

# Live-kommentering av fotbollsmatcher med hjälp av maskininlärning och naturlig språkgenerering

Marcus Grönvall

**I takt med växande möjligheter och utvecklingen av maskininlärning och naturlig språkgenerering (eng. Natural Language Generation, NLG) de senaste åren, har även antalet användningsområden ökat. Bland dessa applikationer finns automatisk live-kommentering av fotbollsmatcher. Olika händelser kan detekteras med hjälp av maskininlärning, därefter händelserna kan kommenteras i text format genom NLG.**

På företaget *Spideo* i Malmö, vars affärsidé bygger på videoanalys av olika sporter, har man identifierat en växande marknad för automatisk live-kommentering av fotbollsmatcher. Speciellt för matcher i lägre divisioner eller ungdomsmatcher, vilka normalt inte kommenteras av mänskliga kommentatorer. Syftet med detta examensarbete var att konstruera en kommentator, vilken detekterar händelser med hjälp av maskininlärning och kommenterar händelserna genom NLG.

Vid träningen av maskininlärningsmodeller, användes data från 33 fotbollsmatcher i de svenska ligorna Allsvenskan och Superettan. Datan bestod av ljudinspelningar och spelarpositioner från matcherna, där ett flertal särdrag extraherades beroende på typ av detektion. Speciellt för klassificering användes taggar, som innehåller information och tidsstämplat för händelserna i matcherna, vilka var givna av Svensk Elitfotboll eller producerade manuellt genom att kolla på matcherna.

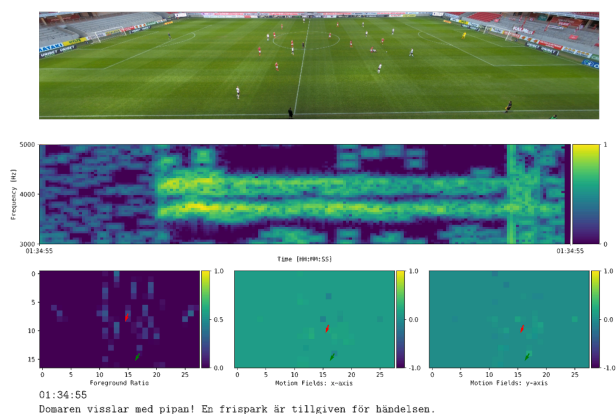
För ljudinspelningarna extraherades spektrogrammet, vilket användes för att detektera domarvisslingar under en match. Speciellt karaktäriseras en vissling av två horisontala linjer i frekvensintervallet 3000 – 5000 Hz, vilka motsvarar visslingens längd i tid. För spelarpositionerna extraherades förgrunden och rörelsefälten, vilka tillsammans användes för att detektera olika typer av händelser under en match. Specifikt användes positionssärdraget för att detektera pågående anfall och för att bestämma händelserna som detekterades som domarvisslingar, vilka kategoriserades som avspark, frispark, omgivningen eller övriga. Förgrunden extraherades genom att varje pixel i en videobild modellerades som sannolikheten att tillhöra ett objekt i rörelse, såsom spelarna, domarna, bollen, etc. Vidare extraherades rörelsefälten genom att beräkna de momentära positionerna och motsvarande hastigheter för varje objekt med hjälp av en spårningsalgoritm. Därefter projicerades sannolikheterna och rörelserna på fält av storleken  $17 \times 28$ , där motsvarande värden för varje område beräknades som medelvärdet. Positionssärdraget motsvarar alltså en tre-dimensionell tensor, där den första dimensionen motsvarar förgrunden och de övriga dimensionerna rörelsefälten i  $x$ - respektive  $y$ -led vid olika tidsstämplat.

Modellerna för maskininlärning bestod av neurala nätverk med konvolutioner (eng. Convolutional Neural Networks, CNNs), där 1D konvolutioner användes för spektrogrammet och 3D konvolutioner för positionssärdraget. För att utvärdera modellerna användes förvirringsmatriser (eng. confusion matrices), ROC- och Precision-Recall-kurvor, där motsvarande optimala trösklar bestämdes genom det maximala geometriska medelvärdet respektive F1-värdet. Givet de optimala modellerna för pågående anfall och domarvisslingar samt bestämning av händelsen som domaren visslar för, konstruerades en detektor. Utöver de angivna modellerna så detekterades även höga ljudvolymmer, där volymtröskeln anpassades manuellt.

För att kommentera händelserna i en match så konstruerades en kommentator, vilken bestod av detektorn och ett NLG-system. Vidare för att öka variationen av kommentarerna, matades så mycket information som möjligt in i NLG-systemet. Däribland själva detektionen, det nuvarande innehållet i matchen och allmän information om matchen. NLG-problemet består av att lösa olika delproblem, där man bestämmer vilken information som ska ingå i texten, den lingvistiska strukturen och slutligen textgenerering. Dessa problem kan lösas på olika sätt beroende på användningsområde, där gränsen mellan olika delproblem kan vara diffus och vissa delproblem kan lösas tillsammans. För händelserna i en fotbollsmatch löstes delproblemen i sekvensiell ordning, där texten genererades genom stokastiska val bland olika alternativ i handgjorda mallar av meningssupplyggnader.

Resultatet visade att detektion av höga ljudvolymmer och domarvisslingar är känsliga för arenan, där detektionströsklarna var tvungna att anpassas manuellt innan en inspelning från en ny arena. Vidare var det svårt att bestämma den specifika händelse som domaren visslade för under en live-inspelning, där prestandan var försämrad jämfört med evalueringen. Däremot kunde pågående anfall detekteras korrekt, om man sorterar bort detektioner där rörelsen på plan är låg och bortser från detektioner då motståndarlaget pressar backlinjen. Givet hur detektorn presterade, kunde beskrivande kommentarer av händelserna kommenteras. Resultatet visade att variationen av kommentarerna är känsliga för den information som kunde extraheras i samband med en detektion, vilket var relativt begränsat.

Slutligen, för att exemplifiera detektorn och motsvarande kommentar genererad av NLG-systemet, ges kommentaren då domaren blåser för en frispark enligt figur 1.



**Figur 1:** Kalmar FF - Örebro SK, Guldfågeln Arena i Kalmar, 6 augusti 2020. Den övre bilden visar videobilden för detektionen. Den mellersta bilden visar spektrogrammet för detektionen av en domarvissling, där man tydligt kan urskilja två horisontala linjer motsvarande visslingens längd i tid. Den nedre bilden visar positionssärdraget för bestämning av händelsen, där den vänstra bilden motsvarar förgrunden och den mellersta och högra rörelsefälten i  $x$ - respektive  $y$ -riktningen. Den röda och gröna pilen motsvarar position och riktningen för masscentrum respektive den maximala rörelsen. Detektionen är korrekt, därav motsvarande kommentar beskriver händelsen korrekt.