer implementation.

- 2.4 The Board is not persuaded that the alleged effect is actually achieved by the invention. There is no link between the word-oriented and handcrafted features, so that the latter reduces the processing involved in the former. The handcrafted features are, rather, a different class of features that the user considers indicative of spam, but which cannot be expressed in terms of the presence of individual words. Simply adding a second class of features to the analysis increases the load rather than reducing it.

Furthermore, the Board does not consider that the de-automation of a computer-implemented method, by making a human perform steps that a computer could do automatically, is a technical solution to a technical problem. Any reduction in computer processing would be a mere consequence of the de-automation.

- 2.5 In the Board's view, handcrafted features relate to information content that is considered as indicative of spam. Including such features in the analysis might, if well chosen, improve the quality of the classification, but the designation of a second class of features does not provide a technical effect.

2.6 It is common ground that the SVM, being a mathematical method, does not, as such, provide a technical effect. The appellant argued, however, that there was a technical effect in the particular combination of an SVM and a sigmoid function, as in claim 1. Performing the method in two stages, first using an SVM, and then applying an adjustable sigmoid function as a threshold to the output of the SVM, reduced the processing load, which reduced the complexity of the computer implementation. Thus, the invention was motivated by technical considerations of the computer implementation.

2.7 The Board is not persuaded by the appellant's arguments on this point. The Board does not find support, anywhere in the application, for the classifier being updated by adjusting the sigmoid parameters alone, without retraining the SVM.

As shown in figure 3B, the generation of parameters for the classifier during the training phase involves two steps (page 39, lines 6 to 15):

1) first the weight vector w is determined by conventional SVM training methods (page 39, line 18 to page 50, line 22);

2) second the optimal sigmoid parameters are calculated by using a maximum likelihood on the training data (page 50, line 24 to page 54, line 2)

There is nothing to suggest that re-training may involve only one of those steps, or that the classifier may be updated by simply adjusting the parameters A and B. On the contrary, it is the teaching of the application that, when the conditions of what is

considered as spam change (e.g. when the user reclassifies a message) the whole classifier is retrained (page 36, lines 14 to 30).

- 2.8 Furthermore, the Board does not consider that reducing the complexity of an algorithm is necessarily a technical effect, or evidence of underlying technical considerations. That is because complexity is an inherent property of the algorithm as such. If the design of the algorithm were motivated by a problem related to the internal workings of the computer, e.g. if it were adapted to a particular computer architecture, it could, arguably, be considered as technical (see T 1358/09, point 5.5, referring to T 258/03 "Auction method/HITACHI", OJ EPO 2004, 578, point 5.8). However, the Board does not see any such motivations in the present case.
- 2.9 Thus, the Board is not persuaded that the use of an SVM in combination with a sigmoid threshold function contributes, technically, to the the computer implementation. The Board rather considers this to be a mathematical method.
- 2.10 In the Board's view, the technical implementation of the method consists in programming the computer to perform the method steps. The Board concurs with the Examining Division that this would have been a routine task for the skilled programmer.
- 2.11 Therefore, the Board concludes that the subject-matter of claim 1 lacks an inventive step (Article 56 EPC).
3. *Auxiliary request I*

- 3.1 The first auxiliary request differs from the main request in that claim 1 defines the sigmoid function as the logistic function

$$f(u) = \frac{1}{1 + e^{Au+B}}$$

This does not make the claimed subject-matter more inventive, because, as concluded with regard to the main request, the sigmoid function is part of the mathematical method, and not of the computer implementation. Thus, the first auxiliary request is no more allowable than the main request.

4. *Auxiliary request II*

- 4.1 Claim 1 of auxiliary request II has neither the handcrafted features, nor the SVM in combination with a sigmoid function. Instead, claim 1 comprises the feature that the classifier is "updated" from a remote server.

The updates include software modules for the classifier, and feature set definitions (page page 55, lines 4 to 10). The updates are effected either manually on demand by the user, or automatically by the server (page 55, lines 10 to 21).

- 4.2 The appellant argued that the updates were weights for the classifier, so that, by updating the classifier from a remote server, it became unnecessary to retrain the classifier to determine new weights. Nevertheless, in oral proceedings, the appellant agreed that claim 1 covered any kind of updates, such as a new version of the software.

4.3 The Board considers that software updates from a remote server were notorious at the date of the invention, and that it would have been obvious to use such updates for the classifier.

The Board furthermore considers that it would have been obvious to determine the weights for the classifier at a central location, in order to save local processing resources. The Board moreover notes that the software vendor would normally have access to a larger training set, and, therefore, it would be natural to determine the weights from that set.

4.4 In conclusion, claim 1 of auxiliary request II does not define any subject-matter which establishes an inventive step over a notoriously known client-server computer system.

5. *Auxiliary requests III and IV*

5.1 Auxiliary requests III and IV differ from the main request and auxiliary request I, respectively, by the definitions of the handcrafted features, the textual components, the word-oriented features, and the support vector machine. According to the appellant, these definitions were clarifying amendments that did not change the situation regarding inventive step. Therefore, the Board considers that the same reasons apply to auxiliary requests III and IV, and concludes that also these requests lack inventive step.

Order

For these reasons it is decided that:

The appeal is dismissed.

The Registrar:

The Chairman:



T. Buschek

W. Chandler

Decision electronically authenticated