December 03, 2018

Crowd-sourced verification of content

Tuna Toksoz

John Dukellis

Crowd-sourced verification of content

ABSTRACT

Damaging or objectionable innuendo can sometimes get associated with otherwise harmless-looking symbols or images. Although machines can recognize a wide range of mature or objectionable material, innuendo is difficult to detect. In certain jurisdictions, innuendo, even if transitory or understood only within limited groups of users, is considered objectionable. A media or communications provider who is unable to detect such objectionable material can be subject to penalty.

This disclosure provides techniques to crowdsource the verification of content, e.g., ads, for the presence of innuendo or other objectionable content. The techniques apply to ads or content that is presented visually or aurally.

KEYWORDS

Content verification, mature content, objectionable material detection, crowdsourcing

BACKGROUND

Ads and media content are generally regulated or rated for the presence of objectionable material. Media providers typically use machine-learning models to check (and warn users as necessary) for the presence of mature content, or content considered objectionable by the community. While machine learning can detect a range of material that is widely agreed upon to be objectionable, it does not do as well to detect messaging that is obliquely or ambiguously objectionable. Innuendo is an example of material that is potentially objectionable in certain communities or jurisdictions. Innuendo is difficult to detect using machines because it can be associated with otherwise harmless-looking symbols (or statements), arises spontaneously, and often exists within limited groups before spreading to a wider audience. In some jurisdictions,

media providers or ad networks that display or broadcast ads or content with innuendo or indirect messaging, even if inadvertently or user-generated, are subject to penalty. Therefore, ad-networks operating in such jurisdictions are careful to review ads prior to selling them through web or app properties. Typically ad networks serving ads in such jurisdictions rely on manual human review of most ads prior to display. Human review, which is often in the form of hired internal employees, is relatively expensive.

DESCRIPTION

This disclosure provides techniques to crowdsource the verification of content, e.g., ads, for the presence of content that is objectionable within a jurisdiction. The vetting of content can be done outside the jurisdiction in question, and those users who assess content can be rewarded.

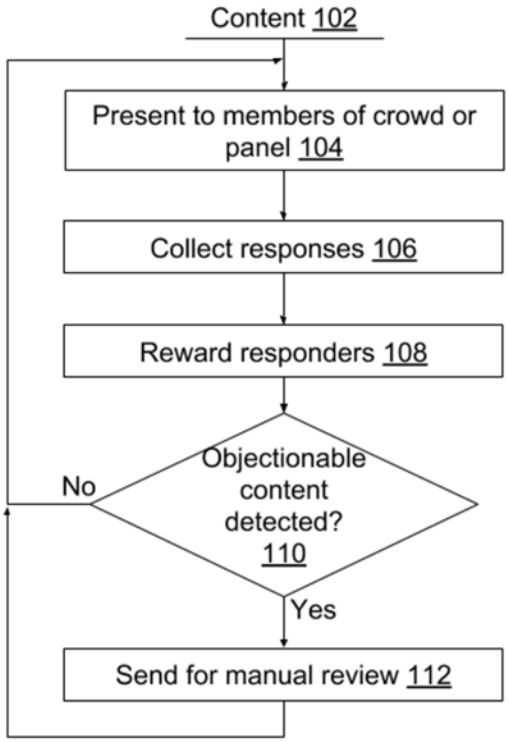

Fig. 1: Crowdsourced verification of content

Fig. 1 illustrates an example of crowdsourced verification of content, e.g., ads, per techniques of this disclosure. Content (102) is presented to members of a panel or crowd (104) with a request for review for objectionable content. The members of the panel may be located outside the jurisdiction where the content is being regulated, although they are familiar with the language and regulations of the jurisdiction. The content may be presented visually to panel members in the form of, e.g., pop-up dialog boxes with yes/no response. Alternately, a member of the panel may be asked to read the content out verbally, so as to verify words as they appear to users. The panel member is asked if objectionable innuendo is present within the content, with response being a yes or a no.

Responses are collected (106) and responders rewarded (108). Rewards can be, e.g., free services, ad-free web/app usage, in monetary form, e.g., one dollar per a predetermined number of content assessments, etc. If objectionable content is detected (110), e.g., if a substantial fraction of panel members affirm the presence of objectionable content, the content is sent for internal manual review (112). The process repeats. In this manner, a substantial fraction of content is vetted in a crowdsourced, e.g., inexpensive, manner, with a relatively small fraction of content being subject to expensive internal review.

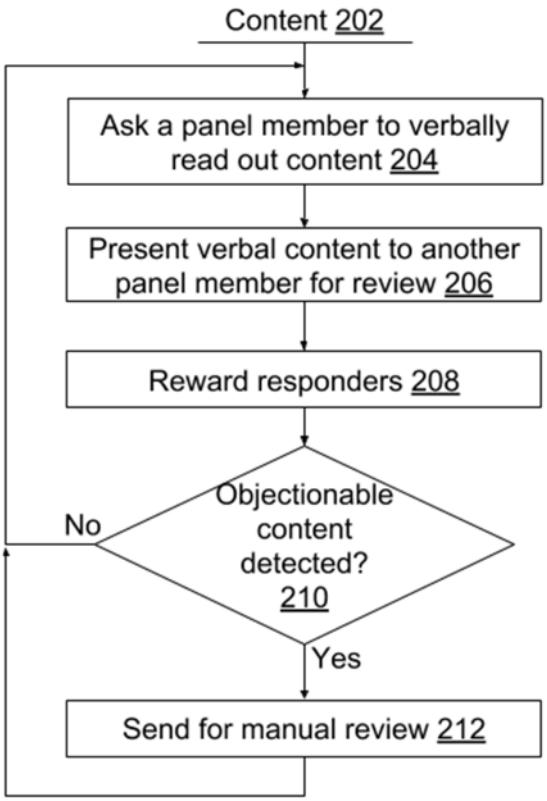

Fig. 2: Aural verification of content

Fig. 2 illustrates an aural-based crowdsourced technique for verification of content, per techniques of this disclosure. Content (202) is presented to a crowd-member or panel-member (204) to be read out verbally. The read-out verbal content is presented to another panel member (206) with a request to review for objectionable content, e.g., innuendo. The presenting of content for review may be done over audio-enabled or screen-free devices, e.g., smart speakers. Alternately, content, e.g., ads, that is directly created in an audio medium is presented to panel members for review. The reviewing panel member is asked if objectionable content, e.g., innuendo, is present within the content, with response being a yes or a no.

Responses are collected (206) and responders rewarded (208). Rewards can be, e.g., free services, ad-free web/app usage, in monetary form, e.g., one dollar per a predetermined number

of content assessments, etc. If objectionable content is detected (210), e.g., if a substantial fraction of panel members affirm the presence of objectionable content, the content is sent for internal manual review (212). The process repeats. In this manner, a substantial fraction of content, including audio-origin content, is vetted in a crowdsourced, e.g., inexpensive, manner, with a relatively small fraction of content being subject to expensive internal review. The technique described herein also has the ability to filter out visual content that reveals innuendo only when verbalized.

Once content, e.g., ads, is vetted using the techniques described herein, it is eligible for presentation to the user base in a manner compliant with local regulations.

CONCLUSION

This disclosure provides techniques to crowdsource the verification of content, e.g., ads, for the presence of objectionable content. The techniques apply to ads or content that is presented visually or aurally.